LONGITUDINAL SURVEYS OF AUSTRALIAN YOUTH

TECHNICAL REPORT 91

2015 cohort user guide

Longitudinal Surveys of Australian Youth

Longitudinal Surveys of Australian Youth (LSAY)

2015 cohort user guide

National Centre for Vocational Education Research

LONGITUDINAL SURVEYS OF
AUSTRALIAN YOUTH

TECHNICAL PAPER 91

Date created: December 2017
Last updated: August 2026
Version: 10.0

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

Publisher's note

Additional information relating to this publication is available from the LSAY website: <www.lsay.edu.au/publications/search-for-lsay-publications/lsay-2015-cohort-user-guide>.

© Commonwealth of Australia, 2026

With the exception of the Commonwealth Coat of Arms, the Department's logo, any material protected by a trade mark and where otherwise noted all material presented in this document is provided under a Creative Commons Attribution 3.0 Australia <creativecommons.org/licenses/by/3.0/au> licence.

The details of the relevant licence conditions are available on the Creative Commons website (accessible using the links provided) as is the full legal code for the CC BY 3.0 AU licence <creativecommons.org/licenses/by/3.0/legalcode>.

The Creative Commons licence conditions do not apply to all logos, graphic design, artwork and photographs. Requests and enquiries concerning other reproduction and rights should be directed to the National Centre for Vocational Education Research (NCVER).

This document should be attributed as NCVER 2026, *Longitudinal Surveys of Australian Youth (LSAY) 2015 cohort user guide*, NCVER, Adelaide.

This work has been produced by NCVER through the Longitudinal Surveys of Australian Youth (LSAY) Program, on behalf of the Australian Government and state and territory governments, with funding provided through the Australian Government Department of Education.

COVER IMAGE: GETTY IMAGES/THINKSTOCK

TD/TNC 130.14

Published by NCVER, ABN 87 007 967 311

Level 4, 191 Pulteney Street (Naylor House), Adelaide, SA 5000
PO Box 8288 Station Arcade, Adelaide SA 5000, Australia

Phone +61 8 8230 8400 **Email** ncver@ncver.edu.au
Web <https://www.ncver.edu.au> <http://www.lsay.edu.au>

Follow us:

<https://www.linkedin.com/company/ncver>

User guide updates

Date	Version	Update
August 2026	10.0	Updated for the latest data release (Wave 11, 2025). Revised to include linked data from the 2024 National VET Provider Collection. See section 'Data linkage' and appendices B and C for details.
October 2025	9.0	Updated for latest data release (wave 10, 2024).
July 2025	8.1	Revised to include data from the 2023 National VET Provider Collection and Higher Education Statistics Collection, as well as linked data for newly consenting respondents. See section 'Data linkage' and appendices B and C for details.
September 2024	8.0	Updated for latest data release (wave 9, 2023). Minor modifications to section 'Sample sizes and response rates'.
May 2024	7.1	Revised to include updates to the linked National VET Provider Collection and Higher Education Statistics Collection data. See section 'Data linkage' and appendices B and C for details.
September 2023	7.0	Updated for latest data release (wave 8, 2022).
May 2023	6.1	Revised to include: references to LSAY-NAPLAN data linkage technical paper; updates to the linked National VET Provider Collection data; and revisions to linked NAPLAN data (to include weights and exclude non-eligible NAPLAN records from 2007). See section 'Data linkage' and appendices B and C for details.
September 2022	6.0	Updated for latest data release (wave 7, 2021).
December 2021	5.1	Revised to include senior secondary linked data, and updates to linked National VET Provider Collection and Higher Education Statistics Collection data. Minor edits to NAPLAN linkage consent and linkage tables (tables 7 and 8) to reflect corrections to value assignments on data file. Correction to sample sizes and response rates notes at table 24.
August 2021	5.0	Updated for latest data release (wave 6, 2020). Updated information relating to use of the ANZSIC codeframe.
June 2021	4.1	Updates to incorporate revisions to linked NAPLAN data.
December 2020	4.0	Updated for latest data release (wave 5, 2019). Includes information about updates to existing and newly linked datasets.
March 2020	3.1	Added information regarding a break in the time series between 2010 and 2011 for the National Assessment Program – Literacy and Numeracy (NAPLAN) writing task.
December 2019	3.0	Updated for latest data release (wave 4, 2018). Includes information regarding the linked datasets.
March 2019	2.0	Updated for latest data release (wave 3, 2017). Includes information regarding the wave 3 (2017) top-up sample.
December 2017	1.0	Original version of user guide. Includes information for the first two survey waves: wave 1 (PISA 2015) and wave 2 (2016).

Contents

User guide updates 3

Tables and figures 6
- Tables 6
- Figures 7

Background 8

Using this guide 9

The Y15 cohort 10

Accessing the data 12
- Subscribe 12
- Unit record files 12
- LSAY QuickStats 15
- Pivot tables 16
- Specific data requests 16

Overview of the questionnaires 18
- Programme for International Student Assessment 18
- The LSAY questionnaire 19

The LSAY data 23
- Programme for International Student Assessment 23
- Variable naming conventions 24
- Derived variables 28
- Classifications and code frames 29

Data linkage 33
- Linked datasets 33
- Methodology 33
- ACARA *My School* data 34
- National Assessment Program – Literacy and Numeracy (NAPLAN) 35
- Senior secondary data 38
- National VET Provider Collection 42
- Higher Education Statistics Collection 45

Supporting documentation 48
- Questionnaires and frequency tables 48
- Data framework 48
- Variable listing and metadata workbook 50
- Online data dictionary 51
- Topic maps 53
- Technical and discussion papers 53

Sample and survey design 54
- PISA sample design 54
- Establishing the LSAY sample 54

Response rates 55
Top-up sample 61
Data collection 62
Sources of error 66
Topic maps 70
Topic map 1: Demographics – Student 72
Topic map 2: Demographics – Parent 74
Topic map 3: Education – School 77
Topic map 4: Education – School transition 89
Topic map 5: Education – Post-school 92
Topic map 6: Employment – Current 106
Topic map 7: Employment – Job history and training 111
Topic map 8: Employment – Seeking employment 113
Topic map 9: Employment – Not in the labour force 116
Topic map 10: Social – Health, living arrangements and finance 117
Topic map 11: Social – General attitudes 117
References 129
Appendix A: Updates to the Y15 data file 130
Appendix B: Linked datasets 143
Appendix C: Updates to the linked data files 145

Tables and figures

Tables

1 PISA technical documents 23

2 Summary of PISA variable naming conventions 26

3 Summary of LSAY non-standard variable naming conventions 27

4 Derived variables 29

5A Summary of classifications and code frames used in the LSAY Y15 data file 30

5B Summary of classifications and code frames used in the linked LSAY Y15 data files 30

6 LSAY ACARA *My School* datasets 34

7 LSAY NAPLAN consent rates 36

8 LSAY NAPLAN linkage rates 36

9 LSAY NAPLAN linked data by academic year level 37

10 LSAY NAPLAN datasets 37

11 LSAY senior secondary consent rates 39

12 LSAY senior secondary linkage rates 39

13 LSAY senior secondary dataset 40

14 LSAY VET consent rates 43

15 LSAY VET linkage rates 44

16 LSAY VET linked datasets 44

17 LSAY higher education consent rates 45

18 LSAY higher education linkage rates 46

19 LSAY higher education linked datasets 46

20 LSAY higher education key analysis variables 47

21 Topic maps 53

22 Sample cleaning outcomes 55

23 Sample sizes and response rates 56

24 Additions and exclusions to the LSAY Y15 sample 56

25 LSAY top-up sample, wave 3 (2017) 57

26 Wave 5 (2019) sample groups 58

27 Weight variables 68

A1 Summary of changes made to the Y15 data file 130

B1 Linked datasets 143

C1 Summary of changes made to the linked NAPLAN data files 145

C2 Summary of changes made to the linked National VET Provider Collection data files 146

C3 Summary of changes made to the linked higher education statistics data files 151

Figures

1 LSAY QuickStats 16
2 Pivot tables 17
3 PISA variable naming convention 25
4 LSAY standard variable naming convention 27
5 LSAY hierarchical levels 48
6 Major topic area 1 – Demographics 49
7 Major topic area 2 – Education 49
8 Major topic area 3 – Employment 49
9 Major topic area 4 – Social 50
10 Identifying related topic areas 52
11 Data dictionary 52

Background

The Longitudinal Surveys of Australian Youth (LSAY) is a research program that tracks young people as they move from school into further study, work and other destinations. It uses large nationally representative samples of young people to collect information about education and training, work and social development.

The LSAY program includes surveys conducted from the mid-1970s through to the mid-1990s: the Youth in Transition (YIT) program; the Australian Longitudinal Survey (ALS); the Australian Youth Survey (AYS); and the current LSAY collection, which began in 1995. Information about earlier versions of LSAY can be found in LSAY technical paper no. 2 (ACER 1997).

Survey participants in the current LSAY collection (collectively known as a 'cohort') enter the study at age 15 years or, as was the case in earlier studies, when they were in Year 9. Individuals are contacted once a year up until the age of 25. Studies began in 1995 (Y95 cohort), 1998 (Y98 cohort), 2003 (Y03 cohort), 2006 (Y06 cohort), 2009 (Y09 cohort) and in 2015 (Y15 cohort). About 14 000 students start out in each cohort.

Since 2003, the initial LSAY survey wave has been integrated with the Organisation for Economic Co-operation and Development's (OECD) Programme for International Student Assessment (PISA).

The LSAY research program provides a rich source of information for enabling a better understanding of young people and their transitions from school to post-school destinations. It also explores their social outcomes, such as wellbeing. Information collected as part of the LSAY program covers a wide range of school and post-school topics including: student achievement, student aspirations, school retention, social background, attitudes to school, work experiences and what students do when they leave school.

LSAY is managed and funded by the Australian Government Department of Education, Skills and Employment with support from state and territory governments. On 1 July 2007, the National Centre for Vocational Education Research (NCVER) was contracted to provide analytical and reporting services for the LSAY program. Between 1995 and 2007, these services were provided by the Australian Council for Educational Research (ACER).

More information can be obtained from the LSAY website, or by contacting the LSAY team at NCVER:

Telephone:	+61 8 8230 8400	Email:	<lsay@ncver.edu.au>
Facsimile:	+61 8 8212 3436	Website:	<www.lsay.edu.au>

Using this guide

This *User guide* has been developed for users of the LSAY data. The guide brings together the resources available for data users and includes information on: how to access the data, the questionnaires, variable naming conventions, derived variables, the classifications and code frames used, the structure of the data (using topic areas, topic maps and data elements), supporting documentation, sample design and weights.

The LSAY data dictionary complements this user guide. It is designed to provide easy access to LSAY metadata using 'topic areas' to group variables into common themes; and 'data elements' to represent variables that are common within and between waves.

Further information about the data dictionary is contained in the 'Supporting documentation' section of this user guide. The data dictionary can be accessed at: <www.lsay.edu.au/data/lsay-data-dictionary>.

Users may also find the variable listing and metadata workbook useful. This workbook has the same information as the data dictionary but it is presented in Excel rather than as an online tool. The variable listing provides a complete list of the variables in the LSAY data files, as well as metadata for each variable, which includes the questionnaire text, base populations and values. The data can be filtered and inspected by cohort, wave/year, questionnaire section, topic area(s) and/or data element.

Further information about the variable listing and metadata is contained in the 'Supporting documentation' section of this user guide. The variable listing can be accessed at: <www.lsay.edu.au/publications/2621.html>.

If you have any feedback or encounter issues finding the information you need in this guide, please do not hesitate to contact the LSAY team at NCVER.

Telephone:	+61 8 8230 8400	Email:	<lsay@ncver.edu.au>
Facsimile:	+61 8 8212 3436	Website:	<www.lsay.edu.au>

The Y15 cohort

In 2015, a nationally representative sample of students aged 15 years was selected to participate in the Programme for International Student Assessment (PISA), conducted by the OECD. This sample became the sixth cohort of the LSAY program. This is referred to as the LSAY 'Y15' cohort, reflecting the year in which the sample was taken.

The PISA sample was constructed by randomly selecting students aged 15 years from a sample of schools designed to represent all states and school sectors. In Australia, 758 schools and 14 530 students participated in PISA. Assessments in mathematical literacy, reading literacy and scientific literacy were administered in schools to provide information on student achievement. Collaborative problem-solving and financial literacy[1] were also assessed as part of PISA 2015.

Students also completed a background questionnaire, which collected information on: their family background; their attitudes towards learning, their habits and life in and outside school; aspects of students' interest, motivation and engagement; and their learning and instruction in science. They also completed an information and communications technology (ICT) questionnaire, which collected information on the availability and use of ICT, students' perceptions of their competence in completing computer tasks and their attitudes towards computer use. A third educational career questionnaire gathered information on whether students had experienced interruptions in their schooling, and their preparedness for their future career (Thomson, de Bortoli & Underwood 2017). Students were also asked to provide their contact details so they could be contacted for a follow-up study (that is, LSAY).

Students providing their contact details as part of PISA were contacted in 2016 for their annual LSAY interview, conducted by the Wallis Social Research (Wallis). The questionnaire for their 2016 interview included questions on school, transitions from school, post-school education and training, engagement in education, careers advice, work, job satisfaction, job-search activity, income, soft skills, personality and wellbeing. Respondents were also asked for their consent to link their National Assessment Program — Literacy and Numeracy (NAPLAN) scores and senior secondary subject results to their LSAY records.

Response rates for the 2016 interviews were lower than expected due to a high rate of missing or unusable contact details provided at the time of PISA. To address this, recruitment of a top-up sample of Year 12 students was conducted in 2017 to ensure that future waves of the survey remain representative of the 15-year-old population in 2015.

As part of respondents' first LSAY interview in 2016, a series of questions regarding consent to undertake linkages to other administrative data sources were introduced for the first time. Linked administrative data is now available for several data sources including the Australian Curriculum, Assessment and Reporting Authority's (ACARA) *My School*, National Assessment Program – Literacy and Numeracy (NAPLAN), the National VET Provider Collection and the Higher Education Statistics Collection.

The unchanging design, scope and collection of the LSAY data since the program began in 1995 has allowed for consistent and comparable data over the life of the surveys. This enables comparisons over time and between cohorts at the same age, providing opportunities to explore the effects of economic, technological and generational changes on Australian youth. Users need to use appropriate methods when undertaking cross-cohort comparisons, particularly when comparing the year level-based cohorts (Y95 and

1 The financial literacy data have not been made publicly available. Interested researchers need to submit an application to ACER to access these data.

Y98) with the age-based PISA cohorts (from Y03 to Y15), given the differences in the sample design for these groups.

Due to both population shifts over time and survey attrition, care must also be taken when comparing individual survey waves from any one LSAY cohort with samples drawn from other populations. For example, it can be misleading to compare the LSAY Y15 wave 2 (2016) information with information about 16-year-olds from other surveys in the same year.

Accessing the data

A number of different data products are available for different types of data users. Unit record data files can be accessed via a formal registration and approval process. The LSAY website also features data tools that provide quick and easy access to summary LSAY data. LSAY data tools and resources are available free of charge and include the pivot tables and LSAY QuickStats.

Information on the latest LSAY data releases is available from the LSAY website: <https://www.lsay.edu.au/data/access>.

Part of NCVER's role is to promote the use of the LSAY data. We encourage data users to provide NCVER with details of their published research to help us to maintain a bibliography of publications that use or reference LSAY, as well share research with colleagues of similar interests and promote LSAY research through our networks and dissemination activities.

If you have any feedback or queries about the data and how to access it, please contact the LSAY team.

Email: <lsay@ncver.edu.au>

Telephone: +61 8 8230 8400

Subscribe

You can also subscribe to the LSAY website to receive email alerts when new data and publications are released using the following link: <www.lsay.edu.au/subscribe.html>.

Unit record files

LSAY data files are deposited annually with the Australian Data Archive (ADA) at the Australian National University in Canberra. Data files are available to researchers and data users free of charge, and permission to use the data and access requirements are managed by the Australian Data Archive. Data access requires authorisation from the Data Archive Manager and NCVER.

The data can be accessed by:

- registering with the ADA Dataverse located at: <https://dataverse.ada.edu.au>
  - select 'Sign Up' from the top right corner and complete the Dataverse registration form. You will need to validate you email address for your registration to be accepted by Dataverse.
  - you can explore your Dataverse account by selecting your user name and heading to 'My Data', 'Notifications' or 'Account Information'.
- requesting access to the LSAY datasets
  - navigate to the LSAY Dataverse located at: <https://dataverse.ada.edu.au/dataverse/lsay> and login to your ADA Dataverse account.
  - navigate to the LSAY cohort you want to access from the list of datasets. Note: If you want access to multiple cohorts at one time you can select this option when filling out the online application form.
  - scroll down to the data files, select the file type/s you wish to access. Click on 'Request Access' and complete the online application form. **Users must comply with the terms and**

conditions outlined in the user undertaking in order to obtain access to the data (see below for details); a confirmation email will be sent from the ADA

- the ADA and NCVER will review your request and you will be notified of the outcome; if your request is approved, you will be able to download the requested files via the LSAY Dataverse.

User undertaking

LSAY is managed by NCVER on behalf of the Australian Government Department of Education, Skills and Employment. LSAY is designed to provide a rich source of information about young people and their pathways to help researchers and policymakers make informed decisions about youth policies.

Access to the nominated dataset is provided only if the individual requesting access undertakes to comply with the terms and conditions. Individuals must accurately identify themselves when interacting with any entity or technology supporting access to LSAY data. Applications must accurately reflect the intended use of the data. NCVER either directly or through an entity delivering services on its behalf, may seek to verify an applicant's identity and/or confirm the intended use of the data.

Terms and conditions

Authorised data users must undertake and agree to take full responsibility for ensuring LSAY unit record data files will be protected according to the following terms and conditions and hereby undertake to:

1 Use all information provided by the Australian Data Archive (ADA) only for the purposes specified in their application and as approved by NCVER.

2 Should the unit record data provided by NCVER be used for data matching/linking activities, comply at all times with the following conditions, and with any reasonable direction given by NCVER with respect to the disclosure, use or storage of matched/linked data. Authorised users must hereby undertake to:

   a. Comply with all applicable laws and regulations (including the Privacy Act 1988 (Cth));

   b. Comply with all NCVER policies, procedures and protocols, including those published on NCVER's website at <www.ncver.edu.au> to the extent that they relate to that party's role in the Project;

   c. Comply with all guidelines published by the Office of the Australian Information Commissioner which are mandatory for that party to comply with when matching/linking data;

   d. Comply, so far as is reasonably practicable, with all guidelines published by the Office of the Australian Information Commissioner which are relevant to that party's role in the data matching/linking activities/project but which are not mandatorily imposed on that party;

   e. And ensure compliance with this clause by the authorised users' employees, agents and subcontractors.

3 Store and protect the data from misuse, interference and loss and from unauthorised access, modification or disclosure, including:

   a. Protecting the privacy of the data and related individuals who may be identifiable in accordance with the Australian Privacy Principles (APPs) under the Privacy Act 1988 (Cth) as amended by the Privacy Amendment (Enhancing Privacy Protection) Act 2012.

   b. Not copying, sending or providing the data to other persons or organisations

c. Not attempting to identify an individual, including matching the information with any other information for the purposes of identifying individuals.

d. Not disclosing information to other persons or organisations on any particular individual or any information in the unit record data where the identity of a particular individual might reasonably be ascertained.

e. Not disclosing information on any particular organisation contained in the unit record data to any other person or organisation without the written permission of the organisation to which the information relates.

f. Not publishing or disclosing the data or research results in a way that would enable any individual or organisation (other than your own) to be identified.

g. Not using the information as a basis for legal, administrative, or other actions that could affect individuals or organisations (other than your own) contained in the unit record data.

h. Relinquishing access and not attempting to access the requested data if no longer working on the specified project/purpose, or upon ceasing employment with the specified organisation.

i. Ensuring that data in all media (CD-ROMs, DVDs, portable storage devices, electronic files, hard copy) are stored securely with access controls.

j. Destroying the data, including any data resulting from matching the unit record file with other datasets, and any copies of it at the conclusion of the specified project/purpose, provided that the organisation may, with NCVER's prior written consent (which will not be unreasonably withheld, but may be granted subject to conditions at NCVER's discretion), retain a single copy of data for archive purposes or to comply with any applicable laws or institutional policy, subject to the organisation continuing to comply with the terms of this undertaking.

k. Attribute the source of the data in any publications resulting from the use of the unit record data.

4 Provide a copy of any final reports and other data products to NCVER unless otherwise agreed by NCVER.

5 Not do anything or permit anything to be done that may cause NCVER to breach its obligations under the Privacy Act 1988 (Cth) or its Privacy Policy (located at www.ncver.edu.au/privacy.html

6 Regarding the linked ACARA *My School* data:

a. Use data in a manner consistent with the ACARA Principles and protocols for reporting on schooling in Australia

b. Provide ACARA at or before time of publication any publicly available document incorporating the data (or any part of the data) to ACARA for reference

c. Include the text 'These data may not be disseminated to external parties' for any document including the data or analysis of the data

d. Not report on individual schools or undertake comparisons between schools as part of the research

e. Not hold NCVER, ACARA, state/territory governments or the Australian Institute of Family Studies (AIFS) responsible for the accuracy, quality or fitness for purpose of the data.

Authorised users must unconditionally and irrevocably indemnify NCVER against any loss incurred by NCVER (including legal costs, on a solicitor own client basis) as a result of any failure by the Indemnifying Party or any of its officers, employees, contractors, agents or representatives to comply with these terms and conditions for any reason. The Indemnifying Party must on demand from NCVER immediately pay or reimburse NCVER's loss.

Authorised users must undertake to adhere to all conditions listed above and understand that any breach of these terms may result in withdrawal of access to the information and/or incur a legal penalty if there is a breach of the Privacy Act or a breach under Common Law through disclosure of an organisation's commercial in confidence information.

Authorised users must represent and warrant that the information set out in their Request is true and correct and acknowledge that NCVER will rely upon and be induced thereby to grant access to data held by the ADA.

LSAY QuickStats

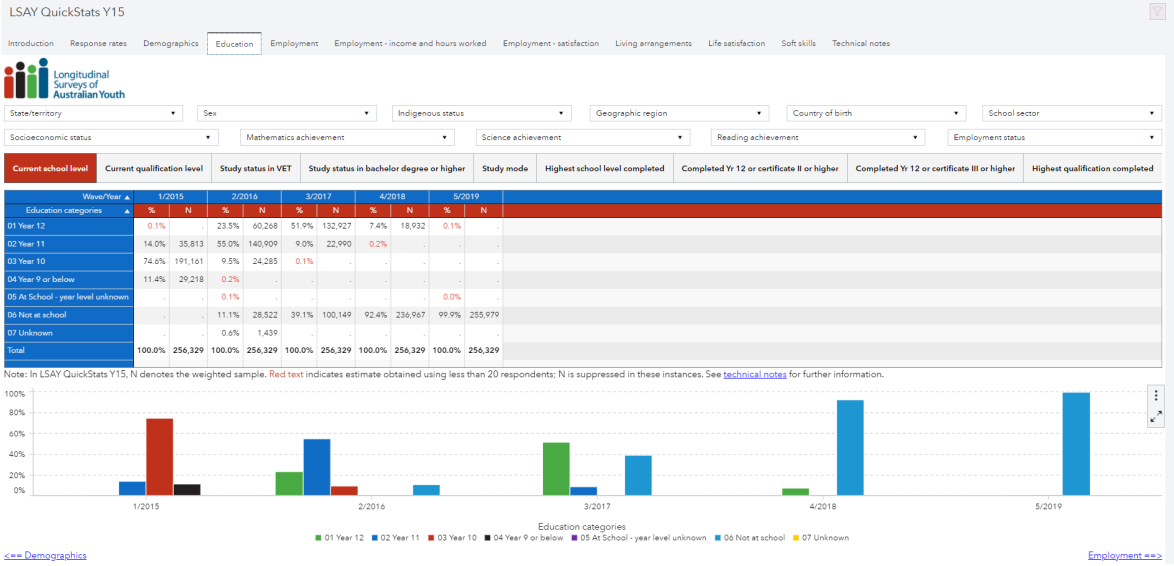
LSAY QuickStats provides access to summary data to enable users to readily explore results from the LSAY surveys. Data are presented as a series of tables and charts and include information on education and employment pathways, as well as social indicators on living arrangements and satisfaction with life.

Data are organised by wave/year, beginning with the first wave of data collection (for example, wave 1/2015) through to the final or most recent wave (for example, wave 6/2020). For those interested in particular groups of young people, data can be filtered by a range of demographic variables.

LSAY QuickStats can be accessed at <https://www.lsay.edu.au/data/lsay-quickstats>.

Note that LSAY QuickStats only includes respondents from the main sample (recruited from PISA) and does not include respondents from the top-up recruitment survey (conducted at wave 3 in 2017). This is because some measures presented are not available for the top-up group (for example, PISA scores). To reduce high proportions of missing data for these measures and ensure comparability with previous cohorts, the top-up group has been omitted.

Figure 1 LSAY QuickStats



Pivot tables

The pivot tables allow users to create their own Excel tables from a selection of variables. Users can drill down into the data and generate time series on the activities of young people, from the first to the most recent survey wave.

Data on key employment, education, study and work, and social indicators are presented. Selected demographics are presented in each of the pivot tables, including: sex, state, geographic location, school sector, country of birth and socioeconomic status.

The pivot tables can be accessed at: <https://www.lsay.edu.au/data/pivot-tables>.

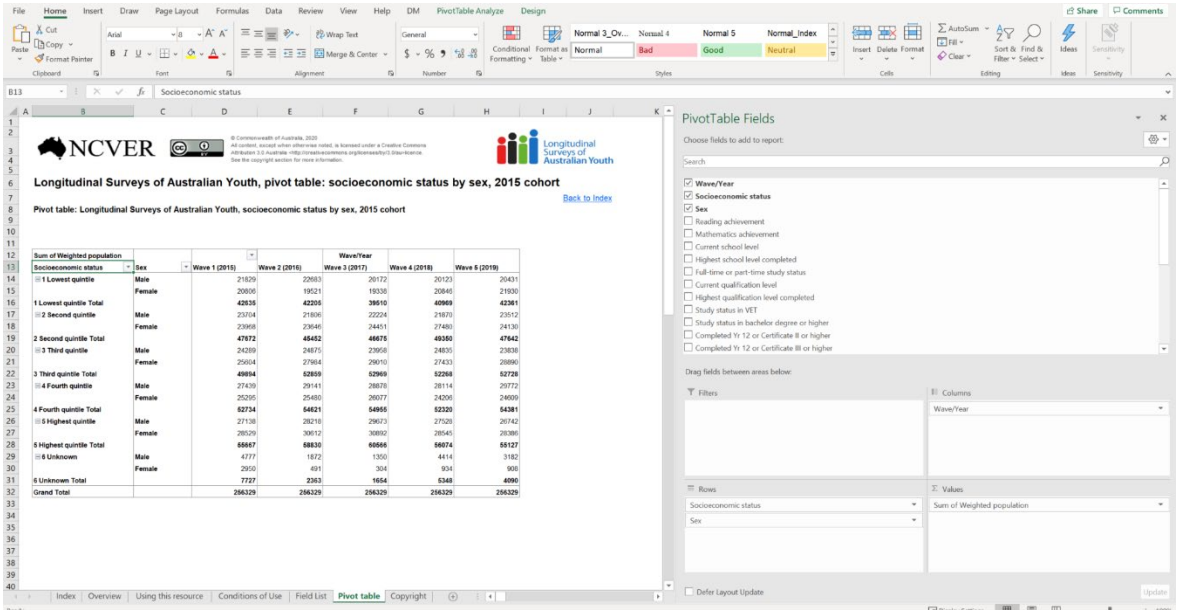
Note that the pivot tables only include respondents from the main sample (recruited from PISA) and does not include respondents from the top-up recruitment survey (conducted at wave 3 in 2017). This is because some measures presented are not available for the top-up group (for example, PISA scores). To reduce high proportions of missing data for these measures and ensure comparability with previous cohorts, the top-up group has been omitted.

Specific data requests

A specific data request allows a user to request specific tables and/or data analysis to be undertaken by NCVER without having to obtain full sets of the data. A specific data request can be made to <lsay@ncver.edu.au>.

There are fees and charges applicable for all data requests. Please refer to NCVER's data access and charging policy: <https://www.ncver.edu.au/support/support/all-support/data-access-and-charges>.

Figure 2 Pivot tables



Overview of the questionnaires

Programme for International Student Assessment

Since 2003, the first wave of LSAY has been integrated with the OECD's Programme for International Student Assessment. It is therefore important to understand the PISA 2015 instruments when using the Y15 cohort data.

As part of PISA 2015, students were assessed in three domains: mathematical literacy, reading literacy and scientific literacy, with the aim of providing information on academic achievement. For each PISA data collection, one of these domains is chosen as a major domain, while the others are considered minor domains. A major domain is tested more thoroughly in the year of collection. The major domain for PISA 2015 was scientific literacy. Students also undertook assessments in collaborative problem-solving and financial literacy as part of PISA in 2015. The PISA 2015 assessments consisted of a self-completion computer-based test. This presented a change from the previous paper-based delivery mode (Thomson, De Bortoli & Underwood 2017).

The *PISA 2015 Assessment and Analytical Framework* presents the guiding principles of the PISA 2015 assessment, which are described in terms of: the skills students need to acquire; the processes that need to be performed; and the context in which knowledge and skills are applied. The framework can be accessed at: <http://www.oecd-ilibrary.org/education/pisa-2015-assessment-and-analytical-framework_9789264281820-en>.

In addition to the cognitive assessment, the following survey questionnaire instruments were used:

- student questionnaire
- ICT questionnaire
- educational career questionnaire
- school questionnaire
- teacher questionnaire.

Parent, teacher, ICT and educational career questionnaires were offered as international options, with Australia participating in all but the parent questionnaire. The teacher questionnaire was offered as an international option for the first time in PISA 2015.

The *student questionnaire* collected information on: family background; aspects of students' lives, such as their attitudes towards learning, their habits and life in and outside of school; aspects of students' interest, motivation and engagement; and learning and instruction in science, including instructional time and class size. Students' attitudes towards science were also measured, assessing students': interest in science and technology; environmental awareness; and valuing scientific approaches to enquiry.

Students also completed an *information and communications technology (ICT) questionnaire*, which collected information on: the availability and use of ICT; students' perceptions of their competence in completing tasks; and their attitudes towards computer use.

A third questionnaire, the *educational career questionnaire*, collected information about schooling interruptions and preparedness for their future career. Students were also asked to provide their contact details so they could be contacted for a follow-up study (that is, LSAY) (Thomson, De Bortoli & Underwood 2017).

The 2015 PISA questionnaires and code books are available from the PISA 2015 database: <http://www.oecd.org/pisa/data/2015database/>.

The LSAY questionnaire

The unchanging design, scope and collection of the LSAY data since the program began in 1995 has allowed for consistent and comparable data over the life of the surveys. However, it also means the survey instrument has not been revised to reflect the changing needs of policy-makers and researchers. Recognising the need for the redevelopment of the survey instrument, the Australian Government Department of Education and Training undertook a review of LSAY in 2014 which included a series of stakeholder consultations.

Stakeholder consultations

The process for developing the wave 2 survey instrument involved a series of submissions and stakeholder consultations. As a part of this process, a background paper to the consultations was prepared. The paper *Future topic areas for the Longitudinal Surveys of Australian Youth — Y15 cohort* can be found at <www.lsay.edu.au/publications/2814.html>.

The major themes of interest emerging from the consultations and submissions were (in order of priority):

- choosing and understanding education and training
- the social environment of young people
- developing skills for work
- health and wellbeing
- the young person's home and personal life
- priority skills (including STEM and digital literacy)
- transitions and pathways
- labour market experiences
- labour market disengagement.

Many of the major themes of interest are already addressed by existing content in the LSAY questionnaires and were incorporated into the Y15 wave 2 survey instrument. However, due to time and space limitations, it is not possible to include *all* of these themes. Factors considered when selecting new modules include: the value of the topic (compared with the overall purpose of LSAY); the length of time required to effectively collect the information; the availability of suitable instruments; the extent of existing content identified for deletion; the sensitive nature of certain questions; and ethical requirements (particularly when collecting information from respondents less than 18 years of age).

As a result, the following four topic areas were addressed through the incorporation of six new modules (identified below in italics) into the questionnaire:

- Choosing and understanding education and training — the module *barriers to completing study or training* was introduced to help understanding of the process of how students choose their education courses and subjects, the factors that affect their decision-making and the limitations on their choices.
- The social environment of young people — the modules *time spent volunteering* and *providing unpaid care* provide insight into the activities of young people and how these activities and responsibilities might strengthen character and develop essential skills.

- Developing skills for work — *soft skills* and *personality* assist in identifying whether young people have the skills required to enter the world of work, alongside the role that non-cognitive skills (for example, confidence, resilience, empathy [personality type and personal attributes]) play in a successful youth transition.
- Transitions and pathways — collecting information on *engagement in education* will assist in understanding more about early school leavers and provide further insights into the factors that affect post-school education participation, attrition and completion.

Wave 2 (2016)

As a result of the 2014 LSAY review, the initial questionnaire for the new Y15 cohort (aged around 16 years) in 2016 (wave 2) was revised. The wave 2 questionnaire retained many of the elements from the previous questionnaires, while incorporating some new modules. The wave 2 questionnaire included the following sections:

- *LA: Last activity* — summary information about activities undertaken over the preceding year, including school activity, other qualifications undertaken, working for pay or profit, volunteering, and providing unpaid care
- *SC: School education* — school subject information, including vocational education and training (VET) subjects, learning in a work environment, engagement in school and reasons for leaving school before graduating
- *PSS: Post-school study* — higher education and training pathways, which includes reasons for changing courses, withdrawing/deferring from study or training, and/or changing employer
- *EE: Engagement in post-school education* — frequency of participating in post-school education activities
- *CAD: Career advice, aspirations and decision-making* — plans to complete Year 12, post-school plans, barriers to completing study or training, careers advice and expected occupation at age 30
- *EMP: Employment* — hours worked, employment status, occupation, job-finding methods and preferred work hours
- *JSA: Job satisfaction* — job satisfaction and whether employed in preferred career job
- *JSE: Job-search activity* — job-search methods, problems looking for work and availability to start work
- *INC: Income* — income from pay and other sources, including government benefits and allowances
- *SS: Soft skills* — self-perception of behaviours demonstrating creativity and innovation, critical thinking and problem-solving skills; and teamwork and communication.
- *PT: Personality* — factors from the Big-Five personality domains measuring extroversion, agreeableness, conscientiousness and openness.[2]
- *WEL*: *Wellbeing* — how respondents feel about life
- *DEM*: *Household and demographics* — living arrangements, household formation and information about schools attended (for the purpose of data linkage)
- *DL: Data linkage* — respondents were asked for their permission to link their LSAY records with their NAPLAN and senior secondary subject results.

2 Neuroticism was not collected at wave 2 because these questions were deemed not suitable to be asked of young people less than 18 years of age.

Waves 3 to 11 (2017 to 2025)

To assist in comparability with previous cohorts, from wave 3, the structure of the questionnaire was changed to reflect the former cohort questionnaire sections, with new modules added to relevant parts of the questionnaire.

New sections included gig work (acknowledging the growing 'gig economy'), volunteering and caring activities, transience (questions regarding homelessness and unstable housing situations) and social support (who respondents could ask for support in a time of crisis).

The data linkage section also expanded to include permission to link LSAY records with any vocational education and training (VET) or higher education undertaken.

From wave 3, the questionnaire includes the following sections:

- Section A: School — graduation, changing schools, subjects, post-school plans, careers advice
- Section B: Transition from school — school leaving, reasons for leaving school and senior secondary results
- Section C: Post-school study — current and past study, university preferences, apprenticeships and traineeships, qualifications, study outcomes, reasons for withdrawing/deferring from study, changes to study (including changes to course, institution, employer, and apprenticeship or traineeship), student engagement, course fees and careers advice
- Section D: Work — employment characteristics, time worked, wages and benefits, starting work, reasons for leaving work and job satisfaction
- Section E: Job history/additional employment — includes past/other employment, employment characteristics, time worked and reasons for leaving work
- Section F: Job search activity — job-seeking behaviour, whether looking for work, job preferences, job-search activity and problems looking for work
- Section FA: Gig work — earned money through gig work, types of gig work, importance of gig work, income received
- Section FB: Volunteering and unpaid care — volunteer activities, months/hours providing unpaid care, who cared for
- Section G: Not in the labour force — main activity
- Section H: Living arrangements and health — living with parents, accommodation type, household formation, moving house, government payments and income, transience and homelessness, and health
- Section J: General attitudes — soft skills, personality, life satisfaction/well-being, social support
- Section K: Interview variables — school history, consent for data linkage[3] and contact details.

Top-up sample recruitment questionnaire

An additional short paper-based recruitment questionnaire was also administered to Year 12 students in schools as part of the 2017 top-up activity (see section 'Sample and survey design') with the goal of collecting key demographic information, matching that collected by PISA at wave 1. This questionnaire collected the following information:

3 Data linkage questions were interwoven with the related linkage questions from wave 4. For example, consent to link to NAPLAN was moved to the end of 'Section A: School'.

- About you: date of birth, gender, school program, country of birth, Indigenous status
- Future plans: post-school plans and careers advice
- Work: hours worked and job expectations
- Family and home: parental education and occupation, household possessions
- Personal skills: soft skills.

LSAY and COVID-19

To allow comparability with previous cohorts and waves, the questionnaire remained largely unchanged during 2020. Some adjustments were made to some questions to allow for new scenarios arising because of the coronavirus pandemic. For example:

- additional response options were created to capture those: whose study or apprenticeship/traineeship was on hold due to COVID-19; who had a job but were not currently working due to COVID-19; and/or had deferred their rent, board or housing repayments due to COVID-19
- a response of zero hours was allowed when reporting hours worked
- JobKeeper and JobSeeker were added to the list of possible government benefits respondents (or their partners) received.

Some other additions included:

- 'current hours worked' was collected in addition to 'usual hours worked'
- a question about changes to a respondent's housing situation
- a measure of loneliness.

The LSAY questionnaires can be accessed from the 'User support and documentation page' on the LSAY website at: <www.lsay.edu.au/publications/user-support-and-documentation>.

The LSAY data

Programme for International Student Assessment

The OECD's 2015 PISA forms the basis of the data for the first LSAY Y15 survey wave, so it is important that users of the LSAY data also take some time to understand PISA 2015 when using the data. The following section briefly describes some features of the PISA data. Users are also encouraged to read the PISA technical documents, as outlined in table 1.

Table 1 PISA technical documents

Technical report/paper	Web address
PISA 2015 technical report	<http://www.oecd.org/pisa/data/2015-technical-report/>
PISA 2015: Reporting Australia's results	<http://research.acer.edu.au/ozpisa/22/>
The role of plausible values in large-scale surveys	<http://www.acer.edu.au/files/plausiblevaluesinsee.pdf>

National options

Countries participating in PISA are able to introduce country-specific questions into PISA questionnaires, referred to as 'national options' questions. In 2015, as part of these national options, Australia included questions about a student's: Indigenous background, language spoken at home and country of birth. For this reason, in addition to the publicly available PISA international data file, a separate national data file is created for Australia, which includes these national options questions. Some variables available from the international data file are omitted from the national data file (for example, country). In addition, some minor differences may exist between the two versions of the data file, for example, the way missing or not applicable values have been assigned to observations, or whether the variables are in numeric or character format.

PISA international school, teacher and student files

The PISA international student, teacher and school data files are available from the PISA 2015 database: <http://www.oecd.org/pisa/data/2015database/>. LSAY data can be matched to the PISA international data files by filtering for Australian records using the country identifiers (CNT, CNTRYID), and using student and school identifiers (CNTSTUID and CNTSCHID), available on both the international and LSAY data files.

For users wishing to use the PISA school or teacher data, or to make international comparisons using PISA data, we recommend downloading the international data file available from the OECD 2015 PISA international database, located at: <http://www.oecd.org/pisa/data/2015database/>.

Plausible values

For previous PISA assessments, to mitigate the effects of students responding differently depending on the position of the item in the test, 13 different test booklets were used, with students randomly assigned to one of the booklets. To counteract any biases resulting from the use of different text booklets, the OECD calculates plausible values. Plausible values allow for the fact that there is measurement error at the individual level (through differing questionnaires), and the determination of these plausible values takes this error into account.

In PISA 2015, the assessment mode was changed from paper-based to computer-based delivery, with 66 different test forms, allowing for an improved test design (Thomson, De Bortoli & Underwood 2017). For each student, 10 plausible values have been calculated for each of the three domains (reading,

mathematics and science) and for the financial literacy and collaborative problem-solving domains. Given that scientific literacy was the major domain in 2015, the following three science scales (and eight subscales) were also reported:

- knowledge scales (included subscales: content; and procedural and epistemic)
- competency scales (included subscales: explain phenomena scientifically; evaluate and design scientific inquiry; and interpret data and evidence scientifically)
- system scales (including subscales: physical; living; and earth and space).

How do I use plausible values?

There are 10 plausible values for each achievement domain. Unbiased estimates of achievement can only be obtained if plausible values are incorporated appropriately. The following are some key points:

- Averaging plausible values over individuals will lead to biased estimates and incorrect standard errors.
- The analysis should be repeated for each plausible value (10 times), and any subsequent estimate (for example, coefficients and/or standard errors) combined in an appropriate way to obtain population estimates.
- Plausible values are correlated within a domain and, as such, an analysis may be undertaken using only a single plausible value, but with an awareness that standard errors may be incorrect.

Users are reminded that plausible values are not equivalent to the achievement scores in the LSAY Y95 and Y98 cohorts, nor are they equivalent to an individual's raw test scores.

Further information about using plausible values is available from the *PISA 2015 technical report*: <http://www.oecd.org/pisa/data/2015-technical-report/>.

Variable naming conventions

PISA variables

The naming conventions for different types of PISA variables are summarised in Table 2. These conventions centre on the PISA 2015 questionnaire instruments, which include:

- the student questionnaire (ST)
- the information and communications technology questionnaire (ICT)
- the education career questionnaire (EC).

The student questionnaire also measured students' attitudes towards science.

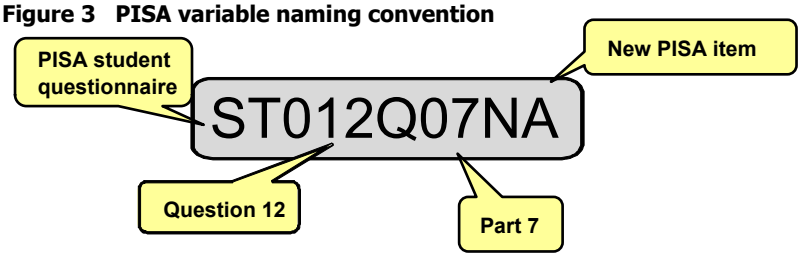
Most PISA variables use the following convention: questionnaire component, question number, question part (where applicable), and end with the two characters 'TA' or 'NA' or 'NB'. The questionnaire items ending in 'NA' or 'NB' refer to items that were new in 2015, and questionnaire items ending in 'TA' refer to trend items.

For example:

- ST006Q01TA is question number 6 (part 1) from the student questionnaire and is a trend item.
- ST012Q07NA is question number 12 (part 7) from the student questionnaire and is a new item.

- IC001Q01TA is question number 1 from the information and communications technology questionnaire and is a trend item.

Figure 3 PISA variable naming convention



Country-specific items are denoted by the character 'C' (rather than the character 'Q') and are suffixed with the characters '_AUS'. For example:

- ST007C01TA_AUS is question number 7 from the student questionnaire and is a country-specific trend item.

Plausible values and replicate weights

Plausible values are used to report student achievement in PISA. There are 10 plausible values for each of the domains and subscales[4] and the PISA student achievement variables reference this information as part of the variable name. For example, the variable:

- PV1MATH is the first plausible value in the maths domain.
- PV4SCIE is the fourth plausible value in the science domain.
- PV3SCEP is the third plausible value in the science subscale: competency — explain phenomena scientifically.
- PV10SKCO is the tenth plausible value in the science subscale: knowledge — content.

Replicate weights have been used to estimate sampling variances for population estimates derived from a complex sample design. The weights are named chronologically, from W_FSTURWT1 to W_FSTURWT80. The variable W_FSTUWT is the final student weight.

Detailed information about plausible values and replicate weights is available from the PISA technical report located at: <http://www.oecd.org/pisa/data/2015-technical-report/>.

Simple and scale indices

Two types of indices are provided in the PISA data file: simple indices and scale indices. Simple indices are constructed by arithmetically transforming or recoding one or more items, for example, age. Scale indices combine several of the answers provided by the students or principals to build a broader, not directly observable, concept. For example, CULTPOSS is a student-level scale index derived from cultural possessions such as classic literature, books of poetry and works of art.

Simple and scale indices appear towards the end of the PISA (wave 1) data and tend to be descriptive rather than bearing a variable naming convention.

4 Students were assessed in three domains: mathematical literacy, reading literacy and scientific literacy. In PISA 2015, scientific literacy was the major domain, and three science scales (competency, knowledge and system) with eight subscales were reported. Students also undertook assessments in collaborative problem solving and financial literacy.

Table 2 Summary of PISA variable naming conventions

PISA variable	Examples of PISA variable names	Description
Standard variables	ST007Q01TA IC001Q03TA EC012Q02NA	The first two characters indicate the questionnaire instrument. The PISA questionnaire instruments include: the student questionnaire (ST), information and communications technology questionnaire (ICT) and educational career questionnaire (EC). The following three digits indicate the question number (e.g. ST007Q01TA is question 7 from the student questionnaire). The final three characters are the question part or sub-section. So EC012Q02NA is part 2 of question 12 from the educational career questionnaire.
Country specific items	ST007C01TA_AUS	Variables with a 'C' at the fifth position and ending with the characters '_AUS' indicate the question is a country-specific item.
Student achievement/ plausible values	PV1SCIE PV4READ PV2SCEP	The first two characters 'PV' indicate that the variable is a plausible value. The next character indicates whether it is the first plausible value up to the tenth plausible value. The next four characters indicate the domain or subscale. • PV1SCIE indicates that the variable is the first plausible value from the science domain. • PV4READ indicates that the variable is the fourth plausible value from the reading domain. • PV9SCID points to the second plausible value in the science subscale 'Competency - Interpret Data and Evidence Scientifically'. *For further information on plausible values, see section,* 'The LSAY data: plausible values'.
PISA weights	W_FSTUR1 W_FSTUR80 W_FSTUWT	Replicate weights are identified using the characters 'W_FSTUR', followed by a chronological number. W_FSTUWT is the final student weight. *For further information on PISA weights, see the* PISA 2015 technical report *available at:* <http://www.oecd.org/pisa/data/2015-technical-report/>.
Indices	AGE HISCED CULTPOSS	Student and school-level simple and scaled indices tend to be descriptive rather than adopting a naming convention.

LSAY variables

Standard LSAY variables

Most LSAY variable names are constructed using four pieces of information: the questionnaire instrument, the survey wave, the questionnaire section and the question number.

- The first character of the variable name is assigned the character 'L' to differentiate the LSAY survey instrument from the PISA survey instrument.
- From wave 2 (when the LSAY survey instrument is first used) a wave identifier is used. The second character of the variable name is allocated a B, the third survey wave is allocated a C, etc.
- The third and fourth characters of the variable name are used to identify the section of the questionnaire.
- The fifth, sixth and seventh characters of the variable name are used to identify the question number.[5]
- The final character is used to identify a question subsection (where applicable).

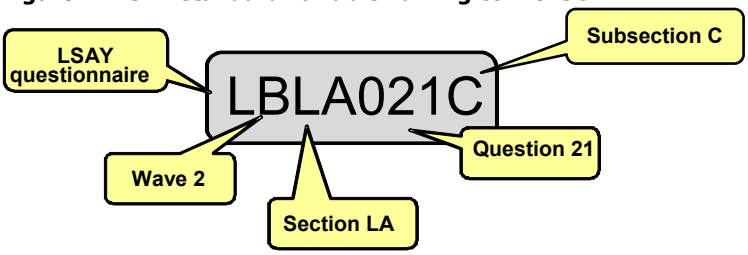
For example, the variable LBLA021C refers to:

- the LSAY survey instrument, denoted by the first character 'L'

5 `For some variables, the question number is collapsed into two rather than three characters (i.e. leading zeros are removed) to allow for additional characters at the end of the variable name to differentiate between data items used multiple times (for cases where sequencing loops are used in the questionnaire).

- wave 2, denoted by the second character 'B'
- section LA (last activity), denoted by the third and fourth characters 'LA'
- question 21, denoted by the fifth, sixth and seventh characters '021'
- question subsection C, denoted by the final character 'C'.

Figure 4 LSAY standard variable naming convention



LSAY non-standard variables

There is a series of other variables that do not adopt the standard variable naming convention. These variables are summarised in the following table.

Table 3 Summary of LSAY non-standard variable naming conventions

Non-standard variable	Examples of non-standard variable names	Description
Demographics	INDIG	Some demographic variables, such as Indigenous status, tend to be descriptive rather than carrying a naming convention.
School characteristics	STATE SECTOR	School characteristics, such as state of the school and school sector, tend to be descriptive rather than carrying a naming convention.
Derived variables	XLFS2015 XCEL2016	Derived variables are constructed across all waves to summarise key information such as labour force status and current education level. *For further information about derived variables see the section,* 'Derived variables'.
IN flag	IN2015 IN2016	IN flags are created for each survey wave to indicate whether a respondent participated in the survey in that year. If the value of the IN flag is equal to 1, the respondent participated in the survey in that year. IN flag variables are denoted by the two characters 'IN', followed by four digits for the survey year.
Interview dates	LBWID LBWIM LBWIY INTDAT15 INTSAS16	Day of interview, month of interview, and year of interview are collected each survey year and consolidated into an interview date variable. Interview date variables use the same variable-naming convention for the first two characters, followed by the two characters 'WI', and then 'D' for day of interview, 'M' for month of interview, or 'Y' for year of interview. The INTDAT and INTSAS variables are the interview date variables (in both character and SAS® date format respectively), followed by two digits for the survey year.
Postcode	PC2016	Respondents' home postcodes are indicated by the first two characters 'PC' followed by the year of interview.

Non-standard variable	Examples of non-standard variable names	Description
Sample and derived items	LBWSAM01 LBWSAM02	Sample and derived items take information from previous years to sequence respondents through appropriate sections of the questionnaire. For example, to ensure the correct list of subjects are displayed for a respondent, the variable LBWSAM02 uses the respondent's state or territory from their PISA data to confirm whether students went to school in that state. Sample items are denoted by: ▪ the first character 'L' (to indicate the LSAY survey instrument was used) ▪ followed by the wave identifier (e.g. 'B') ▪ followed by the character 'W' ▪ followed by the characters 'SAM' ▪ followed by two digits denoting the sample item.
Weights	WT16GEN ACH16WT WT2016 WT16GENP ACH16WTP WT2016P	Weight variables are denoted by the two characters 'WT', either at the beginning or end of the variable name. Two sets of weight variables are produced: the first adds to the sample sizes in each wave, and the second (denoted by 'P' at the end of the variable name) adds to the population size at each wave. *For further information about weights see section,* 'Weights' in the chapter 'Sample and survey design'.

Variables from the top-up recruitment questionnaire

To assist in using and identifying common variables collected as part of the top-up activity at wave 3 (2017) when compared with earlier LSAY waves, variable names have been assigned using one of the following three approaches:

1 **Assigning the same variable name** in cases where the data item collected has the same properties as data items from previous waves (including PISA) and the time period of collection is unlikely to affect that data item. For example, indigenous status was collected as part of both PISA and the recruitment questionnaire, so has been allocated the variable name 'INDIG' as per the PISA data item.

2 **Assigning the same variable number annotated with an '_R'** in cases where the data item collected has the same properties as data items from previous waves (including PISA) but the time period of collection is likely to affect that data item. For example, the question "In the year immediately after you leave school, what do you plan to do?" was asked at wave 2 for the main sample (variable LBCA002). This item was also asked as part of the recruitment questionnaire at wave 3, so is given the variable name LBCA002_R to indicate the recruitment question was also asked as part of the main survey but at a different point in time, i.e. wave 3, 2017.

3 **Assigning a new variable name that corresponds to the recruitment questionnaire** for items that have different properties to data items from previous waves. For example, the data item 'In what country were you born', while asked as part of PISA (variables ST019AQ01T and COBSTD_MIN), has a different coding framework when compared to the coded values for the same data item collected as part of the recruitment questionnaire (variable DEM05A). As such the variable name retains the name that corresponds to the recruitment questionnaire (DEM05A).

The variable listing and metadata workbook and/or online data dictionary can be used to help identify common variables when comparing the recruitment and main survey waves.

Derived variables

Derived variables provide useful indicators for analysis and have been developed to simplify use of the LSAY data. The derived variables focus on the areas of educational attainment, employment, measures of engagement in study and work, and social indicators. Derived variables are created to simplify a complex

combination of data items or instances where information from previous waves is required to determine a certain status in a current wave (for example, highest qualification level).

Derived variables contain data gathered from across successive waves and are updated each year. For respondents who missed an interview but rejoined the survey in a subsequent wave, information from the previous survey wave is used to determine the derivation for the current wave.

Table 4 summarises the series of derived variables available on the Y15 data file. Derived variables are denoted by the character X, followed by three characters uniquely identifying the derived variable, followed by four digits for the survey year.

Detailed technical documentation outlining how the variables are derived, as well as their properties, can be accessed at: <https://www.lsay.edu.au/publications/search-for-lsay-publications/lsay-2015-cohort-derived-variables>.

Table 4 Derived variables

Indicators	Derived variable	Variable name
Education	Current school level	XCSL*YYYY*
	Current qualification level	XCEL*YYYY*
	Highest school level completed	XHSL*YYYY*
	Highest qualification level completed	XHEL*YYYY*
	Study status in VET	XVET*YYYY*
	Study status in bachelor degree or higher	XBAC*YYYY*
	Full-time or part-time study status	XFTS*YYYY*
	Completed Year 12 or certificate II or higher	X122*YYYY*
	Completed Year 12 or certificate III or higher	X123*YYYY*
Employment	Labour force status	XLFS*YYYY*
	Full-time or part-time employment status	XFTP*YYYY*
	Permanent or casual employment	XEMP*YYYY*
	Status in apprenticeship/traineeship	XATR*YYYY*
	Job mobility during last year	XMOB*YYYY*
	Occupation (1-digit ANZSCO first edition)	XOCC*YYYY*
	Average weekly pay	XWKPY*YYY*
	Average hourly pay	XHRP*YYYY*
	Average weekly hours worked	XHRS*YYYY*
	Any spell of unemployment during the year	XUNE*YYYY*
	In full-time employment or full-time education	XFTE*YYYY*
Social	Marital status	XMAR*YYYY*
	Living with parent(s)	XATH*YYYY*
	Living in own home	XOWN*YYYY*
	Number of dependent children	XCHI*YYYY*

Classifications and code frames

A number of variables contained in the LSAY data files are coded using standard classifications or detailed code frames. The information for these variables is usually collected using open-ended questions, and verbatim responses are recorded. These responses are then coded using standard classifications or code frames.

The details of these classifications are not always provided in the documentation because they can be lengthy and summarised in a variety of ways. Table 5A provides a summary of the classifications and code frames used for each survey wave.

Standard classifications are also used to code several variables available as part of the linked datasets. These are summarised at table 5B.

Table 5A Summary of classifications and code frames used in the LSAY Y15 data file

Wave/year	Education	Occupation	Industry[1]	Institution	Subject
1/2015	ISCED 97	ISCO 08	Not applicable	Not applicable	Not applicable
2/2016	ASCED	ANZSCO First Edition	ANZSIC 2006	Institution code frame	Subject code frame
3/2017	ASCED	ANZSCO First Edition	ANZSIC 2006	Institution code frame	Subject code frame
4/2018	ASCED	ANZSCO First Edition	ANZSIC 2006	Institution code frame	Subject code frame
5/2019	ASCED	ANZSCO First Edition	ANZSIC 2006	Institution code frame	Not applicable
6/2020	ASCED	ANZSCO First Edition	ANZSIC 2006	Institution code frame	Not applicable
7/2021	ASCED	ANZSCO First Edition	ANZSIC 2006	Institution code frame	Not applicable
8/2022	ASCED	ANZSCO 2021 revision	ANZSIC 2006	Institution code frame	Not applicable
9/2023	ASCED	ANZSCO 2021 revision	ANZSIC 2006	Institution code frame	Not applicable
10/2024	ASCED	ANZSCO 2021 revision	ANZSIC 2006	Institution code frame	Not applicable
11/2025	ASCED	ANZSCO 2021 revision	ANZSIC 2006	Institution code frame	Not applicable

Notes: ISCED = International Standard Classification of Education; ASCED = Australian Standard Classification of Education; ISCO = International Standard Classification of Occupations; ANZSCO = Australian and New Zealand Standard Classification of Occupations; ANZSIC = Australian and New Zealand Standard Industrial Classification.

1 A special codeframe has been developed to capture additional industry codes and reserved values not included as part of the ANZSIC. See section 'Industry' below for more information.

Table 5B Summary of classifications and code frames used in the linked LSAY Y15 data files

Linked dataset	Education	Occupation	Country	Language	Geolocation
National VET Provider Collection	ASCED National Register of VET	ANZSCO	SACC	ASCL	ASGS, SEIFA
Higher Education Statistics Collection	ASCED		SACC	ASCL	

Notes: ASCED = Australian Standard Classification of Education; ANZSCO = Australian and New Zealand Standard Classification of Occupations; SACC = Standard Australian Classification of Countries; ASCL = Australian Standard Classification of Languages; ASGS = Australian Statistical Geography Standard; SEIFA = Socio-Economic Indexes for Areas

Education

As part of PISA, the International Standard Classification of Education (ISCED) 1997 is used to code parental education levels and expected student educational levels in the first wave of the 2015 cohort.

The ISCED has the following categories:

- ISCED 1 (primary education)
- ISCED 2 (lower secondary, that is, up to Year 10)
- ISCED 3B or 3C (vocational/pre-vocational upper secondary, for example, Year 11 with certificate III)
- ISCED 3A (upper secondary, for example, Year 12)
- ISCED 4 (non-tertiary post-secondary, for example, certificate IV)
- ISCED 5B (vocational tertiary, for example, diploma)
- ISCED 5A or 6 (theoretically oriented tertiary and postgraduate, for example, bachelor degree, postgraduate degree).

Further information about ISCED is available at: <http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-1997-en_0.pdf>.

Area of study variables are coded to the 6-digit Australian Standard Classification of Education (ASCED 2001; ABS 2001) from wave 2 (2016). The ASCED is also used to code variables for the following linked data sources and items:

- the Higher Education Statistics Collection uses ASCED to code the data items: Field of education (E461); Supplementary field of education (E462); and Discipline group (E464)
- the National VET Provider Collection uses ASCED to code data items: Program field of education and Program level of education.

School subjects

Subject code frames were created for the Longitudinal Study of Australian Children (LSAC) by the Australian Institute of Family Studies (AIFS) in collaboration with the ABS and Australian Government Department of Social Services (DSS). The code frame brings together secondary school subjects from school systems across all states and territories.

Secondary school subjects are coded to 10 digits from wave 2 (2016) and include information about: the subject area; and the state/territory and the year level of the subject being delivered.

The subject coder can be downloaded from the LSAC website: <https://growingupinaustralia.gov.au/data-and-documentation/school-subject-coder>.

VET programs

The National VET Provider Collection uses the National Register of VET to code VET programs. Further information about the register is available from <https://training.gov.au>.

Occupation

Occupation is coded to 4-digit International Standard Classification of Occupations (ISCO 08) in the first wave of the 2015 cohort, as part of PISA. Further information about ISCO is available at: <http://www.ilo.org/public/english/bureau/stat/isco/isco08/index.htm>.

Occupational data are coded to the 4-digit Australian and New Zealand Standard Classification of Occupations (ANZSCO 2006; ABS 2006a) from wave 2 (2016).

The National VET Provider Collection also uses ANZSCO to code the intended occupational outcome of the program enrolled in.

Industry

Industry variables are coded to the Australian and New Zealand Standard Industrial Classification (ANZSIC) 2006 from wave 2 (2016). Depending on the level of detail provided by respondents about their employer's business, industries can be coded to the ANZSIC in different ways. There are also differences with the way the industry codes have been stored in the LSAY data files when compared with ANZSIC: the LSAY dataset stores the codes using numeric variables, while ANZSIC uses a combination of 1-, 2- and 3-character codes. To address the discrepancies that arise because of the different data storage methods used, a separate codeframe has been created.

For a complete list of the industry codes used in the LSAY datasets, users are encouraged to view the code frame available at: <https://www.lsay.edu.au/publications/search-for-lsay-publications/industry-code-frame>.

Institution

Non-standard institution code frames have been developed specifically for LSAY for use in coding education institutions. Institution variables are coded to 6-digit from wave 2 (2016) and include information about the campus attended.

The institution code frame can be accessed at: <https://www.lsay.edu.au/publications/search-for-lsay-publications/institution-code-frame>.

Country and language

The Higher Education Statistics Collection uses the ABS Standard Australian Classification of Countries (SACC) and Australian Standard Classification of Languages (ASCL) to code data items Country of birth (E346) and Language spoken at home (E348).

The National VET Provider Collection uses the SACC and ASCL to code data items Country of birth and Language ID.

Geolocation

The National VET Provider Collection uses the Australian Statistical Geography Standard (ASGS) to code client and training delivery location information. Socio-Economic Indexes for Areas (SEIFA) are also available to ascertain clients' relative socio-economic disadvantage.

Data linkage

Data linkage brings together a range of data sources to enhance the quality and breadth of information from a single data source by providing more detailed, accurate and objective information about educational participation and attainment. It can also increase the richness and depth of information by linking to data that might be outside the scope of LSAY.

Linked datasets

LSAY records for the Y15 cohort have now been linked to the following data sources:

- ACARA *My School* data
- National Assessment Program — Literacy and Numeracy (NAPLAN)
- Senior secondary subjects and results
- National VET Provider Collection
- Higher Education Statistics Collection.

Details about the variables contained in the linked datasets are available from the 'Data linkage' worksheet located in the LSAY variable listing and metadata <https://www.lsay.edu.au/publications/search-for-lsay-publications/2621>.

Methodology

LSAY respondents from the Y15 cohort were asked for their consent to link their NAPLAN results to their LSAY records for the first time as part of their wave 2 (2016) survey. Respondents were also asked for their consent to link their senior secondary subjects and results. Respondents who did not participate at wave 2 (2016) are asked for their consent in subsequent survey waves.

From wave 3 (2017), respondents who had engaged in some form of VET activity were asked for their consent to link their VET records from the National VET Provider Collection to their LSAY records. From wave 4 (2018), consent was sought from respondents to link data from the Higher Education Statistics Collection to LSAY.

Separation principle

To protect the identities of respondents, schools and organisations, the data linkage separation principle has been implemented. The separation principle is a mechanism that ensures no one working with the data can view both the linking (identifying) information (such as name, address, date of birth) in combination with the analysis data (such as NAPLAN scores) in a linked dataset.

Under the separation principle, individuals only have access to the information needed to perform their role. Those involved in linking the datasets only see the identifying information needed to create the links between different datasets, while those involved in analysing the integrated data only have access to de-identified data specific to the project requirements.

ACARA *My School* data

ACARA *My School* data is reported at the school level, so data linkage is undertaken at the school rather than student level. As such, consent from respondents is not required to enable school-level information to be linked.

LSAY ACARA *My School* data is compiled using multiple datasets provided by the Australian Curriculum, Assessment and Reporting Authority (ACARA). The name of the schools respondents were attending/attended is collected as part of respondents' LSAY interview and mapped to unique ACARA school identifiers. These ACARA school identifiers are then used to extract the associated schools from the ACARA *My School* datasets.

Datasets

There are 12 datasets available from the LSAY ACARA *My School* data, prepared in long format with multiple observations per school, one for each calendar year reported. The linked datasets are summarised in table 6.

Table 6 LSAY ACARA *My School* datasets

Dataset	File name	Calendar years reported
School profile	lsayacara_schprofile_v1	2008-2017
NAPLAN results	lsayacara_results_v1	2008-2017
NAPLAN results similar schools	lsayacara_results_simsc_v1	2008-2017
Student gain	lsayacara_studgain_v1	2008-2017
Student gain similar schools	lsayacara_gain_simsch_v1	2008-2017
Student gain same starting score	lsayacara_gain_ss_v1	2008-2017
Student attendance	lsayacara_attendance_v1	2014-2017
VET in schools	lsayacara_vet_v1	2008-2016
VET in schools – School-based apprenticeships and traineeships	lsayacara_vet_sbat_v1	2008-2016
Senior secondary outcomes	lsayacara_secsch_v1	2008-2016
Finance	lsayacara_finance_v1	2009-2016
Enrolments by grade	lsayacara_enrolments_v1	2008-2017

Study specific school identifier

To protect the identities of the schools, a study specific school identifier has been used in place of the ACARA school identifier on both the linked ACARA datasets and the main LSAY datafile.

Rolled reporting

ACARA is unable to provide data at campus level for schools. For this reason, all schools with multiple campuses are subject to 'rolled reporting' and ACARA data is available at the main campus level only. This information is flagged using the variable 'RolledReptDesc' available on the linked datasets.

To maximise the number of linked ACARA records, a substitute school identifier is provided in addition to the original school identifier in cases where only data for the main (not individual) campus level is available. As a result, a study specific school identifier is provided for both the campus (denoted by an _C) and main campus (denoted by an _M).

There were 4405 unique ACARA school identifiers available for linkage from the LSAY data file. Of these:

- 178 school identifiers were given a substitute school identifier to enable linkage of school information (for the main campus) to the ACARA data set

- 59 schools could not be linked to the ACARA data set as these schools were deemed inactive because they have merged or closed.

As a result, the number of unique school identifiers (reported at the main campus level) is reduced to 4168. However, an additional 34 new school identifiers were added to the list of main campus identifiers (which weren't available on the original list of ACARA identifiers in the LSAY dataset) as a result of rolled reporting (i.e. to enable ACARA data to be linked for schools requiring a substitute main campus identifier). As a result, there are 4202 unique school identifiers on the linked ACARA data files available for linkage to the main LSAY data file using the study specific school identifier at the main campus level.

Missing data

School information may not be available in every dataset, resulting in varying levels of missing data. This is because some variables may not be relevant for all schools, for example, primary schools will not be reported on in 'Senior secondary outcomes' dataset.

Additional resources

More information on the ACARA *My School* data is available from the following links:

- Technical and statistical information: <https://www.myschool.edu.au/more-information/technical-and-statistical-information>
- Glossary: <https://www.myschool.edu.au/glossary>
- Making fair comparison using ICSEA: <https://www.myschool.edu.au/more-information/information-for-parents/making-a-fair-comparison>.

National Assessment Program – Literacy and Numeracy (NAPLAN)

LSAY respondents from the Y15 cohort were asked for their consent to link their NAPLAN results from Years 3, 5, 7 and 9 to their LSAY records for the first time as part of their wave 2 (2016) survey. Respondents who did not participate at wave 2 (2016) are asked for their consent in subsequent survey waves.

The linkage is undertaken for consenting respondents using the contact details held by Wallis, the LSAY fieldwork contractor.

Linking NAPLAN data to LSAY

To protect the identity of respondents and satisfy the requirements of the separation principle, the following steps were undertaken in the LSAY-NAPLAN linkage process:

1. NCVER creates a unique linking identifier which is mapped to the LSAY identifier and provided to Wallis, the LSAY fieldwork contractor.
2. Wallis sends respondents' contact details of consenting respondents along with the unique linking identifier to the data integration authority, the Australian Institute of Family Studies (AIFS), who undertook the linkage on NCVER's behalf.
3. AIFS sends the contact details to the relevant state/territory data custodians along with the unique linking identifier. The state/territory data custodian uses the contact details to identify the administrative records it holds for each consenting respondent. The state/territory data custodians then remove all contact details but provide NAPLAN scores along with the unique linkage identifier to AIFS. The state/territory data custodian does not keep any personal information passed onto them, and all personal information is destroyed once the data linkages are complete (i.e. within three years of receipt of this information).

4 AIFS collates and validates the information provided by the states and territories and provides the linked NAPLAN data to NCVER along with the unique linking identifier.

5 NCVER adds the NAPLAN scores to the appropriate LSAY records using the unique linkage identifier. NCVER does not have access to any contact detail information and the unique linkage identifier is removed from the datasets.

Consent and linkage rates

Table 7 shows the number of respondents providing their consent from waves 2-5 (2016-2019) and table 8 shows the number of successfully linked NAPLAN records.

Table 7 LSAY NAPLAN consent rates

	Wave 2 (2016)		Wave 3 (2017)		Wave 4 (2018)		Wave 5 (2019)		Total	
	n	%	n	%	n	%	n	%	n	%
Provided consent	4004	85	476	80	728	77	79	66	5287	83
Did not provide consent	700	15	121	20	215	23	41	34	1077	17
Respondents asked for consent	4704		597		943		120		6364	

Note: Excludes those who provided their consent but were removed from the LSAY dataset because they were ineligible as part of the PISA guidelines.

Table 8 LSAY NAPLAN linkage rates

	Wave 2 (2016)		Wave 3 (2017)		Wave 4 (2018)		Wave 5 (2019)		Total	
	n	%	n	%	n	%	n	%	n	%
Provided consent	4004		476		728		79		5287	
Of these: NAPLAN records linked	3842	96	448	94	674	93	75	95	5039	95
Respondents asked for consent	4704		597		943		120		6364	
Of these: NAPLAN records linked	3842	82	448	75	674	71	75	63	5039	79

Note: Excludes those who provided their consent but were removed from the LSAY dataset because they were ineligible as part of the PISA guidelines.

NAPLAN data by academic year level

LSAY respondents were, on average, 15 years old when they undertook PISA and span multiple year levels. On the other hand, NAPLAN data are collected by academic year levels, which span multiple calendar years. As a result, LSAY data is comprised of young people of the same age but different year levels, while the NAPLAN data is comprised of young people of the same year level but different ages. This means the linked NAPLAN results for any one academic year (e.g. Year 3) will span multiple calendar years corresponding to different survey waves.

Results from earlier years proved harder to match across the jurisdictions, with the number of matches for each academic year level increasing with the recency of results. The results data received, categorised by academic year level, is summarised in table 9.

Table 9 LSAY NAPLAN linked data by academic year level

Academic year	Participants with a linked record (n)
Year 3	3850
Year 5	4657
Year 7	4766
Year 9	4865
Total (Years 3, 5, 7 or 9)	**5039**

Datasets

The LSAY NAPLAN dataset has been prepared in long format with multiple observations per participant, one for each assessment/academic year. The LSAY NAPLAN datasets are summarised in table 10.

There were 21 LSAY participants found to have repeated a year level. An additional dataset (lsaynaplan_repeats_v2) has been generated that includes only those who have repeated an academic year level.

Table 10 LSAY NAPLAN datasets

File name	Participants (n)	Variables (n)
lsaynaplan_v3	5039	33
lsaynaplan_repeats_v3	21	33

Customised weights

To account for the issue that only a subset of NAPLAN data is available through linkage, a custom set of weights for the linked LSAY-NAPLAN data has been created. Because respondents completed the NAPLAN assessment in different calendar years, the modal year (the calendar year in which the highest number of respondents completed their NAPLAN assessment for that year level) is used as the reference point for the weightings. For example, most respondents undertook their Year 9 NAPLAN assessment in 2014, their Year 7 NAPLAN assessment in 2012 etc, so these calendar years are used as the basis for creating the weights for each year level.

A unique set of weights for each NAPLAN academic year level (Years 3, 5, 7 and 9) are now available on the linked dataset. A few things to note include:

- the custom weights add to the original population from which the LSAY sample was drawn
- the modal calendar year of the NAPLAN assessment is used as the reference point for the weightings
- custom weights are created for: respondents with linked NAPLAN data; those who undertook the NAPLAN assessment in the modal calendar year; and those from the original LSAY sample drawn from PISA[6]. All other records are assigned zero values.

Further information about how these weights have been created and how they can be used is available from the technical paper 'Understanding and using the linked LSAY-NAPLAN data: issues and considerations' available at <https://lsay.edu.au/publications/search-for-lsay-publications/understanding-and-using-the-linked-lsay-naplan-data-issues-and-considerations>.

6 Those from the top-up group (see 'Sample and survey design') have not been assigned weights as information required to create the weights are not available for this group.

Notes about the data

In 2011, students were required to complete a persuasive writing task for the first time. This is a change from previous years (2010 and prior) when students were required to write a narrative or story. Due to this change in genre, post-2010 writing results should not be compared to previous years and the different writing tasks are identified using different variable names: narrative writing (2007-2010) and writing (2011-2016).

It is possible that NAPLAN data has been supplied for a student who was absent, exempt or withdrawn from their assessment. For these cases, no score, band or National Minimum Standards (NMS) values are available and a value of -1 has been assigned (for scores, bands, and NMS) to represent non-participation.

A value of -9 has been assigned where data are not available, in order to differentiate these unavailable values from true missing data.

Not all jurisdictions were able to consistently provide information about the respondent's school sector. In such instances a value of -9 has been assigned.

While all jurisdictions provided NAPLAN scores, some jurisdictional datasets did not include a flag to indicate participation, position on the national assessment scale (band) or achievement of the National Minimum Standards (NMS). NAPLAN scores were used to impute a participation flag in cases where this flag was missing, and missing bands and NMS values were imputed using the following approaches:

- Score equivalence tables for assessment years 2009-2017 were used to populate band and NMS values for 2009-2016.
- Score equivalence tables for 2007 and 2008 were requested but are not available from ACARA. As such, where available, missing band and NMS values were imputed using equivalent NAPLAN scores and their corresponding band and NMS values.

Some corrections were also made to bands and NMS values that did not correspond to the score equivalence tables.

Additional resources

More information about NAPLAN is available from the following links:

- NAPLAN website: <https://www.nap.edu.au/naplan>
- Guide on interpreting NAPLAN results: <https://www.nap.edu.au/results-and-reports/how-to-interpret>.
- Score equivalence tables: <https://www.nap.edu.au/results-and-reports/how-to-interpret/score-equivalence-tables>.

Senior secondary data

LSAY respondents from the Y15 cohort were asked for their consent to link their senior secondary subject information to their LSAY records for the first time as part of their wave 2 (2016) survey. Respondents who did not participate at wave 2 (2016) were asked for their consent in subsequent survey waves (up until wave 5 in 2019).

As with the NAPLAN linkage, the linkage is undertaken for consenting respondents using the contact details held by Wallis, the LSAY fieldwork contractor.

Linking senior secondary data to LSAY

To protect the identity of respondents and satisfy the requirements of the separation principle, the following steps were undertaken in the LSAY-senior secondary linkage process:

1 NCVER creates a unique linking identifier which is mapped to the LSAY identifier and provided to Wallis, the LSAY fieldwork contractor.

2 Wallis sends respondents' contact details of consenting respondents along with the unique linking identifier to the data integration authority, the Australian Institute of Family Studies (AIFS), who undertook the linkage on NCVER's behalf.

3 AIFS sends the contact details to the relevant state/territory data custodians along with the unique linking identifier. The state/territory data custodian uses the contact details to identify the administrative records it holds for each consenting respondent. The state/territory data custodians then remove all contact details but provide senior secondary subject information along with the unique linkage identifier to AIFS. The state/territory data custodian does not keep any personal information passed onto them, and all personal information is destroyed once the data linkages are complete (i.e. within three years of receipt of this information).

4 AIFS collates and validates the information provided by the states and territories and provides the linked senior secondary data to NCVER along with the unique linking identifier.

5 NCVER adds the senior secondary data to the appropriate LSAY records using the unique linkage identifier. NCVER does not have access to any contact detail information and the unique linkage identifier is removed from the datasets.

Consent and linkage rates

Table 11 shows the number of respondents providing their consent from waves 2-5 (2016-2019) and table 12 shows the number of successfully linked NAPLAN records.

Table 11 LSAY senior secondary consent rates

	Wave 2 (2016)		Wave 3 (2017)		Wave 4 (2018)		Wave 5 (2019)		Total	
	n	%	n	%	n	%	n	%	n	%
Provided consent	3806	81	441	74	708	75	78	65	5033	79
Did not provide consent	898	19	156	26	235	25	42	35	1331	21
Respondents asked for consent	**4704**		**597**		**943**		**120**		**6364**	

Note: Excludes those who provided their consent but were removed from the LSAY dataset because they were ineligible as part of the PISA guidelines.

Table 12 LSAY senior secondary linkage rates

	Wave 2 (2016)		Wave 3 (2017)		Wave 4 (2018)		Wave 5 (2019)		Total	
	n	%	n	%	n	%	n	%	n	%
Provided consent	3806		441		708		78		5033	
Of these: senior secondary records linked	**3037**	**80**	**351**	**80**	**638**	**90**	**72**	**92**	**4098**	**81**
Respondents asked for consent	4704		597		943		120		6364	
Of these: senior secondary records linked	**3037**	**65**	**351**	**59**	**638**	**68**	**72**	**60**	**4098**	**64**

Note: Excludes those who provided their consent but were removed from the LSAY dataset because they were ineligible as part of the PISA guidelines.

Datasets

The LSAY senior secondary dataset has been prepared in long format with multiple observations per participant, one for each subject and academic year. The LSAY senior secondary dataset is summarised in table 13.

Table 13 LSAY senior secondary dataset

File name	Participants (n)	Variables (n)
lsaysnrsec_v1	4080	11

Note: Excludes 18 linked participant records where only year 9 or year 10 subject information was provided

Notes about the data

The following variables relating to senior secondary data were requested from each jurisdiction: Calendar year, Year level, Subject name, Subject area, Subject level (e.g., unit 1, 2, 3 or 4), Subject result, Year 12 completion status, Senior secondary certificate awarded and Australian Tertiary Admission Rank (ATAR).

Not all jurisdictions were able to provide the requested information. In particular, subject level, subject result, Year 12 completion status and ATAR proved to be particularly problematic. In addition, there were differences in the way this information is reported across the jurisdictions, making harmonisation of the variables unfeasible. For this reason, these items have been omitted from the linked data file.

There may be students who left school before entering senior secondary year levels (i.e. years 11 and 12). No matching records could be expected for these students.

There are a small number of records where students have moved interstate and as such have senior secondary records spanning more than one jurisdiction.

It is not possible to systematically determine whether the subjects reported are VET subjects. Some jurisdictions included 'VET' or 'Certificate' in the subject name for some records. However, it is unlikely this flag provides an exhaustive list of VET subjects undertaken.

Scope

Data was restricted to subject information for years 11 and 12 only. Three jurisdictions provided subject information for years 9 and 10 which were omitted from the final linked dataset. Note that the year level for the subject information was not always known; these records have been coded to a value of '-1 Unknown' and may relate to information outside of the scope of the linked dataset (i.e. subject information from years 9 and 10).

Data cleaning

Some data cleaning to remove inconsistencies across subject names (for example, capitalisation and formatting) has been undertaken in an effort to harmonise the subject names.

Missing data and derivations

Not all jurisdictions were able to provide information for all fields. The following list provides a summary of the exceptions with regards to the data submitted for each jurisdiction for each field along with the fields that were derived using the data submitted.

- **Subject information** was only provided for Queensland and ACT for year 12; year 11 subject information was not submitted for these two jurisdictions[7].

7 Queensland did provide a very small number of year 11 subject records. However, given this information was incomplete, these records have been removed from the final dataset.

- **Subject calendar year** (i.e. the calendar year to which the subject information relates) was not provided by Queensland, so this field was derived using the Y*ear 12 calendar year* in combination with the *Subject academic year* (i.e. year level) field [8].

  This field was derived for ACT using the *Year 12 calendar year* where available.

- **Subject academic year** (i.e. the year level to which the subject information relates) was not provided by Victoria, South Australia, the Northern Territory or the ACT, so this field was derived using the *Year 12 calendar year* in combination with the *Subject calendar year* [9].

- **Year 12 calendar year** (i.e. the calendar year to which the year 12 senior secondary certificate information relates) was not available for Western Australia[10].

  This field was only provided for some ACT subject records; as such, Year 12 calendar year is missing for a large proportion of ACT records.

- **Year 12 senior secondary certificate name** was not available for Western Australia[11].

  Victoria was the only jurisdiction that provided information about the International Baccalaureate program. This indicates there may be some underreporting of a senior secondary certificate completed given only the state-based certificate is reported for all jurisdictions (with the exception of Victoria).

  Victoria also provided information regarding students who had completed the Victorian Certificate of Applied Learning (VCAL), however the level of detail varied across the records provided. The VCAL is accredited and issued at three award levels: VCAL Foundation; VCAL Intermediate and VCAL Senior. The data provided classified respondents as having completed: VCAL Intermediate and VCAL Senior, or VCAL (level undefined).

- **Year 12 senior secondary certificate awarded**[12] was not available for Western Australia[13].

  In Victoria, those who had completed a VCAL certificate (and the VCAL level was undefined) were classified by that jurisdiction as having been awarded a Year 12 senior secondary certificate. For cases where the VCAL certificate level has been reported, those who had completed the 'VCAL Intermediate' program have been classified as '3 Other'.

  See also *Year 12 senior secondary certificate name* for further information related to this field.

Recoding subject information

Jurisdictions report subject information using learning areas particular to their state or territory. The subject name and subject area for each subject have been reported using the same nomenclature provided by the jurisdictions[14].

---

8 This derivation may introduce some misclassification errors for students who repeated or skipped a year level.

9 This derivation may introduce some misclassification errors for students who repeated or skipped a year level.

10 WA was unable to provide information about senior secondary certificates completed. As such, all year 12 fields (Year 12 calendar year, Year 12 senior secondary certificate awarded and Year 12 senior secondary certificate name) have all been assigned the value '-9 Data unavailable'.

11 WA was unable to provide information about senior secondary certificates completed. As such, all year 12 fields (Year 12 calendar year, Year 12 senior secondary certificate awarded and Year 12 senior secondary certificate name) have all been assigned the value '-9 Data unavailable'.

12 There are (four) records with more than one outcome for Year 12 senior secondary certificate completed. For three of these records, this is because this data has been reported by multiple jurisdictions for an individual student. For the remaining (one) record, this reflects the data as submitted by the jurisdiction and is likely to reflect a clerical error.

13 WA was unable to provide information about senior secondary certificates completed. As such, all year 12 fields (Year 12 calendar year, Year 12 senior secondary certificate awarded and Year 12 senior secondary certificate name) have all been assigned the value '-9 Data unavailable'.

14 NSW did not provide a subject area for each subject record. There were also instances across most jurisdictions where the subject area was missing or inconsistent when compared with the same subject name for other records. Nevertheless, for

To make it easier for users to work with the subject information across the jurisdictions, subjects have been recoded to the subject code frame created for the Longitudinal Study of Australian Children (LSAC).

The code frame brings together secondary school subjects from school systems across all states and territories and classifies them into 12 subject areas. Each subject area includes a number of 'sub-topics' within each area (for example, 'Arts' includes sub-topics such as 'Drama', 'Music, and 'Creative Arts'). Subjects recoded using the LSAC subject codeframe are denoted with a '_C' at the end of the variable name.

For further information about the codeframe, please refer to the section 'Classifications and code frames'.

Additional resources

For reference, the subject coder can be downloaded from the LSAC website:

- <https://growingupinaustralia.gov.au/data-and-documentation/school-subject-coder>.

National VET Provider Collection

From wave 3 (2017), respondents who had engaged in some form of VET activity were asked for their consent to link their data from the National VET Provider Collection to their LSAY records. The National VET Provider Collection collects data on VET activity delivered by Australian training providers to a nationally agreed standard. It provides information on the number of students (and full year training equivalents), participation rates, program and subject enrolments, program completions and training providers[15].

From 2016, data submitted to the National VET Provider Collection has also been used for individual student unique student identifier (USI) transcripts. This enables the USIs to be utilised for data linkage.

While the USI provides a simple and reliable identifier to undertake the LSAY-VET linkage, it is unlikely respondents will know or be able to conveniently access their USI during their telephone interview. For this reason, only respondents completing their interview online are asked to provide their USI where possible; otherwise contact details held by Wallis (the LSAY fieldwork contractor) are used to match to contact details available from the USI Registry to obtain a valid USI. The USI is then used to gather corresponding VET records from the National VET Provider Collection.

As a result:

- all respondents who interacted with the VET system were asked for their consent to use their USI to retrieve their VET records for linking to their LSAY records
- in cases where a USI is not available (i.e. for all respondents completing their interviews by telephone; and for online respondents unable to provide a valid USI) a subsequent question asks respondents for their consent to use their contact details to retrieve their USI.

Name, date of birth, email address, mobile, home phone number and/or mailing address is used to facilitate the matching exercise (for cases where the USI is not available).

The need to ask for two forms of consent (one for consent to use the USI to undertake the linkage; the second for consent to use contact details to undertake the linkage in cases where the USI is not available) resulted in reduced rates of consent (see table 14).

---

nearly all missing or inconsistent records, sufficient information was provided to allow the subject area to be coded or recoded using corresponding information available from the state and territory board of studies websites.

15 The scope of the linked LSAY-VET data corresponds to what is referred to as total VET activity (TVA) which represents data on nationally recognised training activity delivered by Australian registered training organisations (RTOs). These data are sourced from both the National VET Provider Collection and the VET in Schools Collection with duplicate activity removed. See the section 'Additional resources' for more information.

From 2021 (wave 7), respondents who had previously refused to consent to the linkage were re-asked for their consent. This explains the sudden increase in the number of consenting respondents at wave 7 (see table 14) and has improved the overall consent and linkage rates.

Linking National VET Provider Collection data to LSAY

To protect the identity of respondents and satisfy the requirements of the separation principle, the following steps were undertaken in the LSAY-VET linkage process:

1 NCVER creates a unique linking identifier which is mapped to the LSAY identifier and provided to Wallis, the LSAY fieldwork contractor.

2 Wallis provides USIs (for those completing their interviews online and able to provide a valid USI), contact details (for those completing their interviews by telephone or unable to provide a USI) and unique linking identifier of consenting respondents to the USI Office.

3 The USI Office matches and validates contact details and USIs provided by Wallis with those on the USI Registry and provides corresponding USIs along with the unique linking identifier to NCVER.

4 NCVER uses the validated USIs to link to National VET Provider Collection records and provides these records alongside the unique linking identifier to the LSAY team at NCVER. The USIs are removed from the linked dataset before provision to the LSAY team at NCVER.

5 The LSAY team at NCVER adds the National VET Provider Collection records to the corresponding LSAY records using the unique linkage identifier. NCVER does not have access to any contact detail information and the unique linkage number is removed from the datasets.

Consent and linkage rates

Table 14 shows the number of respondents providing their consent to the VET linkage from waves 3 to 10 and table 15 shows the total number of successfully linked VET records. Data for those who provided their consent at wave 11 will be updated as part of a final linked data release.

Table 14 LSAY VET consent rates

	Wave 3 (2017)		Wave 4 (2018)		Wave 5 (2019)		Wave 6 (2020)		Wave 7 (2021)		Wave 8 (2022)		Wave 9 (2023)		Wave 10 (2024)		Total	
	n	*%*	*n*	*%*	*n*	*%*	*n*	*%*	*n*	*%*	*n*	*%*	*n*	*%*	*n*	*%*	*n*	*%*
Provided consent	1553	79.9	557	73.4	160	57.3	101	78.9	268	59.8	101	44.5	49	26.1	23	59.0	2756	85.4
Of these: records available for linkage	1464	75.3	513	67.6	149	53.4	88	75.4	238	53.1	88	38.8	47	25.0	22	56.4	2564	79.5
Did not provide consent	391	20.1	202	26.6	119	42.7	126	21.1	180	40.2	126	55.5	139	73.9	16	41.0	470	14.6
Total respondents asked for consent	**1944**	**100**	**759**	**100**	**279**	**100**	**227**	**100**	**448**	**100**	**227**	**100**	**188**	**100**	**39**	**100**	**3226**	**100**

Notes

1 Excludes those who provided their consent but were removed from the LSAY dataset because they were ineligible as part of the PISA guidelines.

2 Respondents who previously refused their consent to the linkage were re-asked for their consent from wave 7 (2021) so the sum of the number of respondents asked for consent exceeds the total.

Table 15 LSAY VET linkage rates

	Linked records (n)	Total records (n)	Linkage rate (%)
Linked USIs to LSAY consenting records	2388	2564	93
Linked National VET Provider Collection records to USIs	2338	2388	98
Linked National VET Provider Collection records to LSAY consenting records	2338	2564	91
Linked National VET Provider Collection records to LSAY respondents asked for consent	2338	3226	72

Note: Excludes those who provided their consent but were removed from the LSAY dataset because they were ineligible as part of the PISA guidelines.

Datasets

The linked LSAY-VET datasets have been prepared in long format with multiple observations per student, one for each program and/or subject that they have completed or enrolled in.

Two datasets are available for analysis:

1 Total VET activity (TVA) program completions 2015-24: contains information on students and the program(s) which they have completed between year 2015-24.

2 Total VET activity (TVA) subject enrolments 2015-24: contains information on students and the subject(s) which they have enrolled between year 2015-24.

A student can be included in either one or both datasets. A summary of the number of students and number of variables for which linked VET is available are summarised in table 16.

Further details about the variables are available from the 'Data linkage' worksheet located in the LSAY variable listing and metadata <https://www.lsay.edu.au/publications/search-for-lsay-publications/2621>.

Table 16 LSAY VET linked datasets

Dataset	File name	Students (n)	Variables (n)
TVA program completions	lsayvet_c_v7	1800	47
TVA subject enrolments	lsayvet_s_v7	2336	60

Restricted variables

The following variables are available for analysis but access is restricted:

- Client residential address (suburb, postcode and statistical area level 4)
- Training organisation delivery location (statistical area level 4).

Users can apply to access these restricted variables when submitting their application to access the LSAY data. More information can be found in the section 'Accessing the LSAY data'.

Additional resources

More information regarding the National VET Provider Collection and total VET activity (TVA) is available from the following links:

- About the National VET Provider Collection: <https://www.ncver.edu.au/research-and-statistics/collections/students-and-courses-collection>
- Total VET students and courses: <https://www.ncver.edu.au/research-and-statistics/collections/students-and-courses-collection/total-vet-students-and-courses>
- AVETMISS data elements definitions: edition 2.3: <https://www.ncver.edu.au/rto-hub/statistical-standard-software/avetmiss-data-element-definitions-edition-2.3>

Higher Education Statistics Collection

From wave 4 (2018), respondents who had engaged in higher education were asked for their consent to link their data from the Higher Education Statistics Collection to their LSAY records. Name, date of birth, sex and mailing address are used to match higher education data to LSAY records for consenting respondents.

The Higher Education student data collection encompasses enrolments, equivalent full-time student load (unit of study data) and completions, and includes all higher education institutions that have been approved under the *Higher Education Support Act 2003*.

Linking Higher Education Statistics Collection data to LSAY

To protect the identity of respondents and satisfy the requirements of the separation principle, the following steps were undertaken in the LSAY-higher education linkage process:

1. NCVER creates a unique linking identifier which maps to the LSAY identifier. These two identifiers are provided to Wallis, the LSAY fieldwork contractor.
2. Wallis provides the unique linking identifier and the contact details of consenting respondents to the Australian Government Department of Education, Skills and Employment.
3. The Department matches the contact details provided by Wallis with those in the Higher Education student data collection. The corresponding Commonwealth Higher Education Student Support Number (CHESSN) is then used to extract higher education student data for the matching records. The CHESSN is removed from the linked dataset before provision to the LSAY team at NCVER.
4. The LSAY team at NCVER adds the higher education student data to the corresponding LSAY records using the unique linkage identifier. NCVER does not have access to any contact detail information and the unique linkage number is removed from the datasets.

Consent and linkage rates

Table 17 shows the number of respondents providing their consent to the higher education linkage from waves 4 to 9 and table 18 shows the total number of successfully linked higher education records.

From 2021 (wave 7), respondents who had previously refused to consent to the linkage were re-asked for their consent. This explains the sudden increase in the number of consenting respondents at wave 7 (see tables 17 and 18) and has improved the overall consent and linkage rates.

Table 17 LSAY higher education consent rates

	Wave 4 (2018)		Wave 5 (2019)		Wave 6 (2020)		Wave 7 (2021)		Wave 8 (2022)		Wave 9 (2023)		Total	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Provided consent	1792	73	409	72	128	75	229	49	166	34	87	24	2811	83
Did not provide consent	654	27	160	28	43	25	235	51	329	66	276	76	558	17
Respondents asked for consent	2446		569		171		464		495		363		3369	

Notes

1 Excludes those who provided their consent but were removed from the LSAY dataset because they were ineligible as part of the PISA guidelines.

2 Respondents who previously refused their consent to the linkage were re-asked for their consent from wave 7 (2021) so the sum of the figures by wave exceeds the total.

Table 18 LSAY higher education linkage rates

	Wave 4 (2018)		Wave 5 (2019)		Wave 6 (2020)		Wave 7 (2021)		Wave 8 (2022)		Wave 9 (2023)		Total	
	n	%	n	%	n	%							n	%
Provided consent	1792		409		128		229		166		87		2811	
Of these: higher education records linked	1729	96	394	96	118	92	217	95	151	91	84	97	2693	96
Respondents asked for consent	2446		569		171		464		495		363		3369	
Of these: higher education records linked	1729	71	394	69	118	69	217	47	151	31	84	23	2693	80

Notes

1 Excludes those who provided their consent but were removed from the LSAY dataset because they were ineligible as part of the PISA guidelines.

2 Respondents who previously refused their consent to the linkage were re-asked for their consent from wave 7 (2021) so the sum of the number of respondents asked for consent exceeds the total.

3 The number of successful linkages for earlier waves increased slightly when linking to the 2023 Higher Education Statistics Collection, reflecting updates made to the linked data supplied by the Department of Education.

Datasets

The LSAY higher education datasets have been prepared in long format with multiple observations per student. The following linked LSAY higher education datasets are available for analysis:

1 Student: student enrolments — student course enrolment information; includes demographic and other background information about the student

2 Course: course of study — course information; includes course information on all courses offered and is not limited to courses undertaken by LSAY higher education students

3 Unit of study: unit of study completions — courses and units of study undertaken and the completion status for the units of study

4 Completions: past course completions — includes demographic and other background information about the student

5 Scholarship: Commonwealth scholarships — Indigenous Commonwealth education and accommodation scholarships.

A high-level summary of the datasets is provided at table 19. The key analysis variables summarised at table 20 outline the variables that should be considered when summarising and analysing the higher education data.

A listing of the linked higher education data items is available from the 'Data linkage' worksheet located in the LSAY variable listing and metadata <https://www.lsay.edu.au/publications/search-for-lsay-publications/2621>.

Table 19 LSAY higher education linked datasets

Dataset	File name	Students (n)	Data items (n)
Student	lsayheims_student_v4	2693	22
Course	lsayheims_course_v4	NA	13
Unit of study	lsayheims_unit_of_study_v4	2693	16
Completions	lsayheims_completions_v4	1854	16
Scholarship	lsayheims_scholarship_v4	47	6

Notes

1 The 'Course' dataset contains course information for all courses, and is not limited to courses undertaken by LSAY respondents. Information about the courses undertaken for LSAY respondents can be found in the 'Unit of study' file.

2 Five data items were retired from the 'Student' dataset as part of the 2023 Higher Education Statistics Collection updates.

Table 20 LSAY higher education key analysis variables

Dataset	File name	LSAY ID	Reference year (E550)	Institution code (E306)	Course code (E307)	Major course (E331)	Primary course record indicator (E095)	Unit of study code (E354)
Student	lsayheims_student_v3	✓	✓	✓	✓	✓		
Course	lsayheims_course_v3		✓	✓	✓			
Unit of study	lsayheims_unitofstudy_v3	✓	✓	✓	✓			✓[1]
Completions	lsayheims_completions_v3	✓	✓	✓	✓		✓	
Scholarship	lsayheims_scholarship_v3	✓	✓	✓				

1 Unit of study should be considered when analysing information at the unit of study level but is not required if analysing data at the course level.

Additional resources

More information regarding the Higher Education Statistics Collection is available from:

- Higher Education Statistics student data collection: <https://www.dese.gov.au/higher-education-statistics/student-data>.
- Higher Education Statistics Collection data element dictionary: <https://www.tcsisupport.gov.au/elements>

Supporting documentation

This section outlines the resources available to help understand and work with the LSAY data.

Questionnaires and frequency tables

LSAY questionnaires and frequency tables are available for each survey wave. The frequency tables provide a simple output listing of each data item, along with the (unweighted) values corresponding to each category. Summary statistics are provided for continuous variables, and ranges are provided for variables with a large number of categories (for example, ANZSCO codes).

The questionnaires and frequency tables are available from the 'User support and documentation page' on the LSAY website: <https://www.lsay.edu.au/publications/user-support-and-documentation>.

Data framework

The LSAY data files are large and particularly complex. More than 500 variables are collected across each survey wave, culminating in more than 5000 variables across a data file that spans over ten survey waves. To help users navigate their way around the data, a framework has been developed that groups data items into common themes.

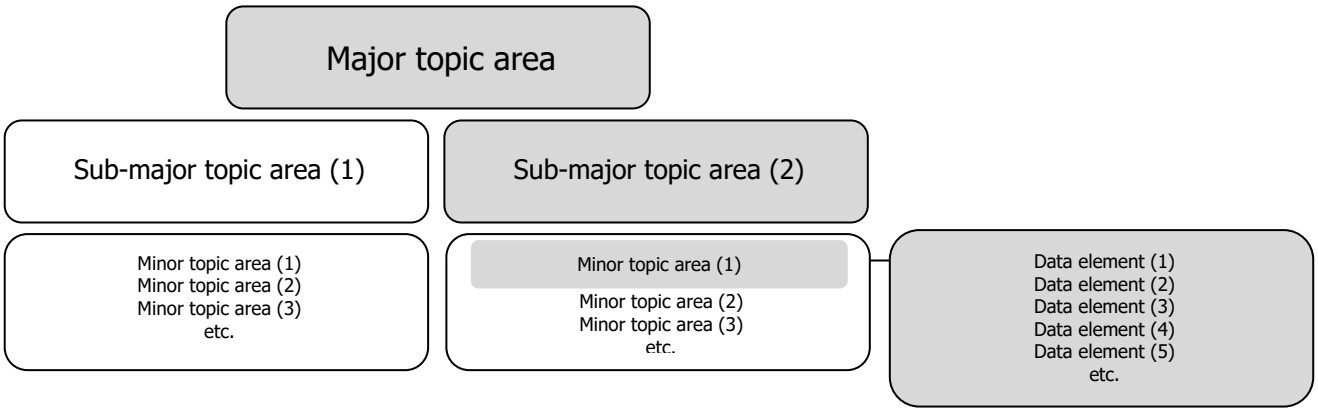
The framework has two main features:

- *Topic areas* use different themes to organise data items into common topics.
- *Data elements* identify data items (or variables) that are common within and across waves.

The framework is comprised of four hierarchical levels: major topic areas, sub-major topic areas, minor topic areas and data elements (as illustrated at figure 5). There are four major topic areas: Demographics, Education, Employment and Social.[16] These major topic areas, which are divided into sub-major topic areas and minor topic areas, are illustrated at figures 6 to 9.

Minor topic areas are further divided into data elements which represent variables that are common within any one survey wave, across multiple survey waves, and across LSAY cohorts. In some instances a data element may represent a single variable (when not collected across multiple waves).

Figure 5 LSAY hierarchical levels



16 Sampling variables (for example, respondent identifiers, interview dates and weight variables) are categorised using a fifth major topic area 'Sampling variables' but are not formally included in the framework.

Figure 6 Major topic area 1 – Demographics

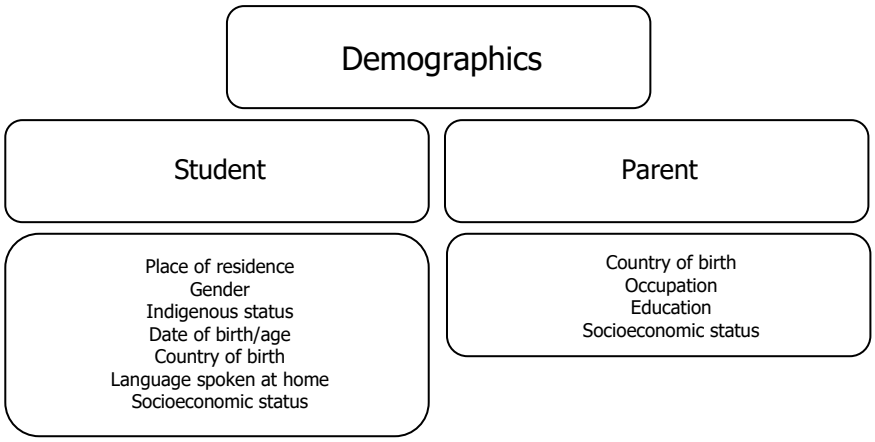


Figure 7 Major topic area 2 – Education

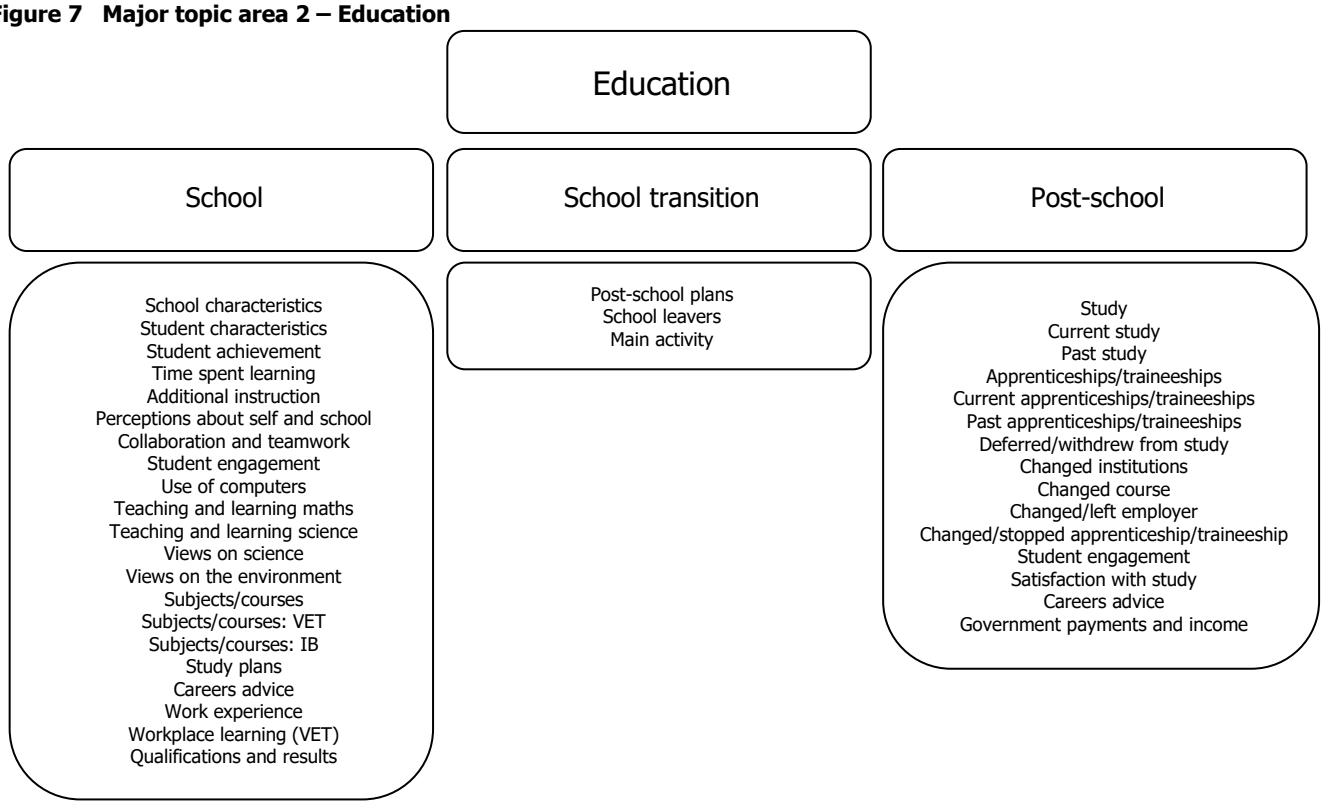


Figure 8 Major topic area 3 – Employment

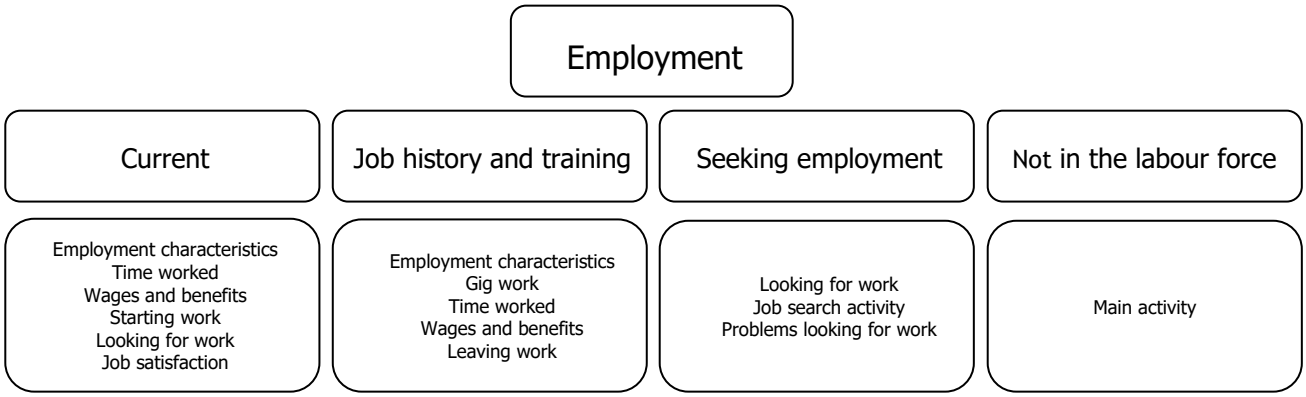
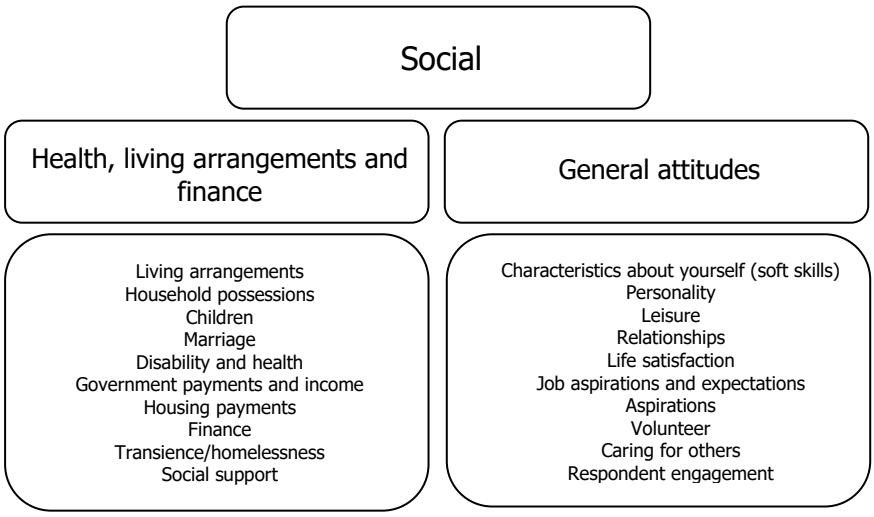


Figure 9 Major topic area 4 – Social



Variable listing and metadata workbook

The variable listing and metadata uses the LSAY data framework to bring together information from the LSAY datasets and questionnaires. The Excel workbook is organised using topic areas and data elements and provides a complete listing of all the variables in the LSAY datasets. Variables are listed in the order they appear in the data files and users can make use of filters and search functions to locate variables of interest.

The following information (or metadata) is available for each variable:

- *variable name* used in the dataset
- *variable label* includes the question number (where applicable) and a short description of the variable(s)
- *wave/year* the information was collected
- *section* refers to the section of the questionnaire from which the variable was collected
- *values* identifies the possible values the variable(s) can take and corresponding formats
- *question* wording for the variable(s) as it appears in the questionnaire
- *base population* provides a description of and the syntax for the respondents required to answer the specific question
- *notes* outline other important information (where applicable).

Two main worksheets are included in the metadata workbook: 'Variables' and 'Values'.

- The 'Variables' worksheet lists the variable name, variable label, question (wording) and base population.
- The 'Values' worksheet lists the applicable values for each variable.

An additional worksheet 'Linked data' provides a summary of the data items available from the linked datasets. For more information, refer to the 'Data linkage' section of this user guide.

The variable listing and metadata workbook can be accessed at: <www.lsay.edu.au/publications/2621.html>.

Variable selection

Data elements group together common variables, making it easier to locate similar variables. However, *it is important to note that not all variables assigned to a data element are directly comparable*, and additional attributes such as question wording, values, classifications used and base populations must be considered when selecting variables and analysing the data.

When using data elements and/or variables, *it is also important to consider other data elements, which may be located in a different topic area*. While data elements are unique *within* a minor topic area, they may not be unique *across* topic areas. For example, the data element 'Study type' exists under the major and sub-major topic area 'Education > Post-school', and under two different minor topic areas: 'Study' and 'Current study'. This is because the data element in the 'Study' minor topic area includes both past and current study because the nature of the study (that is, either past or current) cannot be ascertained until later in the questionnaire, when a question is asked about whether respondents are still undertaking the study.

An example where a data element appears in multiple topic areas is illustrated in figure 10, using an excerpt from the metadata workbook.

Online data dictionary

The online data dictionary uses the LSAY data framework to present information from the LSAY datasets, questionnaires and metadata. The dictionary is organised using topic areas and data elements and indicates the cohorts and waves for which each data element is available.

Making a selection at the topic area or data element level will return the relevant variables in the 'Variables' section of the dictionary. The variable label, question text, base population and notes for that variable are also displayed. Selecting a variable will display the values (or variable formats/labels) in the 'Values' section of the data dictionary.

Users can navigate the data dictionary by:

- browsing the data dictionary using the scroll bars
- using the search function to return results for specific keywords, cohorts or topic areas.

Users can also access the 'how to guide' for tips and tricks on using the online data dictionary. The data dictionary can be accessed at: <www.lsay.edu/data/lsay-data-dictionary>.

Figure 10 Identifying related topic areas

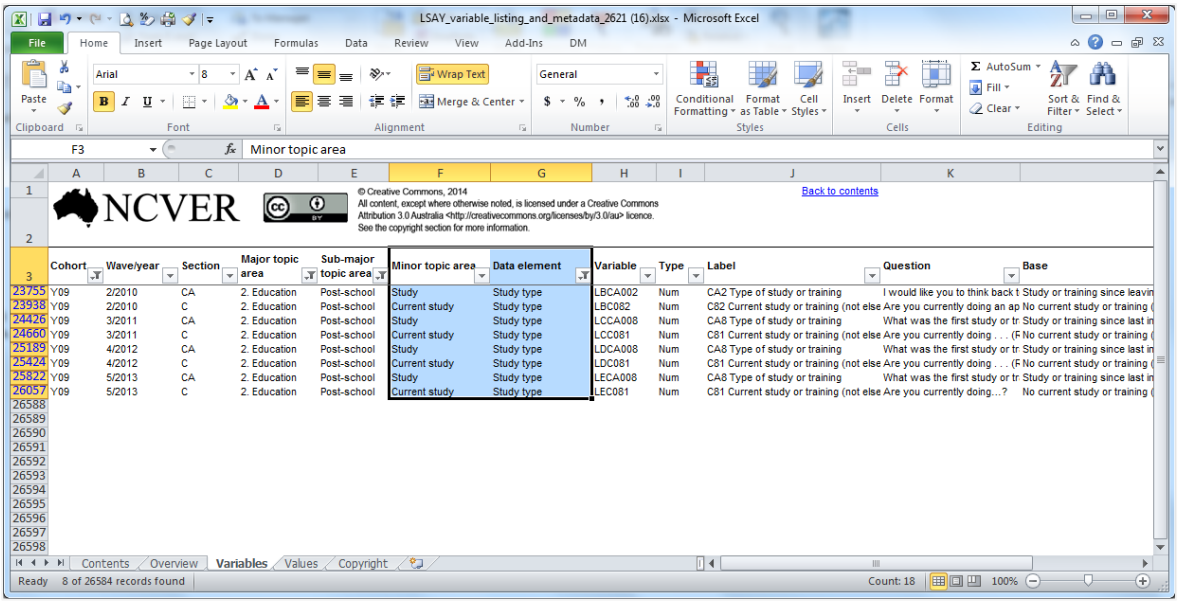
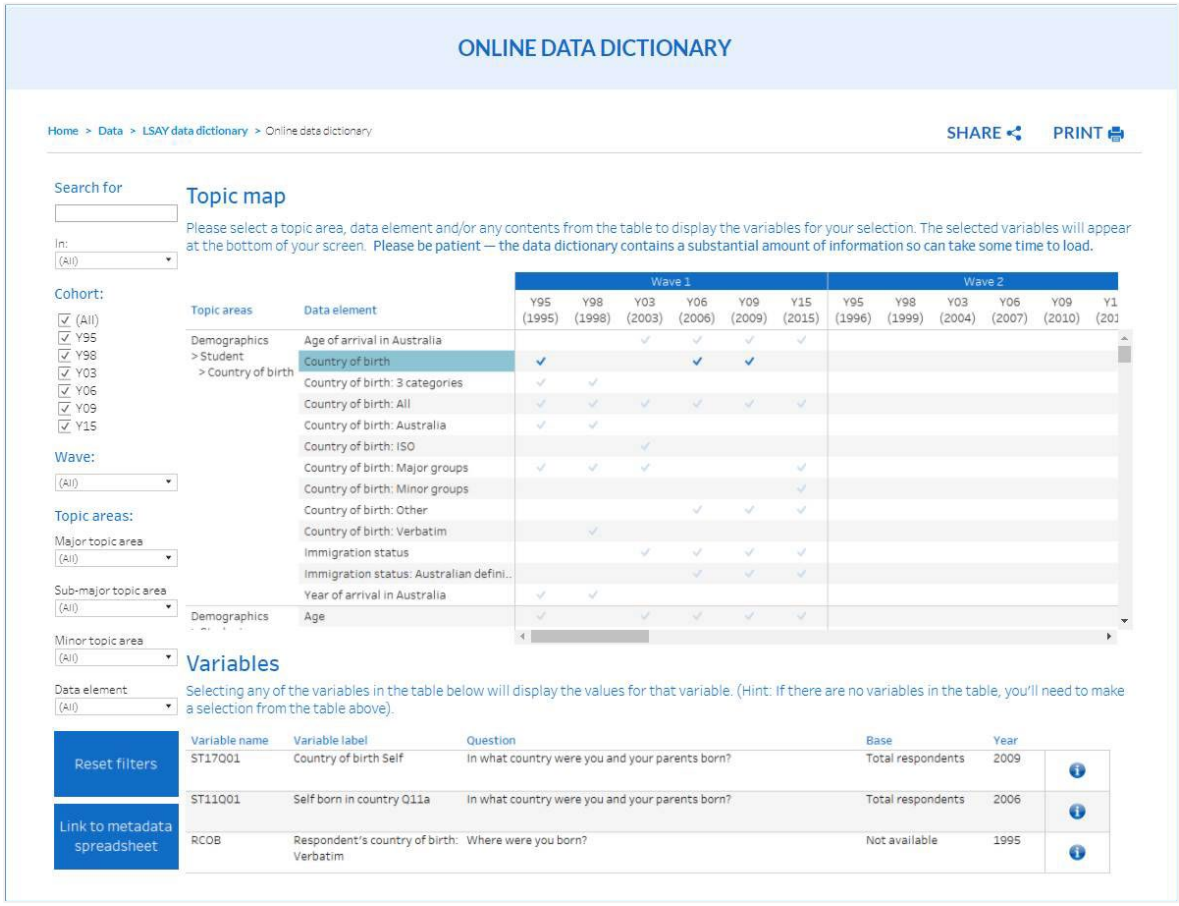


Figure 11 Data dictionary



Topic maps

Similar to the data dictionary, topic maps are organised by topic area and indicate the survey wave and year that the data elements were collected. The number of variables associated with any one data element is displayed in the corresponding wave/year columns. Topic maps are generated for each sub-major topic area. The topic maps are available in the 'Topic map' section of this user guide. A summary of the topic maps are shown in table 21.

Unlike the data dictionary, topic maps are static and only contain information up to the data element level. For more detail at the variable level, users should refer to the online data dictionary or the variable listing and metadata workbook.

Table 21 Topic maps

Major topic area	Topic map	Sub-major topic area
Demographics	1	Student
	2	Parent
Education	3	School
	4	School transition
	5	Post-school
Employment	6	Current
	7	Job history and training
	8	Seeking employment
	9	Not in the labour force
Social	10	Health, living arrangements and finance
	11	General attitudes

Technical and discussion papers

Technical papers assist data users and address topics such as sample design, attrition and weighting. They include papers on specific topics such as weighting and attrition. The papers can be accessed at: <https://www.lsay.edu.au/publications/technical-papers>.

Discussion papers form part of the LSAY technical paper series and focus on methodological issues in research. They can be accessed at: <https://www.lsay.edu.au/publications/discussion-papers>.

Users may also find the PISA technical documents useful (see table 1).

Sample and survey design

PISA sample design

Since 2003, LSAY has been integrated with the OECD's PISA. A nationally representative sample of 15-year-old students[17] was selected to participate in the Australian 2015 PISA assessment, which took place between late July and early September. This group of young people became the sixth LSAY cohort.

PISA uses a two-stage stratified sample. The first stage comprises the sampling of individual schools, while in the second stage individual 15-year-old students in each of the designated schools are sampled. Schools are sampled using probability proportional to enrolment size of 15-year-olds (PPS). Under this sampling scheme, larger schools have a greater chance of selection than smaller schools (Lim 2011).

The PISA sample is designed to be representative of students across Australia, using state/territory, school sector, geographic location, gender and socioeconomic background as strata. Smaller jurisdictions and Indigenous students were oversampled to ensure that reliable results could be produced for these populations.

Improvements were made to the survey design in 2015 to enhance the reliability of the data; this was done by increasing the number of schools and decreasing the number of students within schools. The 2015 PISA sample was comprised of 14 530 students selected from 758 schools from all states and territories. Within each school, 20 students were sampled; 30 students were sampled from ACT schools; 27 students were sampled from NT schools; and all age-eligible Indigenous students were sampled. In schools with fewer than the required number of students, all 15-year-olds were selected.

Further information about the survey design for PISA 2015 can be found from the PISA documentation:

- *PISA 2015 technical report*, which can be accessed at: <http://www.oecd.org/pisa/data/2015-technical-report/>.
- PISA 2015 Australian country report: *PISA 2015 — Reporting Australia's results,* which can be accessed at: <https://research.acer.edu.au/ozpisa/22/>.

Establishing the LSAY sample

Prior to the PISA assessment, schools were provided with information packs to be distributed to students. A brochure containing information on the LSAY program and how students can participate (that is, by providing their contact details when undertaking PISA) was included in the packs. In the November following the assessment, ACER provided students (via the schools) with a certificate of participation. As part of this mail-out, an LSAY flyer was also provided to students, along with a letter to parents introducing them to the program and stressing the importance of the study, and urging them to encourage their child to take part. The LSAY flyer encouraged students to provide or update their contact details, which could be achieved using any of five methods:

- the PISA student questionnaire
- an online form available on the LSAY website

17 The target population was students aged between 15 years and 3 (complete) months and 16 years and 2 (complete) months at the beginning of the assessment period, and who were enrolled and attending an educational institution full-time or part-time. Since the majority of the PISA target population is made up of 15-year-olds, the target population is often referred to as 15-year-olds (Thomson, De Bortoli & Underwood 2017).

- the detachable form at the bottom of the flyer via reply paid mail
- an email to Wallis, the fieldwork contractor
- a toll free phone number.

Students providing their contact details via any of these methods were entered into a prize draw to one win of 50 gift cards, each valued at $100.

ACER provided the names of 14 849 PISA participants to Wallis, of which 319 were later deemed ineligible to be part of PISA and were removed from the final PISA sample. Many participants provided no contact details, or the details provided were found to be unusable for the LSAY sample.

Table 22 Sample cleaning outcomes

	No. of records
Total provided	14 849
No contact details	4 554
Unusable contact details	86
Withdrew prior to fielding	7
Total removed	**4 647**
Total included in sample	**10 202**

For the 10 202 sample members providing usable contact details, the number of contact points given (for example, address, home phone, mobile phone or email) were far fewer when compared with previous cohorts. Compared with the starting sample for the previous (Y09) cohort, the Y15 sample had significantly more records containing only one contact point (11%, compared with 2% for Y09) and far fewer records with four contact points (36%, compared with 53% for Y09). Only 80% of the fielded sample for Y15 had a phone number, compared with 96% of the initial fielded sample for Y09. Close to a quarter of sample members could not be tracked using the contact details provided, and 13.2% of interviews were refused (Wallis Social Research 2017).

Response rates

Table 23 shows the sample sizes and response rates for the LSAY Y15 cohort. Some records supplied by Wallis were later removed from the LSAY datasets as they were deemed ineligible to be part of PISA (see table 24). In addition, respondents recruited as part of the top-activity at wave 3 (table 24) are included in the sample sizes at table 23. Further information about the top-up activity can be found in the sub-section that follows 'Wave 3 (2017): top-up activity'.

Wave 2 (2016)

At the end of the interview period, 4725 interviews had been completed representing a response rate of 46.3% of the total sample with usable contact details. Online was the favoured mode of completion, with over half (55.3%) of the respondents choosing this option, while 39.6% completed the survey with the assistance of a telephone interviewer. A small proportion (5.2%) used a combination of online and telephone modes to complete the survey. The average CATI interview length for the main survey was 26.7 minutes.

Table 23 Sample sizes and response rates

	Wave, year										
	W1, 2015	W2, 2016	W3, 2017	W4, 2018	W5, 2019	W6, 2020	W7, 2021	W8, 2022	W9, 2023	W10, 2024	W11, 2025
Age at June 30	15.7	16.7	17.7	18.7	19.7	20.7	21.7	22.7	23.7	24.7	25.7
PISA sample (n)[1,2]	14 530	4 704	4 603	4 825	3 721	3 759	3 705	3 427	3 301	3 246	3301
% of wave 1	100	32.4	31.7	33.2	25.6	25.9	25.5	23.6	22.7	22.3	22.7
LSAY sample (n)[1,2]	10 202[3]	4 704	4 603	4 825	3 721	3 759	3 705	3 427	3 301	3 246	3301
% of wave 1	100	46.1	45.1	47.3	36.5	36.8	36.3	33.6	32.4	31.8	32.4
% of previous wave		46.1	97.9	104.8	77.1	101.2	98.6	92.5	96.3	98.3	101.7

Notes:
1 Excludes respondents deemed ineligible to be part of PISA
2 Includes respondents recruited as part of the top-up activity conducted at wave 3 (2017) and completed an LSAY interview at each respective survey wave after that
3 Refers to the total number of PISA participants with usable contact details

Table 24 Exclusions and additions to the LSAY Y15 sample

	Wave, year										
	W1, 2015	W2, 2016	W3, 2017	W4, 2018	W5, 2019	W6, 2020	W7, 2021	W8, 2022	W9, 2023	W10, 2024	W11, 2025
PISA exclusions		21	20	22	13	14	13	14	11	12	13
Top-up additions			251	472	341	351	349	316	305	301	310

Wave 3 (2017)

Response rates for the 2016 interviews were lower than expected due to a high rate of missing or unusable contact details provided at the time of PISA. Two approaches were taken to increase the sample size of the Y15 cohort in 2017:

- re-engaging wave 2 non-responders using incentives
- recruiting a top-up sample by drawing a new random sample of school students the same age as the main Y15 cohort, with the goal of collecting a short recruitment questionnaire and integrating these participants into the main LSAY sample at waves 3 or 4.

Re-engaging non-responders

Two separate groups were established for the main Y15 sample at wave 3. One group (the 2016 'responders') completed an interview in 2016; the second group (the 2016 'non-responders') did not complete their 2016 survey — this group did not include anyone who explicitly refused to participate in 2016.

At the end of the interview period, 4,372 interviews had been completed (4,001 responders and 371 from the 'wave 2 non-responder' group). This equated to an overall response rate of 84.7% achieved for the main 'wave 2 responder' group. The re-engaging activity resulted in 9% of wave 2 non-responders (371 participants) completing the survey in wave 3 and an additional 95 wave 2 non-responders updating their details for future contact (but not completing wave 3). The average interview length was 21.8 minutes.

Top-up activity

The top-up activity resulted in 746 students completing the recruitment questionnaire and/or an interview at wave 3 (see table 25). Of these, only 251 completed both the recruitment questionnaire and the wave 3 survey so were eligible to be added to the sample. Further information about the top-up sample can be found in the following section.

Table 25 LSAY top-up sample, wave 3 (2017)

Group	Number of respondents
Recruitment questionnaire only	464
Recruitment questionnaire and wave 3 survey	251
Wave 3 survey only	31
Total	**746**

Survey mode

Online completion was again the preferred mode with 63.3% of wave 2 responders choosing to participate in this way. One-third (33.9%) of the wave 2 responder group who completed the survey online did so within the first few days of receiving their initial invitations via mail and email, requiring no follow up. The wave 2 non-responder group showed a different trend with more than half (59.3%) of interviews completed using CATI, demonstrating the importance of telephone follow-up with this group.

Wave 4 (2018)

As a result of the success of the re-engagement of previous non-responders at wave 3, this approach was retained at wave 4. In addition, those who had participated in the previous year's top-up activity were also approached to complete an interview. To increase the likelihood of a completed interview, certain respondent types were incentivised, resulting in the following five sample groups:

- Group 1: Recruited from the PISA sample and completed wave 4 (not incentivised)
- Group 2: Recruited from the PISA sample and provided a 'soft refusal' to participate at wave 2 (2016) and/or wave 3 (2017) (incentivised)
- Group 3: Recruited from the PISA sample and participated at wave 2 (2016) but not wave 3 (2017) (incentivised)
- Group 4: Recruited from the top-up sample but did not participate at wave 3 (2017) (non-responders from this group were incentivised later in the fieldwork)
- Group 5: Recruited from the top-up sample and participated at wave 3 (2017) (non-responders from this group were incentivised later in the fieldwork).

At the end of the interview period, 4847 interviews had been completed representing a response rate of 49.7% of the total sample with usable contact details (n = 9746). This represents an increase in the number of completed interviews when compared to the previous year (wave 3, 2017) as a result of the additional interviews achieved through targeting of non-responders (788 additional interviews were completed by participants from groups 2 and 3) and additional interview completions from the top-up group (472 interviews were completed by those from groups 4 and 5, of which 278 completed their first LSAY interview at wave 4).

Online was the favoured mode of completion, with more than two-thirds (67.3%) of respondents choosing this option, while 32.7% completed the survey with the assistance of a telephone interviewer. The average interview length was 25.4 minutes via telephone and 21.8 minutes online.

Wave 5 (2019)

To help target communications and incentives, the 2019 sample was comprised of six groups based on: their previous level of engagement with the survey; and their classification in terms of being easy or hard to reach (see table 26).

The easy-to-reach groups (groups 1 and 3) tended to be more engaged with LSAY and have higher levels of literacy. The hard-to-reach groups (groups 2 and 4) were more likely to be from lower equity groups, have difficulty with reading and comprehension, be late responders and/or Indigenous.

All respondents, regardless of the group they were in, were eligible to go into the draw to win 1 of 50 $100 Coles Myer Gift Cards.

Survey invitations and reminders were customised for each group and an additional incentive of a $20 eGift card was provided to the groups most at risk of not participating.

Table 26 Wave 5 (2019) sample groups

Group	Sample group	Prize draw	$20 eGift card
Group 1	• Completed LSAY survey in 2018 • Not hard to reach	✓	×
Group 2	• Completed LSAY survey in 2018 • Hard to reach	✓	✓
Group 3	• Last completed LSAY survey in 2016/17 • Not hard to reach	✓	×[1]
Group 4	• Last completed LSAY survey in 2016/17 • Hard to reach	✓	✓
Group 5	• Did not complete LSAY survey in 2018; generally completed one LSAY survey only	✓	✓
Group 6	• Did not complete LSAY survey in 2018; top-up recruit from 2017	✓	✓

1 This group was incentivised later in the fieldwork period. Initially, only groups 2, 4, 5 and 6 were incentivised in this way. From 5 December 2019, following the end of year mailing, members from group 3 who had not yet completed the 2019 survey became eligible for an incentive.

At the end of the interview period, 3734 interviews had been completed representing a response rate of 41.3% of the total sample with usable contact details (n = 9035). The total number of completions includes an additional 19 respondents from the wave 3 top-up (recruitment) sample who participated in an LSAY interview for the first time in 2019.

More than three-quarters of respondents completed their interviews online (77%), while 23% completed their survey with the assistance of a telephone interviewer. The average interview length was 27.4 minutes via telephone and 18.9 minutes online.

Wave 6 (2020)

To effectively target communications, the 2020 sample was comprised of two sample groups and survey invitations, reminder emails and SMS were customised for each group. Sample members were assigned to the groups according to their most recent year of survey engagement in order to encourage the best possible response rate.

- Group 1: completed their last survey in 2019 (wave 5) and made up about three-quarters of the entire 2020 sample.
- Group 2: completed their last survey in 2018 (wave 4) and made up about one-quarter of the entire 2020 sample.

All respondents who completed the survey automatically received a $20 eGift Card. This is the first year since wave 3 that *all* respondents received an incentive for completing their LSAY interview.

To further encourage early survey completion, a rolling prize draw approach was introduced. There were three prize draws each worth $2,000 with a total prize pool of $6,000. All sample members who completed the survey before 12am the day of each of the three prize draws were given the chance to win a prize. For each of the three prize draws, the prizes were one of:

- 1 x $500 pre-paid Eftpos gift card; or
- 4 x $250 pre-paid Eftpos gift cards; or
- 5 x $100 pre-paid Eftpos gift cards.

At the end of the interview period, 3773 interviews were completed, representing a response rate of 76.9% of the total sample (n = 4905).

The most common mode of completion continues to be online with 83.2% of respondents completing their interviews in this way. The remainder (16.8%) completed their survey over the phone. The average interview length was 27.9 minutes by telephone and 18.9 minutes online.

Wave 7 (2021)

The 2021 sample was comprised of two sample groups. Sample members were allocated according to their most recent year of survey engagement to help encourage the best possible response rate.

- Group 1: completed their last survey in 2020 (wave 6) and made up about three-quarters of the entire 2021 sample.
- Group 2: completed their last survey in 2018 or 2019 (waves 4 or 5) and made up about one-quarter of the entire 2021 sample.

All respondents who completed the survey automatically received a $20 eGift Card, and the rolling prize draw and structure introduced in 2020 was retained.

At the end of the interview period, 3718 interviews were completed, representing a response rate of 78.9% of the total sample (n = 4710).

Most respondents completed their survey online (85.9%), with the remainder completing their survey over the phone (14.1%). The average interview length was 30 minutes by telephone and 17 minutes online.

Wave 8 (2022)

As for 2021, the 2022 sample was comprised of two sample groups:

- Group 1: completed their last survey in 2021 (wave 7) and made up 83.2% of the entire 2022 sample.
- Group 2: completed their last survey in 2018, 2019 or 2020 (waves 4-6) and made up 16.8% of the entire 2022 sample.

All respondents who completed the survey automatically received a $20 eGift Card, and the rolling prize draw and structure introduced in 2020 was retained.

At the end of the interview period, 3441 interviews were completed, representing a response rate of 77.0% of the total sample (n = 4469).

Most respondents completed their survey online (83.0%), with the remainder completing their survey over the phone (17.0%). The average interview length was 31 minutes by telephone and 17 minutes online.

Wave 9 (2023)

As in previous years, the 2023 sample was comprised of two sample groups:

- Group 1: completed their last survey in 2022 (wave 8) and made up 92.7% of the entire 2023 sample.
- Group 2: completed their last survey in 2018, 2019, 2020 or 2021 (waves 4-7) and made up 7.3% of the entire 2023 sample.

All respondents who completed the survey automatically received an eGift Card as a thank you for completing the survey, and the rolling prize draw and structure introduced in 2020 was retained.

The incentive value for Group 1 remained unchanged and respondents received a $20 eGift Card. However, the incentive for participation among Group 2 respondents increased to a $40 eGiftcard to encourage their reengagement and improve the response rate.

At the end of the interview period, 3312 interviews were completed, representing a response rate of 75.5% of the total sample (n = 4386).

Most respondents completed their survey online (87.4%), with the remainder completing their survey over the phone (12.6%). The average interview length was 31 minutes by telephone and 17 minutes online.

Wave 10 (2024)

In 2024, the sample was comprised of four groups:

- Group 1: completed the 2022 (wave 8) and 2023 (wave 9) surveys (72% of the sample)
- Group 2: completed the 2022 (wave 8) survey but not the 2023 (wave 9) survey (8%)
- Group 3: did not complete the 2022 (wave 8) survey but not the 2023 (wave 9) survey (6%)
- Group 4: had not completed a survey since 2021 (wave 7) or earlier (waves 4-6) (14%)

All respondents who completed the survey automatically received an eGift Card, and the rolling prize draw and structure introduced in 2020 was retained. Group 1 received a $20 eGift card. To encourage re-engagement, Groups 2 and 3 were offered a $40 eGift card, while Group 4 respondents, who had not participated for several years, were offered a $60 eGift card, providing a stronger incentive to return to the study.

By the end of the fieldwork period, 3258 interviews were completed, representing a 76.7% response rate. Most respondents completed their survey online (90.5%), with the remainder completing their survey by telephone (9.5%). Average interview lengths were 22 minutes by telephone and 13 minutes online.

Wave 11 (2025)

In 2025, the sample was comprised of five groups:

- Group 1 (core responders): completed the 2022 (wave 8), 2023 (wave 9) and 2024 (wave 10) surveys (66% of the sample)
- Group 2 (recent non-responders): completed the 2022 (wave 8) and 2023 (wave 9) surveys, but not the 2024 (wave 10) survey (6%)
- Group 3 (re-engaged responders): did not complete the 2022 (wave 8) or 2023 (wave 9) surveys, but completed the 2024 (wave 10) survey (8%)
- Group 4 (disengaged responders): did not complete since 2022 (wave 8) or earlier (waves 4-7), but completed the survey in 2024 (wave 10) when offered $60 (3%)

- Group 5 (long-term disengaged): had not participated since 2021 (wave 7) or earlier (waves 4,5,6) (17%).

All respondents who completed the survey automatically received an eGift Card, and the rolling prize draw and structure introduced in 2020 was retained. Group 1 received a $20 eGift card. To encourage re-engagement, Groups 2 and 3 were offered a $50 eGift card, while Group 4 and 5 respondents were offered a $70 eGift card, providing a stronger incentive to return to or encourage ongoing participation in the final year of the study for those who had disengaged.

By the end of the fieldwork period, 3314 interviews were completed, representing a 78.3% response rate. Most respondents completed their survey online (88.3%), with the remainder completing their survey by telephone (11.7%). Average interview lengths were 15 minutes by telephone and 12 minutes online.

Top-up sample

The top-up sample was selected in 2017 using a two-stage sampling design. The sampling frame included all 3,395 secondary schools in Australia providing Year 12 education (this included secondary schools, combined primary-secondary schools and senior secondary schools). Explicit and implicit strata were defined. Explicit stratification variables included state/territory and sector.

Systematic probability proportional to size (PPS) sampling was used to determine which schools to sample within each explicit stratum (based on number of students enrolled in Year 12). Before selecting the sample, units in each explicit stratum were also sorted by a number of implicit stratification variables including geographic zone, school gender composition, school socio-economic level, and school-level NAPLAN scores. This method results in the sample being more representative of the population with respect to the distribution of these variables. Overall 347 schools were selected at stage one. Schools were identified and contacted to participate in the survey in September 2017.

In stage 2, a maximum of 50 students in Year 12 of the sampled schools were selected to participate in the survey. Schools were advised to select one or two Year 12 classes (depending on the size) and ask all students in that classroom to fill out the recruitment questionnaire. No additional efforts were made to oversample Indigenous students in the school. In case of early school refusals, a replacement school from the same stratum was selected to participate.

The recruitment questionnaire was developed to collect key demographic information, matching that collected by PISA at wave 1. To encourage participation, a $20 incentive was provided to students for completion. To test the effectiveness of various incentivising approaches, half of the sample were offered an incentive after completing the recruitment questionnaire and the other half were offered an incentive after completing the wave 3 questionnaire. However, due to the differing requirements from the relevant state/territory authorities who were responsible for approving the project at the jurisdictional level, it was not possible to provide all participants with an incentive. In addition, parental consent was required for students to participate from some states and territories.

Only 28% of the schools (97 schools) agreed to participate with 746 students completing the recruitment questionnaire. School participation was lower than anticipated, which was largely due to the timing of recruitment which occurred towards the end of the school year when students were undertaking their final exams.

On completing the recruitment survey, students were also encouraged to complete their LSAY survey. 282 students completed the main wave 3 survey through the top-up activity, however 31 of those did not complete a recruitment questionnaire therefore key demographic information is not available for this group.

Data collection

PISA participants providing usable contact details were contacted in 2016 for the purpose of participating in follow-up interviews as part of the LSAY program. This interview collected further information on the respondent's school experiences, school and post-school intentions, school leavers and their transitions from school, engagement in education, post-school study, employment, living arrangements, wellbeing, personality and soft skills.

Survey mode

From the second wave, Y15 participants were able to choose between computer-assisted telephone interviews (CATI) and computer-assisted web interviews (CAWI). LSAY telephone interviews take an average of 25 minutes in length. Sample members can access the CAWI through a number of entry points, including the primary approach letter, email invitation, reminder emails, survey website, link via email to respondent during CATI, and linked email to respondent via SMS reminders.

Pilot testing

Pilot testing was conducted in each wave from 2016 onwards. In all cases, interviewing began following a full interviewer briefing, and the pilot survey remained open alongside the main survey to maximise response.

- Wave 2 (2016): fieldwork ran from 15 September 2016 to 12 February 2017. Of 600 sample members, 300 interviews were achieved (50.0%).
- Wave 3 (2017): fieldwork ran from 8 August 2017 to 18 February 2018. Of 500 sample members, 344 interviews were achieved (68.8%).
- Wave 4 (2018): fieldwork ran from 2 August 2018 to 24 February 2019. Of 632 sample members, 349 interviews were achieved (55.2%).
- Wave 5 (2019): fieldwork ran from 12 August 2019 to 13 February 2020. Of 482 sample members, 220 interviews were achieved (45.6%).
- Wave 6 (2020): fieldwork ran from 9 June 2020 to 4 January 2021. Of 279 sample members, 212 interviews were achieved (76.0%).
- Wave 7 (2021): fieldwork ran from 22 June 2021 to 10 January 2022. Of 269 sample members, 206 interviews were achieved (76.6%).
- Wave 8 (2022): fieldwork ran from 16 June 2022 to 4 January 2023. Of 254 sample members, 186 interviews were achieved (73.2%).
- Wave 9 (2023): fieldwork ran from 16 June 2023 to 4 January 2024. Of 252 sample members, 176 interviews were achieved (69.8%).
- Wave 10 (2024): fieldwork ran from 18 June 2024 to 6 January 2025. Of 245 sample members, 173 interviews were achieved (70.6%).
- Wave 11 (2025): fieldwork ran from 12 June 2025 to 5 January 2026. Of 245 sample members, 182 interviews were achieved (74.3%).

Survey materials and sample engagement

Social media campaign

In 2015 NCVER developed a social media campaign carrying the theme Five Million Minutes (5mm) and in 2016 all survey materials built on this concept. The theme focuses on the longevity of LSAY, demonstrating what can occur in the 10 years of the life of the survey (which works out to be five million minutes) and explaining that the few minutes participants spend each year participating in LSAY can help to shape Australia's future. The social media campaign included:

- a promotional video
- Facebook and YouTube advertising targeted at 15 to 16-year-olds
- a mobile responsive website.

In 2018 a new campaign concept was developed called the 'Missing Piece' which was incorporated into all survey communications and materials. The theme played on the idea that, in order to conduct an accurate survey, we need all participants to continue to engage with the LSAY each year. Everyone's information is valuable and contributes to the bigger picture. Without them the information collected would be skewed, and they are an important piece of this puzzle.

Letter to parents

The first contact with this cohort in 2016 was directed to parents/caregivers. Parents were mailed a letter and information sheet. The aim of the letter was to:

- introduce LSAY to the parents
- explain why their child had been chosen
- gain their support in encouraging their child to take part
- provide contact details of Wallis and NCVER if parents wanted more information.

The information sheet was included to ensure that parents and sample members were fully informed before giving their consent to participate. The letters were addressed to the parent/guardian of the sample member where postal addresses were provided (8433 letters), on 3 October 2016. Following the mailing, 675 of these letters were returned to Wallis due to incorrect addresses.

'Update details' campaign

Prior to the commencement of the wave 4 fieldwork, a letter was sent in June 2018 to all respondents with a postal address encouraging them to update their details, at the same time introducing them to the new campaign theme the 'Missing piece'. The letter instructed respondents to either update the details on the form and return it to Wallis in the post, or alternatively log onto the LSAY website using their unique ID and update their details online.

As the various wave 4 groups within the cohort have different experiences with the LSAY survey, the wording of the letter changed slightly to suit the respondents LSAY experience. For instance, the letter for respondents in groups 4 and 5 who were new to the survey (having only been recruited in the previous year) contained more detail about what the LSAY survey is and why it is so important to participate. Shortly after the hardcopy letter mailing, an email was sent out to all respondents with an active email address.

This 'update your details' initiative has continued to be incorporated into the LSAY fieldwork campaign for every subsequent survey wave.

Primary approach letter

Sample members are contacted before the commencement of the annual fieldwork. Sample members are sent a mailer to reinforce their involvement in LSAY and to advise them of their forthcoming interview. The mailer included a tearaway panel for updating their contact details and in 2016 respondents were sent a data linkage brochure and information sheet.

Key information provided in the mailer includes: the 1800 LSAY toll-free contact number, the LSAY email address and the LSAY website for updating or correcting their details. The mailer also reminds respondents that they are able to complete their interview online and provides a description and a link to the LSAY website.

Primary invitation email

Where email addresses are available, an email invitation is sent a few days before telephone interviews are scheduled to commence and includes:

- a personalised link to the survey
- a link to the LSAY website
- a link to the LSAY Facebook page.

Reminder emails and SMS

Throughout the survey period reminder emails and SMS are sent to non-respondents whose email addresses and/or mobile numbers have been provided. All SMS reminders are sent with the LSAY alphatag and included the respondent's unique survey link.

End of year mailer, email and SMS

Towards the end of the fieldwork, a mailer and email are sent to thank participants for their involvement in LSAY and as a reminder for those who hadn't yet completed their survey. SMS reminders are sent only to those who had not yet completed an interview.

Incentives

Wave 2 (2016)

Fifty respondents who had completed their 2016 survey by 19 January 2017 were randomly selected to win a $100 Coles Group & Myer Gift Card. The total prize pool was $5000 worth of gift cards.

In December 2016 a promotional 'screen buddy' (a Five Million Minutes themed screen cleaner mounted on a LSAY branded business card) was also included in the end of year mail-out. The screen buddy was used as a way to thank respondents for completing their interview, and to engage and motivate sample members who had not yet done so.

Wave 3 (2017)

Due to the low response rates in 2016, incentives in the form of a $20 eGift card were offered to all participants who completed their 2017 survey. In addition, any Y15 non-responders who provided, confirmed or updated their mailing address, email address and phone number received a $20 eGift card of their choice, however they did not receive a second $20 eGift card if they went on to complete the survey. A $20 incentive was also provided to students selected as part of the top-up sample for completing the recruitment questionnaire. However, due to the differing requirements from the relevant state/territory authorities who were responsible for approving the project at the jurisdictional level, it was not possible to provide all participants with an incentive.

Wave 4 (2018)

Incentives were again provided to selected groups at wave 4 (2018). All sample members completing the survey before the day of the prize draw were given the chance to win one of 50 $100 Coles Myer Gift Cards. In addition, selected sample members were provided with an incentive in the form of a $20 eGift card. As outlined above in the above 'Response rates' section, the 2018 sample was comprised of five groups. Broadly speaking, those who completed an interview at wave 3 (2017) were not offered an incentive (groups 1, 4 and 5); and an incentive was offered to sample members reintroduced from the original sample (groups 2 and 3).

Later in the wave 4 fieldwork, members of group 4 who had not yet completed the 2018 survey became eligible for an incentive; and sample members from group 1 with low levels of literacy and numeracy who had not yet completed the 2018 survey also became eligible for the incentive (to reduce the bias in the survey completions due to low responses from these sample members).

Wave 5 (2019)

Incentives were again provided to selected groups at wave 5 (see table 26). All respondents, regardless of the group they were in, were eligible to go into the draw to win 1 of 50 $100 Coles Myer Gift Cards. An additional incentive of a $20 eGift card was provided to the groups most at risk of not participating.

Waves 6-8 (2020-2022)

Since 2020 (wave 6), to further encourage survey completion, all respondents who completed the survey automatically received a $20 eGift Card, regardless of their respondent group. 2020 was the first year since wave 3 that *all* respondents received an incentive for completing their LSAY interview. In addition, multiple prize draws were introduced to help encourage early completion and participation over the entire fieldwork period. There were three prize draws each worth $2,000 with a total prize pool of $6,000. All sample members who completed the survey before 12am the day of each of the three prize draws were given the chance to win a prize.

Waves 9-11 (2023-2025)

As in previous years, all respondents who completed the survey automatically received an eGift Card as a thank you for completing the survey, and the rolling prize draw and structure introduced in 2020 was retained.

In 2023, the incentive value for Group 1 (i.e. those that had completed their last survey in 2022) remained unchanged and respondents received a $20 eGift Card. However, the incentive for participation among Group 2 respondents (i.e. those that had completed their last survey prior to 2022) increased to a $40 eGiftcard to encourage their reengagement and improve response rates.

In 2024, the incentive value for Group 1 (i.e. those that had completed the 2022 and 2023 surveys) remained unchanged and respondents received a $20 eGift card. To encourage re-engagement, Groups 2 and 3 (i.e. those who had missed one of the previous two survey waves) were offered a $40 eGift card. Group 4 respondents (i.e. those who had not participated since 2021 or earlier) were offered a $60 eGift card to strengthen the incentive to return to the study.

In 2025, the incentive value for Group 1 (i.e. those that had continually completed the survey year on year) remained unchanged, and respondents received a $20 eGift card. To encourage re-engagement, Groups 2 and 3 (i.e. those who had missed one of the previous two survey waves) were offered a $50 eGift card. Groups 4 and 5 (i.e. longer-term disengaged responders) were offered a $70 eGift card to complete the survey.

1800 and email helpdesk

Wallis maintains an LSAY free-call 1800 number throughout the year and a dedicated LSAY mailbox is also checked daily. A team of Wallis staff respond to general queries, often within minutes, and in some cases requests to be removed from the survey. Approximately 70 emails were received from respondents and their parents; this included a mix of opt-outs, contact details updates, and requests for assistance to complete the survey online.

Sources of error

Estimates based on sample surveys have two major sources of error: non-sampling and sampling error. A brief description of the two types and an outline of what can be done to overcome the effects of these errors are given below.

Non-sampling error

Non-sampling error arises from inaccuracies in collecting, recording and processing the data. Some common examples of non-sampling error include: non-response, incorrect responses, missing responses, and interviewer and processing error. Non-sampling error can be accounted for, in part, by using weighted estimates to adjust for non-response. However, there are no statistical measures to accurately adjust for other types of non-sampling error. Nevertheless, other types of non-sampling error can be minimised through good questionnaire design, training and monitoring of interviewers, the use of computer-assisted interviews and effective data-checking and processing procedures.

Non-response and attrition

All surveys suffer from error related to non-response. Non-response is a form of non-sampling error and can be taken into account in the analysis of survey data. There are typically two forms of survey non-response:

- *Item non-response* occurs when a respondent does not answer all the questions in the survey.
- *Unit non-response* occurs when not all respondents answer the survey due to, for example, refusal to participate or inaccurate contact details.

Item non-response can be minimised with the use of CATI, which can forward-feed information from previous interviews. Item non-response is generally treated using imputations. There are currently no imputed data for missing values in LSAY. However, data users can apply a number of techniques to assist in making the data more complete. The use of statistical modelling techniques, such as multiple imputation, allows data users to estimate item non-response, along with their respective standard errors.

Unit non-response (also called attrition) can lead to biased population estimates and incorrect standard errors, particularly if certain groups of the sample drop out at differing rates. Survey attrition is counteracted by attempting to maximise the year-on-year response rate, appropriate statistical modelling techniques, and/or the application of appropriate survey weights.

Weights

In order for the LSAY sample to more accurately represent the population of Australian 15-year-olds in 2015, the collected sample must be weighted to account for differences in the sampling distributions from the original population distribution, which may have arisen during the sampling process.

In 2010, NCVER reviewed the weighting methodology used for the LSAY Y03 cohort. As a result of this review, a logistic regression approach to weighting has been adopted. This methodology is consistent with the approach taken to calculate the Y06, Y09 and Y15 weights. These weights are provided in the data files deposited with the Australian Data Archive.

Further information on the current weighting methodology used is available from technical paper no. 61, *Weighting the LSAY PISA cohorts*, which can be accessed at: <www.lsay.edu.au/publications/2429.html>.

There are two weighting procedures applied to the LSAY data:

- *Sample weights* reflect the original sample design and ensure that the sample matches the population distribution from which the original sample was drawn. In the Y15 cohort, two sampling weights were created. The first weights sum to the sample size for that given wave. For example, the sample weights add to 14 530 in wave 1, 4704 in wave 2 etc. In the second set of weights, the sum of the weights equals the original population from which the sample was drawn (256 329). Students from states and territories with smaller numbers of 15-year-olds are over-sampled and students from jurisdictions with larger numbers of 15-year-olds are under-sampled. In order for the sample to more accurately represent the population of Australian 15-year-olds, the sample is weighted so that the sample sizes within strata are proportional to the original population sizes of the states and territories (that is, strata).
- *Attrition weights* are used to address unit non-response by ensuring that the distribution of the sample matches the distribution of the sample population. The attrition weights used in LSAY account for wave-on-wave attrition from the first wave.

In calculating attrition weights, a non-response analysis was undertaken to determine the factors that contributed to attrition. The use of attrition weights ensures that distributions in each wave match those obtained in PISA (for the factors identified as contributing to attrition). Logistic regressions have been used to calculate attrition weights. The response variable of whether or not a respondent replied to the survey in a given year was regressed against a series of factors that may contribute to non-response. The inverse of the predicted probability of responding then forms the attrition weights.

The final LSAY weights for each wave combine both the sample and attrition weights. Two sets of final weights are produced. The first reproduces the sample sizes in each wave, and the second reproduces the population size (256 329) at each wave. In both cases, the distributions in each wave match those obtained in the original population.

Users must be aware that bias that results from survey attrition may not be fully accounted for in the weighting strategies used. To allow users to determine the effectiveness of the attrition weights, data in the LSAY QuickStats demographic tables are presented both weighted and unweighted. LSAY QuickStats can be accessed at: <www.lsay.edu.au/data/lsay-quickstats>. Researchers are encouraged to determine their own weighting or analysis methodology to counteract attrition; this may include using methods of multiple imputations for missing values.

Table 27 shows the three different types of available weights and the variable naming convention for each, where *YY* or *YYYY* denotes the survey year at two or four digits respectively. Weights that sum to the (PISA) population size are denoted by 'P' at the end of the weight variable. A new set of weights have been added to the dataset at wave 3 (2017) to adjust the weights to account for the top-up sample, denoted by an '_R' indicating they completed the recruitment survey. Further details can be found in the following sub-section 'Adjustments to include the top-up sample'.

Adjustments to include the top-up sample

Various methods can be used to incorporate top-up samples into an ongoing longitudinal panel. Two main ways to integrate independent surveys together include: 'combining estimates' and 'pooling samples' (Watson, 2014).

The first approach includes forming a weighted average of the estimates from each of the surveys, with weights that minimize the variance of the weighted average. The second approach has a more multi-

purpose application, and pools data from two or more surveys into a single database that can be analysed in a standard way. For this approach, weights need to be derived for the units in the pooled data base and therefore only one set of weights is created.

Pooling samples is more efficient than combining estimates especially for those surveys that oversample certain parts of the population, have highly differential non-response, or have a greater difference in the size of the main sample compared to the top-up sample. This is because when pooling samples, the sample members with similar characteristics (as far as they can be modelled) would have similar weights which would reduce the overall coefficient of variation of the weights.

It is important to note that the pooled approach should be used only if it can be assumed that the population characteristics as well as the domains of interest are similar from one survey to the next. Although the LSAY top-up and main samples were taken three years apart, and some changes in the Australian youth population would have occurred between the two time points (due to death, migration etc.), it is safe to assume that the target Australian youth population has not changed much in the time period between the two surveys.

The main advantage of the pooled approach is that once a set of suitable weights is found for the pooled sample, they can apply to many different estimates including totals, proportions and coefficients (Watson, 2014). The disadvantage is that a minimum estimated variance for all estimates may not be achieved.

Table 27 Weight variables

Weight	Variables	Sum
Sample weight	WTYYGEN	Sample size in YY
Sample weight (N)	WTYYGENP	Population size (256 329)
Attrition weight	ACHYYWT	Sample size in YY
Attrition weight (N)	ACHYYWTP	Population size (256 329)
Final weight	WTYYYY	Sample size in YYYY
Final weight (N)	WTYYYYP	Population size (256 329)
Final weight (recruitment)	WTYYYY_R	Sample size in YYYY – includes top-up sample
Final weight (recruitment - N)	WTYYYYP_R	Population size (256 329)

Sampling error

Users of the LSAY data must consider the size of the sampling error when deriving or interpreting estimates obtained from LSAY. Sampling error arises because estimates are obtained from the use of a sample rather than from measuring the entire population. It is possible to select many different individual samples from a single population; each of these would provide a different population estimate. An estimate obtained from a sample is subject to sample-to-sample variation (sampling error). In random (probability) sampling, the size of the sampling error (for a given sample) is measured using the standard error of the estimate.

It is important that users take into consideration the reliability of the estimates obtained from survey data. Standard errors, confidence intervals and relative standard errors (RSEs) can be calculated to determine the reliability of the estimate(s).

The greatest contributor to standard error is the sample size. Small sample sizes generally result in higher standard errors and wider confidence intervals. The relative standard error enables a comparison of the accuracy of two different estimates. An estimate with a high relative standard error or wide confidence interval should be used with caution, and users are advised against relying on estimates obtained from sample sizes of fewer than five, or estimates that have a relative standard error of greater than 25%.

Standard errors

The standard error of an estimate indicates the accuracy with which that estimate approximates the true population parameter. There are multiple methods for calculating the standard errors in complex surveys. One method commonly used is the Taylor series expansion.[18] This technique has been applied to obtain estimates of standard errors for the LSAY cohort reports. These standard errors can then be used to calculate confidence intervals and relative standard errors.

Confidence intervals

The confidence interval is an interval estimate of the population parameter. Sample estimates with high standard errors will have wide confidence intervals. In a confidence interval, the range of values above and below the sample estimate is called the margin of error. The smaller the margin of error, the higher our confidence in the estimate is.

The mathematical derivation of a 95% confidence interval for a proportion is:

$$\hat{p} \pm 1.96 \times se(\hat{p})$$

where $\hat{p}$ is the estimate obtained from the sample, $se(\hat{p})$ is the standard error of the estimate and $1.96 \times se(\hat{p})$ is the margin of error. The standard error and margin of error are typically obtained from a statistical analysis package.

Relative standard errors

The relative standard error is a standardised measure that enables a comparison between different estimates in terms of their reliability. The relative standard error is derived by dividing the standard error of the estimate by the estimate itself, expressed as a percentage:

$$RSE(\hat{p}) = \frac{se(\hat{p})}{\hat{p}} \times 100$$

18 For further information on this technique, users should consult William Cochran, *Sampling techniques*, 3rd edn, John Wiley and Sons, New York, 1977, sections 11.18, 11.19, 11.20.

Topic maps

Topic maps list the data elements for each sub-major topic area by minor topic area. The digits within the tables indicate the:

- survey waves in which this data element exists
- number of times the data element appears within a wave. This is equivalent to the number of variables that correspond to the data element in a single wave.

The following is a list of the topic maps:

- *'Topic map 1: Demographics — Student'* includes place of residence, gender, Indigenous status, date of birth and age, country of birth, language spoken at home, and socioeconomic status.
- *'Topic map 2: Demographics — Parent'* includes country of birth, occupation and education levels of a respondent's mother and father and their socioeconomic status.
- *'Topic map 3: Education — School'* includes school characteristics, student characteristics, student achievement, time spent learning, additional instruction, perceptions about self and school, collaboration and teamwork, student engagement, use of computers, learning maths and science, views on science, views on the environment, subjects and courses, study plans, careers advice, work experience, workplace learning and qualifications and results.
- *'Topic map 4: Education — School transition'* includes post-school plans, intentions and reasons for leaving school, and school leaver's main activity since leaving school.
- *'Topic map 5: Education — Post-school'* includes study (including current and past study, apprenticeships and traineeships), university preferences, qualifications obtained, reasons for withdrawing/deferring from study, changes in study status and/or details (including changes to course, institution, employer and apprenticeship/traineeship), student engagement, satisfaction with study careers advice and government payments and income.

  It is worth noting that within the following minor topic areas:

  - 'Study' may refer to past and/or current study, as well as apprenticeships and traineeships.
  - 'Current study' may refer to apprenticeships and traineeships.
  - 'Past study' may refer to apprenticeships and traineeships.
  - 'Apprenticeship/traineeships' may refer to past and/or current apprenticeships.
- *'Topic map 6: Employment — Current'* includes current employment, characteristics of employment, time worked (including preferred hours), wages and benefits, when started work, whether looking for (additional) work and job satisfaction.
- *'Topic map 7: Employment — Job history and training'* additional or past employment, characteristics of additional employment, gig work, time worked, wages and benefits, and main reason for leaving work.
- *'Topic map 8: Employment — Seeking employment'* includes job-search activity, preferred hours and problems looking for work.
- *'Topic map 9: Employment — Not in the labour force'* main activity while not in the labour force.

- *'Topic map 10: Social — Health, living arrangements and finance'* includes living arrangements, household possessions, marriage, disability and health, government payments and income, transience/homelessness and social support.
- *'Topic map 11: Social — General attitudes'* includes self-perception of soft skills (e.g. creativity and innovation, critical thinking, problem-solving and teamwork), personality, leisure activities, relationships with parents, life satisfaction, job expectations, volunteer work, and caring for others.

Topic map 1: Demographics – Student

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
Place of residence	State			1	1	1	1	1	1	1	1	1
	Postcode		1	1	1	1	1	1	1	1	1	1
Gender	Sex	2		1								
Indigenous status	ATSI	1		4								
Date of birth/ age	Age	1										
	Date of birth: Month	1	1	1								
	Date of birth: Year	1	1	1								
	Date of birth	1	1	1								
	Date of birth: SAS date	1										
Country of birth	Country of birth: Major groups	1										
	Country of birth: Minor groups	1		1								
	Country of birth: Other	1										
	Country of birth: All	1										
	Immigration status	1										

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
	Immigration status: Australian definition	1										
	Age of arrival in Australia	1										
Language spoken at home	Language spoken at home: Major groups	1										
	Language spoken at home: 3 digit	1										
	Language spoken at home: English/other	1										
Socioeconomic status	Household possessions (index)	1										
	Wealth (index)	1										
	Economic social and cultural status (index)	1										
	Expected occupation (ISEI)	1										

Topic map 2: Demographics – Parent

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
Country of birth	Mother's country of birth: Major groups	1										
	Mother's country of birth: Minor groups	1		1								
	Mother's country of birth: Other	1										
	Mother's country of birth: All	1										
	Father's country of birth: Major groups	1										
	Father's country of birth: Minor groups	1		1								
	Father's country of birth: Other	1										
	Father's country of birth: All	1										
Occupation	Mother's occupation (ISCO)	2										
	Mother's occupation (ANZSCO)			1								
	Father's occupation (ISCO)	2										
	Father's occupation (ANZSCO)			1								
Education	Mother's schooling	2		1								
	Mother's qualifications: Post-secondary training certificate	1		1								
	Mother's qualifications: Post-secondary training qualification	1		1								

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
	Mother's qualifications: University	1		1								
	Mother's qualifications: Doctorate (PhD) or equivalent	1		1								
	Mother's highest education level (ISCED)	1										
	Mother's qualifications: None			1								
	Mother's qualifications: Don't know			1								
	Father's schooling	2		1								
	Father's qualifications: Post-secondary training certificate	1		1								
	Father's qualifications: Post-secondary training qualification	1		1								
	Father's qualifications: University	1		1								
	Father's qualifications: Doctorate (PhD) or equivalent	1		1								
	Father's highest education level (ISCED)	1										
	Father's qualifications: None			1								
	Father's qualifications: Don't know			1								
	Parents' highest education level (ISCED)	1										
	Parents' highest education level (years)	1										
Socioeconomic status	Mother's ISEI score	1										
	Father's ISEI score	1										

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
	Parents' ISEI score	1										

Topic map 3: Education – School

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
School characteristics	Geographic location	1										
	School state	1	3	3	3	1	1					
	School state (last interview)		1									
	School postcode	1										
	School sector	1		1								
	School identifier	1	3	4	2							
Student characteristics	Student identifier	1										
	At school		2	2	2	2	2					
	At school (at last interview)			1								
	Year level	2	1	2	1							
	Year level (at last interview)			1	1	1						
	Study program	2		1								
	ISCED level	1										
	ISCED program	1										
	ISCED orientation	1										
	Studying for IB		1	1	2	1	1					
	Age commenced kindergarten/pre-school	1										
	Age commenced primary school	1										

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
	Duration in early childhood education and care (index)	1										
	School attended in Year 3		4	4	4	4						
	School attended in Year 5		5	5	5	5						
	School attended in Year 7		5	5	5	5						
	School attended in Year 9		5	5	5	5						
	School attended in 2015		4		4	4						
	Repeated year level: primary	1										
	Repeated year level: lower secondary	1										
	Repeated year at school	1										
	Repeated year level: Year 11 or 12	1										
	Changed schools (primary)	1										
	Changed schools (secondary)	1										
	Changed schools/same school		1	1	1							
	Changed schools	1	1									
	Changed schools/study program	1										
	Changed schools: Month		1	1	1							
	Changed schools: Year		1	1	1							
	Current school level (derived variable)	1	1	1	1	1	1	1	1	1	1	1
	Changed study program	1										

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
Student achievement	Plausible value in maths	10										
	Plausible value in reading	10										
	Plausible value in science	10										
	Plausible value in evaluating and designing scientific enquiry	10										
	Plausible value in explaining phenomena scientifically	12										
	Plausible value in interpreting data and evidence scientifically	8										
	Plausible value in science knowledge: Content	10										
	Plausible value in science knowledge: Procedural & epistemic	10										
	Plausible value in science system: Earth & science	10										
	Plausible value in science system: Living	10										
	Plausible value in science system: Physical	10										
Time spent learning	Minutes per class	1										
	Minutes per week	3										
	Minutes of classes (week)	1										
	Number of classes (week)	1										

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
	Number of English classes (week)	1										
	Number of maths classes (week)	1										
	Number of science classes (week)	1										
	Out-of-school	8										
	Help with homework	7										
	Truancy	3										
Additional instruction	Out of school: Primary	2										
	Out of school: Secondary	1										
	Years of additional instruction	1										
	Number of subjects with additional instruction	1										
	Number of science subjects with additional instruction	1										
	Subjects covered	6										
	Content covered	4										
	Location of instruction	7										
	Type of instruction	23										
	Instructor	12										
	Number of hours (English)	1										
	Number of hours (mathematics)	1										

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
	Reasons for attending additional instruction (mathematics)	12										
	Reasons for not attending additional instruction (mathematics)	13										
	Number of hours (science)	1										
	Reasons for attending additional instruction (science)	12										
	Reasons for not attending additional instruction (science)	13										
	Number of hours (social sciences)	1										
	Number of hours (foreign language)	1										
	Number of hours (music)	1										
	Number of hours (visual arts)	1										
	Number of hours (performing arts)	1										
	Number of hours (sports)	1										
	Number of hours (other)	1										
	Number of hours (total)	1										
Perceptions about self and school	Life at school: Outsider	1										
	Life at school: Make friends	1										
	Life at school: Belonging	1										

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
	Life at school: Awkward	1										
	Life at school: I'm liked	1										
	Life at school: Feel lonely	1										
	Teachers: Treat me fairly	6										
	Belonging	1										
	Achievement motivation	9										
	Anxiety	6										
	Bullying	8										
Collaboration and teamwork	Enjoy cooperation	5										
	Value cooperation	5										
Student engagement	Pay attention in class		1									
	Interested in schoolwork		1									
	Check schoolwork		1									
	Follow rules		1									
	Non-compulsory study		1									
	Feel bored		1									
Use of computers	Age first used computers	1										

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
	Age first used digital devices	1										
	Age first used internet	1										
	At school	10										
	Resources: home	13										
	Resources: school	11										
	Entertainment	13										
	Frequency: Internet	3										
	ICT in social interaction	6										
	At home for school work	13										
	Attitudes	7										
	Self-efficacy	12										
Teaching and learning maths	Comparison: Structuredness (mathematics)	10										
	Comparison: Support (mathematics)	7										
	Comparison: Teacher-student relation (mathematics)	6										
Teaching and learning science	Course: Applied sciences and technology	2										
	Course: Biology	2										
	Course: Chemistry	2										

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
	Course: Earth and space	2										
	Course: General, integrated, or comprehensive science	2										
	Course: Physics	2										
	Teaching and learning	10										
	Adaption of instruction	3										
	Adaption of instruction (index)	1										
	Disciplinary climate	5										
	Disciplinary climate (index)	1										
	Feedback	5										
	Feedback (index)	1										
	Motivation	4										
	Motivation (index)	1										
	Strategies	4										
	Strategies (index)	1										
	Teacher support	5										
	Teacher support (index)	1										
	Comparison: Structuredness (science)	10										
	Comparison: Support (science)	7										
	Comparison: Teacher-student relation (science)	6										

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
Views on science	Science activities	9										
	Science activities (index)	1										
	Science enjoyment	5										
	Science enjoyment (index)	1										
	Science interest	5										
	Science interest (index)	1										
	Science self-efficacy	8										
	Science self-efficacy (index)	1										
	Science: Epistemological beliefs	6										
	Science: Epistemological beliefs (index)	1										
Views on the environment	Future	7										
	Future (index)	1										
	Informed	7										
	Informed (index)	1										
Subjects/ courses	School subject information		12	10	10							
	Subject choice	3										
	Science subject	1										

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
	Health and physical education	1										
	Number of subjects			1	1							
Subjects/ courses: VET	VET subjects			1	1							
	VET subjects at school		1									
	VET subjects at TAFE		1									
	VET subjects at ACE		1									
	VET subjects at other training organisation		1									
	TAFE subjects		1									
	TAFE subjects part of apprenticeship/traineeship		1									
	None of the above		1									
	Prefer not to say		1									
	Awarded VET certificate		1									
Subjects/ courses: IB	School subject information		6	6	6							
Study plans	Complete Year 12		1	1	1							
	Expected education (ISCED)	1										

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
Careers advice	Talked with family			1	1							
	Talked with friends			1	1							
	Talked to a careers guidance officer		1	1								
	Talked to a teacher/careers guidance officer			1	1							
	Talked with person in desired job		1	2	1							
	Questionnaire		1	2	1							
	Visited workplace		1	2	1							
	University information session		1	2	1							
	TAFE information session		1	1	1							
	Used internet site/computer program		1	2	1							
	Taught to develop formal plan		1	2	1							
	None		1									
	Prefer not to say		1									
	Source: Other		1	2	1							
	Most useful careers advice			1	1							
Work experience	Work experience		1	2	1							
Workplace learning (VET)	Workplace learning		1									

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
	Number of days (actual)		2									
	Number of hours (actual)		1									
Qualifications and results	Awarded certificate			1	1	1	1					
	Received (state-specific) score			1	1	1	2					
	Result			3	3	3	3					
	Highest school level completed (derived variable)	1	1	1	1	1	1	1	1	1	1	1
	Completed Year 12 or certificate II or higher (derived variable)	1	1	1	1	1	1	1	1	1	1	1
	Completed Year 12 or certificate III or higher (derived variable)	1	1	1	1	1	1	1	1	1	1	1

Topic map 4: Education – School transition

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
Post-school plans	Student plans		1	1								
	Student plans (immediate)			1	1							
	Student plans (eventual)			1	1							
	Parents' plans		1	1								
	Friends' plans		1	1								
	Study plans		1	1								
	Study plans: Type		1									
	Likelihood of completion		1									
	Barriers		10									
School leavers	Left school before completing Year 12			1	1	1						
	Month left school		1	1	1	1	1					
	Year left school		1	1	2	2	2					
	Year level left school			2	1	1	1					
	Reason: Have job/apprenticeship		1	1	1	1	1					
	Reason: Wanted job/apprenticeship/traineeship		1	1	1	1	1					
	Reason: Not doing very well at school		1	1	1	1	1					

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
	Reason: Study/training not available at school		1	1	1	1	1					
	Reason: Didn't like school		1	1	1	1	1					
	Reason: Financially difficult		1	1	1	1	1					
	Reason: Teachers		1	1	1	1	1					
	Reason: Earn own money		1	1	1	1	1					
	Reason: Parents		1	1	1	1	1					
	Reason: Year 12 wouldn't help get a job		1	1	1	1	1					
	Reason: Year 12 wouldn't help with further study/training		1	1	1	1	1					
	Reason: Caring responsibilities		1	1	1	1	1					
	Reason: Travel/time away		1	1	1	1	1					
	Reason: Parental responsibility		1	1	1	1	1					
	Reason: Concentrate on other areas of life		1	1	1	1	1					
	Reason: Other		1	1	1	1	1					
	Reason: Prefer not to say		1									
	Reason: Main reason		1	1	1	1	1					
	Feelings about having left school		1									
	Main reason returned to school			1	1	1	1					
	Month returned to school			1	1	1	1					
	Year returned to school			1	1	1	1					

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
Main activity	Main activity			1	1	1	1	1	1	1	1	1

Topic map 5: Education – Post-school

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
Study	Study status (at last interview)			1	1	1	1	1	1	1	1	1
	Still studying		1	5	5	5	5	5	5	5	5	5
	Study completed/withdrawn/deferred/changed		3									
	Confirmation of study			1	1	1	1	1	1	1	1	1
	Confirmation of deferred study			1	1	1	1	2	2	2	2	2
	Resumption of deferred study			1	1	1	1	1	1	1	1	1
	Commenced study		5	1	1	1	1	1	1	1	1	1
	Study type		3	1	1	1	1	1	1	1	1	1
	Qualification		3	2	2	2	2	2	2	2	2	2
	Qualification (at last interview)			1	1	1	1	1	1	1	1	1
	Main area of study		6	2	2	2	2	2	2	2	2	2
	Institution		2	4	4	4	4	4	4	4	4	4
	Full-time or part-time study		1									
	Month started study		3	1	1	1	1	1	1	1	1	1
	Year started study		3	1	1	1	1	1	1	1	1	1
	Contact hours		1									
	Likelihood of completion		2									
	Applied for university place			1	1	1						

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
	First preference			1	1	1						
	First preference: Offered place			1	1	1						
	First preference: Institution			1	1	1						
	First preference: Accepted place			1	1	1						
	First preference: Reason did not take up place (taking break/holiday/travelling)			1	1	1						
	First preference: Reason did not take up place (required leaving home)			1	1	1						
	First preference: Reason did not take up place (need Youth Allowance)			1	1	1						
	First preference: Reason did not take up place (considering options)			1	1	1						
	First preference: Reason did not take up place (course costs)			1	1	1						
	First preference: Reason did not take up place (financial)			1	1	1						
	First preference: Reason did not take up place (prefer to work)			1	1	1						
	First preference: Reason did not take up place (prefer to study at TAFE)			1	1	1						
	First preference: Reason did not take up place (other)			1	1	1						
	First preference: Reason did not take up place (main reason)			1	1	1						

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
	University: Offered place			1	1	1						
	University: Institution			1	1	1						
	University: Accepted place			1	1	1						
	University: Reason did not take up place (taking break/holiday/travel)			1	1	1						
	University: Reason did not take up place (required leaving home)			1	1	1						
	University: Reason did not take up place (need Youth Allowance)			1	1	1						
	University: Reason did not take up place (considering options)			1	1	1						
	University: Reason did not take up place (course costs)			1	1	1						
	University: Reason did not take up place (financial)			1	1	1						
	University: Reason did not take up place (prefer to work)			1	1	1						
	University: Reason did not take up place (prefer to study at TAFE)			1	1	1						
	University: Reason did not take up place (other)			1	1	1						
	University: Reason did not take up place (main reason)			1	1	1						

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
	Study status in bachelor degree or higher (derived variable)	1	1	1	1	1	1	1	1	1	1	1
	Study status in VET (derived variable)	1	1	1	1	1	1	1	1	1	1	1
	Number of courses/apprenticeships/traineeships		1									
Current study	Study type		1	1	1	1	1	1	1	1	1	1
	Qualification			1	1	1	1	1	1	1	1	1
	Main area of study			2	2	2	2	2	2	2	2	2
	Institution			10	10	10	10	10	10	10	10	10
	Contact hours			2	2	2	2	2	2	2	2	2
	Full-time or part-time study			4	4	4	4	4	4	4	4	4
	Month started study			2	2	2	2	2	2	2	2	2
	Year started study			2	2	2	2	2	2	2	2	2
	Current qualification level (derived variable)	1	1	1	1	1	1	1	1	1	1	1
	Full-time or part-time study status (derived variable)	1	1	1	1	1	1	1	1	1	1	1
Past study	Study completed/withdrawn/deferred/changed			2	2	2	2	2	2	2	2	2
	Main area of study			2	2	2	2	2	2	2	2	2

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
	Institution			2	2	2	2	2	2	2	2	2
	Contact hours			2	2	2	2	2	2	2	2	2
	Full-time or part-time study			3	3	3	3	3	3	3	3	3
	Month stopped study		3	3	3	3	3	3	3	3	3	3
	Year stopped study		3	3	3	3	3	3	3	3	3	3
	Completed study		1									
	Highest qualification level completed (derived variable)	1	1	1	1	1	1	1	1	1	1	1
Apprenticeships/ traineeships	Still studying			2	2	2	2	2	2	2	2	2
	Confirmation of apprenticeship/traineeship			1	1	1	1	1	1	1	1	1
	Confirmation of deferred apprenticeship/traineeship							1	1	1	1	1
	Resumption of deferred apprenticeship/traineeship							1	1	1	1	1
	Month started study			1	1	1	1	1	1	1	1	1
	Year started study			1	1	1	1	1	1	1	1	1
	Full-time or part-time study		1									
	Qualification			1	1	1	1	1	1	1	1	1
	Main area of study			1	1	1	1	1	1	1	1	1

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
	Employer type		1	1	1	1	1	1	1	1	1	1
	Classes/off-the-job training at TAFE			1	1	1	1	1	1	1	1	1
	Provider of off-the-job training		2	1	1	1	1	1	1	1	1	1
	Status in apprenticeship/traineeship (derived variable)	1	1	1	1	1	1	1	1	1	1	1
Current apprenticeships/ traineeships	Full-time or part-time study			2	2	2	2	2	2	2	2	2
	Employer type		1	2	2	2	2	2	2	2	2	2
	Classes/off-the-job training at TAFE			1	1	1	1	1	1	1	1	1
	Provider of off-the-job training			1	1	1	1	1	1	1	1	1
Past apprenticeships/ traineeships	Study completed/withdrawn/time out/other		1	1	1	1	1	1	1	1	1	1
	Employer type			1	1	1	1	1	1	1	1	1
	Reason apprenticeship/traineeship ended			1	1	1	1	1	1	1	1	1
	Month stopped study			1	1	1	1	1	1	1	1	1
	Year stopped study			1	1	1	1	1	1	1	1	1
Deferred/ withdrew from study	Reason: Problems juggling study and work commitments		1	1	1	1	1	1	1	1	1	1

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
	Reason: Wanted job/apprenticeship/traineeship		1	1	1	1	1	1	1	1	1	1
	Reason: Financially difficult		1	1	1	1	1	1	1	1	1	1
	Reason: Lost interest		1	1	1	1	1	1	1	1	1	1
	Reason: Never wanted to study		1	1	1	1	1	1	1	1	1	1
	Reason: Course was not what you wanted		1	1	1	1	1	1	1	1	1	1
	Reason: Wouldn't have led to good job/career		1	1	1	1	1	1	1	1	1	1
	Reason: Poor results		1	1	1	1	1	1	1	1	1	1
	Reason: Study load		1	1	1	1	1	1	1	1	1	1
	Reason: Never intended to complete the course		1	1	1	1	1	1	1	1	1	1
	Reason: Because of problems with access or transport		1	1	1	1	1	1	1	1	1	1
	Reason: Because of health or personal reasons		1	1	1	1	1	1	1	1	1	1
	Reason: Main reason		1	1	1	1	1	1	1	1	1	1
	Reason: Other		1	1	1	1	1	1	1	1	1	1
	Reason: Parental responsibility		1		1	1	1	1	1	1	1	1
	Reason: Prefer not to say		1									
	Reason: Problems juggling study and family commitments		1	1	1	1	1	1	1	1	1	1

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
	Study completed/withdrawn/deferred/changed		1									
Changed institutions	Same institution			7	7	7	7	7	7	7	7	7
Changed course	Same course		1	2	2	2	2	2	2	2	2	2
	Reason: Completed first course		1									
	Reason: Course costs		1	2	2	2	2	2	2	2	2	2
	Reason: The first course was a pre-requisite for the second course		1	2	2	2	2	2	2	2	2	2
	Reason: Didn't like the first course		1	2	2	2	2	2	2	2	2	2
	Reason: Turned out to be not what you wanted		1	2	2	2	2	2	2	2	2	2
	Reason: Better career prospects		1	2	2	2	2	2	2	2	2	2
	Reason: Poor results		1	2	2	2	2	2	2	2	2	2
	Reason: Study load		1	2	2	2	2	2	2	2	2	2
	Reason: Preferred to do second course		1	2	2	2	2	2	2	2	2	2
	Reason: Because of health or personal reasons		1	2	2	2	2	2	2	2	2	2
	Reason: Flexibility		1	2	2	2	2	2	2	2	2	2
	Reason: Issues with instructor		1	2	2	2	2	2	2	2	2	2
	Reason: Did not fit in		1	2	2	2	2	2	2	2	2	2

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
	Reason: Lonely/socially isolated		1	2	2	2	2	2	2	2	2	2
	Reason: Other		1	2	2	2	2	2	2	2	2	2
	Reason: Prefer not to say		1									
	Reason: Main reason		1	2	2	2	2	2	2	2	2	2
	Contact hours			2	2	2	2	2	2	2	2	2
Changed/ left employer	Same employer		1	2	2	2	2	2	2	2	2	2
	Circumstances of changing employer		1	1	1	1	1	1	1	1	1	1
	Reason: Someone offered you a better job		1	1	2	2	2	2	2	2	2	2
	Reason: Boss/other people at work		1	1	2	2	2	2	2	2	2	2
	Reason: Pay		1	1	2	2	2	2	2	2	2	2
	Reason: On-the-job training		1	1	2	2	2	2	2	2	2	2
	Reason: Feeling rushed and pressed for time		1									
	Reason: Because of problems with travelling or transport		1	1	2	2	2	2	2	2	2	2
	Reason: Problems juggling work and family commitments		1									
	Reason: Parental responsibility		1									
	Reason: Because of health or personal reasons		1	1	2	2	2	2	2	2	2	2

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
	Reason: Didn't use skills/abilities		1									
	Reason: Didn't have necessary skills/abilities		1									
	Reason: Other		1	1	2	2	2	2	2	2	2	2
	Reason: Prefer not to say		1									
	Reason: Main reason		1	1	3	3	3	3	3	3	3	3
	Month changed employer		1	2	2	2	2	2	2	2	2	2
	Year changed employer		1	2	2	2	2	2	2	2	2	2
Changed/ stopped apprenticeship/ traineeship	Reason: Offered better job		1	1	1	1	1	1	1	1	1	1
	Reason: Pay		1	1	1	1	1	1	1	1	1	1
	Reason: Job prospects		1	1	1	1	1	1	1	1	1	1
	Reason: Type of work		1	1	1	1	1	1	1	1	1	1
	Reason: Boss/other people at work		1	1	1	1	1	1	1	1	1	1
	Reason: On-the-job training		1	1	1	1	1	1	1	1	1	1
	Reason: Off-the-job training		1	1	1	1	1	1	1	1	1	1
	Reason: Study/training too difficult		1	1	1	1	1	1	1	1	1	1
	Reason: Feeling rushed and pressed for time		1	1	1	1	1	1	1	1	1	1

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
	Reason: Because of problems with travelling or transport		1	1	1	1	1	1	1	1	1	1
	Reason: Problems juggling work and family commitments		1	1	1	1	1	1	1	1	1	1
	Reason: Parental responsibility		1	1	1	1	1	1	1	1	1	1
	Reason: Because of health or personal reasons		1	1	1	1	1	1	1	1	1	1
	Reason: Didn't have necessary skills/abilities		1									
	Reason: Didn't use skills/abilities		1									
	Reason: Other		1	1	1	1	1	1	1	1	1	1
	Reason: Prefer not to say		1									
	Reason: Main reason		2	1	1	1	1	1	1	1	1	1
Student engagement	Feedback		1									
	Frequency of class presentations		1									
	Frequency of group work		1									
	Frequency of discussions about coursework		1									
	Frequency of discussions with instructor		1									
	Frequency of extracurricular activities		1									

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
Satisfaction with study	Overall satisfaction		1									
	Improved career prospects		1									
	Helped make contacts		1									
Careers advice	Careers guidance officer			1	1	1	1	1	1			
	Questionnaire			1	1	1	1	1	1			
	Job application assistance			1	1	1	1	1	1			
	Information about further study			1	1	1	1	1	1			
	Online tool			1	1	1	1	1	1			
	None			1	1	1	1	1	1			
Government payments and income	Sources of income: Study payment (Youth Allowance/ABSTUDY/AUSTUDY)			1	1	1	1	1	1	1	1	1
	Amount of study payment received (fortnight)			1	1	1	1	1	1	1	1	1
	Course fees: None			1	1							
	Course fees: Respondent			1	1							
	Course fees: Parents/family			1	1							
	Course fees: Employer			1	1							
	Course fees: Government			1	1							
	Course fees: Other			1	1							

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
	Commonwealth supported (HECS)			1	1	1						
	Commonwealth supported (HECS)/full-fee paying			1	1	1						
	Full-fee paying: FEE-HELP			1	1	1						
	Full-fee paying: Up-front			1	1	1						
	Full-fee paying: Payment scheme			1	1	1						
	Full-fee paying: Employer			1	1	1						
	Full-fee paying: Scholarship			1	1	1						
Qualifications completed	Certificate I											1
	Certificate II											1
	Certificate III											1
	Certificate IV											1
	Certificate level unknown											1
	Traineeship											1
	Apprenticeship											1
	VET/TAFE diploma											1
	VET/TAFE advanced diploma/associate degree											1
	University diploma											1
	University advanced diploma/associate degree											1

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
	Bachelor degree											1
	Bachelor degree with Honours											1
	VET/TAFE graduate diploma/graduate certificate											1
	University graduate diploma/graduate certificate											1
	Masters degree											1
	PhD or other doctorate											1
	Other											1
	None											1
	Don't know											1

Topic map 6: Employment – Current

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
Employment characteristics	Work in job/business/farm		1	2	1	1	1	1	1	1	1	1
	Still have job		3									
	Still have job (reported at last interview)			1	1	1	1	1	1	1	1	1
	Away from job			1	1	1	1	1	1	1	1	1
	School holiday/seasonal job			2	2	2	2	1	1	1	1	1
	More than one job			1	1	1	1	1	1	1	1	1
	Number of other jobs had			1	1	1	1	1	1	1	1	1
	Wages/salary/self-employed		1	1	1	1	1	1	1	1	1	1
	Kind of work (ANZSCO)		2	1	1	1	1	1	1	1	1	1
	Employer's main kind of business (ANZSIC)		2	1	1	1	1	1	1	1	1	1
	Employment status		1	1	1	1	1	1	1	1	1	1
	Labour force status (derived variable)	1	1	1	1	1	1	1	1	1	1	1
	Permanent or casual employment (derived variable)	1	1	1	1	1	1	1	1	1	1	1
	Occupation (derived variable)	1	1	1	1	1	1	1	1	1	1	1
	In full-time employment or full-time education (derived variable)	1	1	1	1	1	1	1	1	1	1	1

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
	Job mobility during last year (derived variable)	1	1	1	1	1	1	1	1	1	1	1
	Any spell of unemployment during the year (derived variable)	1	1	1	1	1	1	1	1	1	1	1
Time worked	Hours worked per week (present job)			1	1	1		1	1	1	1	1
	Hours worked per week: prior to COVID-19 (present job)						1					
	Hours worked per week: current (present job)						1					
	Hours worked per week (main job if more than one)		4	1	1	1		1	1	1	1	1
	Hours worked per week: prior to COVID-19 (main job if more than one)						1					
	Hours worked per week: current (main job if more than one)						1					
	Hours worked per week (all jobs if more than one)		2	1	1	1		1	1	1	1	1
	Hours worked per week: prior to COVID-19 (all jobs if more than one)						1					
	Hours worked per week: current (all jobs if more than one)						1					
	Hours worked per week (job reported at last interview)			1	1	1		1	1	1	1	1

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
	Hours worked per week: prior to COVID-19 (job reported at last interview)						1					
	Hours worked per week: current (job reported at last interview)						1					
	Hours worked per week (weekdays)			1								
	Hours worked per week (weekend)			1								
	Prefer different hours		1	1	1	1	1	1	1	1	1	1
	Preferred weekly hours		2	1	1	1	1	1	1	1	1	1
	Months worked			22	22	22	21	21	21	21	21	21
	Full-time or part-time employment status (derived variable)	1	1	1	1	1	1	1	1	1	1	1
	Average weekly hours worked (derived variable)	1	1	1	1	1	1	1	1	1	1	1
Wages and benefits	Frequency of pay		2	1	1	1	1	1	1	1	1	1
	Pay type			1	1	1	1	1	1	1	1	1
	Gross pay		2									
	Take home pay		2									
	Pay (unknown tax type)			1	1	1	1	1	1	1	1	1
	Tax type			1	1	1	1	1	1	1	1	1
	Annual/sick leave			1	1	1	1	1	1	1	1	1
	Average weekly pay (derived variable)	1	1	1	1	1	1	1	1	1	1	1

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
	Average hourly pay (derived variable)	1	1	1	1	1	1	1	1	1	1	1
Starting work	Month began job		1	1	1	1	1	1	1	1	1	1
	Year began job		1	1	1	1	1	1	1	1	1	1
	How found job		11	1	1	1	1	1	1	1	1	1
Looking for work	Looking for work			1	1	1	1	1	1	1	1	1
	Looking for work: Additional or to change jobs			1	1	1	1	1	1	1	1	1
	Looking for work: In the last 4 weeks		1	1	1	1	1	1	1	1	1	1
	Looking for full-time or part-time work		1									
Job satisfaction	Like job as career		1			1	1	1	1	1	1	1
	Kind of work		1	1	1	1	1	1	1	1	1	1
	Utilise skills/experience			1	1	1	1	1	1	1	1	1
	Immediate boss/supervisor		1	1	1	1	1	1	1	1	1	1
	Other people			1	1	1	1	1	1	1	1	1
	Pay		1	1	1	1	1	1	1	1	1	1
	Opportunities for training		1	1	1	1	1	1	1	1	1	1
	Tasks assigned					1	1	1	1	1	1	1
	Recognition		1	1	1	1	1	1	1	1	1	1

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
	Opportunities for promotion		1	1	1	1	1	1	1	1	1	1
	Flexibility		1	1	1	1	1	1	1	1	1	1
	Hours worked		1									
	Overall satisfaction		1	1	1	1	1	1	1	1	1	1

Topic map 7: Employment – Job history and training

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
Employment characteristics	Employed last year		1									
	Number of jobs		1									
	Work in job/business/farm (at last interview)			1	1	1	1	1	1	1	1	1
	Re-definition of second job as main job			1	1	1	1	1	1	1	1	1
	Kind of work (ANZSCO)		2									
	Kind of work: Other/second job (ANZSCO)		2	1	1	1	1	1	1	1	1	1
	Kind of work: Other/third job (ANZSCO)		2									
	Employer's main kind of business: Other/second job (ANZSIC)			1	1	1	1	1	1	1	1	1
	Wages/salary/self-employed: Other/second job			1	1	1	1	1	1	1	1	1
Gig work	Earned money by taking on gig jobs			1	1	1		1	1	1	1	1
	Already reported gig job			1	1	1		1	1	1	1	1
	Gig jobs or tasks			5	5	5		6	7	7	7	7
	Importance: Income earned			1	1	1		1	1	1	1	1
	Reasons for taking on gig jobs			6	6	6		6	6	6	6	6

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
	Pay			1	1	1		1	1	1	1	1
	Frequency of pay			1	1	1		1	1	1	1	1
Time worked	Hours worked per week: Other/second job			1	1	1		1	1	1	1	1
	Hours worked per week: Other/second job: prior to COVID-19						1					
	Hours worked per week: Other/second job: current						1					
	Months worked		63									
Wages and benefits	Frequency of pay: Other/second job			1	1	1	1	1	1	1	1	1
	Pay type: Other/second job			1	1	1	1	1	1	1	1	1
	Pay (unknown tax type): Other/second job			1	1	1	1	1	1	1	1	1
	Tax type: Other/second job			1	1	1	1	1	1	1	1	1
Leaving work	Main reason left job		3	2	2	2	2	2	2	2	2	2
	Month left/finished job			2	2	2	2	2	2	2	2	2
	Year left/finished job			2	2	2	2	2	2	2	2	2

Topic map 8: Employment – Seeking employment

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
Looking for work	Looking for work		1									
	Prefer full-time work		1									
	Available for work last week		1									
Job search activity	Looked for work			1	1	1	1	1	1	1	1	1
	Months looking for work		21	21	21	21	20	20	20	20	20	20
	Preferred weekly hours			1	1	1	1	1	1	1	1	1
	Registered with Centrelink/jobactive provider			1	1	1	1	1	1	1	1	1
	Checked/registered with any other employment agency			1	1	1	1	1	1	1	1	1
	Checked/registered with Centrelink/another employment agency		1									
	Looked at advertisements in newspaper/on noticeboards/on the internet		1	1	1	1	1	1	1	1	1	1
	Answered advertisements in newspaper/on noticeboards/on the internet		1	1	1	1	1	1	1	1	1	1
	Contacted friends or relatives		1	1	1	1	1	1	1	1	1	1

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
	Written/phoned/approached an employer about a job		1	1	1	1	1	1	1	1	1	1
	Had interview with an employer		1	1	1	1	1	1	1	1	1	1
	Asked school or another organisation for advice		1	1	1	1	1	1	1	1	1	1
	Advertised/tendered for work		1	1	1	1	1	1	1	1	1	1
	Taken steps to purchase or start your own business			1	1	1	1	1	1	1	1	1
	Refused a job offer			1	1	1	1	1	1	1	1	1
	Reason for refusing job offer			1	1	1	1	1	1	1	1	1
Problems looking for work	Health problems or some disability		1	1	1	1	1	1	1	1	1	1
	Problems with childcare		1	1	1	1	1	1	1	1	1	1
	Don't have suitable transport		1	1	1	1	1	1	1	1	1	1
	Not enough of the right kind of education		1	1	1	1	1	1	1	1	1	1
	Don't have enough work experience		1	1	1	1	1	1	1	1	1	1
	Not enough jobs available		1	1	1	1	1	1	1	1	1	1
	Age, gender or other discrimination		1	1	1	1	1	1	1	1	1	1
	Need better reading and writing skills		1	1	1	1	1	1	1	1	1	1
	Don't have good interview skills		1	1	1	1	1	1	1	1	1	1
	Lack confidence		1	1	1	1	1	1	1	1	1	1

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
	Not good with numbers		1	1	1	1	1	1	1	1	1	1
	Poor language or communication skills		1	1	1	1	1	1	1	1	1	1
	Work hours not suitable		1	1	1	1	1	1	1	1	1	1
	None		1									
	Prefer not to say		1									

Topic map 9: Employment – Not in the labour force

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
Main activity	Main activity		1	1	1	1	1	1	1	1	1	1

Topic map 10: Social – Health, living arrangements and finance

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
Living arrangements	Type of accommodation			1	1	1	1	1	1	1	1	1
	Live with parents			1	1	1	1	1	1	1	1	1
	Number of (other) people in household		2	1	1	1	1	1	1	1	1	1
	Relationship to respondent		13	15	15	15	15	15	15	15	15	15
	Moved house		2	2	2	2	2	2	2	2	2	2
	Husband/wife/partner currently working				1	1						
	Husband/wife/partner other activity				1	1						
	Husband/wife/partner works full-time or part-time				1	1						
	Husband/wife/partner current occupation (ANZSCO)				1	1						
	Housing affected by COVID-19						4					
	Living with parent(s) (derived variable)	1	1	1	1	1	1	1	1	1	1	1
	Living in own home (derived variable)	1	1	1	1	1	1	1	1	1	1	1
	Number of dependent children (derived variable)	1	1	1	1	1	1	1	1	1	1	1
Household possessions	Desk	1		1								
	Own room	1		1								
	Quiet study place	1		1								

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
	Computer	1										
	Software	1		1								
	Internet	1		1								
	Literature	1		1								
	Poetry	1		1								
	Art	1		1								
	Books on art, music, or design	1		1								
	Textbooks	1		1								
	Technical reference books	1		1								
	Dictionary	1		1								
	Educational resources	1										
	Espresso machine	1		1								
	Home gym/gym membership	1		1								
	Solar panels (on roof)	1		1								
	Cultural possessions	1										
	Number of mobile phones	1		1								
	Number of TVs	1		1								
	Number of computers	2		2								
	Number of e-book readers	1		1								
	Number of cars	1		1								
	Number of rooms with bath/shower	1		1								

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
	Number of books	1		1								
	Number of musical instruments	1		1								
Children	Number of children				1	1	1	1	1	1	1	1
	Age of children				5	5	5	5	5	5	5	5
Marriage	Marital status				1	1	1	1	1	1	1	1
	Marital status (at last interview)					1	1	1	1	1	1	1
	Month married				1	1	1	1	1	1	1	1
	Year married				1	1	1	1	1	1	1	1
	Marital status (derived variable)	1	1	1	1	1	1	1	1	1	1	1
Disability and health	General health				1	1	1	1	1	1	1	1
	Disability/health problem limits amount or type of work					1		1	1		1	1
	Disability/health problem(s): Arms/legs/hands/feet/back or neck					1		1	1		1	1
	Disability/health problem(s): Seeing					1		1	1		1	1
	Disability/health problem(s): Hearing					1		1	1		1	1
	Disability/health problem(s): Skin/allergies					1		1	1		1	1

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
	Disability/health problem(s): Breathing/asthma/bronchitis					1		1	1		1	1
	Disability/health problem(s): Heart/blood pressure					1		1	1		1	1
	Disability/health problem(s): Stomach/liver/kidney/digestive problem(s)					1		1	1		1	1
	Disability/health problem(s): Diabetes					1		1	1		1	1
	Disability/health problem(s): Mental health, nervous or emotional condition					1		1	1		1	1
	Disability/health problem(s): Epilepsy					1		1	1		1	1
	Disability/health problem(s): Dyslexia/other learning problem(s)					1		1	1		1	1
	Disability/health problem(s): Chronic fatigue/post-viral syndromes					1		1	1		1	1
	Disability/health problem(s): Other problem(s) or disabilities					1		1	1		1	1
	Disability/health problem(s): Prefer not to say					1		1	1		1	1
	You felt nervous						1	1	1			1
	You felt hopeless						1	1	1			1
	You felt restless or fidgety						1	1	1			1
	You felt that everything was an effort						1	1	1			1
	You felt so sad that nothing would cheer you up						1	1	1			1

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
	You felt worthless						1	1	1			1
	You felt lonely						1	1	1			
	Eat breakfast	1										
	Eat dinner	1										
	Parental support	5										
Government payments and income	Sources of income: Government payment		1									
	Sources of income: Other		1									
	Sources of income: None		1									
	Sources of income: Prefer not to say		1									
	Government payment			1	1	1	1	1	1	1	1	1
	Youth Allowance/ABSTUDY			1	1	1						
	Youth Allowance/JobSeeker Payment						1	1	1	1	1	1
	Living Away from Home Allowance			1	1	1						
	Disability Support Pension			1	1	1						
	Rent Assistance			1	1	1	1	1	1	1	1	1
	JobKeeper Payment						1					
	Other government payment			1	1	1	1	1	1	1	1	1
	Amount received: Youth Allowance/ABSTUDY			1	1	1						

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
	Amount received: Youth Allowance/JobSeeker Payment						1	1	1	1	1	1
	Amount received: Living Away from Home Allowance			1	1	1						
	Amount received: Disability Support Pension			1	1	1						
	Amount received: Rent Assistance			1	1	1	1	1	1	1	1	1
	Amount received: Other government payment		2	1	1	1	1	1	1	1	1	1
	Amount received: Other		2									
	Frequency of receiving Youth Allowance/ABSTUDY			1	1	1						
	Frequency of receiving Youth Allowance/JobSeeker Payment						1	1	1	1	1	1
	Frequency of receiving: Living Away from Home Allowance			1	1	1						
	Frequency of receiving Disability Support Pension			1	1	1						
	Frequency of receiving Rent Assistance			1	1	1	1	1	1	1	1	1
	Frequency of receiving other government payment			1	1	1	1	1	1	1	1	1
	Frequency of payment		2									
Housing payments	Amount of housing payments				1	1	1	1	1	1	1	1

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
	Frequency of housing payments				1	1	1	1	1	1	1	1
Finance	Use of credit card					1	1	1	1	1	1	1
	Use of Buy Now Pay Later service						1	1	1	1	1	1
	Frequency of clearing debt on credit card					1	1	1	1	1	1	1
	Missed a Buy Now Pay Later payment						1	1	1	1	1	1
	Able to save money					1	1	1	1	1	1	1
	Frequency of saving money					1	1	1	1	1	1	1
	Managing financially					1	1	1	1	1	1	1
	Shortage of money: Sold something because you needed money					1	1	1	1	1	1	1
	Shortage of money: Went without meals					1	1	1	1	1	1	1
	Shortage of money: Had to ask family or friends for money					1	1	1	1	1	1	1
	Shortage of money: Had to borrow money					1	1	1	1	1	1	1
	Shortage of money: Didn't get medicines or go to a doctor					1	1	1	1	1	1	1
	Shortage of money: Couldn't buy text books or other study materials					1	1	1	1	1	1	1
	Shortage of money: Couldn't buy other things you needed					1	1	1	1	1	1	1

Minor topic area	Data element	Wave/year										
		1/2015	2/2016	3/2017	4/2018	5/2019	6/2020	7/2021	8/2022	9/2023	10/2024	11/2025
	Shortage of money: Couldn't pay electricity, gas or telephone bills					1	1	1	1	1	1	1
	Shortage of money: Couldn't pay mortgage/rent on time					1	1	1	1	1	1	1
	Shortage of money: Couldn't afford to heat your home					1	1	1	1	1	1	1
Transience/ homelessness	Experienced homelessness			1	2	2		2	2		2	
	Circumstance due to homelessness			13	13	13		13	13		13	
	Number of homelessness experience(s)			1	1	1		1	1		1	
	Transition to a permanent place to live			1	1	1		1	1		1	
	Month/year transitioned to a permanent place to live			1	1	1						
	Duration of homelessness			1	1	1		1	1		1	
	Services sought assistance from			14	14	14		14	14		14	
	Reasons for not seeking assistance from services			9	9	9		9	9		9	
	Helpfulness of services			12	12	12		12	12		12	
Social support	Access to social support			1	1	1	1		1		1	1
	Type of social support			9	10	11	11		11		11	11

Topic map 11: Social – General attitudes

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
Characteristics about yourself	Soft skills		14	28		14				14		14
Personality	Agreeableness		2		2						2	
	Conscientiousness		2		2						2	
	Extroversion		2		2						2	
	Openness		2		2						2	
	Neuroticism				2						2	
Leisure	Play computer or video games	2										
	Play sport or do exercise	4			1	1	1	1	1	1	1	1
	Use the Internet	2										
	Home duties/care for others	2										
	Meet friends or talk to friends on phone	2										
	Read books, newspapers or magazines	2										
	Watch TV/DVD/Video	2										
	Work for pay	2										
Relationships	Talk to parents	2										

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
Life satisfaction	The work you do				1	1	1	1	1	1	1	1
	What you do in your spare time				1	1	1	1	1	1	1	1
	How you get on with people				1	1	1	1	1	1	1	1
	The money you get each week				1	1	1	1	1	1	1	1
	Your social life				1	1	1	1	1	1	1	1
	Your independence				1	1	1	1	1	1	1	1
	Your career prospects				1	1	1	1	1	1	1	1
	Your future				1	1	1	1	1	1	1	1
	Your life at home				1	1	1	1	1	1	1	1
	Your standard of living				1	1	1	1	1	1	1	1
	The way the country is run							1	1	1	1	1
	The state of the economy							1	1	1	1	1
	Where you live				1	1	1	1	1	1	1	1
	Your life as a whole	1	1	1	2	2	2	2	2	2	2	2
	Things in life are worthwhile		1									
Job aspirations and expectations	Expect to have job at age 30		1	1		1			1			1
	Type of job expect at age 30 (ISCO)	2										
	Type of job expect at age 30 (ANZSCO)		1			1			1			1

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
	Wages/salary/self-employed								1			1
	Main activity					1			1			1
Aspirations	Personal goal						1					1
Volunteer	Canvassing/campaigning/fundraising				1	1		1			1	
	Unpaid member of board or committee				1	1		1			1	
	Provide information				1	1		1			1	
	Help organise activities				1	1		1			1	
	Coaching/teaching				1	1		1			1	
	Collect, serve or deliver food				1	1		1			1	
	Provide health care/support/counselling				1	1		1			1	
	Other volunteer activities				1	1		1			1	
	Volunteer		1	1								
	Website/social media/other online support					1		1			1	
	No volunteer activities				1	1		1			1	
	Frequency of voluntary work					1		1			1	
	Hours spent doing unpaid/volunteer work					1		1			1	
Caring for others	Cared for others		1	1	1	1	1		1	1	1	1

		Wave/year										
Minor topic area	**Data element**	**1/2015**	**2/2016**	**3/2017**	**4/2018**	**5/2019**	**6/2020**	**7/2021**	**8/2022**	**9/2023**	**10/2024**	**11/2025**
	Who cared for		7	6	6	6	6		6	6	6	6
	Hours spent caring for others		2	1	1	1	1		1	1	1	1
	Months spent caring for others		21	21	21	21	20		20	20	20	20
Respondent engagement	Survey participation					1						

References

ABS (Australian Bureau of Statistics) 2001, *Australian Standard Classification of Education (ASCED)*, cat.no.1272.0, Canberra.

——2006a, *Australian and New Zealand Standard Classification of Occupations*, 1st edn, cat.no.1220.0, ABS, Canberra.

——2006b, *Australian and New Zealand Standard Industrial Classification,* cat.no.1292.0, ABS, Canberra.

ACER (Australian Council for Educational Research) 1997, *Overview of the Longitudinal Surveys of Australian Youth program*, LSAY technical report no.2, Camberwell, viewed 9 November 2017, <http://www.lsay.edu.au/publications/1934.html>.

Chew, E, Parvazian, S & Semo, R 2023, *Understanding and using the linked LSAY-NAPLAN data: issues and considerations*, NCVER, Adelaide

Lim, P 2011, *Weighting the LSAY Programme of International Student Assessment cohorts*, NCVER, Adelaide, viewed 2 November 2017, <https://www.lsay.edu.au/publications/search-for-lsay-publications/2429>.

OECD (Organisation for Economic Co-operation and Development) 2016, *PISA 2015 technical report*, OECD Publishing, Paris, viewed 9 November 2017, <http://www.oecd.org/pisa/data/2015-technical-report/>.

NCVER (National Centre for Vocational Education Research) 2015, *Future topic areas for the Longitudinal Surveys of Australian Youth — Y15 cohort*, NCVER, Adelaide, viewed 2 November 2017, <https://www.lsay.edu.au/publications/search-for-lsay-publications/2814>.

Thomson, S, De Bortoli, L & Underwood, C, 2017, PISA 2015: Reporting Australia's results, Australian Council for Educational Research, Camberwell viewed 9 October 2017, <http://research.acer.edu.au/ozpisa/22/>.

Wallis Market & Social Research, 2017, 'Longitudinal Surveys of Australian Youth Y15 2016 methodological report', unpublished.

——, 2018a, 'Longitudinal Surveys of Australian Youth Y15 2017 methodological report', unpublished.

——, 2018b, 'LSAY schools top-up activity: methodology and learnings', unpublished.

Watson, N, 2014. 'Evaluation of weighting methods to integrate a top-up sample with an ongoing longitudinal sample', *Survey Research Methods* vol 8 no 3, p. 195-208, <https://ojs.ub.uni-konstanz.de/srm/article/view/5818>.

Appendix A: Updates to the Y15 data file

The following table tracks updates made to the Y15 data file deposited with the Australian Data Archive. Users are encouraged to download the most recent version of the data file to ensure all updates are included.

Table A1 Summary of changes made to the Y15 data file

Wave/ year	Version	Date published	Variable	Variable name	Description	Number of records affected
Waves 1 to 11 (2015 to 2025)	V10	August 2026	*Wave 11 (2025) variables added to data file.*			All
			Modifications to variables			
			Months worked	LJD033B	Months worked variables updated for consistency with LJD033U: *worked during every month up to the current month*. Where respondents reported that they had (or had not) worked every month since their last interview, the relevant monthly indicators were updated.	1051
				LJD033C		1724
				LJD033D		1933
				LJD033E		2165
				LJD033F		2206
				LJD033G		2226
				LJD033H		2226
				LJD033I		2226
				LJD033J		2226
				LJD033K		2226
				LJD033L		2226
				LJD033M		2226
				LJD033N		2186
				LJD033O		994
				LJD033P		326
				LJD033Q		170
				LJD033R		66

Wave/ year	Version	Date published	Variable	Variable name	Description	Number of records affected
				LJD033S		31
				LJD033U	LJD033U: *worked during every month up to the current month* updated for consistency with the relevant months worked variables (LJD033B–LJD033S). Where the monthly indicators showed that respondents had (or had not) worked every month, LJD033U was updated.	42
			Months looking for work	LJF002A	Months looking for work variables updated for consistency with LJF002T: *looking for work during every month up to the current month*. Where respondents reported that they had (or had not) been looking for work every month since their last interview, the relevant monthly indicators were updated.	130
				LJF002B		65
				LJF002C		46
				LJF002D		12
				LJF002E		10
				LJF002T	LJF002T: *looking for work during every month up to the current month* updated for consistency with the relevant months looking for work variables (LJF002A–LJF002E). Where the monthly indicators showed that respondents had (or had not) been looking for work every month, LJF002T was updated.	2
				XUNE2024	Updated due to change to months worked (LJD003B-S, LJD003U) and months looking for work (LJF002A-E, LJF002T).	918
Waves 1 to 10 (2015 to 2024)	V9	October 2025	*Wave 10 (2024) variables added to data file.*			All
			Modifications to variables			
			Occupation codes	LIE002X LIE005 LIE006	Respondents missing a value for occupation have been corrected and assigned a value.	9
			No employment since last interview	LID033A-U	One respondent who should not have responded to the question series had their responses recorded. This has been corrected and their values set to missing.	1
			Modifications to formats			

Wave/ year	Version	Date published	Variable	Variable name	Description	Number of records affected
			Full-time or part-time employment status	XFTP*YYYY*	Label for value 98 changed from 'Unknown employment status' to 99 'Unknown labour force status' to better reflect the category and maintain consistency across employment variables.	
			Permanent or casual employment	XEMP*YYYY*	As above	
Waves 1 to 9 (2015 to 2023)	V8	September 2024	*Wave 9 (2023) variables added to data file.*			All
			Modifications to variables			
			Frequency of pay	LHD016	Frequency of pay values which were overwritten at wave 8 (2023) to facilitate the recoding of (2) outliers for weekly and hourly pay derivations have been reinstated to their original values.	2
			ANZSCO occupation codes	LHD022	Several occupations were previously coded using ANZSCO 2013 (version 1.3). These codes have now been updated using the ANZSCO 2021.	9
				LHE022X		5
				LHJ005A		3
			Modifications to derived variables			
			Study status in VET	XVET2022	Some respondents were incorrectly assigned a VET study status of 'Commenced, but did not complete'. These records have been corrected and are now assigned a VET study status of 'Completed'.	8
			Average weekly hours worked	XHRS2022	Outliers for average weekly hours worked have been recoded to the reserved value '999 Unknown'.	5
			Full-time or part-time employment status	XFTP2022	Change due to modification to 'Average weekly hours worked'	5
			Average weekly pay	XWKP2022	Outliers for average weekly pay have been recoded to the reserved value '99999 Unknown' (2 records) Outliers for average weekly pay have been derived using a corrected frequency of pay period (11 records)	13
			Average hourly pay	XHRP2022	Outliers for average weekly pay have been recoded to the reserved value '9999 Unknown' (3 records) Outliers for average weekly pay have been derived using a corrected frequency of pay period (11 records) Change due to modification to 'Average weekly hours worked' (5 records)	19
			In full-time employment or full-time education	XFTE2022	Change due to modification to 'Full-time or part-time employment status'	4
			Any spell of unemployment during the year	XUNE2019		3

Wave/ year	Version	Date published	Variable	Variable name	Description	Number of records affected
				XUNE2022	The previous derivation did not consider every month not worked since the last interview. The derivation has been revised to ensure all months not worked have been used in the derivation.	3
			Number of dependent children	XCHI2018 XCHI2020 XCHI2021 XCHI2022	The previous derivation did not consider all household information provided at variable H6 (Relationship to respondent). This has been revised and this derivation now accounts for information supplied at both H6 and H7 (Number of children).	5
Waves 1 to 8 (2015 to 2022)	V7	September 2023	*Wave 8 (2022) variables added to data file.*			All
Waves 1 to 7 (2015 to 2021)	V6	September 2022	*Wave 7 (2021) variables added to data file.*			
			Modifications to derived variables			
			Labour force status	XLFS2016	The labour force status for apprentices and trainees who had not indicated they were working were previously assigned a labour force status of 'Unknown', 'Unemployed' or 'Not in the labour force' depending on their circumstances. For consistency, this has been modified and these records are now classified as 'Employed'.	16
			Average weekly hours worked in main job	XHRS2016	Change due to modification to ‘Labour force status’	14
			Full-time or part-time employment status	XFTP2016	Change due to modification to ‘Labour force status’	16
			Permanent or casual employment	XEMP2016	Change due to modification to ‘Labour force status’	16
			Average weekly pay	XWKP2016	Change due to modification to ‘Labour force status’	14
			Average hourly pay	XHRP2016	Change due to modification to ‘Labour force status’	14
			Occupation (1 digit ANZSCO First Edition)	XOCC2016	Change due to modification to ‘Labour force status’	16
			In full-time employment or full-time education	XFTE2016	Change due to modification to ‘Labour force status’	2
			Job mobility during last year	XMOB2017	Change due to modification to ‘Labour force status’ in 2016	13

Wave/ year	Version	Date published	Variable	Variable name	Description	Number of records affected
			Status in apprenticeship/traineeship	XATR2019	Two respondents who had withdrawn from an apprenticeship/traineeship and started another one were previously assigned an apprenticeship/traineeship status of 'Commenced, but did not complete'. This has been modified and these records are now classified as 'Currently undertaking'.	2
				XATR2020	Two respondents who had completed an apprenticeship/traineeship and then went on to withdraw from a subsequent apprenticeship/traineeship were previously assigned an apprenticeship/traineeship status of 'Commenced, but did not complete'. This has been modified and these records are now classified as 'Completed'. Three respondents who had withdrawn from an apprenticeship/traineeship and started another one were previously assigned an apprenticeship/traineeship status of 'Commenced, but did not complete'. This has been modified and these records are now classified as 'Currently undertaking'.	5
			Marital status	XMAR2018	For consistency, respondents previously assigned a marital status of 'Prefer not to say' have now been classified as 'Unknown'.	291
				XMAR2019	As above	185
				XMAR2020	As above	142
			Current qualification level	XCEL2020	Respondents who were not currently studying but later indicated their studies were paused due to COVID-19 were previously assigned a qualification level of 'Not studying for a qualification'. This has been modified and these records have now been assigned the qualification level of the paused study.	2
			Full-time or part-time study status	XFTS2020	As above, respondents who were not currently studying but later indicated their studies were paused due to COVID-19 were previously assigned a study status of 'Not studying'. This has been modified and these records have now been assigned the study status of the paused study.	2
			Study status in bachelor degree or higher	XBAC2020	Change due to modification to 'Current qualification level'	2
			Completed Year 12 or certificate II or higher	X1222017	Respondents who had not completed a certificate II or above and whose year 12 completion status was unknown were previously assigned a status of 'Did not complete year 12 or certificate II or higher'. This has been modified and these records are now classified as 'Unknown'.	6
				X1222018	As above	3
				X1222019	As above	4
				X1222020	As above	1

Wave/ year	Version	Date published	Variable	Variable name	Description	Number of records affected
			Completed Year 12 or certificate III or higher	X1232017	Respondents who had not completed a certificate III or above and whose year 12 completion status was unknown were previously assigned a status of 'Did not complete year 12 or certificate III or higher'. This has been modified and these records are now classified as 'Unknown'.	7
				X1232018	As above	4
				X1232019	As above	4
				X1232020	As above	2
Waves 1 to 6 (2015 to 2020)	V5	August 2021	*Wave 6 (2020) variables added to data file.*			All
			Modifications to variable names for wave 3 top-up recruitment sample			
			CAD10 Have you done any of the following to help you make decisions about your future?	LBCA010A-LBCA010F, LBCA010H-LBCA010J	Top-up sample records removed and assigned to separate variables LBCA010A_*R*-LBCA010F_*R*, LBCA010H_*R*-LBCA010J_*R* to maintain the variable naming convention for recruitment questionnaire items and records.	548
			Modifications to value assignments for data linkage permission questions			
			NAPLAN data linkage permission by state wave 4 (2018)	LDDL001C	Records with a value of 2 (equal to 'No') were recoded to the value 0 as per the binary coding approach used in LSAY. However, for these variables, the value 0 is used to code those who responded to the data linkage permission question but were not from the specified state (0 = 'Not asked for this state'). As such, records which had a value of 2 and were recoded to 0 have been recoded back to 2.	28
				LDDL001D		27
				LDDL001E		20
				LDDL001F		13
				LDDL001G		5
				LDDL001H		16
			Senior secondary subject data linkage permission by state wave 4 (2018)	LDDL002A	As above	76
				LDDL002B		45
				LDDL002C		30
				LDDL002D		25
				LDDL002E		25
				LDDL002F		13
				LDDL002G		7
				LDDL002H		20
			NAPLAN data linkage permission by state wave 5 (2019)	LEDL001A	As above	19
				LEDL001B		10

Wave/ year	Version	Date published	Variable	Variable name	Description	Number of records affected
				LEDL001C		4
				LEDL001D		7
				LEDL001E		5
				LEDL001F		3
				LEDL001H		1
				LEDL001A		19
			Senior secondary data linkage permission by state wave 5 (2019)	LEDL002A	As above	14
				LEDL002B		11
				LEDL002C		6
				LEDL002D		5
				LEDL002E		5
				LEDL002G		1
			Corrections, modifications and changes to derived variables			
			Current qualification level	XCEL2019	Incorrect variable used in deriving current qualification for those who had resumed their deferred studies. This has now been corrected.	14
			Full-time or part-time study status	XFTS2019	Change due to correction to 'Current qualification level'.	5
			Study status in bachelor degree or higher	XBAC2019	As above.	13
			Study status in VET	XVET2019	As above.	8
			In full-time employment or full-time education	XFTE2019	As above.	4
			Marital status	XMAR2018	Due to the high incidence of 'Don't know/prefer not to say' responses, in addition to relationship status information, this derivation has been modified to include household formation information.	16
				XMAR2019	As above.	22
Waves 1 to 5 (2015 to 2019)	V4	December 2020	*Wave 5 (2019) variables added to data file; includes an additional 19 respondents from the wave 3 top-up (recruitment) sample who participated in an LSAY interview for the first time in 2019.*			All
			Corrections to assignment of anonymised ACARA school ID			
			DEM07 Study specific school campus ID	LBDE007A_C		2
			DEM07 Study specific school main campus ID	LBDE007A_M		2

Wave/ year	Version	Date published	Variable	Variable name	Description	Number of records affected
			K1(x) Study specific school campus ID	LCK001A_C		1
			K1(x) Study specific school main campus ID	LCK001A_M		1
			K1 Study specific school campus ID	LDK001A_C		1
			K1 Study specific school main campus ID	LDK001A_M		1
			Corrections to assignment of ACARA school state			
			State derived from DEM13 school	LBDE013B		2
			State derived from K1 school	LCK001X		1
			State derived from K7 school	LCK007X		2
			State derived from K1 school	LDK001X		4
			Modifications to 'in survey' flags and/or derived variable assignments for: respondents from the top-up group; respondents who undertook a VET qualification while at school; and respondents who were unemployed or not in the labour force.			
			Current school level	XCSL2017	Respondents from the top-up group are only incorporated into the LSAY sample if they complete the main LSAY interview. Respondents from the top-up group who did not complete the main LSAY interview in 2017 were previously assigned an 'in survey' flag. This has been modified and these respondents are no longer assigned an 'in survey' flag or 'Current school level' in 2017.	278
			Highest school level completed	XHSL2017	As above. These respondents are no longer assigned an 'in survey' flag or 'Highest school level completed' in 2017.	278
			Current qualification level	XCEL2017	As above. These respondents are no longer assigned an 'in survey' flag or 'Current qualification level' in 2017.	278
			Highest qualification level completed	XHEL2017	As above. These respondents are no longer assigned an 'in survey' flag or 'Highest qualification level completed' in 2017. In addition, respondents who had undertaken a VET qualification while at school were previously recorded as having completed a VET qualification. To maintain consistency with the way this measure is derived for other cohorts, only VET qualifications completed post-school are recorded and these respondents are now recorded as not having completed a VET qualification.	328
				XHEL2018	Change due to change to 'Highest qualification level completed' in previous year.	35

Wave/ year	Version	Date published	Variable	Variable name	Description	Number of records affected
			Study status in bachelor degree or higher	XBAC2017	Respondents from the top-up group are only incorporated into the LSAY sample if they complete the main LSAY interview. Respondents from the top-up group who did not complete the main LSAY interview in 2017 were previously assigned an 'in survey' flag. This has been modified and these respondents are no longer assigned an 'in survey' flag or 'Study status in bachelor degree or higher' in 2017.	278
			Study status in VET	XVET2017	As above. These respondents are no longer assigned an 'in survey' flag or 'Study status in VET ' in 2017. In addition, respondents who participated in VET while at school were recorded as having undertaken a VET qualification. To maintain consistency with the way this measure is derived for other cohorts, only VET qualifications undertaken post-school are recorded and these respondents are now recorded as not having undertaken a VET qualification.	334
				XVET2018	Change due to change to 'Study status in VET' in previous year.	37
			Completed Year 12 or certificate II or higher	X1222017	Respondents from the top-up group are only incorporated into the LSAY sample if they complete the main LSAY interview. Respondents from the top-up group who did not complete the main LSAY interview in 2017 were previously assigned an 'in survey' flag. This has been modified and these respondents are no longer assigned an 'in survey' flag or 'Completed Year 12 or certificate II or higher' in 2017. Change due to change to 'Highest qualification level completed'.	280
			Completed Year 12 or certificate III or higher	X1232017	Respondents from the top-up group are only incorporated into the LSAY sample if they complete the main LSAY interview. Respondents from the top-up group who did not complete the main LSAY interview in 2017 were previously assigned an 'in survey' flag. This has been modified and these respondents are no longer assigned an 'in survey' flag or 'Completed Year 12 or certificate III or higher' in 2017.	278
			Labour force status	XLFS2017	Respondents who were not working at the time of their interview, had looked for work since their last interview and had not looked for work in the last 4 weeks were incorrectly assigned a labour force status of unemployed. This has been corrected and these respondents are now assigned a labour force status of not in the labour force.	265
				XLFS2018	As above.	214

Wave/ year	Version	Date published	Variable	Variable name	Description	Number of records affected
			Status in apprenticeship/ traineeship	XATR2017	Respondents from the top-up group are only incorporated into the LSAY sample if they complete the main LSAY interview. Respondents from the top-up group who did not complete the main LSAY interview in 2017 were previously assigned an 'in survey' flag. This has been modified and these respondents are no longer assigned an 'in survey' flag or 'Status in apprenticeship/ traineeship' in 2017. In addition, respondents who participated in an apprenticeship or traineeship while at school were recorded as having undertaken an apprenticeship/traineeship. To maintain consistency with the way this measure is derived for other cohorts, only apprenticeships and traineeships undertaken post-school are recorded and these respondents are now recorded as not having undertaken an apprenticeship/traineeship.	303
				XATR2018	Change due to change to 'Status in apprenticeship/traineeship' in previous year.	17
			Job mobility during last year	XMOB2017	Respondents from the top-up group are only incorporated into the LSAY sample if they complete the main LSAY interview. Respondents from the top-up group who did not complete the main LSAY interview in 2017 were previously assigned an 'in survey' flag. This has been modified and these respondents are no longer assigned an 'in survey' flag or 'Job mobility during last year' in 2017. Change due to change to 'Labour force status'.	298
				XMOB2018	Change due to change to 'Labour force status'.	92
			Occupation (1 digit ANZSCO First Edition)	XOCC2016	Respondents with a labour force status of unknown were previously assigned an occupation of 'Unknown or not classifiable'. This has been modified and these respondents are now assigned to the category 'Unknown labour force status'.	357
			Any spell of unemployment during the year	XUNE2017	Change due to change to 'Labour force status'.	10
				XUNE2018	Change due to change to 'Labour force status'.	14
			Marital status	XMAR2017	Respondents from the top-up group are only incorporated into the LSAY sample if they complete the main LSAY interview. Respondents from the top-up group who did not complete the main LSAY interview in 2017 were previously assigned an 'in survey' flag. This has been modified and these respondents are no longer assigned an 'in survey' flag or 'Marital status' in 2017.	278
			Number of dependent children	XCHI2017	As above. These respondents are no longer assigned an 'in survey' flag or 'Number of dependent children' in 2017.	278

Wave/ year	Version	Date published	Variable	Variable name	Description	Number of records affected
Waves 1 to 4 (2015 to 2018)	V3	December 2019			Wave 4 (2018) variables added to data file; includes an additional 278 respondents from the wave 3 top-up (recruitment) sample who participated in an LSAY interview for the first time.	All
			CAD09 ANZSCO - What kind of job do you expect to have when you are about 30 years old? (2017 Recruitment questionnaire)	LBCA009_R	LBCA009_R renamed to LBCA009A_R to maintain consistency with the variable naming convention for recruitment questionnaire questions.	
			School Identifier (ACARA) (2017 Recruitment questionnaire)	ACARA	ACARA was renamed SCHOOLID_R to distinguish between PISA school identifiers (CNTSCHID) and ACARA school identifiers (SCHOOLID).	
			School identifiers	SCHOOLID_R LBDE007A LBDE009A LBDE011A LBDE013A LBLA002C LBLA009C LCA008_C LCK001A LCK003A LCK005A LCK007A LDA008 LDK001A LDK002A LDK004A LDK006A LDK008A	To enable data linkage with the ACARA *My School* data while protecting the identities of schools, a study specific school identifier has been created in place of the ACARA school identifier. Variables have been appended with: ▪ '_C' to denote the campus level school identifier ▪ '_M' to denote the main campus school identifier (used for linking the main LSAY data file with the LSAY-ACARA linked datasets). Further information is available from the 'Data linkage' section of this user guide.	

Wave/ year	Version	Date published	Variable	Variable name	Description	Number of records affected
			Weight variables	WT2016 WT16GEN WT2016P WT16GENP WT2017 WT17GEN WT2017P WT17GENP WT2017_R WT2017P_R	Minor adjustments to weight variables from previous waves to account for variation in the sample as a result of respondents from the top-up sample completing an LSAY interview for the first time in wave 4 (2018).	
			Status in apprenticeship/ traineeship	XATR2017	A number of respondents who commenced an apprenticeship or traineeship while at school and then left school and continued their apprenticeship/traineeship were previously derived as having never commenced an apprenticeship/traineeship.	124
			Study status in VET	XVET2017	A number of respondents who commenced VET study while at school and then left school and continued their VET study were previously derived as having never commenced VET.	62
			Current qualification level	XCEL2017	A number of respondents who commenced VET study while at school and then left school and continued their VET study were previously derived as not studying for a qualification.	11
			Highest qualification level completed	XHEL2017	Respondents who completed a VET qualification while at school were recorded as having completed a VET qualification. To maintain consistency with the way this measure is derived for other cohorts, only VET qualifications completed post-school are recorded and these respondents are no longer recorded as having completed a VET qualification.	51
			Full-time or part-time study status	XFTS2017	Change results from the change to 'Current qualification level'.	11
			Completed Year 12 or certificate II or higher	X1222017	Change results from the change to 'Highest qualification level completed'.	3
			Completed Year 12 or certificate III or higher	X1232017	Change results from the change to 'Highest qualification level completed'.	1
Waves 1 to 3 (2015 to 2017)	V2	March 2019			Wave 3 (2017) variables added to data file; includes data items from 251 respondents from the top-up (recruitment) sample	All
			LSAY identifier	LSAYID	New identifier variable 'LSAYID' added to account for the top-up sample being added to the dataset. The original 'STUDENID' only includes PISA respondents.	All

Wave/ year	Version	Date published	Variable	Variable name	Description	Number of records affected
			Current qualification level	XCEL2016	At wave 2 (2016), the derivations for these measures included (non-school) qualifications undertaken while at school. For consistency with the other LSAY cohorts, these measures have been modified to only include (non-school) qualifications undertaken while not at school.	309
			Study status in VET	XVET2016		505
			Study status in bachelor degree or higher	XBAC2016		3
			Completed Yr 12 or certificate II or higher	X1222016		172
			Completed Yr 12 or certificate III or higher	X1232016		73
			Highest qualification completed	XHEL2016		200
			Status in apprenticeship/traineeship	XATR2016		172
			Any spell of unemployment during the year	XUNE2016	Some respondents who reported that they did not have a job and were looking for work at the time of their interview were categorised as not having experienced a spell of unemployment during the year, due to reporting having worked at some point during the month of their interview. This has been corrected such that anyone unemployed at the time of their interview (i.e. not working and looking for work) is now classified as having experienced a spell of unemployment during the year.	59
Waves 1 to 2 (2015 to 2016)	V1	Dec 2017			Data file created incorporating data from wave 1 (PISA 2015) and wave 2 (2016).	All

Appendix B: Linked datasets

The following table provides a summary of the linked datasets deposited with the Australian Data Archive. More information about the linked data is available in the section 'Data linkage'. Updates to the linked datasets can be found at appendix C.

Details about the variables contained in the linked datasets are available from the 'Data linkage' worksheet located in the LSAY variable listing and metadata: <https://www.lsay.edu.au/publications/search-for-lsay-publications/2621>.

Table B1 Linked datasets

Data source	Version	Date published	Dataset	Dataset name
ACARA *My School*	V1	December 2019	School profile	lsayacara_schprofile_v1
			NAPLAN results	lsayacara_results_v1
			NAPLAN results similar schools	lsayacara_results_simsc_v1
			Student gain	lsayacara_studgain_v1
			Student gain similar schools	lsayacara_gain_simsch_v1
			Student gain same starting score	lsayacara_gain_ss_v1
			School attendance	lsayacara_attendance_v1
			VET in schools	lsayacara_vet_v1
			VET in schools – School-based apprenticeships and traineeships	lsayacara_vet_sbat_v1
			Senior secondary outcomes	lsayacara_secsch_v1
			Finance	lsayacara_finance_v1
			Enrolments by grade	lsayacara_enrolments_v1
Senior secondary	V1	December 2021	Senior secondary	lsaysnrsec_v1
NAPLAN	V3	May 2023	NAPLAN scores	lsaynaplan_v3
			NAPLAN scores – including repeats	lsaynaplan_repeats_v3
	V2	June 2021	NAPLAN scores	lsaynaplan_v2
			NAPLAN scores – including repeats	lsaynaplan_repeats_v2
	V1	December 2019	NAPLAN scores	lsaynaplan_v1
			NAPLAN scores – including repeats	lsaynaplan_repeats_v1
National VET Provider Collection	V7	August 2026	Completions	lsayvet_c_v7
			Subject enrolments	lsayvet_s_v7
	V6	July 2025	Completions	lsayvet_c_v6
			Subject enrolments	lsayvet_s_v6
	V5	May 2024	Completions	lsayvet_c_v5
			Subject enrolments	lsayvet_s_v5
	V4	May 2023	Completions	lsayvet_c_v4
			Subject enrolments	lsayvet_s_v4
	V3	December 2021	Completions	lsayvet_c_v3
			Subject enrolments	lsayvet_s_v3
	V2	December 2020	Completions	lsayvet_c_v2
			Subject enrolments	lsayvet_s_v2
	V1	December 2019	Completions	lsayvet_c_v1
			Subject enrolments	lsayvet_s_v1
Higher Education Statistics Collection	V4	July 2025	Student	lsayheims_student_v4
			Course	lsayheims_course_v4

Data source	Version	Date published	Dataset	Dataset name
			Unit of study	lsayheims_unit_of_study_v4
			Completions	lsayheims_completions_v4
			Scholarship	lsayheims_scholarship_v4
	V3	May 2024	Student	lsayheims_student_v3
			Course	lsayheims_course_v3
			Unit of study	lsayheims_unit_of_study_v3
			Completions	lsayheims_completions_v3
			Scholarship	lsayheims_scholarship_v3
	V2	December 2021	Student	lsayheims_student_v2
			Course	lsayheims_course_v2
			Unit of study	lsayheims_unitofstudy_v2
			Completions	lsayheims_completions_v2
			Scholarship	lsayheims_scholarship_v2
	V1	December 2020	Student	lsayheims_student_v1
			Course	lsayheims_course_v1
			Unit of study	lsayheims_unit_v1
			Completions	lsayheims_completions_v1
			Scholarship	lsayheims_scholarship_v1

Appendix C: Updates to the linked data files

The following table tracks updates made to the Y15 linked data files deposited with the Australian Data Archive. Users are encouraged to download the most recent version of the data file to ensure all updates are included.

Details about the variables contained in the linked datasets are available from the 'Data linkage' worksheet located in the LSAY variable listing and metadata <https://www.lsay.edu.au/publications/search-for-lsay-publications/2621>.

Table C1 Summary of changes made to the linked NAPLAN data files

Version	Date published	Data file/s	Variables	Description
V3	May 2023	NAPLAN scores and NAPLAN scores – including repeats	*Data updated to include weights*	Custom weights added to dataset. See sub-section 'National Assessment Program – Literacy and Numeracy' for more information.
			Data revised to remove 2007 scores	The Queensland Curriculum and Assessment Authority provided literacy and numeracy test scores for 2007 which were based on test items from the Queensland curriculum. These scores have been removed from the dataset as they are not part of the NAPLAN assessment which commenced in 2008.
V2	June 2021	NAPLAN scores and NAPLAN scores – including repeats	*Data updated to include respondents who consented at waves 4 and 5 (2018 and 2019).*	
			'Writing' data items replaced by 'Narrative writing' data items	Writing data items are now separated into 'Narrative writing' and 'Writing' data items to reflect the difference in the writing assessments since 2011. 'Narrative writing' was assessed from 2007-2010 and 'Persuasive writing' (referred to in the documentation as 'Writing') was assessed from 2011-2016.
			NARRATIVEWRITING_PARTICIPATION, NARRATIVEWRITING_SCORE, NARRATIVEWRITING_BAND, NARRATIVEWRITING_NMS	'Narrative writing' items added and populated with 'Writing' data for the years 2007-2010 to reflect the difference in the writing assessments between 2007-2010 and 2011-2016.
			WRITING_PARTICIPATION, WRITING_SCORE, WRITING_BAND, WRITING_NMS	'Writing' data for the years 2007-2010 removed and added to 'Narrative writing' to reflect the difference in the writing assessments between 2007-2010 and 2011-2016.
			Variables assigned reserved missing values	

Version	Date published	Data file/s	Variables	Description
			SECTOR, SPELLING_PARTICIPATION, SPELLING SCORE, SPELLING_BAND, SPELLING_NMS, GRAMMARPUNCT_PARTICIPATION, GRAMMARPUNCT_SCORE, GRAMMARPUNCT_BAND, GRAMMARPUNCT_NMS	For cases where data is unavailable and a reserved value was not previously assigned, '-9 Data unavailable' or '9 Not stated or unknown' is assigned in place of a missing or blank value.
			Variables imputed	
			SPELLING_BAND, SPELLING_NMS, GRAMMARPUNCT_BAND, GRAMMARPUNCT_NMS, READING_BAND, READING_NMS, NUMERACY_BAND, NUMERACY_NMS	Variables imputed for cases where bands or NMS values are missing using: the NAPLAN score equivalence tables; or equivalent NAPLAN scores and their corresponding band and NMS values.
			Variables corrected	
			WRITING_BAND, WRITING_NMS, SPELLING_BAND, GRAMMARPUNCT_BAND, GRAMMARPUNCT_NMS, READING_BAND, , READING_NMS, NUMERACY_BAND, NUMERACY_NMS	Corrections made to bands and NMS values for cases where scores did not correspond to NAPLAN score equivalence tables.
			Variables removed	
			SEX	Sex removed due to some inconsistencies with the data provided by the jurisdictions. This information is available and should be sourced from the main LSAY data file instead.
V1	December 2019	NAPLAN scores and NAPLAN scores – including repeats		NAPLAN data 2007-2016

Table C2 Summary of changes made to the linked National VET Provider Collection data files

Version	Date published	Data file	Variable	Description
V7	August 2026	*Data updated to include: 2024 training activity; and data for newly consenting respondents.*		
		Note: Data from previous collections may differ from the current collection as the data is updated and revised annually. More information is available from the 'Total VET students and courses 2024: explanatory notes' available at <https://www.ncver.edu.au/research-and-statistics/publications/all-publications/total-vet-students-and-courses-2024>.		
V6	July 2025	*Data updated to include: 2023 training activity; and data for newly consenting respondents.*		
		Note: Data from previous collections may differ from the current collection as the data is updated and revised annually. More information is available from the 'Total VET students and courses 2023: explanatory notes' available at <https://www.ncver.edu.au/research-and-statistics/publications/all-publications/total-vet-students-and-courses-2023>.		
		Completions	*Variable names updated as per source data*	
			CLIENT_SEIFA_IRSD_QUINTILE_DERIVED	Replaced by CLIENT_SEIFA_IRSD_DERIVED

Version	Date published	Data file	Variable	Description
			Values updated	
			Various	Victoria previously over-stated the number of program completions over the years 2015 to 2022. This has been remediated as part of the 2023 time-series processing, and the records for two (2) respondents have been corrected.
		Subjects	*Variable names updated as per source data*	
			CLIENT_SEIFA_IRSD_QUINTILE_DERIVED	Replaced by CLIENT_SEIFA_IRSD_DERIVED
V5	May 2024	*Data updated to include: 2022 training activity; and data for newly consenting respondents.*		
V4	May 2023	*Data updated to include: 2021 training activity; and data for newly consenting respondents.*		
		Completions	*Variable names updated as per source data*	
			CLIENT_STATE_RESIDENCE_DERIVED	Replaced by CLIENT_RES_STATE_ID_DERIVED
			CLIENT_RESIDENTIAL_SA4	Replaced by CLIENT_RESIDENTIAL_SA4_DERIVED
			CLIENT_REMOTENESS_ID_DERIVED	Replaced by CLIENT_REMOTE_ID_DERIVED
			HEAD_OFFICE_STATE	Replaced by HEAD_OFFICE_STATE_ID
			Variables added	
			DISAB_TYPE_FG_11	Disability type ID 11: Hearing/Deaf
			DISAB_TYPE_FG_12	Disability type ID 12: Physical
			DISAB_TYPE_FG_13	Disability type ID 13: Intellectual
			DISAB_TYPE_FG_14	Disability type ID 14: Learning
			DISAB_TYPE_FG_15	Disability type ID 15: Mental Illness
			DISAB_TYPE_FG_16	Disability type ID 16: Acquired Brain Injury
			DISAB_TYPE_FG_17	Disability type ID 17: Vision
			DISAB_TYPE_FG_18	Disability type ID 18: Medical Condition
			DISAB_TYPE_FG_19	Disability type ID 19: Other
			DISAB_TYPE_FG_99	Disability type ID 99: Not Specified
			LANG_OTHER_THAN_ENG_ID	Language other than English spoken at home
			DEL_LOC_STATE_ID	Training organisation delivery location state
			PROGRAM_APPRENTICESHIP_FG	Program undertaken as part of an apprenticeship/traineeship
			PROGRAM_COMMENCING_FG	Commencing program status
			STUDY_MODE	Study mode
		Subjects	*Variable names updated as per source data*	

Version	Date published	Data file	Variable	Description
			CLIENT_STATE_RESIDENCE_DERIVED	Replaced by CLIENT_RES_STATE_ID_DERIVED
			CLIENT_RESIDENTIAL_SA4	Replaced by CLIENT_RESIDENTIAL_SA4_DERIVED
			CLIENT_REMOTENESS_ID_DERIVED	Replaced by CLIENT_REMOTE_ID_DERIVED
			HEAD_OFFICE_STATE	Replaced by HEAD_OFFICE_STATE_ID
			DELIVERY_LOCATION_STATE	Replaced by DEL_LOC_STATE_ID
			DELIVERY_LOCATION_SA4	Replaced by DEL_LOC_SA4
			Variables added	
			DISAB_TYPE_FG_11	Disability type ID 11: Hearing/Deaf
			DISAB_TYPE_FG_12	Disability type ID 12: Physical
			DISAB_TYPE_FG_13	Disability type ID 13: Intellectual
			DISAB_TYPE_FG_14	Disability type ID 14: Learning
			DISAB_TYPE_FG_15	Disability type ID 15: Mental Illness
			DISAB_TYPE_FG_16	Disability type ID 16: Acquired Brain Injury
			DISAB_TYPE_FG_17	Disability type ID 17: Vision
			DISAB_TYPE_FG_18	Disability type ID 18: Medical Condition
			DISAB_TYPE_FG_19	Disability type ID 19: Other
			DISAB_TYPE_FG_99	Disability type ID 99: Not Specified
			LANG_OTHER_THAN_ENG_ID	Language other than English spoken at home
			TRAINING_ORG_OFFSHORE_FG	Training organisation offshore flag
			PROGRAM_COMMENCING_FG	Commencing program status
			SUBJECT_FOE_ID	Subject field of education 6-digit ID
			STUDY_MODE	Study mode
V3	December 2021	*Data updated to include: 2020 training activity; and data for newly consenting respondents.*		
		Completions	*Variable name updated as per source data*	
			INDIGENOUS_ID	Replaced by INDIGENOUS_STATUS_ID
			REMOTENESS_SCORE_ID_DERIVED	Replaced by CLIENT_REMOTENESS_ID_DERIVED
			SEIFA_IRSD_QUINTILE_DERIVED	Replaced by CLIENT_SEIFA_IRSD_QUINT_DERIVED
			RTO_KEY	Replaced by RTO_ID_ENCRYPT
		Subject enrolments	*Variable name updated as per source data*	
			INDIGENOUS_ID	Replaced by INDIGENOUS_STATUS_ID
			REMOTENESS_SCORE_ID_DERIVED	Replaced by CLIENT_REMOTENESS_ID_DERIVED

Version	Date published	Data file	Variable	Description
			SEIFA_IRSD_QUINTILE_DERIVED	Replaced by CLIENT_SEIFA_IRSD_QUINT_DERIVED
			RTO_KEY	Replaced by RTO_ID_ENCRYPT
			MEASURE_PROGRAM_NOM_HOURS	Replaced by PROGRAM_NOMINAL_HOURS
			MEASURE_SUBJECT_NOM_HOURS	Replaced by SUBJECT_NOMINAL_HOURS
			REPORTING_HOURS_MEASURE	Replaced by SUBJECT_REPORTING_HOURS
			Variable added	
			STUDENT_COMMENCING_FG	Commencing student status
V2	December 2020	*Data updated to include: 2018 and 2019 training activity; and data for newly consenting respondents.*		
		Completions	*Variables removed as per source data*	
			TOTAL_VET_SCOPE_FG	No longer required; NATIONALLY_RECOGNISED_RTO_FG is instead used to report nationally recognised VET delivered by Australian registered training organisations.
			ENGLISH_PROFICIENCY_ID	No longer collected.
			Variables added	
			CLIENT_RESIDENTIAL_SA4	Complements existing client location variables using Statistical Area Level 4 (SA4) geographical areas (restricted).
			TRAINING_ORG_OFFSHORE_FG	Training organisation offshore flag
			PROGRAM_TYPE_OF_TRAINING_ID	Program type of training
			PROGRAM_TRAINING_PACKAGE_ID	Program training package ID
			DATE_PROGRAM_COMPLETED	Date program completed
			Variables replaced as per source data	
			PROGRAM_COMPLETED_YR	Replaced by YR_PROGRAM_COMPLETED
			SEX_ID	Replaced by GENDER_ID
			CLIENT_STATE_RESIDENCE	Replaced by CLIENT_STATE_RESIDENCE_DERIVED
			CLIENT_SUBURB	Replaced by CLIENT_SUBURB_DERIVED (restricted)
			CLIENT_POSTCODE	Replaced by CLIENT_POSTCODE_DERIVED (restricted)
			CLIENT_REMOTENESS_ID	Replaced by REMOTENESS_SCORE_ID_DERIVED
			CLIENT_SEIFA_IRSD	Replaced by SEIFA_IRSD_QUINTILE_DERIVED
		Subject enrolments	*Variables removed as per source data*	
			TOTAL_VET_SCOPE_FG	No longer required; NATIONALLY_RECOGNISED_RTO_FG is instead used to report nationally recognised VET delivered by Australian registered training organisations.
			ENGLISH_PROFICIENCY_ID	No longer collected.
			YR_HIGHST_SCHL_LVL_COMPLTD_ST	No longer collected.

Version	Date published	Data file	Variable	Description
			COMMENCING_PROGRAM_ID	No longer collected.
			Variables added	
			SCHOOL_TYPE_ID	Client home school type identifier
			CLIENT_RESIDENTIAL_SA4	Complements existing client location variables using Statistical Area Level 4 (SA4) geographical areas (restricted).
			HEAD_OFFICE_STATE	Training organisation state of head office
			PROGRAM_TYPE_OF_TRAINING_ID	Program type of training
			PROGRAM_TRAINING_PACKAGE_ID	Program training package ID
			DELIVERY_LOCATION_SA4	Location of training delivery using Statistical Area Level 4 (SA4) geographical areas (restricted).
			APPRENTICESHIP_FG	Apprenticeship/traineeship flag
			REPORTING_HOURS_MEASURE	Subject hours of delivery
			Variables replaced as per source data	
			PROGRAM_COMPLETED_YR	Replaced by YR_PROGRAM_COMPLETED
			SEX_ID	Replaced by GENDER_ID
			CLIENT_STATE_RESIDENCE	Replaced by CLIENT_STATE_RESIDENCE_DERIVED
			CLIENT_SUBURB	Replaced by CLIENT_SUBURB_DERIVED (restricted)
			CLIENT_POSTCODE	Replaced by CLIENT_POSTCODE_DERIVED (restricted)
			CLIENT_REMOTENESS_ID	Replaced by REMOTENESS_SCORE_ID_DERIVED
			CLIENT_SEIFA_IRSD	Replaced by SEIFA_IRSD_QUINTILE_DERIVED
			DEL_MODE_ID	Replaced by DELIVERY_MODE_ID
			SUBJECT_NOMINAL_HRS	Replaced by MEASURE_SUBJECT_NOM_HOURS
			PROGRAM_NOMINAL_HRS	Replaced by MEASURE_PROGRAM_NOM_HOURS
V1	December 2019			National VET Provider Collection Data 2015-2017

Table C3 Summary of changes made to the linked higher education statistics data files

Version	Date published	Data file	Variable	Description
V4	July 2025	*Data updated to include: 2023 activity; and data for newly consenting respondents*		
		Student	*Variables removed*	
			E465, E562, E563, E564, E565	These elements have been retired form the Higher Education Statistics Collection.
		Completions	*Records updated*	
			E328	Updates to one record for element 'E328 – Course commencement date'.
			Disability	Updates to 25 records for element 'Disability'.
			E463_multiple	Updates to 439 records for element 'E463 – Specialisation code (multiple)'.
V3	May 2024	*Data updated to include: 2021 and 2022 activity; and data for newly consenting respondents*		
		Completions	*Variable added*	
			e463_multiple	This element 'e463_multiple: Specialisation code – multiple' has been added to identify multiple specialisations for a course
			Records updated	
		Student	E319	Updates to two records for element 'E319 – Term address postcode'.
			E330	Updates to one record for element 'E330 – Type of attendance code'.
		Completions	disability	Updates to 19 records for element 'disability'.
			E463	Updates to six records for element 'E463 – Specialisation code'.
V2	December 2021	*Data updated to include: 2020 activity; and data for newly consenting respondents*		
		Student	*Variables removed*	
			E561	This element 'Credit basis code' relates to VET study only so has been removed.
		Student, Completions	*Variables changed*	
			E386, E386A	These elements have changed and been replaced by the element 'disability' which uses a one-character code to indicate whether a student has a disability, impairment or long-term medical condition.
V1	December 2020			Higher education statistics data 2015-2019

Australian Government
Department of Education

National Centre for Vocational Education Research Ltd
Level 4, 191 Pulteney Street, Adelaide, South Australia
PO Box 8288, Station Arcade, SA 5000 Australia
Telephone +61 8 8230 8400
Web www.ncver.edu.au Email ncver@ncver.edu.au