

Equity Frontiers QED

Trial Protocol

“The impact of enabling programs on student’s initial academic retention and success in higher education: A propensity score modelling approach”

Table of Contents

1	Overview.....	3
2	Background and objective.....	5
2.1	Background	5
2.2	Initiative	6
2.3	Objectives.....	6
3	QED design	8
3.1	Outcomes	8
3.2	QED design	8
3.3	Detailed QED process	9
3.3.1.	Context Mapping	9
3.3.2.	Data framework.....	9
3.3.3.	Data extraction	10
3.3.4.	University data collation and anonymisation.....	12
3.3.5.	Creating a counterfactual and match testing	13
3.3.6.	Variable construction	14
3.3.7.	Impact estimation	15
3.3.8.	Interpretation	15
3.3.9.	Reporting.....	16
3.4	Participants.....	16
3.5	Participant withdrawal criteria and procedures.....	17
4	Data management and statistical analysis	18
4.1	Statistical methods	18
4.2	Sample cleaning	18
4.3	Sample inclusion/exclusion.....	19
4.4	Data management	20
4.5	Additional analyses.....	20

4.6	Quality control & quality assurance.....	20
5	Publication	22
6	References	23
7	Appendices.....	25
7.1	Shared cross-institution data specification and user guide for enabling program data collation	25



1 Overview

Title:	The impact of enabling programs on student's initial academic success and retention in higher education: A propensity score modelling approach		
Start date:	Oct 2025	QED Finish Date	Dec 2026
Principal investigator name (and protocol contributors):	Stuart Levy , Patricia Vermillion Peirce, Mahid Yousuf, Michele Morris, Andrew Gunstone, Emma-Jaye Gavin, Karen Seary, Sarah Hattam, Suzi Syme, Jim Nyland, George Lambrinidis, Noraisha Oyama, Charmaine Davis, Fiona Navin, Ange Jones, Joanne Lisciandro, Anna Sexton, Peter Reurich, Catherine Madigan, Erin Crandon, Ilse Hogendorf, Wim Schoeman, Le Hoa Phan, Helen Wakely		
Principal investigator contact details:	stuart.levy@federation.edu.au		
Sponsor (and participating universities):	Federation University , Murdoch University, University of Southern Queensland, Central Queensland University, Adelaide University, Charles Darwin University, Southern Cross University, University of Tasmania, Edith Cowan University, Deakin University, Monash University		
QED summary (<300 words)	Enabling programs in Australia are fee-free, government-funded preparatory pathways designed to equip students with the academic skills and confidence required to succeed in higher education. Legislatively, enabling programs prepare students for study leading to a higher education award, without awarding a qualification themselves. There are 48 enabling programs nationally, with 32,579 student enrolments in 2020 (Habel et al., 2016; Department of Education, 2022). Programs vary by institution but share a focus on foundational mathematics, academic writing, study skills, and discipline-specific knowledge, with broadly comparable learning outcomes and teaching practices (Davis et al., 2023). Key elements of enabling programs include low-stakes, no-cost access and flexible study options, personalised study planning to clarify academic and career goals, enabling pedagogy and academic support, and provision of foundational academic skills. These features are intended to foster goal commitment to higher education and support students' integration into university norms, thereby improving retention and unit completion in the first and second years of study. This study uses propensity score matching to estimate the association of enabling programs on student retention and unit success, under assumptions that support causal interpretation. Domestic students who transition into higher education following enabling are matched to comparable domestic students who did not participate in enabling. Matching draws on administrative data routinely collected at application and enrolment, including variables empirically associated with program participation and		



	academic success, such as age, socio-economic status, parental education, first-in-family status, Indigenous identity, disability, alternative pathways, and Australian Tertiary Admission Rank (ATAR) or equivalent. The primary outcomes are (1) the proportion of units successfully completed within the first two years of undergraduate study and (2) retention in the first and second years of undergraduate study. Analyses of these outcomes across and between institutions, and, where sample sizes permit, by levels of cumulative disadvantage.
Ethics approval (grantor, date, reference ID)	Federation University Research Ethics 2025/240; reciprocal ethics at each participating institution.



2 Background and objective

2.1 Background

In Australia, “enabling programs” (also called enabling pathways) are fee-free, government-funded preparatory programs designed to help students gain the academic skills and confidence required to succeed in university study. Legislatively, enabling programs are provided to a person for the purpose of enabling that person to undertake a program leading to a higher education award. They are preparatory programs, do not award a higher education qualification, and are designed to equip students with skills for success in higher education.

There is some evidence available on enabling programs in Australia, and these data show some small, positive outcomes but mostly from observational studies (Lisciandro, 2022; Syme et al., 2021; Pitman et al., 2017). These studies did not compare enabling program students with others who share a similar profile of multiple, complex characteristics. This is a notable gap, given the intersectional disadvantages and heterogeneity believed often present in the enabling program’s student cohort. Consequently, there remains a need to better understand the effects of enabling programs by comparing these students to others with similar observed characteristics. Methods such as Propensity Score Matching can help address this need by matching students in enabling programs with students not in enabling programs who have a similar probability of participating, based on a set of observed variables. This approach allows for a more balanced comparison group.

Impact evaluations undertaken overseas using experimental and quasi-experimental designs have shown significant benefits from “bridging courses”. Bridging courses are similar to enabling programs in Australia insofar as they support students who may not have the requisite capabilities to succeed easily in their university study, and these courses prepare for university-level study. In the United States of America (USA), Barnett et al. (2012) found that the bridging program, which included courses in math, reading, and writing, did not lead to greater persistence in studies over a two-year period. However, over the subsequent three semesters following the bridging course, significantly more bridging students passed college-level math (increases between 5.7 to 9.4 percentage points) and writing (increases between 4.1 to 5.3 percentage points) compared to their peers. These differences were no longer statistically significant by the end of the two years, possibly due to a range of factors. Similarly in South Africa, Joynt (2022) examined a four-day bridging course in introductory accounting presented before the start of a student’s first year, which aimed to supplement students’ lack of prior accounting knowledge. Using propensity score matching, they found a positive and significant association between attending the bridging course and performance in the first assessment in accounting. In Denmark, Rosholm et al. (2019) used propensity score matching to examine the effects of a bridging program that consisted of classroom training, educational internships, and mentoring for a group of young people not in employment, education, or training. The results showed significant benefits for students in relation to increasing educational enrolment and successful completions, and these effects were particularly large for those that were assessed as “not ready for education” and with diagnosed psychological disorders.



2.2 Initiative

There are 48 enabling programs across Australia (Habel et al., 2016). In 2024, there were 16,718 students enrolled in Australian enabling programs (Department of Education, 2024). Each institution offering enabling programs (i.e. university) determines the specific age and entry requirements of students (NAEEA 2024, typology of enabling programs Australia 2023). These are broadly aligned, with a minimum age requirement (17 or 18 years old, or older) and English language proficiency.

Each institution also offers different subjects within their enabling programs, with a broad focus on academic writing, foundational mathematics, and study skills, and discipline-specific knowledge. Federation University, for instance, prepares students for degree-level study, offering introduction to tertiary studies, understanding university learning, academic writing, and a choice of mathematics subjects. Murdoch University introduces students in enabling programs to university requirements and numeracy, and the remaining elective subjects align to the students preferred disciplinary pathways (e.g. students intending to apply for the Bachelor of Nursing should complete specific electives). Although study options vary across institutions, enabling programs were found to have highly comparable program learning outcomes in foundational mathematics, academic communications, and study preparation, as well as teaching and learning practices (Davis et al., 2023).

This study examines enabling programs across eight universities, with the following core set of enabling programs' principles:

- Provides low stakes, no cost access, and flexible study options, allowing students to engage in enabling programs and “try out” higher education.
- Provides personalised study planning to each student upon enrolment, encouraging students to clarify their academic and career goals.
- Adopts an enabling pedagogy and support, equipping students to experience success in higher education.
- Provides foundational academic skills, preparing students for success in higher education while easing them into university expectations.

These core elements of enabling programs' design, consistent across these and many Australian universities, aim to help students test and confirm if a higher education pathway is relevant to them, and if so, set students up for success once they transition into an undergraduate qualification.

2.3 Objectives

Theory of change

The low stakes, no cost access, and flexible study options in enabling programs are expected to help students try higher education, and therefore make informed decisions about whether (or not) to transition to higher education. The personalised study planning and the opportunity to try higher education, is expected to test students' commitment to their goals, and result in increased retention among those who complete an enabling program and follow a higher education pathway.

Student “goal commitment” is one of the four key elements to student retention, as set out in Tinto's model of student retention (Tinto, 1975, 102): “As suggested by a number of researchers, once the individual's ability is taken into account, it is his [sic] commitment to



the goal of college completion that is most influential in determining college persistence". By providing personalised study planning and allowing students to "try out" higher education through enabling programs, individuals can commit to these goals before they enter higher education, leading to higher retention.

Further, Tinto suggested that academic integration is another key element to student retention. This may include students' ability to meet academic standards and identification of the norms of the academic system. Those students who successfully complete an enabling program and enrol in university credit-bearing programs are expected to be set up for success through the provision of foundational skills, easing them into university expectations and norms, and the enabling pedagogy supporting their self-efficacy. It is expected that this knowledge, and the acquired skills and confidence will equip them to be successful in their subsequent university qualification.

Trial purpose and objectives

Understanding the impact of enabling programs' education is essential to assess whether government-funded initiatives are delivering on their intended equity goals. The Australian Government allocates a significant amount of public funding to enabling programs to support students, particularly those from underrepresented backgrounds. As highlighted in the Australian Universities Accord Final Report (2024), improving not just access but also student success and completion is central to building a more inclusive and effective higher education system. Accordingly, knowing the impact of enabling programs on equity student success within higher education, for instance, helps guide universities in maximising and/or refining support strategies for their equity groups.

Hypothesis and Key Evaluation Questions

We hypothesise that participation in enabling programs allows students to test and confirm their commitment to a higher education pathway. At the same time, enabling programs support integration into higher education norms and expectations, build foundational academic skills, and provide support to these students. We therefore expect that students who complete an enabling program are more likely than similar non-participants to remain on their higher education pathway, and succeed at greater rates in their undergraduate studies.

Key Evaluation Questions (KEQ)

KEQ1a: What is the impact of post-school enabling programs on student pass-rates in their first two years of subsequent undergraduate study?

KEQ1b: What is the impact of post-school enabling programs on student retention in their first two years of subsequent undergraduate study?

KEQ2a: What is the impact of in-school enabling programs on student pass-rates in their first two years of subsequent undergraduate study?

KEQ2b: What is the impact of in-school enabling programs on student retention in their first two years of subsequent undergraduate study?

The evaluation will also examine variation in impact through the following secondary evaluation questions (EQs):

EQ3a: How do the impacts of enabling programs differ across institutions?

EQ3b: How might variations in delivery or contextual factors explain these differences?

EQ4: How do the impacts of enabling programs differ for students from different equity cohorts and levels of cumulative disadvantage?



3 QED design

3.1 Outcomes

The primary outcome is measured as:

1. Improvement in student success

Measure: Number of units successfully completed, divided by the number of units enrolled by census date (% pass).

Instrument: University administrative data, captured directly from the university student management system.

Approach: The student course data is extracted from each university's student management system, for each year up to two years following enrolment in an undergraduate qualification (AQF level 5-8).

The secondary outcome is measured as:

2. Improvement in student retention

Measure: Student continued enrolment at the same institutions at the end of calendar year (Yes/No)

Instrument: University administrative data, captured directly from the university student management system.

Approach: The student enrolment data is extracted from each university's student management system, for each subsequent following enrolment in the undergraduate qualification, year up to two years.

3.2 QED design

The trial design uses propensity score matching (PSM) to estimate the impact of enabling programs on retention and subject pass rates. Given the diversity of students and the non-random selection into enabling programs, PSM provides an appropriate quasi-experimental approach in this context. The approach compares enabling program graduates who progress to higher education with a matched group of similar students who did not participate in an enabling program. Analyses are conducted within institutions (stratified matching), so students are compared only with peers from the same institution. Matching variables include factors consistently associated with both enrolment in Enabling programs and subsequent outcomes, notably: classification as low socio-economic status (SES), first-in-family, disability, indigeneity; age; pathway into higher education; and ATAR score (or equivalent). Separate analyses are conducted for in-school enabling programs (KEQ2) and post-school enabling programs (KEQ1), recognising that these two groups are unique cohorts, and estimating differences of enabling programs on the student success once enrolled in undergraduate qualification. These estimates rely on the assumption that all noted cofounders are observed and sufficiently measured.



3.3 Detailed QED process

The trial follows a series of sequential steps, which are each undertaken in Semester 2 2025 and 2026.

3.3.1. *Context Mapping*

Contextual factors at each partner institution are systematically identified and documented to support structured comparison across university Enabling programs. Particular attention is given to variations in enabling program that are expected, such as:

- study mode options (full-time and/or part-time availability)
- delivery approach (online, on-campus, blended)
- program duration (e.g., three months to one year)
- focus areas (e.g. specific careers, literacy and numeracy requirements)
- requirements for successful completion of enabling program (e.g., number of subjects required to pass)
- availability of additional university academic or student supports
- availability of alternative university access pathways.

Other potential points of difference are explored with university representatives.

Contextual information is collected through a combination of semi-structured interviews with institutional representatives responsible for enabling programs, structured reviews of institutional documentation (e.g., online program guidelines), and review of relevant literature benchmarking enabling program provision across institutions.

These institutional and program characteristics that may plausibly influence student success within enabling programs are specified in advance. Where possible, these characteristics are translated into defined categories to enable comparability across institutions, qualitatively grouping universities by similar characteristics to examine potential patterns of student success.

3.3.2. *Data framework*

A coherent cross-institutional data framework is developed to ensure consistency and comparability across partner institutions. This process is led by the Federation University Business Intelligence team, and then built in collaboration with data teams at each partner institution.

An initial list of variables is defined below. These variable definitions and categories are reviewed and harmonised through structured reviews and meetings with partner institutions. Differences in terminology, coding, and classification are identified and reconciled.

A shared cross-institutional data dictionary is developed. The data dictionary specifies required variables, category definitions, coding conventions, and collation procedures, and sets out the data requirements.



3.3.3. Data extraction

The undergraduate (AQF level 5–8) qualification enrolment data from each university enrolment records are extracted in alignment with agreed definitions and coding structures, to include for each enrolment:

Undergraduate qualification enrolment

1. Undergraduate qualification commencement year: The formal undergraduate qualification commencing year identifies the reference cohort year in which a student commenced an undergraduate qualification program. It is recorded as a four-digit calendar year (e.g. 2022, 2023) and represents the commencement used to define cohort membership for the analysis. This variable is mandatory and is used to anchor all time-relative outcome measures, including Year 1 and Year 2 success and retention.
2. Undergraduate completion year: The academic year in which the student completed their undergraduate qualification, based on institutional program completion records. NULL indicates that no completion has been recorded. This variable enables identification of completion timing and supports flexible analysis, including distinguishing completion from continued enrolment and enabling optional exclusion from retention-based analyses where required.
3. Undergraduate qualification broad field of study: The high-level disciplinary classification of the student's commencing undergraduate qualification, based on the Australian Standard Classification of Education (ASCED) Broad Field of Education (2-digit level), used to enable consistent cross-institutional comparison.
4. AQF level: The Australian Qualifications Framework (AQF) level of the student's commencing undergraduate qualification (e.g., AQF Level 8—Honours, AQF Level 7—Bachelor Degree, AQF Level 6—Associate Degree, AQF Level 5—Diploma). This variable is used to support comparability and matching across different qualification levels.

Exposure to enabling programs

5. Prior enabling program enrolment: This indicator identifies whether a domestic student (including recent migrants) undertook an enabling program prior to commencing their undergraduate study, at any time. Students are classified as having an enabling enrolment if they have at least one enabling program commencement dated before their undergraduate commencement. Where multiple enabling program enrolments exist, the most recent prior enrolment is used. This variable distinguishes students with enabling pathway exposure from those who entered formal qualification directly.
6. Enabling program completion status: Enabling program completion status describes the extent to which a student completed their enabling study prior to undergraduate commencement. Local institutional completion measures are mapped to a standardised set of categories: full completion, partial completion, or NA (for those who did not enrol in enabling).
7. Type of enabling program enrolment: Students are classified as in-school enabling if their enabling program participation occurred concurrently with secondary school enrolment, and as post-school enabling if participation occurred after they had exited secondary school. This variable distinguishes two unique types of enabling programs,



and those who were exposed to in-school or post-school enabling, which are contrasted with those who entered formal qualification directly.

Outcomes (and related variables)

8. Year 1 attempted course load: Total valid attempted study load in the first academic year following commencement, reported in Equivalent Full-Time Student Load (EFTSL). Where EFTSL is not directly stored, an equivalent institutional load measure is to be converted to EFTSL using the institution's official full-time annual load definition.
9. Year 1 success rate: The Year 1 success rate represents the proportion of attempted undergraduate study successfully completed within the first year following undergraduate commencement. It is calculated as the ratio of passed load to attempted load, expressed as a percentage. Where a student has no eligible attempted load in Year 1, the success rate is recorded as missing and is not interpreted as failure.
10. Year 2 attempted course load: Total valid attempted study load in the second academic year following commencement, reported in EFTSL. Where EFTSL is not directly stored, an equivalent institutional load measure is to be converted to EFTSL using the institution's official full-time annual load definition
11. Year 2 success rate: The Year 2 success rate is calculated using the same method as the Year 1 success rate, applied to the second year following formal qualification commencement. This measure captures academic progression beyond the transition year and supports longitudinal comparisons of performance over time.
12. Year 1 retention indicator: The Year 1 retention indicator identifies whether a student remained enrolled at the same institution in the academic year following their commencement year. Students are coded as retained if they have any qualification enrolment at the same institution in the subsequent academic year and not retained if they do not. Where enrolment continuity cannot be determined from available data, the indicator is recorded as NULL, indicating that the retention outcome is not observable rather than confirmed non-retention.
13. Year 2 retention indicator: The Year 2 retention indicator captures continued enrolment at the same institution in the academic year following the first post-commencement year. It applies the same logic as the Year 1 retention indicator and reflects persistence beyond the initial transition period. Where enrolment continuity cannot be determined from available data, the indicator is recorded as NULL, indicating that the retention outcome is not observable rather than confirmed non-retention.

Note on completion and retention interpretation: Students who complete their qualification prior to the relevant retention observation period may not have subsequent enrolment records. For analytical purposes, these cases will be identified using the undergraduate completion year variable, which captures the timing of completion, and may be treated separately where required.

Matching variables

14. Low SES: This indicator identifies students classified as coming from a low socio-economic background according to the institution's local SES measure, or not. Local measures are mapped to a binary indicator using the students' first provided address



(address at application to the institution). Missing values indicate that SES information was not captured.

15. First-in-family: The first-in-family indicator identifies students recorded as the first in their immediate family to attend university. Students are coded positively where this status is explicitly recorded, and negatively otherwise. Missing values reflect institutions where this information is not collected.
16. Disability status: The disability indicator reflects whether a student has disclosed a disability through institutional reporting mechanisms. Local disclosure categories are mapped to a binary indicator where required. Missing values may reflect non-collection or non-disclosure rather than absence of disability.
17. Indigenous background: This indicator identifies students who have recorded Aboriginal and/or Torres Strait Islander status. It is derived from official student records and coded as present or not present, with missing values indicating that status was not recorded or disclosed.
18. Student age: Student age captures the age of the student at the time of formal qualification commencement. It is typically derived from date of birth and recorded in completed years at commencement. This variable provides contextual information on life stage at entry to higher education and supports analysis of differences in pathway participation, progression, and outcomes across age groups.
19. Alternative pathways into higher education: Alternative pathways into higher education identify students who entered formal qualification pathway through a non-ATAR admission route, such as enabling programs, vocational education and training pathways, or other recognised alternative entry schemes.
20. ATAR score or equivalent: ATAR score or equivalent records the Australian Tertiary Admission Rank, or a formally recognised equivalent measure, used for admission to undergraduate study. Where available, the score reflects the value recorded at the time of admission. For students admitted through pathways that do not generate an ATAR or equivalent ranking, this variable is recorded as missing. This measure provides a standardised indicator of prior academic attainment for students entering via school-based admission routes.

Note on attempted course load variables: The Year 1 attempted course load and Year 2 attempted course load variables are not outcomes in themselves, but are the denominators used in the calculation of the success rate outcomes (Year 1 success rate, Year 2 success rate). For clear interpretation of the results relating to success rates, the distribution of the attempted course load variables for the students participating in the enabling programs are examined to ensure they are statistically similar to the distribution of attempted course load variables in the matched comparison group. To ensure similarity and depending upon the results, it may be necessary to include Year 1 attempted course load and Year 2 attempted course load as additional matching variables.

These data are collated and anonymised (i.e. no university-specific student IDs, names, or personal identifiers). However, a university code is retained for each record to enable institution-level validation, anomaly checking, and stratified analysis. Definitions are checked for consistency with the shared data dictionary prior to collation.

3.3.4. University data collation and anonymisation

The collated cross-institutional dataset is received by the Business Intelligence Team, as a csv file via the AARNet-hosted secure research data transfer service—AARNet FileSender.



The dataset is reviewed for data quality and consistency. Data checks include assessment of missing covariates, completeness of follow-up outcome data, internal consistency of coding, and alignment with the shared data dictionary. Data anomalies or inconsistencies are documented and clarified with the relevant institution prior to final dataset confirmation.

Once combined, the university code is included for each student; while remaining anonymous, this allows the analysis to match students within the same institution.

3.3.5. Creating a counterfactual and match testing

A comparison group is constructed by selecting students at each institution who are enrolled in the same higher education qualification field of study and enrolment year as the enabling programs' participants in their first year of undergraduate study of an AQF5–8 award. This approach is preferred; however, if sufficient matches cannot be identified, the restriction may be relaxed to include students in the same type of qualifications or with similar credits or success rates.

All observations are pooled for analysis, but stratified matching is conducted within institutions so that students are compared only with peers from the same institution, preserving institutional context. Ideally, matching is also stratified by enrolment year so that participating cohorts are compared only to non-participating students in the same year, mitigating year-dependent influences on outcomes. Feasibility of this stratification is assessed based on available data and sample sizes. If sample size limitations arise, exact matching conditions may be relaxed, and institutional differences are controlled for by including institution as a covariate and/or interaction term in the PSM.

The propensity score is estimated using logistic regression. The dependent variable indicates whether a student participates in an Enabling program, and the independent variables include demographic factors, academic preparation, and other relevant covariates described in the Matching variables section above. The logistic regression produces a propensity score for each enabling program participant and each potential comparison student, which is then used to match participants to non-participants.

Nearest neighbour matching with replacement is the preferred matching algorithm with a restriction applied so that the propensity scores of comparison students fall within a specified distance (the "caliper") of a participating student's propensity score. The caliper width and number of nearest neighbours are determined based on match quality. If the caliper is too large, students who are less comparable in terms of participation likelihood may be included. If the caliper is too small, matched students are more similar but fewer in number, reducing statistical precision and potentially excluding a high proportion of participating students.

The aim of the matching step is to match each enabling program student to one or more comparable student who did not participate in the enabling program. Students in the comparison pool can be matched to more than one enabling program student. The matching step provides matching weights to be used in the subsequent steps to assess the quality of the match and to estimate the impact of the Enabling program. The matching weights for each enabling program student are one and the matching weights for comparison students depend on the number of times a comparison student is matched to an enabling program student.

The quality of the match is assessed using multiple diagnostics. Standardised mean differences of covariates are checked to ensure balance. The propensity score model is rerun using the matching weights to confirm that the participation coefficient is no longer significant. Diagnostic graphs, such as common support plots, are examined to ensure comparability.



If covariate imbalance is detected, the matching process is adjusted by first, modifying the propensity score model (e.g., including interaction terms), next adjusting matching parameters (calliper width or number of nearest neighbours), or relaxing matching stratification. The process is repeated until covariates are balanced and a sufficiently high proportion of participating students are included, ensuring representativeness of impact estimates.

Students who cannot be matched due to insufficient comparability are excluded from the analysis. The overlap between the participating and comparison groups defines the common support. Larger overlap increases confidence that impact estimates are representative of the full enabling programs group. The final report will include summary statistics of covariates before and after matching and diagnostic plots to demonstrate the quality of the matching process. It will also include the number of students matched/excluded and will summarise the steps taken to achieve a successful match.

Unobserved confounders cannot be directly tested. However, if omitted variables that influence participation are correlated with included covariates, the effect of unobserved factors is partially mitigated. The most challenging scenario occurs when unobserved factors influence participation and outcomes but are uncorrelated with included variables. This limitation motivates matching within the same institutional and temporal context.

The matching process is applied, where numbers allow, to the following groups:

- all students participating in post-school enabling programs
- all students participating in in-school enabling programs
- post-school enabling program students by levels of cumulative disadvantage, and for each equity cohort where numbers allow
- in-school enabling program students by levels of cumulative disadvantage, *and for each equity cohort where numbers allow*
- post-school enabling program students by institution
- in-school enabling students by institution.

These assessments inform a meeting with each institution's data team to discuss data quality and recommend the final analytic approach. If sample sizes are insufficient to achieve high-quality matches for some institutions, the sample may be augmented by grouping similar institutions together, for example using clusters identified in the Context Mapping stage.

3.3.6. Variable construction

A robust and transparent method is developed to define “cumulative disadvantage” (see EQ4) in a manner that aligns with both sector-relevant practice and statistical validity.

The first step is to count the number of equity group memberships for each student, ranging from 0 to 4 equity groups. If the number of students in higher-count categories is too small for reliable analysis, these counts are aggregated (e.g., combining 4 or more equity groups). The threshold for aggregation is determined based on the distribution of students across counts and the minimum cell sizes required for stable estimates.

Sensitivity analyses are conducted to assess whether alternative definitions of cumulative disadvantage materially affect the findings. In addition, the most common groupings across the five equity groups are identified to determine whether students cluster into specific



combinations (e.g., low SES and first-in-family). Where sample sizes permit, the impact is estimated for these common groupings of disadvantage.

This process ensures that the cumulative disadvantage measure is both analytically robust and meaningful for equity-focused decision-making.

3.3.7. Impact estimation

Once the matching process is complete, the average treatment effect on the treated (ATT) is estimated by comparing outcomes of students who participate in the enabling program (the treated) with outcomes of the matched comparison group of non-participants.

Four outcomes are examined:

- student retention at the university at the end of the first year (yes/no)
- student retention at the university at the end of the second year (yes/no)
- proportion of successful units completed at the end of the first year
- proportion of successful units completed at the end of the second year.

Second-year outcomes are not available for the 2024 undergraduate entry-cohort.

For continuous successful unit completion outcomes, the program impact is estimated by comparing the weighted mean proportion of successful units for enabling programs' participants with the weighted mean proportion for the matched comparison group using the matching weights from the matching step. For binary outcomes, the program impact is estimated by comparing the weighted mean proportion of students retained at the end of the year between the enabling programs and matched groups.

Impact estimation is conducted by fitting a weighted linear regression model, with the outcome as the dependent variable, the participation flag as the independent variable, and the matching weights as weights. The coefficient on the participation flag provides the estimated program impact.

For binary retention outcomes, program impact may also be expressed using relative risk or odds ratios rather than risk differences. In these cases, the outcome model is specified as a log-binomial or logistic regression, and g-methods are applied to estimate marginal effects, rather than directly interpreting the coefficient on the participation flag.

Impact is estimated separately for post-school enabling programs and in-school enabling programs. Where feasible, analyses are also stratified by equity group (as numbers allow), levels of cumulative disadvantage and by institution or institution cluster, depending on sample sizes and match quality.

Standard errors and confidence intervals are calculated using a bootstrapping approach, as the use of robust clustered standard errors is complicated by replacement and the inclusion of multiple comparison students in the matching process.

3.3.8. Interpretation

The reporting process is designed to minimise interpretation bias by ensuring that results are considered within the appropriate institutional and student-context frame. Each institution's findings are first presented in a tailored set of PowerPoint slides and discussed in sense-making sessions with that institution.



Separate results are presented for in-school and post-school enabling programs (KEQs 1–2), as well as for equity students and students experiencing cumulative disadvantage (EQ4). For each institution, the discussion:

- Explores how the institution’s enabling program approach supported or hindered outcomes for these groups.
- Identifies context-specific mechanisms that may explain differential impacts.

Each interpretation session, one with each contributing university (n=8), focus on interpreting results against the key evaluation questions. A consistent facilitation approach is applied across all institutions to ensure questions and prompts are uniform, reducing the risk of inconsistent framing or biased interpretation. This process includes inviting the relevant co-investigators to a university-specific discussion, summarising the intention of the discussion, and reiterating the discussion and their data are shared only with them; reviewing the study approach (as a reminder), and then going over the findings relevant to their institution, against each KEQ. The co-investigators will work together to interpret the findings in light of their specific enabling program. The investigation team then come together as a group and share their interpretations of the findings for the report. By combining quantitative analysis with a structured interpretation process, alongside the pre-defined context mapping, the approach reduces the risk of confirmation bias, selective reporting, or over-attribution of effects to program participation.

Qualitative insights gathered during these sessions are synthesised and shared with the National Association of Enabling Educators of Australia (NAEEA), alongside the collated quantitative findings, to enable sector-wide interpretation of the implications for equity groups.

3.3.9. Reporting

The study is reported using the ACSES report template, and adopting the ABS statistical disclosure control practices, results will be reported where there are five or more individuals per cell. This threshold applies to reporting outputs only.

3.4 Participants

The research population includes all domestic students commencing undergraduate enrolment at partner universities in 2022, 2023, and 2024. This includes students entering programs (AQF 5–8 undergraduate diplomas, associate degrees, and undergraduate degrees) through post-school and in-school enabling pathways, as well as students who do not participate in enabling programs but are otherwise enrolled in the same courses and terms. The population captures students across all institutional contexts, program types, and modes of study (full-time, part-time, online, and on-campus), providing a comprehensive basis for constructing matched comparison groups and estimating the impact of enabling programs on student retention and success outcomes.

Power and sample size

The study is sufficiently powered to detect meaningful changes in the primary outcome, namely the percentage of units successfully completed—namely “student success” in KEQ 1—and the priority focus on post-school secondary enabling programs. Previous research has indicated that bridging programs can improve subject pass rates by approximately five percentage points (Barnett et al., 2012). The team conservatively assumed pooled standard deviation of 20 percentage points, resulting in an estimated effect size of $d = 0.25$. Using G*Power, a priori power analysis ($\alpha = 0.05$, 80% power) indicates the need for approximately



400 matched students (200 enabling program students and 200 matched comparators) under simple random sampling assumptions. The matching procedure is expected to induce variance inflation (a PSM design effect) due to the unequal weights. To account for this in a priori planning, we applied a conservative design effect of 1.5, increasing the required sample size to approximately 600 students. Final estimates use bootstrap methods that capture the variance included by the matching procedures. Missing outcome data are not expected, as analyses rely solely on standard administrative data.

Students who complete an enabling program and enrol in an award program form the intervention group. Other students with similar characteristics are statistically matched to these participants to form the control group within each institution. The numbers available for this study are presented below.

3.5 Participant withdrawal criteria and procedures

The study uses only de-identified administrative data, with no direct contact or intervention, and no new data collection. Should a university choose to withdraw from the study then their administrative data may be withdrawn. University level data may be withdrawn up to the point where counterfactuals are created, and match testing is undertaken.



4 Data management and statistical analysis

4.1 Statistical methods

The average treatment effect on the treated (ATT) is estimated by comparing outcomes of participants in enabling programs with those of matched non-participants.

For continuous outcomes, the weighted mean proportion of successful units is compared between participants in enabling programs and the matched comparison group.

For binary outcomes, the weighted proportion of students retained at the end of the year is compared between Enabling participants and the matched group. Impact may also be expressed as risk differences, relative risk, or odds ratios; log-binomial or logistic regression and g-methods are applied if marginal effects are needed.

Statistical inference is based on point estimates and 95% confidence intervals, calculated using bootstrapping. Bootstrapping is used for the confidence intervals because it provides a pragmatic way to quantify uncertainty across both stages of the estimation process (the propensity score model and the calculation of ATT). Another option is to use robust clustered standard errors from the ATT estimation step, however this ignores the uncertainty associated with the PSM and it is also complicated by replacement and the inclusion of multiple comparison students in the matching process.

The bootstrapping involves drawing random samples with replacement from the original dataset and running through the full estimation process multiple times (the number of replications will be determined based on the stability of the confidence intervals). The PSM used in the bootstrapping process will be the final version (including stratified matching) that resulted in a balanced matched sample using the original dataset.

Statistical significance is assessed at $\alpha = 0.05$.

4.2 Sample cleaning

The sample is systematically cleaned prior to analysis to ensure accuracy, completeness, and consistency. Cleaning procedures are applied to all student enrolment and outcome data, following these steps:

- Identification of missing data: All variables are screened for missing values. The proportion and pattern of missingness are documented for each variable. If missing data are detected in key covariates or outcomes, the source is investigated with the relevant institutional data team to determine whether the data can be recovered.
- Identification of false data: Records are checked for implausible or inconsistent values, including duplicate entries, impossible dates, or contradictory demographic or enrolment information. Any spurious records are flagged and verified with the relevant institution before being excluded or corrected.
- Consistency checks: Variables are examined for consistency across related fields (e.g., enrolment status, course codes, unit outcomes). Discrepancies are resolved in consultation with institutional data teams.

All cleaning steps, decisions, and exclusions are logged in a transparent data cleaning record to support reproducibility and auditability. Following this process, the resulting dataset is deemed complete, accurate, and ready for the construction of matched comparison groups and subsequent impact analysis.



4.3 Sample inclusion/exclusion

Our sample includes all student enrolments into an undergraduate qualification at the university in 2022, 2023, and 2024. This includes students that enrolled via enabling and non-enabling pathways.

About 14,000 enabling program enrolments are available, and we will include only those that have later enrolled in a formal qualification; this is expected to be approximately 40% of these students in enabling programs. Although our challenge will be in finding the match for these students in enabling programs amongst the broader population of enrolments, we are confident that the numbers of enrolments across this number of universities will provide more than ample data.

Table 1: Student enrolments in enabling programs (head count) in undergraduate studies at each university (rows) by cohort enrolment years and enabling pathway (columns), (with predicted numbers of equity student in brackets).

Institution	Post-school enabling programs (KEQ1)		In-school enabling programs (KEQ2)	
	2022–2023 cohorts	2022–2024 cohorts	2022–2023 cohorts	2022–2024 cohorts
Federation University	171 (85)	291 (145)		
Murdoch University	630 (315)	991 (495)	905 (452)	1138 (569)
Central Queensland University	903 (451)	1342 (671)		
Charles Darwin University	291 (145)	400 (200)		
University of South Australia	791 (495)	1118 (559)		
University of Southern Queensland	3221 (1610)	4802 (2401)	214 (107)	475 (237)
Southern Cross University	566 (283)	813 (406)		
Edith Cowan University	1651 (825)	2604 (1302)	149 (74)	268 (134)
Total	8224 (4209)	12361 (6180)	1268 (633)	1881 (940)



4.4 Data management

Data for this study are extracted locally by authorised personnel at each participating institution in accordance with institutional governance processes and the approved Data Specification Guide. Extraction is limited to the defined 2022–2024 commencing undergraduate cohorts and to variables required for the evaluation.

At the point of collection within institutional systems, data are re-identifiable (e.g., student IDs). Prior to transfer, each institution de-identifies its dataset by removing direct identifiers and replacing original student IDs with randomly generated unique study IDs. Federation University does not receive any re-identification keys. A university code is retained to support institution-level validation and stratified analysis.

Data are transmitted from participating universities to Federation University using AARNet-hosted secure research data transfer service that provides encrypted transmission and controlled access. Upon receipt, datasets are stored within Federation University's secure SharePoint environment (Microsoft 365 infrastructure).

Access to collated data is restricted through role-based permissions. Federation University named investigators and the lead analyst have access to the full collated dataset. Co-investigators from partner institutions have access only to their institution's data. SharePoint administrators have system-level access solely for platform security and permission management and are not involved in the research.

Data preparation and coding for analysis are conducted on the secure, password protected device, ensuring restricted access and using reproducible scripts. Derived variables (e.g., age at commencement, success rates, retention indicators, enabling program exposure classifications) are constructed in accordance with the shared data dictionary. SharePoint version control and usage history provide an audit trail of data handling and transformation.

During analysis, data are treated as non-identifiable. Findings are reported in aggregate form only. Small cell sizes ($n < 5$) are suppressed in reporting to minimise any residual risk of re-identification.

De-identified datasets will be securely retained for a minimum of five years following publication of project outputs, consistent with Federation University's Research Data Management Procedure and the National Statement on Ethical Conduct in Human Research (2023). After this period, data will be permanently erased from SharePoint in accordance with approved institutional deletion processes.

4.5 Additional analyses

Where sample size allows, impact estimates are stratified to examine:

- variation by disadvantage (indigeneity, low SES, first-in-family, disability)
- variation in levels of cumulative disadvantage (EQ4)
- variation across institutions or clusters of institutions (EQ3)
- differences between online and in-person delivering modes of enabling programs

These stratified analyses support exploration of heterogeneous program effects and help identify context-specific mechanisms influencing outcomes.

4.6 Quality control & quality assurance



All datasets will undergo structured verification prior to sending the dataset to the analyst, including checks for completeness, consistency, range errors, duplicates, and logical inconsistencies. A documented data cleaning log will record all transformations, exclusions, and derived variables. Final analytic datasets will be locked prior to outcome analysis.

All data management and statistical analyses are conducted using scripted code to ensure full reproducibility. This final code will be published with the results. Version control processes will maintain an audit trail of analytic decisions and revisions.

Adherence to the approved protocol will be monitored throughout. Any deviations will be documented and justified.



5 Publication

Dissemination will include a report for the ACSES Trials Registry, presentations at relevant conferences (e.g. the Student Success Conference, NAEAA), and a peer-reviewed article detailing the findings of the study.



6 References

- Andrewartha, L., & Harvey, A. (2014). Willing and enabled: The academic outcomes of a tertiary enabling program in regional Australia. *Australian Journal of Adult Learning*, 54(1), 50–68.
- Araque, F., Roldán, C., & Salguero, A. (2009). Factors influencing university drop out rates. *Computers & Education*, 53(3), 563–574.
- Australian Centre for Student Equity and Success. (n.d.). Interactive Equity Data Tool. Retrieved June 25, 2025, from <https://www.acses.edu.au/data/interactive-tool/>
- Barnett, E. A., Bork, R. H., Mayer, A. K., Pretlow, J., Wathington, H. D., & Weiss, M. J. (2012). *Bridging the Gap: An Impact Study of Eight Developmental Summer Bridge Programs in Texas*. National Center for Postsecondary Research.
- Davis, C., Syme, S., Cook, C., Dempster, S., Duffy, L., Hattam, S., Lambrinidis, G., Lawson, K., Levy, S. (2023). Report on Benchmarking of Enabling Programs across Australia to the National Association of Enabling Educators of Australia (NAEEA). ISBN: 978-0-646-87548-4. <https://enablingeducators.org/wp-content/uploads/2023/03/Report-on-Benchmarking-of-Enabling-Programs-Across-Australia-2023.pdf>
- Department of Education (2024). Key findings from the 2024 Higher Education Student Statistics. In chapter: Domestic commencements in 2024 trending back towards pre-pandemic levels, <https://www.education.gov.au/higher-education-statistics/student-data/selected-higher-education-statistics-2024-student-data/key-findings-2024-higher-education-student-statistics>
- Joynt, C. (2022). How to assess the effectiveness of accounting education interventions: Evidence from the assessment of a bridging course before introductory accounting. *Meditari Accountancy Research*, 30(7), 237–255.
- Habel, C., K. Whitman, and J. Stokes. 2016. Exploring the Experience of Low-SES Students via Enabling Pathways. Perth: National Centre for Student Equity in Higher Education.
- Lassibille, G., Navarro Gómez, M.L. (2009) Tracking students' progress through the Spanish university school sector. *High Educ* 58, 821–839. <https://doi.org/10.1007/s10734-009-9227-8>.
- Lisciandro, J. G., & Gibbs, G. (2016). OnTrack to university: Understanding mechanisms of student retention in an Australian pre-university enabling program. *Australian Journal of Adult Learning*, 56(2), 198–224.
- Müller, S., & Schneider, T. (2013). Educational pathways and dropout from higher education in Germany. *Longitudinal and Life Course Studies*, 4(3), 218–241.
- O’Kane, M., Behrendt, L., Glover, B., Macklin, J., Nash, F., Rimmer, B., & Wikramanayake, S. (2024). *Australian Universities Accord: Final report*. Australian Government Department of Education. <https://www.education.gov.au/australian-universities-accord/resources/final-report>
- Pitman, T., Trinidad, S., Devlin, M., Harvey, A., Brett, M., & McKay, J. (2016). *Pathways to higher education: The efficacy of enabling and sub-bachelor pathways for disadvantaged students*. National Centre for Student Equity in Higher Education.



- Rosholm, M., Mikkelsen, M. B., & Svarer, M. (2019). Bridging the gap from welfare to education: Propensity score matching evaluation of a bridging intervention. *PLOS ONE*, 14(5), e0216200.
- Syme, S., Roche, T., Goode, E., & Crandon, E. (2021). Transforming lives: the power of an Australian enabling education. *Higher Education Research & Development*, 41(7), 2426–2440. <https://doi-org.ezproxy.scu.edu.au/10.1080/07294360.2021.1990222>
- Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research*, 45(1), 89–125. <https://doi.org/10.3102/00346543045001089>



7 Appendices

Please note that appendices may use the term “Enabling” when referring to enabling programs and enabling pathways.

7.1 Shared cross-institution data specification and user guide for enabling program data collation

1. Rationale

The purpose of this document is to provide guidance to participating institutions on how to prepare and submit data for the **Enabling Programs Impact Evaluation**, and to provide a detailed data dictionary for required and recommended data fields.

This guide supports the Shared Cross-Institution Data Specification by:

- explaining how data should be structured for submission,
- clarifying common questions about cohort scope and outcome derivation,
- standardising definitions across institutions,
- ensuring alignment with the approved Trial Protocol and Propensity Score Matching (PSM) design.

This document is intended for institutional data analysts, business intelligence teams, and reporting units responsible for data extraction and preparation.

2. Enabling Programs Impact Evaluation – Data Collation Overview

The Enabling Programs Impact Evaluation examines the impact of Enabling participation on:

- Academic success (success rate)
- Retention in undergraduate study (continuation or completion)

The design uses Propensity Score Matching (PSM) to compare:

- Students who participated in Enabling prior to undergraduate commencement
- Matched students who did not participate in Enabling

Each participating institution is required to submit a single analytical dataset with:

- One row per student.
- Defined undergraduate commencing cohorts (2022, 2023, 2024)
- Enabling exposure and completion data
- Academic outcomes (Year 1 and Year 2)
- Matching variables (demographic and admission covariates)



3. Cohort Definition:

Unless otherwise specified, the reference cohort includes **domestic students commencing an undergraduate qualification (AQF Levels 5-8) in 2022, 2023, or 2024.**

A **valid commencement** is defined as enrolment retained beyond the census date of the first enrolled teaching period.

A student is in scope if they meet all of the following criteria:

Criterion	Specification
Academic level	Undergraduate qualification (AQF levels 5-8) offered by the university
Commencement	First valid enrolment in an undergraduate qualification in 2022, 2023 or 2024.
Student type	Domestic students only
Delivery	Excludes partner-delivered programs
Reference year	Cohort enrolment year – 2022, 2023 or 2024

4. Data Submission Structure

This section describes the analytical structure and design of the dataset. The dataset represents a student-level analytical file constructed for the purposes of the Enabling Program Impact Evaluation.

Where applicable, variables are conceptually aligned with TCSI elements to support consistency across institutions. However, this dataset represents a derived analytical dataset rather than a direct extract of TCSI records.

Some variables (particularly academic outcomes such as success rates and retention indicators) are analytically derived using underlying TCSI-aligned data elements (e.g. EFTSL and unit completion status) and may not correspond directly to individual TCSI elements. This approach ensures alignment with sector-standard definitions while allowing flexibility for institution-specific derivation where required.

Analytical Unit:

- One row per student.
- Each student is represented once only, based on their earliest valid undergraduate (AQF Levels 5-8) commencement within 2022–2024.

Cohort Anchoring:

- Each student is assigned to a single cohort year (2022, 2023 or 2024), based on their first valid undergraduate commencement.
- All outcomes (Year 1 and Year 2) are calculated relative to that commencement year.

Dataset Domains:

The submitted dataset is expected to include variables grouped across the following domains:

1. Identifiers & Cohort



- De-identified Student ID
- Undergraduate Commencement Year
(aligned to TCSI element E534)
- Undergraduate Commencement Term
(standardised term type description [e.g. Semester 1, Semester 2]; supports alignment with commencement timing)
- AQF Level of Commencing Qualification
(aligned with AQF classification standards)
- Broad Field of Education
(ASCED 2-digit level, derived from TCSI element E461)

2. Enabling Exposure & Completion

- Prior Enabling Enrolment Indicator
(derived using institutional enrolment history; may be informed by TCSI elements such as E310 [Course Type] and E327 [Basis for Admission])
- Type of Enabling Enrolment (in-school/post-school)
(institution-derived classification)
- Enabling Program Completion Status (full/partial)
(institution-derived)

3. Academic Outcomes

- Undergraduate Completion Year
(derived from institutional completion records; conceptually aligned with TCSI completion reporting but not a direct TCSI element)
- Year 1 Attempted Course Load (EFTSL)
(derived using TCSI elements E339 [EFTSL] and E355 [Unit of Study Completion Status])
- Year 1 Success Rate
(derived using EFTSL and unit completion status (E355); no direct TCSI equivalent)
- Year 2 Attempted Course Load (where available)
(same derivation as Year 1 using E339 and E355)
- Year 2 Success Rate (where available)
(same derivation as Year 1; no direct TCSI equivalent)
- Year 1 Retention Indicator
(derived using institution-level retention/attrition logic informed by TCSI-aligned enrolment data; no direct TCSI element)
- Year 2 Retention Indicator (where available)
(same logic as Year 1)

Academic outcome variables are analytically derived using underlying TCSI-aligned data elements (e.g. EFTSL and unit completion status) and institution-level derivation logic. While these variables do not correspond directly to individual TCSI elements, their construction is informed by TCSI definitions and reporting standards, supporting consistency and comparability across institutions.

4. Matching Variables (PSM Covariates)

- Equity Indicators
(derived from TCSI-aligned demographic fields, e.g. E316 [Indigenous Status], SES indicators; institution-level mapping applied)



- Student Age
(derived from date of birth; consistent with TCSI age reporting conventions but not a direct element)
- Alternative Pathway Indicator
(aligned to TCSI element E327)
- ATAR Score
(aligned to TCSI element E632)

The dataset may be implemented as a single flattened student-level file. Separate internal preparation files may be used locally; however, the submitted file should reflect the analytical structure described above.

5. Frequently Asked Questions

Question	Answer
Which year is this collection for?	This collection includes students who commenced undergraduate study (AQF Levels 5–8) in 2022, 2023, or 2024. Each student is assigned to a cohort based on their commencement year. Year 1 corresponds to the undergraduate commencement year, and Year 2 corresponds to the following academic year. Year 2 outcomes for the 2024 cohort are not yet observable. Academic year refers to the institution's reporting year of enrolment (e.g. 2022, 2023) and is used as the common reference point for all students. It does not represent a rolling 12-month study period from the commencement date.
Do we need to include international students?	No. The reference cohort focuses on domestic students only, in line with the evaluation design.
What if a student has multiple enrolments into formal qualifications?	To ensure consistency across institutions, only the earliest valid undergraduate commencement within the 2022–2024 cohort period should be retained. This ensures each student appears once in the dataset.
What counts as an enabling program?	Enabling programs are preparatory or pathway programs that support transition into undergraduate study and do not confer a higher education qualification. Where applicable, enabling enrolments may be identified using TCSI elements (e.g. E310 – Course Type, E327 – Basis for



Question	Answer
	Admission); however, institutional coding practices may vary. Institutions should apply their local definition of enabling programs and document the mapping in the Local Mapping Worksheet, including classification as in-school or post-school where relevant.
Does the timing between enabling and undergraduate study matter?	Prior enabling enrolment includes any qualifying enabling participation that occurred before undergraduate commencement, regardless of the time elapsed. The dataset does not explicitly capture the time gap between enabling and undergraduate study.
What if we cannot supply all demographic variables?	If a variable is not available in institutional systems, it may be recorded as NULL. Any such limitations can be documented in the Local Mapping Worksheet to support transparency in the analysis.
Do we need to provide raw unit or EFTSL data?	No. The dataset is intended to include student-level derived outcomes rather than unit-level records. These outcomes include: • Year 1 and Year 2 attempted load • Year 1 and Year 2 success rate • Year 1 and Year 2 retention indicators Outcomes are calculated using EFTSL as the standard measure of load. If EFTSL is not directly stored, an equivalent institutional load measure may be converted to EFTSL using the institution's official definition of a full-time annual load.

6. Data File Requirements

This section outlines the **technical and operational requirements** for file preparation and submission.

File Format

CSV (.csv) is the preferred format. Excel (.xlsx) may be accepted where required, provided no formatting or data type issues are introduced.

File Content

- One row per student
- De-identified Student ID only
- No direct identifiers (e.g., name, student number, date of birth)
- Cohort anchored to first valid undergraduate commencement (2022–2024)
- Outcomes derived at the student level (not unit level)



- Column names aligned with the Data Dictionary

De-identification

At the point of extraction within institutional systems, data are re-identifiable (e.g. student IDs). Prior to transfer, each institution de-identifies its dataset by removing direct identifiers and replacing original student identifiers with randomly generated unique study IDs.

Federation University does not receive any re-identification keys. Re-identification keys remain securely held within the originating institution. A university code is retained for each record to support institution-level validation, anomaly checking, and stratified analysis, while maintaining student anonymity.

Coding Conventions

Binary variables are coded consistently as:

1 = Yes / presence of the attribute

0 = No / absence of the attribute

Context-specific meanings (e.g. retained, non-ATAR) are defined within each field description.

NULL values represent data that are unavailable, not applicable, or not observable, depending on the variable context.



Handling of Missing Data

If a variable is not available within institutional systems, it may be recorded as NULL. Any limitations can be documented in the Local Mapping Worksheet to support transparency during analysis.

Secure Transfer

Datasets are transmitted from participating universities to Federation University via the AARNet-hosted secure research data transfer service - [AARNet FileSender](#), which provides encrypted transmission and controlled access. Most Universities have the ability to log in to this portal and send files securely.

Upon receipt, datasets are stored within Federation University's secure Microsoft 365 SharePoint research environment, with access managed through role-based permissions in accordance with institutional research data governance processes.

7. Field Groups

Variables are grouped into the following domains:

1. **Cohort & Qualification Information**
2. **Enabling Exposure & Completion**
3. **Academic Outcomes (Load, Success, Retention)**
4. **Matching Variables (PSM Covariates)**

These domains reflect the layered analytical structure used in the evaluation design.



8. Spreadsheet 1 – Cohort, Enabling and Academic Outcomes Data

Spreadsheet 1 captures the core analytical dataset used for impact estimation. It contains one row per student and includes cohort identifiers, enabling exposure and completion, and derived academic outcomes.

Field Group	Field Name	Definition / Description
Cohort & Qualification Information	[De-identified_Student_ID]	De-identified student identifier (no direct identifiers included)
	[Undergraduate_Qualification_Commencement_Year]	Undergraduate (AQF Levels 5–8) commencement year (2022, 2023 or 2024)
	[Undergraduate_Qualification_Commencement_Term]	Standardised term type descriptions (e.g. Semester 1, Semester 2, Trimester 1, Winter, Summer)
	[Undergraduate_Qualification_AQF_Level]	AQF Level of the commencing undergraduate qualification (e.g., AQF 5, 6, 7, 8)
	[Undergraduate_Qualification_Broad_Field_of_Study]	Australian Standard Classification of Education (ASCED) 2-digit broad field classification code of the commencing qualification
Enabling Exposure & Completion	[Prior_Enabling_Enrolment_Indicator]	1 = Prior enabling enrolment before UG commencement; 0 = No enabling enrolment
	[Type_of_Enabling_Enrolment]	In-School / Post-School / NULL (if no enabling enrolment or not available)
	[Enabling_Program_Completion_Status]	Full completion / Partial completion / NULL (if no enabling enrolment or



Field Group	Field Name	Definition / Description
		completion status not available)
Academic Outcomes (Load, Success, Retention, Completion)	[Undergraduate_Completion_Year]	Formal year in which the student completed their undergraduate qualification, derived from the effective completion/progression date recorded in institutional systems. This may differ from the academic year of the retention record and may fall outside the Year 1 / Year 2 observation window. NULL if completion is not observed.
	[Y1_Attempted_Course_Load]	Total valid attempted load in Year 1 (reported in EFTSL)
	[Y1_Success_Rate]	$(\text{Passed Load} \div \text{Attempted Load}) \times 100$ for Year 1; NULL if no eligible attempted load
	[Y2_Attempted_Course_Load]	Total valid attempted load in Year 2 (EFTSL), where available
	[Y2_Success_Rate]	Calculated as per Year 1, applied to Year 2
	[Y1_Retention_Indicator]	Indicates whether the student was retained following the commencement year (Year 1). 1 = Retained (i.e. continues their study following the commencement year or completes their qualification) 0 = Not retained (i.e. did not continue and did not complete)



Field Group	Field Name	Definition / Description
		NULL = Not observable
	[Y2_Retention_Indicator]	Indicates whether the student was retained following the second academic year (Year 2). 1 = Retained (i.e. continues their study following the second academic year or completes their qualification) 0 = Not retained (i.e. did not continue and did not complete) NULL = Not observable



9. Spreadsheet 2 – Matching Variables (PSM Covariates)

Spreadsheet 2 captures demographic, equity, and admission-related variables used in propensity score modelling. Where a variable is not available within institutional systems, it may be recorded as NULL and documented in the Local Mapping Worksheet.

Field Group	Field Name	Definition / Description
Matching Variables (PSM Covariates)	[Low_Socioeconomic_Status]	1 = Classified as low SES under institutional measure; 0 = Not low SES; NULL = Not captured
	[First_in_Family]	1 = Recorded as first in family to attend university; 0 = Not first in family; NULL = Not captured
	[Disability_Status]	1 = Disability disclosed; 0 = No disclosure; NULL = Not captured
	[Indigenous_Background]	1 = Aboriginal and/or Torres Strait Islander; 0 = Non-Indigenous; NULL = Not stated / unknown or not captured
	[Student_Age]	Age in years at undergraduate commencement
	[Alternative_Pathway_Into_HE_Flag]	1 = Non-ATAR pathway; 0 = ATAR entry; NULL = Not recorded
	[ATAR_Score_or_Equivalent]	Recorded ATAR (or equivalent) at admission; NULL if non-ATAR pathway or not available

10. Data Specification

Spreadsheets 1 and 2 together form the required dataset. A comprehensive, field-level data dictionary—covering definitions, permissible values, and implementation guidance for each requested field—is provided in the technical appendix at the end of this document.

Participating institutions are encouraged to:

- Apply definitions consistently across their submission
- Use standardised coding as outlined in the data dictionary
- Document any local derivation logic or assumptions in the Local Mapping Worksheet



11. Data Quality and Validation

Participating institutions are encouraged to undertake reasonable internal validation checks prior to submission to support consistency, comparability, and analytical integrity across the multi-institution dataset. Where applicable, institutions are encouraged to align variable derivations with TCSI definitions and ensure that post-census enrolment logic is consistently applied.

Structural Validation

Institutions are encouraged to confirm that:

- Each spreadsheet contains one row per student.
- No duplicate student identifiers are present.
- All requested fields are included in the submission, even where values are recorded as NULL.

Coding and Value Consistency

To support cross-institution comparability, institutions are encouraged to:

- Apply binary coding consistently (e.g. 1 = Yes / 0 = No).
- Record missing or unavailable values as NULL rather than leaving fields blank.
- Ensure that numerical fields fall within reasonable ranges (e.g. ATAR values, age, success rates between 0 and 100).
- Confirm that attempted load values are non-negative and reported in EFTSL.
- EFTSL represents study load rather than a simple count of courses and accounts for unit weighting and intensity of study.

Logical Consistency Checks

Basic internal consistency checks are encouraged prior to submission. For example:

- Enabling completion status should align with the enabling enrolment indicator.
- Success rates should be derived from the reported attempted load values.
- Retention indicators should align with the institution's year-level retention or attrition framework used for the submission.
- ATAR scores would typically be present for ATAR-entry students and recorded as NULL for non-ATAR pathways.

These checks are intended to support interpretability and reduce the need for post-submission clarification.

Documentation of Local Derivations

Where institutional systems require local derivations (e.g. calculation of attempted load, age at commencement, or institution-specific mapping of enabling pathways), institutions are encouraged to document:

- Any assumptions used
- Any deviations from the suggested derivation approach



- Known limitations in data capture

This documentation can be recorded in the Local Mapping Worksheet to support transparency during pooled analysis.

12. Contact Details & Version Control

For questions regarding data preparation or interpretation of this guide, institutions should contact:

Enabling Programs Study – Data & Insights Team

Email: m.yousuf@federation.edu.au

CC: stuart.levy@federation.edu.au ;
p.vermillionpeirce@deakin.edu.au

Version	Author	Release Date	Notes
V1.0	Enabling Programs Study Team	27/02/2026	Initial user input guide and data dictionary
V2.1	Enabling Programs Study Team	27/03/2026	Updated Data Specification and User guide



13. Technical Appendix

Field-Level Data Dictionary and Implementation Guidance

This appendix provides detailed field-level specifications for each variable included in Spreadsheets 1 and 2. It outlines definitions, expected values, derivation guidance, and implementation notes to support consistent interpretation across participating institutions. Where local systems require alternative derivation approaches, these may be documented in the Local Mapping Worksheet.



Appendix A: Spreadsheet 1 – Cohort, Enabling and Academic Outcomes

A1. Cohort & Qualification Information

Field Name	Description	Format	Expected Values	Implementation Notes
De-identified_Student_ID	De-identified student identifier generated by the institution prior to submission.	Text	Unique value per student	Institutions are to generate a de-identified unique student identifier and retain the mapping key securely. The Enabling Programs Study team will receive only de-identified data; any re-identification will be undertaken by the originating institution using the retained key.
Undergraduate_Qualification_Commencement_Year	Academic year in which the student commenced their undergraduate qualification.	Integer	2022, 2023, 2024	Aligned to TCSI element E534 (Course Commencement Year). Refers to first valid enrolment in any undergraduate qualification (AQF Levels 5–8). A valid undergraduate commencement is defined as enrolment retained beyond the census date of the first enrolled teaching period (i.e. post-census enrolment), resulting in eligible attempted load. Students who enrol but withdraw prior to census (and therefore generate no eligible attempted load) should be excluded from the commencement cohort.
Undergraduate_Qualification_Commencement_Term	The teaching period in which the student commenced their undergraduate qualification, reported as a term type description (e.g. Semester 1, Semester 2, Trimester 1, Winter, Summer).	Integer / Text	Standardised term type descriptions (e.g. Semester 1, Semester 2, Trimester 1, Winter, Summer)	Derived from institutional commencement fields. Institutions should map local term codes to standardised term type descriptions to support cross-institution comparability.

Field Name	Description	Format	Expected Values	Implementation Notes
				Used alongside commencement year to define cohort entry timing (e.g. Semester 1 vs Semester 2). No direct TCSI equivalent.
Undergraduate_Qualification_AQF_Level	AQF Level of the commencing undergraduate qualification	Integer	5, 6, 7, 8	Aligned with AQF classification standards. Included to support comparability across qualification types. Not a direct TCSI element but consistent with sector reporting frameworks.
Undergraduate_Qualification_Broad_Field_of_Study	Broad field of education for the commencing qualification.	Text / Code	ASCED 2-digit classification (01–12)	Derived from TCSI element E461 (Field of Education – ASCED), aggregated to the 2-digit broad field classification for consistency across institutions. Where internal qualification classifications differ, they are to be mapped to the relevant ASCED 2-digit Broad Field code prior to submission.

A2. Enabling Exposure & Completion

Field Name	Description	Format	Expected Values	Implementation Notes
Prior_Enabling_Enrolment_Indicator	Indicates whether the student had any enabling enrolment prior to undergraduate commencement.	Integer	1 = Yes; 0 = No	Derived using institutional enrolment history prior to undergraduate commencement. “Prior enabling” refers to any enabling program enrolment that occurred before the student’s first undergraduate (AQF Level 5-8) commencement within the study cohort. This may include enrolment in a previous



Field Name	Description	Format	Expected Values	Implementation Notes
				academic year or earlier within the same year. No specific time window is applied — all prior enabling enrolment is considered, regardless of how long before undergraduate commencement it occurred. Enrolment must be retained beyond census to be considered valid. Where multiple enabling enrolments exist, only enrolment occurring prior to undergraduate commencement is considered. May be informed by TCSI elements such as E310 (Course Type) and E327 (Basis for Admission), where applicable
Type_of_Enabling_Enrolment	Classification of enabling pathway	Text	In-School / Post-School / NULL	Institution-derived classification. In-School refers to enabling undertaken concurrently with secondary schooling; Post-School refers to enabling undertaken after secondary completion. NULL where no enabling enrolment exists.
Enabling_Program_Completion_Status	Indicates whether a student fully or partially completed their enabling program prior to undergraduate commencement.	Text	Full completion / Partial completion / NULL	Derived from institutional program completion data. Full completion reflects satisfaction of all requirements; partial completion reflects enrolment without completion. NULL where no enabling enrolment exists.

A3. Academic Outcomes (Load, Success, Retention, Completion)

Year 1 and Year 2 are defined relative to the student's undergraduate commencement year:



- Year 1 = the undergraduate commencement year
- Year 2 = the following academic year

Academic year refers to the institution's reporting year of enrolment (e.g. 2022, 2023), regardless of when within the year a student commenced. It does not represent a rolling 12-month period from the student's start date.

These are cohort-based time periods and are not based on curriculum structure or a standard full-time study load (e.g. 1 EFTSL).

Retention is assessed at the academic-year level and reflects whether a student continues their study in the following academic year or completes their qualification. The outcome is assigned to the year being measured (Year 1 or Year 2), based on what occurs after that year.

Field Name	Description	Format	Permissible Values	Implementation Notes
Undergraduate_Completion_Year	Formal year in which the student completed their undergraduate qualification.	Integer	Valid academic year of completion within the institution's observable data period (2022-2026) or NULL	Derived from institutional completion records. Represents formal completion year rather than enrolment academic year. May fall outside Year 1 / Year 2 observation window. NULL indicates no completion observed within available data. Used to support interpretation of retention outcomes (completion vs continuation). Conceptually aligned with TCSI completion reporting but not a direct TCSI element. Validation analysis indicates that the formal completion year aligns with the academic year of completion in the majority of cases, with minor variation (typically within one year) attributable to administrative processing timelines.
Y1_Attempted_Course_Load	Total EFTSL from all attempted units within the student's undergraduate	Decimal (EFTSL)	≥ 0 or NULL	Derived using TCSI elements E339 (EFTSL load) and E355 (Unit of Study Completion Status).



Field Name	Description	Format	Permissible Values	Implementation Notes
	enrolment in the undergraduate commencement year (Year 1).			Includes units with status: 1 (withdrawn without academic penalty), 2 (failed), 3 (completed/passed), and 6 (withdrawn due to medical reasons), where enrolment is retained beyond census. Units withdrawn prior to census are excluded. Work-integrated learning (WIL) units are excluded where identifiable. NULL indicates no eligible attempted load or outcome not observable.
Y2_Attempted_Course_Load	Total EFTSL from all attempted units within the student's undergraduate enrolment in the academic year following undergraduate commencement (Year 2).	Decimal (EFTSL)	≥ 0 or NULL	Same derivation as Year 1 using TCSI elements E339 and E355. NULL indicates no enrolment, no eligible attempted load, or outcome not yet observable.
Y1_Success_Rate	Measure of Year 1 success rate.	Decimal	0–100 or NULL	Success rate for each year is calculated as: $(\text{Passed EFTSL} \div \text{Attempted EFTSL}) \times 100.$ Passed load derived from units with E355 = 3 (Completed/passed). Attempted load includes E355 ∈ (1,2,3,6). NULL indicates no eligible attempted load and is not interpreted as failure.



Field Name	Description	Format	Permissible Values	Implementation Notes
				Derived using TCSI-aligned EFTSL and completion data.
Y2_Success_Rate	Measure of Year 2 success rate.	Decimal	0–100 or NULL	Same calculation as Year 1. NULL indicates no eligible attempted load or outcome not observable
Y1_Retention_Indicator	Retention is defined as whether a student continues their study at the same institution or completes their qualification following the commencement year (Year 1).	Integer	1 = Retained (continued study or completed); 0 = Not retained (did not continue and did not complete); NULL = Not observable	Derived from the institution's year-level retention or attrition framework for the student's undergraduate commencement year. Where a year-level attrition indicator is used, retention may be derived via inverse mapping (e.g. attrition = 0 mapped to retention = 1; attrition = 1 mapped to retention = 0). Students continuing at the same institution (including course changes) are classified as retained. Students who transfer out or do not return are classified as not retained. Students who complete their qualification following the relevant academic year are also classified as retained. NULL indicates that the retention outcome is not observable.
Y2_Retention_Indicator	Retention is defined as whether a student continues their study at the same institution or completes their qualification following	Integer	1 = Retained (continued study or completed); 0 = Not retained (did not continue	Derived from the institution's year-level retention or attrition framework for the student's undergraduate study. Students are classified as retained if they continue their study at the same institution



Field Name	Description	Format	Permissible Values	Implementation Notes
	the second academic year (Year 2).		and did not complete); NULL = Not observable	in the academic year following Year 2, or if they complete their qualification. Students who do not return in the following academic year and have not completed are classified as not retained. Students who transfer to another institution are also classified as not retained. Where a year-level attrition indicator is used, retention may be derived via inverse mapping (e.g. attrition = 0 mapped to retention = 1; attrition = 1 mapped to retention = 0). NULL indicates that the retention outcome is not observable.

Note on completion and retention interpretation:

Retention indicators in this specification are derived using institution-level, academic-year retention logic. Students are classified as retained if they either continue their study in the following academic year or complete their qualification.

The Undergraduate Completion Year variable is included to distinguish, for analysis purposes, between students who are retained through continuation and those retained through completion.



Appendix B: Spreadsheet 2 – Matching Variables (PSM Covariates)

B1. Equity and Demographic Indicators

Field Name	Description	Format	Permissible Values	Implementation Notes
Low_Socioeconomic_Status	Low socio-economic status indicator	Integer	1 = Yes, 0 = No, NULL	Based on the ABS SEIFA IRSAD classification, which assigns socio-economic status according to the Statistical Area Level 1 (SA1) of a person's usual residential address. For this analysis, the SA1 used to derive the SEIFA IRSAD score is based on the student's first submitted home address recorded in the source system. Students in the lowest national quartile (below the 25th percentile) are classified as Low SES (1). Above this threshold = 0. NULL where SES information is not captured.
First_in_Family	First in Family student indicator	Integer	1 = Yes, 0 = No, NULL	Indicates students whose parents/guardians have not attained a higher education qualification, based on information recorded at undergraduate course commencement. Classified as 1 where neither parent/guardian has attained a higher education qualification. Classified as 0 where at least one parent/guardian holds a higher education qualification. NULL where information is not collected or unknown.
Disability_Status	Disability disclosure indicator	Integer	1 = Yes, 0 = No, NULL	Based on student self-reported disability disclosure. 1 = disclosed, 0 = not disclosed, NULL where not captured.
Indigenous_Background	Aboriginal and/or Torres Strait Islander indicator	Integer	1 = Yes, 0 = No, NULL	Derived from TCSI element E316 (Indigenous Status). Mapped as: 1 = Aboriginal and/or Torres Strait Islander; 0 = Non-Indigenous; NULL = Not stated / unknown. Institutions should map their local Indigenous status codes to these categories.



Field Name	Description	Format	Permissible Values	Implementation Notes
Student_Age	Age in years at undergraduate commencement	Integer	≥ 15 or NULL	Age is to be calculated in years at undergraduate commencement. It may be derived using date of birth and commencement date, provided the result reflects completed years at commencement.

B2. Admission Pathway Indicators

Field Name	Description	Format	Permissible Values	Implementation Notes
Alternative_Pathway_Into_HE_Flag	Indicates whether admission to the undergraduate qualification occurred via a non-ATAR pathway.	Integer	1 = Non-ATAR pathway, 0 = ATAR-based admission, NULL = Not observable	Classified based on the admission basis recorded at undergraduate commencement, aligned with TCSI element E327 (Basis for Admission). Students admitted via ATAR-only admission (codes 41 and 42 – ATAR only / ATAR plus) are coded 0 (ATAR-based admission). All other recognised entry pathways (e.g. enabling programs, VET qualifications, mature-age entry, special consideration) are coded 1 (Non-ATAR pathway). Where admission basis is not recorded, the value should be NULL.
ATAR_Score_or_Equivalent	Recorded ATAR (or recognised equivalent ranking score) at undergraduate admission.	Decimal	0.00–99.95 or NULL = Not observable	Recorded ATAR (or equivalent) at admission, derived from TCSI element E632. This field is applicable only for ATAR-based entry students; for non-ATAR pathways, the value should be recorded as NULL.



