Introduction

This document provides the technical details about the calculations used for the graduate satisfaction sections of the ComparED website, relating to data derived from the Course Experience Questionnaire (CEQ). It is intended for an audience with some technical background who wish to understand the statistical details of the calculations.

Data sources, variables and coverage

Course Experience Questionnaire (CEQ)

The CEQ is an annual Australian survey covering the attitudes of higher education coursework graduates towards their courses and the skills they acquired while undertaking tertiary education. The CEQ is administered to graduates of coursework level degrees, including undergraduate and postgraduate coursework levels. It is attached to the Graduate Outcomes Survey.

The following CEQ indicator is used on the ComparEd website:

- *graduate overall satisfaction*: the proportion of coursework graduates who were satisfied with the overall quality of their courses (taken from responses to Question 49 of the CEQ).

The CEQ based indicator is calculated from two years of pooled data. This incorporates the most recent year of published data and the immediately preceding year. For example, the indicator released in association with the 2021 GOS was based on results from the 2021 and 2020 surveys. In this paper these years are notated as Y1 and Y2, where Y1 is the most recent year of published data.

Indicators are calculated separately for undergraduate and postgraduate coursework level graduates.

Coverage

The variables that were used to filter the data for the CEQ as attached to the GOS can be found in Table 1. The coverage for each variable is applied before the calculation of the indicators and the SAS code used is provided in brackets after each variable in the table. The code to create the indicators is available from the Social Research Centre (SRC) on request.

Table 1: Data coverage for the coursework levels indicator from the CEQ-GOS

	Undergraduate	Postgraduate coursework
Variables (coverage)	**Overall satisfaction**	**Overall satisfaction**
Undergraduate level graduates only **(if E310 in (8,9,10,13,20,21,22))**	X	
Post-graduate coursework graduates only **(if E310 in (4,5,6,7,11,12,14))**		X
Two records for double degree students with different study areas **(if ANALYSIS in (1,2))**	X	X
Online responses only **(if SURVEY = 1)**	X	X
Overall satisfaction answered **(if osi49 in (1, 2, 3, 4, 5))**	X	X
Total minimum sample size of 25 across the pooled data years **(if n ≥ 25)**	X	X

X Indicates that the restriction is applied to the data before a particular indicator is calculated.

Standard errors and confidence intervals

Standard errors, and therefore the 90% confidence intervals, were calculated using the Finite Population Correction (FPC) to account for the relatively large samples when compared with the population. The FPC is generally used in cases where the sampling fraction, the proportion of the population sampled, exceeds 5%.

In order to calculate the standard errors for the survey estimates, no non-response bias was assumed and the Agresti-Coull method for confidence intervals for proportions was used.

The general formula used for confidence intervals for proportions was:

$$CI\ bound(\hat{p}) = \tilde{p} \pm z_{\frac{\propto}{2}} \times FPC\ \times SE(\tilde{p}) = \tilde{p} \pm z_{0.05} \times \sqrt{\frac{N-n}{N-1}}\sqrt{\frac{\tilde{p}(1-\hat{\tilde{p}})}{\tilde{n}}}$$

$$\tilde{p} = \frac{\tilde{y}}{\tilde{n}}$$

$$\tilde{y} = y + \frac{z^2{}_{\frac{\alpha}{2}}}{2} = y + \frac{z^2{}_{0.05}}{2}$$

$$\tilde{n} = n + z^2{}_{\frac{\alpha}{2}} = n + z^2{}_{0.05}$$

Where:

$\hat{p}$ is the estimated proportion from the survey data

$\tilde{p}$ is an adjusted estimated proportion used only in confidence interval calculations

FPC is the finite population correction which adjusts the estimated survey standard error based on how large the sample was compared with the population

$z_{0.05}$ is the 95th quantile from the standard Normal distribution ~ N(0,1)

n is the number in the sample in the relevant strata over the pooled relevant number of years

N is the estimated population in the relevant strata over the pooled relevant number of years

y is the number with the characteristic in question in the sample in the relevant strata over the pooled relevant number of years

As mentioned above, the FPC was used to calculate the 90% confidence intervals and the standard errors, therefore an estimate of the population was required. The Student Data Collection Completions data file for the relevant years was used and the majors of the courses were used to determine the study area. For any year of CEQ data, the corresponding population is drawn from Completions data file from the immediately preceding year. For example, for the 2018 CEQ the population is drawn from the 2017 Completions data file.

The variables that were used to filter the population data can be found in Table 2. The coverage for each variable is applied before the calculation of the population and the SAS code used is provided in brackets after each variable in the table.

Table 2: Data coverage for the graduate satisfaction indicator populations

	Undergraduate level	**Postgraduate coursework level**
Variables (coverage)	Overall satisfaction	Overall satisfaction
Undergraduate level graduates only **(if E310 in (8,9,10,13,20,21,22))**	**X**	
Post-graduate coursework graduates only **(if E310 in (4,5,6,7,11,12,14))**		**X**

However, this population estimate will not match perfectly with the survey data, as the CEQ-GOS data uses major as listed by the student, major as listed by the institution or course field of education based on availability in that priority order while the Completions file uses the majors as determined by the graduate's institution. To determine the population numbers for the indicators, all records for students by major were used to allow for a mismatch in order of major on the Student Data Collection when compared with the order used in the CEQ-GOS. This means that the population estimates will potentially be an over-estimate in some study areas, which will mean that some standard errors will be over-estimated.

In addition, the discrepancies between the two files can also lead to a survey sample size greater than or equal to the population estimate in an institution by study area level. To account for this in the calculation of standard errors, a minimum FPC of 0.2237 was set which is equivalent to a sample fraction of 95% of the population. There are also cases where there is no population figures on the completions file that match an institution by study area group, i.e. the population estimate is missing as

the student(s) has written a major not coded by the graduate's institution in the Completions file. In this case the FPC was set to 0.2237 as well.

When the estimated proportion is either 0% or 100% the formula does not allow for the calculation of standard errors but, unless the entire population has been sampled, there would be some error in the estimate. Therefore, for the calculation of standard errors to be used in the confidence intervals, the estimate was perturbed slightly, 0% was changed to 0.04% and 100% was changed to 99.96%. This was only done for the calculation of standard errors and not for the reporting of the estimate.

Calculation of indicators and confidence intervals

Overall Satisfaction

The overall satisfaction indicator is defined as the two year combined proportion of graduates who indicated agreement to overall satisfaction with the quality of their course. This indicator refers to coursework graduates only (either undergraduate or postgraduate coursework separately) and can be expressed as 'the proportion of coursework graduates who were satisfied with the overall quality of their course'.

The overall satisfaction indicator is calculated as follows:

$$\text{OSATIS}_{\text{pooled}} = \frac{\text{Number of graduates with agreement to overall satisfaction}_{\text{Y2–Y1}}}{\text{Number of graduates with a valid response}_{\text{Y2–Y1}}}$$

Where:

Number of graduates with agreement to overall satisfaction$_{Y2–Y1}$ is the total number of coursework graduates who responded with a 4 or 5 to the item 'Overall, I was satisfied with the quality of my course', over the two years combined, after filters are applied, with double degree students potentially counted twice if their study areas are different

Number of graduates with valid responses$_{Y2–Y1}$ is the total number of coursework graduates who responded to the overall satisfaction item, over the relevant two years combined, after filters are applied, with double degree students counted twice if their study areas are different

The 90% confidence interval for the overall satisfaction indicator is calculated as follows:

$$90\%CI_{OSATIS} = \widetilde{OSATIS}_{pooled} \pm z_{\frac{\propto}{2}} \times FPC \times SE_{\widetilde{OSATIS}}$$

$$= \widetilde{OSATIS}_{pooled} \pm z_{0.05} \times \sqrt{\frac{N-n}{N-1}} \times \sqrt{\frac{\widetilde{OSATIS}_{pooled} \times \left(1 - \widetilde{OSATIS}_{pooled}\right)}{\tilde{n}}}$$

$$\widetilde{OSATIS}_{pooled} = \frac{\tilde{y}}{\tilde{n}}$$

$$\tilde{y} = \text{Number of graduates rating the quality of their course positively}_{\text{Y2–Y1}} + \frac{z^2{}_{0.05}}{2}$$

$$= \text{Number of graduates rating the quality of their course positively}_{\text{Y2–Y1}} + \frac{1.645^2}{2}$$

$$\tilde{n} = \text{Number of graduates with a valid response}_{\text{Y2–Y1}} + z^2{}_{0.05}$$

$$= \text{Number of graduates with a valid response}_{\text{Y2–Y1}} + 1.645^2$$

Where:

$\widetilde{OSATIS}_{pooled}$ is an adjusted estimated proportion used only in confidence interval calculations

FPC is the finite population correction which adjusts the estimated survey standard error based on how large the sample was compared with the population

N is the population in the institution study area, or institution, as relevant, over the two years combined

n is the number of graduates with a valid response in the institution study area or institution as relevant, over the two years combined

$z_{0.05}$ is the 95th quantile from the standard Normal distribution ~ N(0,1)

The restrictions for this indicator can be found in Tables 1 and 2.