Description of student outcome and experience measures used in OfS regulation

Definition of measures and methods used to construct and present them

Reference OfS 2022.55

Enquiries to providermetrics@officeforstudents.org.uk

Publication date 30 September 2022

Last updated 3 September 2024

Contents

Introduction

1. The Office for Students (OfS) constructs and publishes a standard set of student outcome and experience data measures for use in our regulation. They inform our regulatory judgements for the following purposes:

   a. Regulating access and participation through registration condition A1.[1]

   b. Regulating student outcomes through registration condition B3, and for risk-based monitoring of quality and standards more generally.[2]

   c. Assessments through the Teaching Excellence Framework (TEF).

2. We construct data indicators as numerical measures that help us to understand the outcomes and experiences that a provider delivers for its students at different stages of the student lifecycle in higher education. The same measures are reported on as key performance measures for the OfS, and within sector-level analyses of student outcomes, experiences, or student demographic groups:

   a. access to higher education study

   b. continuation in, and completion of, the study of higher education qualifications

   c. student views and perceptions of different aspects of their higher education experience

   d. achievement and the awards made to higher education students at the end of their studies

   e. progression into the labour market and other destinations after leaving higher education.

3. Student outcome and experience indicators are produced in the same way for each provider we regulate, using available national datasets and consistent definitions and approaches to data. They provide one part of the evidence used in our regulatory processes. Any judgements that we make about a provider's performance will take into account the context of that provider.

4. We have published interactive data dashboards and associated data files. These use data definitions and approaches which follow from our 2022 consultation on the construction of the student outcome and experience measures used in our regulation.[3]

   To date, these include:

[1] The OfS registration conditions are described in the regulatory framework for higher education in England, and its amendments, at www.officeforstudents.org.uk/publications/regulatory-framework-for-higher-education-in-england/.

[2] As set out in the revised ongoing conditions of registration B1, B2, B4 and B5, which came into effect from 1 May 2022, and the revised initial and ongoing condition of registration B3, which came into effect from 3 October 2022.

[3] See www.officeforstudents.org.uk/publications/student-outcomes-and-teaching-excellence-consultations/outcome-and-experience-data/.

a. The student outcomes data dashboard showing the measures of continuation, completion and progression outcomes used to inform our regulation of condition B3.[4]

b. The TEF data dashboard showing the measures of student experience, and continuation, completion and progression outcomes which were used to inform TEF 2023 assessments.[5]

c. An updated TEF data dashboard with updated student outcome and experience measures, which may be used in future TEF assessments and to inform ongoing provider enhancement activity.[6]

d. A data dashboard showing the sector distributions of student outcome and experience measures.[7]

e. A data dashboard showing information about the size and shape of each provider's student population.[8]

f. The access and participation data dashboard.[9]

5. We expect to update each of the data resources listed in paragraph 4 with the most recent data as it becomes available. This means that we may publish one or more updates each year, typically:

Data resources	Typical update schedule
Student outcomes	Update continuation and completion measures in spring to incorporate the most recent designated data body (DDB) student return and Individualised Learner Record (ILR) student record used in their construction. Update progression measures in summer to incorporate the most recent Graduate Outcomes survey responses used in their construction.
Sector distributions	Update continuation and completion measures in spring to incorporate the most recent DDB and ILR student records used in their construction. Update progression measures in summer to incorporate the most recent Graduate Outcomes survey responses used in their construction. Update student experience measures in summer/autumn to incorporate the most recent National Student Survey responses used in their construction.
Size and shape	Update in spring to incorporate the most recent DDB and ILR student records used in the construction of size and shape of provision data.
Access and participation	Update access, continuation, completion and attainment measures in spring to incorporate the most recent DDB and ILR student records used in their construction.

[4] See www.officeforstudents.org.uk/data-and-analysis/student-outcomes-data-dashboard/.

[5] See www.officeforstudents.org.uk/data-and-analysis/data-used-in-tef-2023/.

[6] See www.officeforstudents.org.uk/data-and-analysis/tef-data-dashboard/.

[7] See www.officeforstudents.org.uk/data-and-analysis/sector-distribution-of-student-outcomes-and-experience-measures-data-dashboard/.

[8] See www.officeforstudents.org.uk/data-and-analysis/size-and-shape-of-provision-data-dashboard/.

[9] See www.officeforstudents.org.uk/data-and-analysis/access-and-participation-data-dashboard/.

Data resources	Typical update schedule
	Update progression measures in summer to incorporate the most recent Graduate Outcomes survey responses used in their construction.
TEF	Update continuation, completion, progression and student experience measures in summer or autumn once all contributing data sources are available.

6. This document sets out the descriptions and definitions of student outcome and experience data indicators that we have published, or expect to publish, during the 2024 calendar year. It describes the methods we use to construct the data indicators listed in paragraph 2, which cover student outcomes and experiences at all of the different stages of the student lifecycle in higher education. In doing so, it assumes that all of the data sources required for the construction of different measures are already available and all of the data resources listed in paragraph 4 have been updated. Prior to this becoming true, measures which are yet to be updated will remain available based on the previous year's data descriptions and definitions, defined according to that year's edition of this document.[10] For the purposes of publications during 2024 this means:

    a. Due to delays in collecting the 2022-23 DDB Student (22056) record, and the need for enhanced assessments of data quality, the first release of 2024 was one in which we shared a provider's own data with it as a set of indicative statistics. This first release was for the purpose of further checking the data quality, with providers invited to share with us any data quality issues they identify, to inform the approach we have taken to publishing the data.

    b. Following changes to the NSS introduced in the 2023 survey, and subsequent consultation, we have published student experience measures incorporating 2023 and 2024 NSS survey responses for the first time in the TEF and student experience dashboards.[11] As described in the NSS quality report for 2024, we reviewed our approach to publication response rate thresholds, benchmarking, and to the number of themes that questions are grouped into and have not made any changes to these areas for NSS 2024 (compared to NSS 2023).[12] The student experience measures based on 2023 and later NSS responses align with these approaches.

7. This document also includes information on data quality issues relating to the 2022-23 DDB Student return. There are some areas where data users are advised to take extra caution. See paragraphs 20-22 for more information.

[10] Available in the 'Archive' section at www.officeforstudents.org.uk/publications/description-and-definition-of-student-outcome-and-experience-measures/.

[11] See www.officeforstudents.org.uk/publications/consultation-on-the-approach-to-publication-of-the-nss-analysis-of-responses-and-decisions/.

[12] See www.officeforstudents.org.uk/data-and-analysis/national-student-survey-data/nss-data-quality-report/.

Who is this document for and what does it cover?

8. This document is intended to aid providers and other users of our student outcome and experience measures to understand the definitions and approaches we have used in our publication of the interactive data dashboards described in paragraph 4. It sets out:

    a. What the student outcome and experience measures are, and how different experiences and outcomes contribute to those measures.

    b. What data we have used to construct the indicators, and what courses and students are covered by them.

    c. How the indicators are structured, including 'split' indicators and different views of a provider's student populations.

    d. How the indicators are presented to users, including our approach to rounding and suppression of data.

    e. The method we use to 'benchmark' the indicators to take account of the mix of courses and students at a provider.

    f. Data about the size and shape of provision, that accompanies the indicators.

9. This document is structured to first summarise some of the key features of our approach in broad terms, intended to aid the understanding of all readers. We then explain each of our definitions and approaches, providing more information on those topics for readers who are seeking to better understand how students contribute to our student outcome and experience measures, and how the measures are presented. Later sections then provide a more comprehensive discussion of those topics, intended to support readers seeking a more in-depth understanding of our use of the student data collected by the DDB and the Education and Skills Funding Agency (ESFA), or of the statistical methods we use in the presentation and contextualisation of the indicators. Readers can navigate through this document using the clickable links provided in the contents page and throughout the document.

Related guidance

10. The information provided in this document supplements guidance about our regulatory approaches. It is one of a series of technical documents that provide details of the definitions and methods that we use to construct student outcome and experience indicators. Readers may want to consider this document alongside these documents on our regulatory approaches in particular:

    - regulatory notice 1: access and participation plan guidance[13]
    - regulatory advice 20: regulating student outcomes[14]

[13] See www.officeforstudents.org.uk/publications/regulatory-notice-1-access-and-participation-plan-guidance/.

[14] See www.officeforstudents.org.uk/publications/regulatory-advice-20-regulating-student-outcomes/.

- regulatory advice 22: guidance on the TEF.[15]

11. We have published dashboard user guides within and alongside each of our interactive data dashboards, as well as a series of frequently asked questions. These resources are intended to support users to navigate and interact with the data dashboards efficiently and effectively. The explanations they include are consistent with those given in this document and readers who have some familiarity with the data definitions may find it helpful to engage with those explanations in the immediate context of the dashboard in question.

12. To understand their own student data, we have released data resources to providers. This includes individualised student data files and information about student outcomes associated with higher education provision delivered through partnership arrangements. Student outcomes data related to the partnerships view of a provider's student population is published within our data dashboards for the first time this year.

13. We have published our data definitions in algorithm form, instructions for rebuilding our indicators from individualised student data, and the sector average outcomes that are used in benchmarking calculations.[16] For the avoidance of doubt, it is the formulation of student outcome and experience measures in algorithm form that underpins our construction of the indicators and split indicators we have published. Readers seeking an in-depth understanding may wish to consider these resources when reading through this document. Table 2 in the 'Instructions for rebuilding our indicators from individualised student data' document describes which data has been used to construct the data resources published at any given time. It will be updated as data resources are updated throughout each year.

Enquires and feedback

14. For enquiries regarding the definitions and methods described in this document, and to give feedback, contact providermetrics@officeforstudents.org.uk.

[15] See regulatory advice 22: Guidance on the Teaching Excellence Framework 2023 at www.officeforstudents.org.uk/publications/regulatory-advice-22-guidance-on-the-teaching-excellence-framework-2023/.

[16] See www.officeforstudents.org.uk/data-and-analysis/institutional-performance-measures/technical-documentation/.

Summary of key features of our approach

This section of the document is aimed at all readers and summarises some of the key features of our approach to support a broad understanding of the information we use and publish as student outcome and experience measures.

What student outcome and experience measures does the OfS construct?

- We construct a range of different measures, which are used in different combinations by each of our regulatory functions.
- **Access to higher education measures** report on the profile of entrants to higher education.
  - They are used in the access and participation data dashboard.
  - They are not used in our assessments of condition B3 or through the TEF, nor for risk-based monitoring of quality and standards more generally.

For further information about the access measures see Indicator definitions: Access to higher education measures.

- **Continuation measures** report the proportion of students that were observed to be continuing in the study of a higher education qualification (or that have gained a qualification) one year and 15 days after they started their course (two years and 15 days for part-time students).
  - They are used in all the regulatory functions described at paragraph 1.

For further information about the continuation measures see Indicator definitions: Continuation measures.

- **Completion measures** report the proportion of students that were observed to have gained a higher education qualification (or were continuing in the study of a qualification) four years and 15 days after they started their course (six years and 15 days for part-time students).
  - They are used in all the regulatory functions described at paragraph 1.

For further information about the completion measures see Indicator definitions: Completion measures.

- **Student experience measures** use responses to the National Student Survey (NSS) to report on the views of students on different aspects of their higher education experience. They report

the level of positivity when responding to a range of questions that comprise a theme of the survey as indicated among final year undergraduates.

  - They are used in our assessments through the TEF and for risk-based monitoring of quality and standards more generally.

  - They are not used in our assessments of condition B3 or in the access and participation data dashboard.

For further information about the student experience measures see Indicator definitions: Student experience measures.

- **Degree outcomes measures** report the proportion of qualifiers awarded a first or upper second classification of a first degree.

  - They are used in the access and participation data dashboard and for risk-based monitoring of quality and standards more generally.

  - They are not used in our assessments of condition B3 or through the TEF.

For further information about the degree outcomes measures see Indicator definitions: Degree outcomes measure.

- **Progression measures** use responses to the Graduate Outcomes (GO) survey to report on qualifiers' labour market and other destinations 15 months after they left higher education. They report the proportion of qualifiers that identify managerial or professional employment, further study, or other positive outcomes among the activities that they were undertaking at the GO survey census date.

  - They are used in all the regulatory functions described at paragraph 1.

For further information about the progression measures, see Indicator definitions: Progression measures.

What data the OfS has used to construct the indicators, and what courses and students are covered by them

- The indicators are constructed based on individualised student data returned by providers to the Designated Data Body (DDB) student record or Individualised Learner Record (ILR) collections on an annual basis. Some measures use students' responses to the GO and NSS survey instruments.

- The coverage of the indicators generally extends to all students who are:

- Reported with a qualification aim for their course which refers to a higher education qualification. This includes all qualifications at Level 4 and above, whether or not they are courses recognised for OfS funding, and whether or not they are studied as part of an apprenticeship.

- Studying wholly or mainly in the UK for their whole programme of study, or through UK-based distance learning, including international students where appropriate.

For further information about the exceptions to this general coverage, see Coverage of student populations.

- The student outcome and experience measures each make use of a number of years of data, so the coverage of each measure is influenced by the available years and coverage of the data it relies on. Our measures are reported as an aggregate of those years, as well as through a time series of the individual years. The most recent years of available data correspond to different academic years depending on the measure in question.

- Table 1 shows the **most recent four years** of available data that inform the student outcome measures reported in the student outcomes data dashboard.

Table 1: Four-year time series for each measure in the student outcomes and TEF data dashboards

Measure	Year 1 (least recent)	Year 2	Year 3	Year 4 (most recent)
Continuation: full-time and apprenticeship	2018-19 entrants	2019-20 entrants	2020-21 entrants	2021-22 entrants
Continuation: part-time	2017-18 entrants	2018-19 entrants	2019-20 entrants	2020-21 entrants
Completion: full-time and apprenticeship	2015-16 entrants	2016-17 entrants	2017-18 entrants	2018-19 entrants
Completion: part-time	2013-14 entrants	2014-15 entrants	2015-16 entrants	2016-17 entrants
Progression: full-time, part-time, and apprenticeship	2018-19 qualifiers	2019-20 qualifiers	2020-21 qualifiers	2021-22 qualifiers
Student experience: full-time, part-time, and apprenticeship	2023 NSS	2024 NSS	Not available	Not available

- The 2023 NSS survey was the first with revised questions, following consultation in 2022. As such, there are only two years of data for student experience measures. Every year, another year will be added to the time series until there are four years within the time series.

- Table 2 shows the most recent six years of available data that inform the construction of student outcome and experience measures informing the access and participation data dashboard. For the progression measures, data for 2017-18 qualifiers was the first year of available data. As such, there are only five years of data for the progression measure. Every year, another year of data will be added to the time series until there are six years within the time series.

Table 2: Six-year time series for each measure in the access and participation data dashboard

Measure	Year 1 (least recent)	Year 2	Year 3	Year 4	Year 5	Year 6 (most recent)
Access: full-time, part-time, and apprenticeship	2017-18 entrants	2018-19 entrants	2019-20 entrants	2020-21 entrants	2021-22 entrants	2022-23 entrants
Continuation: full-time and apprenticeship	2016-17 entrants	2017-18 entrants	2018-19 entrants	2019-20 entrants	2020-21 entrants	2021-22 entrants
Continuation: part-time	2015-16 entrants	2016-17 entrants	2017-18 entrants	2018-19 entrants	2019-20 entrants	2020-21 entrants
Completion: full-time and apprenticeship	2013-14 entrants	2014-15 entrants	2015-16 entrants	2016-17 entrants	2017-18 entrants	2018-19 entrants
Completion: part-time	2011-12 entrants	2012-13 entrants	2013-14 entrants	2014-15 entrants	2015-16 entrants	2016-17 entrants
Degree outcomes: full-time, part-time, and apprenticeship	2017-18 qualifiers	2018-19 qualifiers	2019-20 qualifiers	2020-21 qualifiers	2021-22 qualifiers	2022-23 qualifiers
Progression: full-time, part-time, and apprenticeship	Not available	2017-18 qualifiers	2018-19 qualifiers	2019-20 qualifiers	2020-21 qualifiers	2021-22 qualifiers

For further information about the years of data used and the coverage of each measure, see:

Indicator definitions: Access to higher education measures

Indicator definitions: Continuation measures

Indicator definitions: Completion measures

Indicator definitions: Student experience measures

Indicator definitions: Degree outcomes measure

Indicator definitions: Progression measures.

How are the indicators structured? What are indicators and split indicators?

- For each student outcome and experience measure, **indicators** are constructed for the unique combinations of a student's mode and level of study, based on multiple years of data.
- Those indicators are then broken down further, by student and course characteristics, to show a series of **split indicators**. For the same mode and level covered by the corresponding indicator, split indicators consider one or more of:
  - A time series of individual years contributing to the overall indicator
  - Subject studied
  - Student characteristics, including personal characteristics which are protected under the Equality Act 2010, and other measures of students' backgrounds
  - Specific course types
  - Partnership arrangements through which different providers have responsibilities for registering, teaching, or awarding the qualifications.
- For the access and participation data dashboards, indicators represent both an aggregate and the individual years of a time series of multiple years of data. This means that the reporting structure involves additional indicators and split indicators which accommodate intersections of year with each of the different student characteristics, as well as a limited selection of intersections between student characteristics, in order to support our regulatory objectives for access and participation.
- Student outcome and experience measures are reported for different **views of a provider's student populations**, based on their responsibilities for teaching, registering, or awarding the qualifications of different students.
- The views of a provider's student populations are used in different combinations by each of our regulatory functions:
  - The student outcomes data dashboard reports separately on the populations of students who are:
    - either taught or registered at the provider (or both)
    - taught at the provider
    - associated with the provider through partnership arrangements (sub-contractual partnerships, or validation-only).
    - The TEF data dashboard is based on all students who are either taught or registered at the provider (or both).
    - The access and participation data dashboard is based on all students registered at the provider.

For more information about how we define indicators, split indicators, and views of a provider's student populations, see structure and reporting.

For a detailed specification of the split indicators, see Annex B: Further information about the definition of split indicators.

How are the measures presented? Why are some measures not available for some providers?

- Student outcome and experience measures are presented through interactive data dashboards and accompanying data files. The presentation we use in the data dashboards has been designed to help users interpret a provider's performance, taking account of the concept of statistical uncertainty.[17]
- We are committed to effectively communicating our statistics so that users can have confidence in their use and interpretation of them. This means we aim to use meaningful and effective ways to understand the potential extent of statistical uncertainty within our indicators and split indicators.
- We show the value of each indicator and split indicator and, for some uses, its difference from the provider's benchmark. 'Shaded bars' are used in our presentation of data to inform our regulation of student outcomes, the TEF and access and participation, to communicate the statistical uncertainty associated with each of those values.
- For each indicator or split indicator, we also show:
  - the denominator
  - the numerator (for access indicators)
  - the benchmark value (where appropriate)
  - the provider's contribution to its own benchmark (where a benchmark exists)
  - the survey response rate (if the measure is based on a survey instrument).
- Figures 1 and 2 provide an illustration of the shaded bars we use in our presentation of the student outcomes and TEF data dashboards. These are differentiated by colour and aim to represent the continuous spread (or distribution) of statistical uncertainty around the different values that we have calculated to understand a provider's performance. The green shaded bar shows statistical uncertainty associated with the indicator value. The bar shaded blue shows the difference between indicator and benchmark values.

[17] See Annex C: Further information about how we calculate and present statistical uncertainty.

Figure 1: Example of green shaded bars showing spread of statistical uncertainty around indicator values in the student outcomes and TEF data dashboards

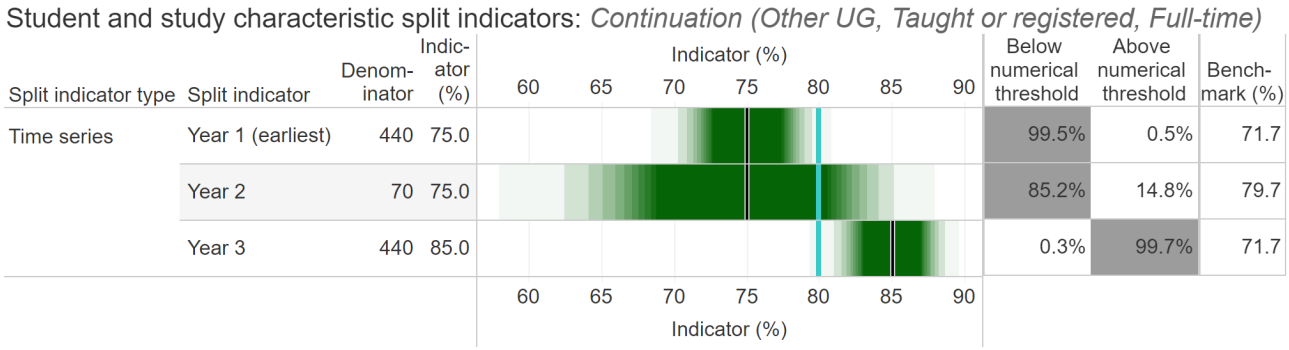
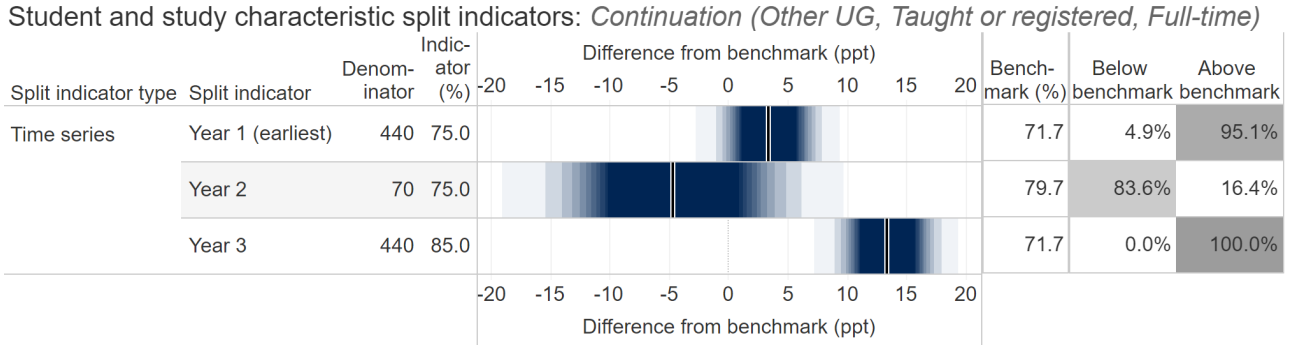


Figure 2: Example of blue shaded bars showing spread of statistical uncertainty around difference between indicator and benchmark in the student outcomes and TEF data dashboards



- To support consistency and transparency of interpretation, Figure 1 and Figure 2 show that we also include summary figures in the table to the right of the shaded bars. These describe the proportion of the distribution of statistical uncertainty, represented by the shaded bar that falls above or below certain values (the minimum numerical thresholds for condition B3, or the provider's own benchmark value).[18] They are highlighted where they show that at least 75 per cent of the distribution falls above or below those values, but users can use the shaded bars to make other interpretations of a provider's performance.

- Figures 3 and 4 provide an illustration of the shaded bars we use in our presentation of the access and participation data dashboards. These aim to represent the continuous spread (or distribution) of statistical uncertainty around the different values that we have calculated to understand a provider's performance. The time series information included in access and participation indicators and split indicators mean that the shaded bars are presented vertically. In order to differentiate between multiple split indicators, the colour of the shaded bars always matches the colour of the indicator. Figure 3 shows how statistical uncertainty associated with the indicator values is presented. Figure 4 shows how we present the statistical uncertainty associated with the percentage point gap we have calculated between two split indicators.

[18] For TEF purposes, the summary figures for the blue shaded bars show the proportions of the statistical uncertainty distribution which fall above, below or between a pair of 'guiding lines' which illustrate where the indicator value could be considered as materially above or below the benchmark value. These guiding lines are positioned at +/- 2.5 percentage points difference between the indicator and benchmark values.

Figure 3: Example of shaded bars showing spread of statistical uncertainty around indicator values in the access and participation dashboard

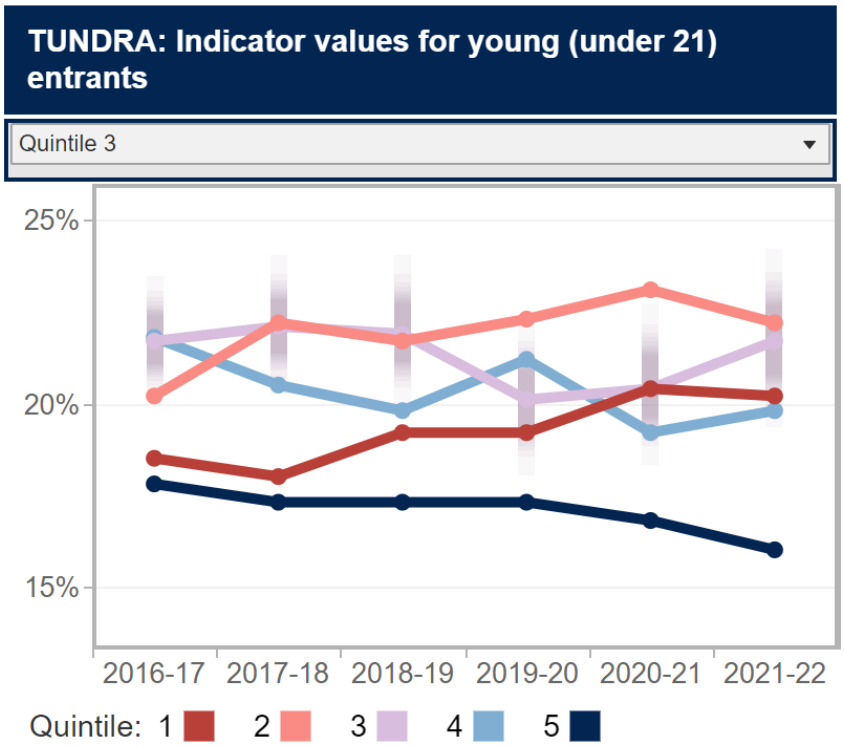
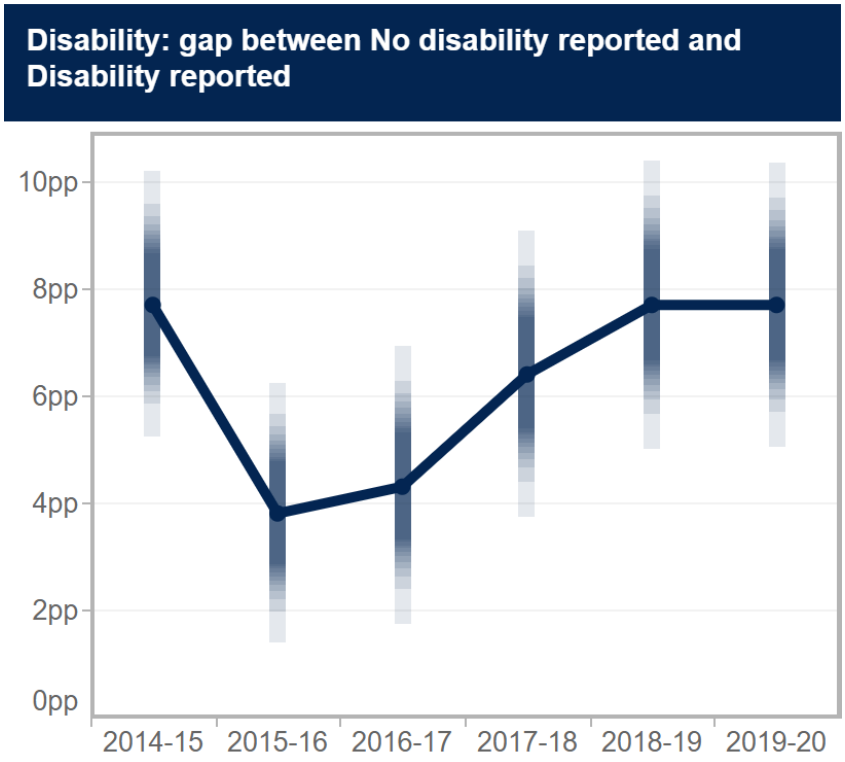


Figure 4: Example of shaded bars showing spread of statistical uncertainty around gaps in the access and participation dashboard



- The shading of the bars used in the student outcomes, TEF and access and participation dashboards indicates the changing likelihood that underlying provider performance takes

different values. The darkest shading represents the range in which there is the greatest likelihood that true provider performance might lie, as the shading lightens in both directions, it represents a lower likelihood that true underlying performance falls at that point. Wider shaded bars mean we need to consider the potential for the provider's true performance falling within a wider range of values around the point estimate that has been observed.

- We will not report an indicator or split indicator in certain circumstances, including:
  - Where there are fewer than 23 students in the denominator.
  - Where an indicator or split indicator based on the Graduate Outcomes survey has a survey response rate below 30 per cent.
  - Where an indicator or split indicator based on the National Student Survey has a survey response rate below 50 per cent.
  - Where data has been suppressed for data protection reasons.
- In the reporting of gaps and ratios in the access and participation dashboard, where any of the constituent indicators or split indicators have been suppressed (according to any of the reasons set out above) the gap or ratio will not be reported.
- In addition, where there is unknown information about one or more benchmarking factors for at least 50 per cent of relevant students we would not report the benchmark or the difference between the indicator and the benchmark.

For more information about the information included in our presentation of the data, see Elements included in the presentation of student outcome and experience measures.

For more information about rounding and suppression of the data, see Rounding and suppression.

For more information about the concepts of statistical uncertainty and multiple comparison adjustments, see Statistical uncertainty and visualisation of this.

For further technical detail about the statistical methods and calculations underpinning our presentation of statistical uncertainty, see Annex C: Further information about how we calculate and present statistical uncertainty .

What is benchmarking and how does it work?

- Benchmarking is the method we use to take account of the mix of courses and students at a provider and indicate how well that provider has performed compared with performance for similar types of students on similar types of courses in the higher education sector as a whole.
- We calculate benchmarks for each provider's indicators and split indicators based on the characteristics of courses and students that we have selected as benchmarking factors.

- The benchmark is calculated as a weighted sector average which represents the outcomes that would have been achieved by the provider if it retained its mix of students and courses, but its outcomes across the benchmarking factors were replaced by the sector-overall rates for those student groups. It represents the performance of similar types of students on similar types of courses to that of the provider. Our approach means that a provider is not being compared with a pre-set group of providers, but rather the outcomes for a provider's students are compared with the outcomes of similar students across the entirety of the higher education sector.

- We use benchmarking factors that, across the sector as a whole, are most correlated with the outcomes and experiences we are measuring once other factors have been controlled for, where we consider it would not be undesirable to control for those factors. These factors relate to characteristics of courses (such as subjects, and level of study) and students (such as their age or the qualifications they held on entry to higher education).

Table 3: Summary of benchmarking factors used for each student outcome and experience measure

Benchmarking factor	Continuation	Completion	Progression	Student experience
Year of survey	×	×	✓	✓
Level of study	✓	✓	✓	✓
Subject of study	✓	✓	✓	✓
Entry qualifications	✓	✓	✓	×
Course length	✓ (Part-time only)	✓ (except for apprenticeships)	×	×
ABCS quintiles	✓	✓	✓	×
Geography of employment quintiles	×	×	✓	×
Individual student characteristics	×	×	×	✓ Age Disability Ethnicity Sex (full-time only)

For more information about the benchmarking method, and the factors and groupings used in benchmarking, see Benchmarking and Annex G: Definitions of entry qualifications and subject areas of study groupings used in benchmarking.

For a worked example of our benchmarking method, see Annex F: Worked example of benchmarking calculations.

For further technical detail about the statistical methods and calculations underpinning our benchmarking method, see Annex H: Technical detail about benchmarking calculations.

What is the data that accompanies the indicators?

- We include data about the size and shape of provision alongside the indicators and split indicators for each provider. Its purpose is to help the users understand:
  - A provider's size in terms of student numbers.
  - The type of courses it offers and its mix of subjects.
  - The characteristics of its students, including their personal characteristics and backgrounds prior to starting higher education study.
  - Information on the numbers of students in each type of teaching partnership arrangement.

For more information about the summaries we provide in relation to the size and shape of provision, see Data about the size and shape of higher education provision.

- We include data about the reporting of interim study activities to the GO survey. It is intended to help users understand the potential influence of these interim activities on a provider's performance in relation to the student outcomes reported by the progression measure.
- We report two separate figures, both based on students who counted negatively towards the progression indicator: those who reported in their GO response that they had undertaken **any** interim study since completing their higher education course, and those who reported undertaking **significant** interim study.

For more information about the data we provide in relation to interim study activities, see Data about the reporting of interim study activities.

Data sources and coverage

This section is aimed at readers who are seeking to better understand how students contribute to our student outcome and experience measures, and how the measures are presented.

The data sources and coverage described in this section are relevant to all indicators and therefore to the **access and participation data dashboard, student outcomes data dashboard,** and the **TEF data dashboard**.

15. All of the student outcome and experience measures make use of the student data returns collected by the DDB and the ESFA, which are linked as appropriate to the following data sources:

    a. Responses to the **Graduate Outcomes** survey, to construct measures of progression outcomes. GO is a population survey of almost all graduates of higher education in the UK, in a given academic year, which collects information about their outcomes and destinations 15 months after completing a higher education qualification.

    b. Responses to the **National Student Survey** to construct student experience measures. The NSS is a UK-wide survey that collects feedback from final year students about their higher education experience. Information is collected about a range of factors including the teaching on their course, assessment and feedback, academic support, and how well courses were organised.

    c. Information drawn from the Department for Education (DfE) **national pupil database** (NPD) about individuals included in the DDB and ESFA data. The NPD holds information about pupils such as free school meal entitlement and educational attainment. It is used to construct split indicators and benchmarking factors.

    d. **Classifications produced by the OfS and other bodies**, to construct split indicators and benchmarking factors. In particular classifications of employment outcomes and occupations, deprivation measures, higher education participation, and outcomes propensity.

16. The student outcome and experience measures each make use of a number of years of the data sources listed in paragraph 15. They are reported as an aggregate of those years, as well as through a time series of the individual years.

17. The most recent years of available data correspond to different academic years, depending on the measure in question. This was shown in Table 1 (which described the **most recent four years** of available data that inform measures reported through the student outcomes data dashboards and the TEF data dashboard) and Table 2 (which described the **most recent six years** of available data that inform the access and participation data dashboard).

18. The student data we use has been submitted and signed off by providers. While we take care to process the data as accurately as possible, the data quality of our outputs depends on the data quality of the data submitted to the DDB and ESFA.

19. Paragraph 15 noted that the DfE's NPD dataset is used to source information for the construction of split indicators and classifications that are used in benchmarking student outcome and experience indicators.[19] It should be noted that the NPD census for key stage 4 (KS4) covers pupils attending maintained and independent schools in England, and censuses for academic years from 2009-10 to the latest have been matched to DDB and ILR student records. From academic year 2013-14, the NPD data includes local authority-maintained Pupil Referral Units and alternative provision academies, including alternative provision free schools. Since pupils are generally 15 years old in their last year of KS4, the academic year 2014-15 is the earliest academic year that a full cohort of young entrants (under 21 on entry) can be tracked back to the NPD. The NPD population is restricted to state-funded mainstream schools, so includes the following school types:

    - academies (16-19 converter, 16-19 sponsor-led, converter, sponsor-led)
    - free schools (16-19, mainstream, studio school, UTC)
    - voluntary aided and voluntary controlled schools
    - community and foundation schools
    - city technology colleges and further education sector institutions.

Data quality

20. Some higher education providers are required to submit student data to the designated data body (Jisc). For the 2022-23 student data return Jisc introduced a new data model and a new data platform known as 'Data Futures'[20], leading to delays in data collection. Consequently, additional risks for the quality of data were tolerated in some areas of the 2022-23 data. We are assessing the impacts of those risks for each of our data outputs on a case-by-case basis.

21. For the student outcome and experience data outputs, our data quality assessment has included releasing an indicative version of these outputs to providers that submitted a 2022-23 student data return to Jisc. In doing so, we invited these providers to further check the quality of data before its publication, because we considered that providers themselves were best placed to identify any anomalies in the data.

22. Through this process, several providers have reported quality issues affecting the data published by the OfS, and we have also identified other significant issues affecting some providers' data. We have provided a page of explanations where users may need to exercise some caution when interpreting the data of particular dashboards for the affected providers.[21] The designated data body is also expecting to publish data intelligence about the 2022-23 student data return and this will also be a useful source of information on data quality.

[19] The Department for Education does not accept responsibility for any inferences or conclusions derived from the NPD data by third parties.

[20] See https://www.hesa.ac.uk/innovation/data-futures

[21] See www.officeforstudents.org.uk/data-and-analysis/student-outcome-and-experience-measures/data-quality/

Coverage of student populations

23. Student outcome and experience measures cover all students who are:

    a. Reported with a qualification aim for their course which refers to a higher education qualification. This includes all qualifications at Level 4 and above, whether or not they are courses recognised for OfS funding, and whether or not they are studied as part of an apprenticeship.[22]

    b. Studying wholly or mainly in the UK for their whole programme of study, or through UK-based distance learning, including international students where possible and meaningful.

24. The student outcome and experience indicators do not cover:

    a. Students reported with a qualification aim which refers to a module of higher education provision or, in the case of degree awarding and progression measures, gaining awards of higher education credit.

    b. Non-UK based students including those studying through transnational education (TNE) arrangements, and incoming visiting and exchange students.

    c. Students leaving their programme of study within the 14 days following their commencement date without gaining an award.

    d. Students on Subject Knowledge Enhancement (SKE) courses.

    e. Overall apprenticeship standards, as identified by student records in the ILR that refer to an apprenticeship standard 'wrapper' programme aim (which are not records of student activity). Individual higher education qualifications at Level 4 or above that are studied as part of an apprenticeship are included in the coverage of student outcome and experience measures.

    f. Student records which have been duplicated across different student returns, which are removed to avoid double counting. This mainly affects apprenticeships reported to both the DDB and the ESFA, where we will normally use the record submitted to the DDB.

    g. ILR records which have been closed to correct a learning planned end date are excluded to avoid double counting as the new, corrected record will report the relevant activity.

25. Coverage of the access and participation data dashboard is restricted throughout to UK-domiciled students studying for undergraduate qualifications.

26. Coverage of the TEF data dashboard is restricted throughout to students studying for undergraduate qualifications.

[22] Qualifications which are not eligible to be included in the OfS funding calculations for Approved (fee cap) providers may include those that are regulated by the Office of Qualifications and Examinations Regulation (listed on the Register of Regulated Qualifications, and for which students may be entitled to Advanced Learner Loans). See Paragraphs 1-2 of Annex B of 'Higher Education Students Early Statistics Survey 2022-23 (HESES22)' at www.officeforstudents.org.uk/publications/heses22.

27. Coverage of the student outcomes data dashboard includes all students studying for undergraduate and postgraduate qualifications.

28. Certain student outcome and experience measures make further restrictions specific to the coverage of the measure in question or the data source it is relying on.

Further information about the restrictions which are specific to the measure in question are available at:

Indicator definitions: Access to higher education measures

Indicator definitions: Continuation measures

Indicator definitions: Completion measures

Indicator definitions: Student experience measures

Indicator definitions: Degree outcomes measure

Indicator definitions: Progression measures.

29. In all cases, we report on student populations in headcount terms, at person level. This means that the data we generate expresses student populations as counts of full-person equivalents (FPE),[23] and a student who was actively studying multiple instances of higher education at the same registering provider, at the same broad level of study (undergraduate or postgraduate) in the same reporting period, will only count once per year.[24]

Coverage of entrant cohorts

30. Measures of continuation and completion are constructed and reported for the **entrants** student population. We also report data about the size and shape of provision at each provider with reference to the entrants population.

31. A cohort of entrants includes all students with a course commencement date between 17 July and the following 16 July. The following exclusions apply:

    a. Students mainly in the UK and aiming for a qualification but are dormant or sabbatical.

    b. Students who did not start their study in the relevant year.

[23] The count of FPE involves apportioning each individual student headcount according to the proportion of their course in each subject and helps to preserve an accurate overall headcount when reporting data about a student's subject studied. For example, a student who is studying a joint course with equal amounts of mathematics and English is apportioned across the two subjects and represented as 0.5 FPE in each. A student who is studying a course involving mathematics (50 per cent) with English (40 per cent) and history (10 per cent), is apportioned across the three subjects and represented as 0.5 FPE in mathematics, 0.4 FPE in English and 0.1 FPE in history.

[24] This means, for example, that a student who is concurrently studying two postgraduate certificates on a part-time basis with the same provider would only count in that year's indicator once. A student who was studying those same certificates across two different providers, or at the same provider but one on a full-time basis, would count twice, once in each provider or mode. A student who was studying concurrently for both an undergraduate and a postgraduate qualification at the same provider would count twice, once for each level of study.

32. If a student is reported as actively studying at the same registering provider, at the same broad level of study (undergraduate, postgraduate taught, or postgraduate research), in two successive years, they are counted as an entrant only once, in their first engagement with the provider.

33. Postgraduate research students who are engaged in sequential collaborative provision (primarily within doctoral training programmes) are included in the entrant cohort of each provider that they register with, counting as an entrant at the point at which their registration with that provider commences rather than at their point of entry to that higher education course overall. This approach does not apply to concurrent collaborative provision, or sequential collaborative provision at other levels of study.

Coverage of qualifier cohorts

34. Measures of degree outcomes and progression are constructed and reported for the **qualifiers** student population. We also report data about the size and shape of provision at each provider with reference to the qualifiers population.

35. A cohort of qualifiers includes students reported to have been awarded a higher education qualification.

Indicator definitions: Access to higher education measures

This section is aimed at readers who are seeking to better understand how students contribute to our student outcome and experience measures, and how the measures are presented.

Access to higher education measures are used in the **access and participation data dashboard.**

36. Access to higher education measures report on the profile of **entrants** to higher education. They express the number of entrants with a particular attribute as a percentage of all entrants. For example, they report the percentage of entrants who were female.

Further information about data sources and coverage

37. The access measures use the same Data sources and coverage as applies across all of the student outcome and experience measures used in our regulation. Table 4 provides further information about the data sources and coverage which are specific to the access measures. For further information on the years and population restrictions used to rebuild our indicators, see the 2024 'Rebuilding student outcome and experience measures used in OfS regulation' document.[25]

Table 4: Data sources and coverage for the access to higher education measures

Data source(s)	Individualised DDB and ILR student records	
Years of data returns used	2017-18 to 2022-23	Access measures cover students entering higher education between: 17 July 2017 and 16 July 2018 (defines Year 1 of the time series) 17 July 2018 and 16 July 2019 (Year 2) 17 July 2019 and 16 July 2020 (Year 3) 17 July 2020 and 16 July 2021 (Year 4) 17 July 2021 and 16 July 2022 (Year 5) 17 July 2022 and 16 July 2023 (Year 6).
Population restrictions	UK-domiciled undergraduate entrants registered at the higher education provider in question.	As with all indicators, access measures exclude the students described at paragraph 24. They also exclude: • non-UK domiciled entrants

[25] See https://www.officeforstudents.org.uk/data-and-analysis/student-outcome-and-experience-measures/documentation/

		• postgraduate entrants • students on clinical medical, dental, or veterinary science qualifications who take an intercalating year (whether at the same provider or different).[26]
Variable definitions by mode of study?	No	

Further information about presentation of the access measures

38. In addition to reporting indicator values which represent the profile of higher education entrants in proportional terms (as described in paragraph 36 above), the access measures also report gaps and ratios between proportions of entrants at the provider from quintile 5 and quintile 1 for POLAR4, English IMD, TUNDRA and ABCS quintiles.

Information about population estimates

39. We have previously published population estimate data as part of the access and participation data dashboard, reporting on the proportions of the 18-year-old population with various characteristics. These estimates have been based on published resources from public bodies such as the Office for National Statistics (ONS), derived from the 2011 census.

40. The ONS are currently in the process of reviewing and reconciling their published mid-year population estimates methodologies in light of data from the 2021 census.[27] The recent census has also led to changes in the availability and release timelines for population estimate data formats that have previously been used to derive population estimate data within the access and participation data dashboard. For these reasons, we have not included population estimates in the access and participation data dashboard released in summer 2024. We will keep under review the developments in the availability of appropriate population estimate data and the latest ONS methodologies, and assess how we can best utilise such information to support our policy aims in relation to equality of opportunity in future.[28] We welcome feedback from users on the value and utility of including population estimates within the access an participation data dashboard in future, including any comments on the most useful formats and reporting structure for their uses of this data.

[26] Intercalation involves an additional year of study on top of a medical, dental or veterinary degree programme and an opportunity to develop knowledge and skills, and gain a standalone qualification, in a new area which may or may not be related to their main degree study.

[27] See https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/articles/reconciliationofmidyearpopulationestimateswithcensus2021englandandwales/2023-02-28

[28] See https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/internationalmigration/methodologies/dynamicpopulationmodelimprovementstodatasourcesandmethodologyforlocalauthoritiesenglandandwales2011to2022

Indicator definitions: Continuation measures

This section is aimed at readers who are seeking to better understand how students contribute to our student outcome and experience measures, and how the measures are presented.

Continuation measures are used in the **access and participation data dashboard**, **student outcomes data dashboard,** and the **TEF data dashboard.**

41. Continuation outcomes are measured by identifying a cohort of **entrants** to higher education qualifications at the provider and following them through the early stages of their course to track how many continue in active study, or qualify, in subsequent years.

42. The continuation measure tracks students from the date they enter a higher education provider to their activity on a census date.

    a. For full-time continuation outcomes, and apprenticeship continuation outcomes, the census date is one year and 15 days after their commencement date.

    b. For the part-time continuation outcomes, the census date is two years and 15 days after their commencement date.

Further information about data sources and coverage

43. The continuation measures use the same Data sources and coverage as applies across all the student outcome and experience measures used in our regulation. Table 5 provides further information about the data sources and coverage which are specific to the continuation measures.

Table 5: Data sources and coverage for the continuation measures

Data source(s)	Individualised DDB and ILR student records.	
Years of data returns used: for continuation measures informing TEF and regulation of student outcomes.	2017-18 to 2022-23.	The **full-time** continuation and **apprenticeship** continuation measures cover students entering higher education between: 17 July 2018 and 16 July 2019 (defines Year 1 of the time series) 17 July 2019 and 16 July 2020 (Year 2) 17 July 2020 and 16 July 2021 (Year 3) 17 July 2021 and 16 July 2022 (Year 4). The **part-time** continuation measures cover students entering higher education between: 17 July 2017 and 16 July 2018 (defines Year 1 of the time series) 17 July 2018 and 16 July 2019 (Year 2)

		17 July 2019 and 16 July 2020 (Year 3) 17 July 2020 and 16 July 2021 (Year 4).
Years of data returns used: for continuation measures in the access and participation data dashboard.	2015-16 to 2022-23.	The **full-time** continuation and **apprenticeship** continuation measures cover students entering higher education between: 17 July 2016 and 16 July 2017 (defines Year 1 of the time series) 17 July 2017 and 16 July 2018 (Year 2) 17 July 2018 and 16 July 2019 (Year 3) 17 July 2019 and 16 July 2020 (Year 4) 17 July 2020 and 16 July 2021 (Year 5) 17 July 2021 and 16 July 2022 (Year 6). The **part-time** continuation measures cover students entering higher education between: 17 July 2015 and 16 July 2016 (defines Year 1 of the time series) 17 July 2016 and 16 July 2017 (Year 2) 17 July 2017 and 16 July 2018 (Year 3) 17 July 2018 and 16 July 2019 (Year 4) 17 July 2019 and 16 July 2020 (Year 5) 17 July 2020 and 16 July 2021 (Year 6).
Population restrictions	Entrants registered at the higher education provider in question.	As with all measures, continuation measures exclude the students described at paragraph 24. They also exclude students on clinical medical, dental, or veterinary science qualifications who take an intercalating year at the same provider (students who take an intercalating year at a different provider are not excluded).[29] For access and participation data they also exclude: • non-UK domiciled entrants • postgraduate entrants. For TEF data, they also exclude postgraduate entrants.
Variable definitions by mode of study?	Yes.	Definition varies with respect to census date: one year and 15 days after commencement for

[29] Intercalation involves an additional year of study on top of a medical, dental, or veterinary degree programme and an opportunity to develop knowledge and skills, and gain a standalone qualification, in a new area which may or may not be related to their main degree study.

		full-time and apprenticeship modes of study, two years and 15 days for part-time.

44. Postgraduate research students who are engaged in sequential collaborative provision (such as doctoral training programmes) are included in the entrant cohort of each provider that they register with, counting as an entrant at the point at which their registration with that provider commences rather than at their point of entry to higher education overall.

45. Information shown in Table 5 about the years of data returns used refers to entrant cohorts which have been defined to align with the census date periods for continuation and completion measures. An entrant year cohort is defined based on those students starting courses between 17 July and the following 16 July because this allows us to determine the activity of all students in that cohort on their census date in the following data reporting period.

Student outcomes counting positively towards continuation measures

46. The continuation measures count as positive outcomes for those students who have either:

    a. Gained a higher education qualification from the provider at which they were previously identified as an entrant, on or before the relevant census date.

    b. Been recorded as actively studying for a higher education qualification at the same provider on the census date.

47. To be counted as a positive outcome in the circumstances described by paragraph 46, a student must have a record in the relevant DDB or ILR student returns that identified these outcomes.

48. Whether students can be identified with these outcomes on the basis of the DDB or ILR student returns can be established using the definitions provided in the 'Core algorithms' document and the instructions given in the 'Rebuild instructions' document. In particular readers may wish to consider the variables named IPENTRANTEXCL and IPCONINDFULL_YX.

49. There are certain circumstances in which it is useful to note the treatment of student outcomes for the purposes of defining the continuation measures:

    a. Students who are recorded in the ILR datasets as having partially completed their qualification, or whose results are recorded as not yet known, are counted as a positive continuation outcome.

    b. Postgraduate research students who (in the year of student data returns in which the relevant census date falls) were recorded as being awarded a qualification from a dormant mode of study are counted as a positive continuation outcome.

    c. Postgraduate research students who are engaged in sequential collaborative provision and transfer (on or before the census date) to a new provider as part of that arrangement are counted as a positive continuation outcome for the provider they transferred from.

d. Other students who transfer (on or before the census date) to study a higher education qualification at a different provider than the one at which they were identified as an entrant are treated as a neutral continuation outcome. This means that they are not included in the population of the continuation indicator and are excluded from both the numerator and denominator involved in its calculation.

e. Students who are recorded in the DDB student data with results not yet known, are not counted as a positive continuation outcome (because their qualification outcome will be reported in a subsequent year of DDB student data returns).

f. Students who transfer (on or before the census date) to the study of higher education modules, for credit only, or to the study of further education qualifications, are counted as a negative continuation outcome. This applies whether the transfer occurs within the same provider or involves the student moving to a different provider.

Indicator definitions: Completion measures

This section is aimed at readers who are seeking to better understand how students contribute to our student outcome and experience measures, and how the measures are presented.

Completion measures are used in the **access and participation data dashboard**, **student outcomes data dashboard,** and the **TEF data dashboard**.

50. Completion outcomes are measured by identifying a cohort of **entrants** to higher education qualifications at the provider and following them through subsequent years of their course to track how many continue in active study, or qualify in subsequent years. The methodology is consistent with the definition of the continuation indicator but based on different census points.

51. The completion measure tracks students from the date they enter a higher education provider and considers their completion outcomes at a census date.

    a. For full-time completion outcomes, and apprenticeship completion outcomes, the cohort-tracking census date is four years and 15 days after their commencement date.

    b. For part-time completion outcomes, the cohort-tracking census date is six years and 15 days after their commencement date.

Further information about data sources and coverage

52. The completion measures use the same Data sources and coverage as applies across all of the student outcome and experience measures used in our regulation. Table 6 provides further information about the data sources and coverage which are specific to the completion measures.

Table 6: Data sources and coverage for the completion measures

Data source(s)	Individualised DDB and ILR student records	
Years of data returns used: for completion measures informing TEF and regulation of student outcomes	2013-14 to 2022-23	The **full-time** and **apprenticeship** completion measures cover students entering higher education between: 17 July 2015 and 16 July 2016 (defines Year 1 of the time series) 17 July 2016 and 16 July 2017 (Year 2) 17 July 2017 and 16 July 2018 (Year 3) 17 July 2018 and 16 July 2019 (Year 4). The **part-time** completion measures cover students entering higher education between: 17 July 2013 and 16 July 2014 (defines Year 1 of the time series)

		17 July 2014 and 16 July 2015 (Year 2) 17 July 2015 and 16 July 2016 (Year 3) 17 July 2016 and 16 July 2017 (Year 4).
Years of data returns used: for completion measures in the access and participation data dashboard	2011-12 to 2022-23	The **full-time** and **apprenticeship** completion measures cover students entering higher education between: 17 July 2013 and 16 July 2014 (defines Year 1 of the time series) 17 July 2014 and 16 July 2015 (Year 2) 17 July 2015 and 16 July 2016 (Year 3) 17 July 2016 and 16 July 2017 (Year 4) 17 July 2017 and 16 July 2018 (Year 5) 17 July 2018 and 16 July 2019 (Year 6). The **part-time** completion measures cover students entering higher education between: 17 July 2011 and 16 July 2012 (defines Year 1 of the time series) 17 July 2012 and 16 July 2013 (defines Year 2 of the time series) 17 July 2013 and 16 July 2014 (Year 3) 17 July 2014 and 16 July 2015 (Year 4) 17 July 2015 and 16 July 2016 (Year 5) 17 July 2016 and 16 July 2017 (Year 6).
Population restrictions	Entrants registered at the higher education provider in question	As with all measures, completion measures exclude the students described at paragraph 24. They also exclude students on clinical medical, dental or veterinary science qualifications who take an intercalating year at the same provider (students who take an intercalating year at a different provider are not excluded).[30] For access and participation data they also exclude: • non-UK domiciled entrants • postgraduate entrants. For TEF data they also exclude postgraduate entrants.

[30] Intercalation involves an additional year of study on top of a medical, dental, or veterinary degree programme and an opportunity to develop knowledge and skills, and gain a standalone qualification, in a new area which may or may not be related to their main degree study.

Variable definitions by mode of study?	Yes	Definition varies with respect to census date: four year and 15 days after commencement for full-time and apprenticeship modes of study, six years and 15 days for part-time.

53. Postgraduate research students who are engaged in sequential collaborative provision (such as doctoral training programmes) are included in the entrant cohort of each provider that they register with, counting as an entrant at the point at which their registration with that provider commences rather than at their point of entry to higher education overall.

54. Information shown in Table 6 about the years of data returns used refers to entrant cohorts which have been defined to align with the census date periods for continuation and completion measures. An entrant year cohort is defined based on those students starting courses between 17 July and the following 16 July because this allows us to determine the activity of all students in that cohort on their census date in the following data reporting period.

Student outcomes counting positively towards completion measures

55. The completion measures count as positive outcomes those students who have either:

    a. Gained a higher education qualification from the provider at which they were previously identified as an entrant, on or before the relevant census date.

    b. Been recorded as actively studying for a higher education qualification at the same provider on the census date.

56. To be counted as a positive outcome in the circumstances described by paragraph 55, a student must have a record in the relevant DDB or ILR student returns that identified these outcomes.

57. Whether students can be identified with these outcomes on the basis of the DDB or ILR student returns can be established using the definitions provided in the 'Core algorithms' document and the instructions given in the 'Rebuild instructions' document. In particular readers may wish to consider the variables named IPENTRANTEXCL and IPCONINDFULL_YX.

58. There are certain circumstances in which it is useful to note the treatment of student outcomes for the purposes of defining the completion measures:

    a. Students who are recorded in the ILR datasets as having partially completed their qualification, or whose results are recorded as not yet known, are counted as a positive completion outcome.

    b. Postgraduate research students who (in the year of student data returns in which the relevant census date falls) were recorded as being awarded a qualification from a dormant mode of study are counted as a positive completion outcome.

c. Postgraduate research students who are engaged in sequential collaborative provision and transfer (on or before the census date) to a new provider as part of that arrangement are counted as a positive completion outcome for the provider they transferred from.

d. Other students who transfer (on or before the census date) to study a higher education qualification at a different provider than the one at which they were identified as an entrant are treated as a neutral completion outcome. This means that they are not included in the population of the completion indicator and are excluded from both the numerator and denominator involved in its calculation.

e. Students who are recorded in the DDB student data with results not yet known, are not counted as a positive completion outcome (because their qualification outcome will be reported in a subsequent year of DDB student data returns).

f. Students who transfer (on or before the census date) to the study of higher education modules, for credit only, or to the study of further education qualifications, are counted as a negative completion outcome. This applies whether the transfer occurs within the same provider or involves the student moving to a different provider.

Indicator definitions: Student experience measures

Due to the changes to the NSS introduced by the 2023 survey, the student experience measures in the TEF dashboard (published for the first time in summer 2024) are not directly comparable to previous ones constructed from responses to earlier years of the NSS. For this reason, this dashboard only presents student experience measures calculated from the 2023 survey onwards.

As described in the NSS quality report for 2024, we reviewed our approach to publication response rate thresholds, benchmarking, and to the number of themes that questions are grouped into and have not made any changes to these areas for NSS 2024 (compared to NSS 2023). The student experience measures published in the TEF dashboard (based on 2023 and later NSS responses) align with these approaches.[31]

The previous student experience measures (derived from NSS responses through to 2022 and used in TEF 2023) are described in full in the 'Description and definition of student outcome and experience measures' document published in September 2022.

This section is aimed at readers who are seeking to better understand how students contribute to our student outcome and experience measures, and how the measures are presented.

Student experience measures are used in the TEF data dashboard.

59. The student experience indicators are based on responses to the National Student Survey (NSS). They report on the extent to which students (most of which are in their final year) were positive when responding to questions about their experience in higher education.

60. Table 7 describes the student experience indicators calculated and the NSS question responses used in deriving the positivity measure for each theme. A full list of the NSS questions is available on the OfS website[32].

Table 7 NSS questions used in the calculation of the student experience measures

Student experience measure	NSS questions used
The teaching on my course	Q1, Q2, Q3, Q4
Learning opportunities	Q5, Q6, Q7, Q8, Q9
Assessment and feedback	Q10, Q11, Q12, Q13, Q14
Academic support	Q15, Q16

[31] See https://www.officeforstudents.org.uk/data-and-analysis/national-student-survey-data/nss-data-quality-report/

[32] See Annex A at https://www.officeforstudents.org.uk/publications/national-student-survey-2024/

Student experience measure	NSS questions used
Organisation and management	Q17, Q18
Learning resources	Q19, Q20, Q21
Student voice	Q22, Q23, Q24

Further information about data sources and coverage

61. The student experience measures use the same data sources and coverage as applies across all the student outcomes and experience measures used in our regulation. Table 8 provides further information about the data sources and coverage which are specific to the student experience measures.

Table 8: Data sources and coverage for the student experience measures

Data source(s)	National Student Survey	
Years of data returns used	2023 to 2024	The NSS-based indicators cover students surveyed during the spring of: 2023 (defines Year 1 of the time series) 2024 (Year 2)
Population restrictions	The NSS is targeted at final year undergraduates. Students on flexible provision or who change their study plans may also be included by participating providers.	The following exclusions apply: • Students who did not respond to the NSS. • Students who did not reach the final year of their course in any of the two most recent years. • Students not aiming for an undergraduate level qualification. • Students on courses which were not recognised for OfS funding. • Students aiming for a qualification consisting of one year of full-time equivalent study or less. • Students on clinical medical, dental, or veterinary science qualifications on an intercalating year (whether at the same provider or different) will not be included in respect to the course they

		are studying during the intercalating year.[33] • Students whose NSS responses have been suppressed as a result of the process for investigating concerns that students have been inappropriately influenced.
Variable definitions by mode of study?	No	
Response rate requirements	Measures constructed on the basis of response rates below 50 per cent are suppressed	Response rates are calculated separately with respect to the population informing calculation of each indicator and split indicator.

62. Responses to the National Student Survey are linked to the DDB, HESA and ILR student data returns which correspond to the year the survey target list is drawn from.

63. Data returns for the academic year that corresponds to the year prior to the academic year in which the survey is conducted are used to:

    a. Create the survey target list.[34]

    b. Establish the student and course characteristics of those individuals included on the target list, including which categories of level of study, subject of study and other personal characteristics that relate to split indicators they fall into.

64. For example, students surveyed in the spring of 2024 (i.e. during the 2023-24 academic year) would have been identified for the survey target list by using the DDB and ILR student data returns for the 2022-23 academic year to establish whether the student was in their penultimate year of study. Information about their subject of study (and other characteristics) in 2022-23 determines which categories the respondent will contribute to for the purposes of summarising NSS responses, including through the construction of split indicators.

65. There are two special cases where information from earlier academic years is used:

    a. Where students were dormant in the year the survey target list is drawn from.

    b. Where students on clinical medical, dental, or veterinary science qualifications take an intercalating year in the year the survey target list is drawn from. This typically happens in

[33] Intercalation involves an additional year of study on top of a medical, dental, or veterinary degree programme and an opportunity to develop knowledge and skills, and gain a standalone qualification, in a new area which may or may not be related to their main degree study.

[34] Providers can add or remove students from the NSS target list where their circumstances have changed from those observed from the student data returns.

the penultimate year of study and so commonly coincides with the year the target list is drawn from.

66. In both cases, to obtain all other student and course characteristics used in the calculation of the student experience measures, we link to the record for the same instance of study in the last academic year in which that student instance was active (up to two years prior), rather than using the information available from the year of the target list, which may be partial or unrepresentative. For example, the intercalating students will typically be associated with the subject of study from the year prior to the intercalation (medicine and dentistry or veterinary sciences) rather than the subject or subjects associated with the intercalation year.

NSS responses counting positively towards student experience measures

67. The NSS asks a range of individual questions, each of which has four possible valid answers, two positive and two negative. Wording varies by question, but in each case within the four-point scale we will consider the first two response options as positive, as set out through Proposal 2 of the 2023 consultation on the approach to publication of results of the NSS.[35]

68. The NSS questions are organised into several groups, known as themes. Across the questions that make up a given theme, the measure of each student's positivity is calculated as the percentage of their responses that were among the two positive options for the question. Questions with an answer of 'this does not apply to me', or which were not answered, are ignored.

69. The student experience measure for a provider is then calculated as the sum of the percentage of positive responses for each respondent divided by the number of respondents. This method ensures that each individual student has the same weight in calculating the student experience measure for the theme, regardless of how many questions they answered, as long as at least one was answered with a response other than 'this does not apply to me'.

70. The worked example in Table 9 illustrates how we would calculate the 'Learning Resources' student experience measure for a provider. To simplify, the example assumes that there were only five respondents. The overall indicator for the provider would be 66.7 per cent (the sum of the percentages divided by the number of respondents: (100+100+0+66.7+66.7=33.4) ÷ 5 = 66.7). The 'this does not apply to me' option is excluded when calculating the positivity for respondent B.

[35] See Consultation on the approach to publication of results of the National Student Survey - Office for Students

Table 9: Worked example for calculation of a student experience measure

Respondent	Response to: How well have the IT resources and facilities supported your learning?	Response to: How well have the library resources supported your learning?	Response to: How easy is it to access subject specific resources when you need them?	Positivity measure for the respondent (%)
A	Very well	Well	Easy	100
B	Well	Well	This does not apply to me	100
C	Not very well	Not at all well	Not very easy	0
D	Well	Not very well	Easy	66.7
E	Very well	Not very well	Very easy	66.7

Indicator definitions: Degree outcomes measure

This section is aimed at readers who are seeking to better understand how students contribute to our student outcome and experience measures, and how the measures are presented.

The degree outcomes measures are used in the **access and participation data dashboard**, where they are also referred to as the **attainment** measures.

71. The degree outcomes measure expresses the number of **qualifiers** from Level 6+ undergraduate degrees who were awarded 1st class or 2:1 degree classifications as a percentage of all those qualifiers from Level 6+ undergraduate degrees who were awarded classified degrees.

72. Whether students can be identified as being awarded a first or upper second degree classifications can be established on the basis of DDB or ILR data returns, using the definitions provided in the 'Core algorithms' document and the instructions given in the 'Rebuild instructions' document. In particular readers may wish to consider the variables named DFAPAPPEXCL, IPDOQUALPOP, and IPDODEGCLASS.

73. Due to changes in the DDB's Student record, the small proportion of qualifiers who were previously recorded as having achieved a 'pass - degree awarded without honours following an honours course' classification for their qualification can no longer be counted in this measure. These qualifiers will continue to be counted negatively in 2021-22 and before, but for 2022-23 onwards will be excluded from the indicator population.

Further information about data sources and coverage

74. The degree outcome measures use the same Data sources and coverage as applies across all of the student outcome and experience measures used in our regulation. Table 10 provides further information about the data sources and coverage which are specific to the degree outcome measures.

Table 10: Data sources and coverage for the degree outcomes measures

Data source(s)	Individualised DDB and ILR student records	
Years of data returns used	2017-18 to 2022-23	Degree outcomes measures cover qualifiers leaving higher education in academic year between: 1 August 2017 and 31 July 2018 (defines Year 1 of the time series) 1 August 2018 and 31 July 2019 (Year 2) 1 August 2019 and 31 July 2020 (Year 3) 1 August 2020 and 31 July 2021 (Year 4)

		1 August 2021 and 31 July 2022 (Year 5) 1 August 2022 and 31 July 2023 (Year 6).
Population restrictions	UK domiciled qualifiers from Level 6+ undergraduate degrees	As with all measures, degree outcomes measures exclude the students described at paragraph 24. They also exclude: • Non-UK domiciled qualifiers. • Students who were not awarded an undergraduate Level 6+ degree qualification. • Students on clinical medical, dental, or veterinary science qualifications who take an intercalating year which does not result in a qualification being recorded.[36] • Students who were awarded Level 6+ degrees without an honours classification. In the event that a student is identified as receiving more than one undergraduate degree qualification from the same provider in the same reporting period, this means that we will select the best classification outcome reported.
Variable definitions by mode of study?	No	

[36] Intercalation involves an additional year of study on top of a medical, dental, or veterinary degree programme and an opportunity to develop knowledge and skills, and gain a standalone qualification, in a new area which may or may not be related to their main degree study.

Indicator definitions: Progression measures

This section is aimed at readers who are seeking to better understand how students contribute to our student outcome and experience measures, and how the measures are presented.

Progression measures are used in the **access and participation data dashboard**, **student outcomes data dashboard,** and **TEF data dashboard**.

75. Progression measures are constructed from data reported through the Graduate Outcomes survey. They report on the proportion of **qualifiers** from higher education qualifications who have been included on the Graduate Outcomes target list, responded to the survey, and reported that they have progressed to professional or managerial employment, further study, or other positive outcomes, 15 months after gaining their qualification. [37]

Further information about data sources and coverage

76. The progression measures use the same Data sources and coverage as applies across all of the student outcome and experience measures used in our regulation. Table 11 provides further information about the data sources and coverage which are specific to the progression measures.

Table 11: Data sources and coverage for the progression measures

Data source(s)	Individualised DDB and ILR student records, Graduate outcomes survey	
Years of data returns used	2017-18 to 2021-22	Progression measures cover students who qualified and have been linked to their responses to the Graduate Outcomes survey. For the student outcomes data dashboard, they are linked to academic year: 2018-19 (defines Year 1 of the time series) 2019-20 (Year 2) 2020-21 (Year 3) 2021-22 (Year 4). For the access and participation data dashboard they are linked to academic year: 2017-18 (defines Year 2 of the time series) 2018-19 (Year 3)

[37] See www.hesa.ac.uk/collection/c21071/coverage

		2019-20 (Year 4) 2020-21 (Year 5) 2021-22 (Year 6).
Population restrictions	UK domiciled qualifiers	As with all measures, progression outcomes measures exclude the students described at paragraph 24. They also exclude: • Non-UK domiciled qualifiers. • Students who did not achieve a higher education qualification after following a higher education course intended to lead to the award of a qualification. • Students on clinical medical, dental, or veterinary science qualifications who qualify from an intercalating year (whether at the same provider or different).[38] • Students who were recorded in the ILR datasets as having partially completed their qualification, or whose results are recorded as not yet known. • Students on courses which were not recognised for OfS funding. • Students who did not respond to the Graduate Outcomes survey. For TEF and access and participation data they also exclude postgraduate qualifiers. Progression outcome measures are reported separately for qualifiers at each mode and level of study. The level of study reflects the level of the student's qualification aim at the commencement of their studies.
Variable definitions by mode of study?	No	
Response rate requirements	Measures constructed on the basis of response	Response rates are calculated separately with respect to the population informing

[38] Intercalation involves an additional year of study on top of a medical, dental, or veterinary degree programme and an opportunity to develop knowledge and skills, and gain a standalone qualification, in a new area which may or may not be related to their main degree study.

	rates below 30 per cent are suppressed.	calculation of each indicator and split indicator.

Student outcomes counting positively towards progression measures

77. The progression measures count as positive outcomes those students who report in their response to the Graduate Outcomes survey, 15 months after gaining their qualification:

    a. Managerial or professional employment (defined as employment in an occupation which falls within major groups 1 to 3 of the Office for National Statistics (ONS) Standard Occupational Classification 2020).

    b. Further study at any level of study.

    c. Travelling, caring for someone else or retirement.

78. To be counted as a positive outcome in the circumstances described by paragraph 77, a student must have reported that they were undertaking at least one of these activities during the census week for the Graduate Outcomes survey.

79. Whether students can be identified with these outcomes on the basis of the Graduate Outcomes survey can be established using the definitions provided in the 'Core algorithms' document and the instructions given in the 'Rebuild instructions' document. In particular, readers may wish to consider the variables named IPEMPEXCL, IMPEMPRESPONSE, IPEMPINDPOP, IPEMPIND, and IPEMPINDNUM.

80. There are certain circumstances in which it is useful to note the treatment of student outcomes for the purposes of defining the progression measures:

    a. Students who reported working in self-employment, voluntary or unpaid roles are treated in the same way as those in paid employment, meaning that students count towards a positive progression outcome if the information they provide about their job or employer names and duties identifies it as a managerial or professional occupation.

    b. Students who report that they were 'doing something else' count as a positive progression outcome only if this is reported in combination with another activity that counts as a positive outcome (e.g. professional employment or further study).

    c. Students who report that they are due to start a job or studying in the next month count as a positive progression outcome only if this is reported in combination with another activity that counts as a positive outcome.

    d. Students who report that they are unemployed at the census date, or not otherwise engaged in activities that count as a positive outcome, but that they were previously employed or had undertaken study since completing their higher education course do not count as a positive outcome towards the progression measures.

    e. Students who submit partial responses to the Graduate Outcomes survey but completed the first two questions of the survey are counted as responses for the purposes of

calculating response rates and constructing student outcome measures. Where a partial response leads to the student identifying that they were in employment at the Graduate Outcomes census date, but not providing information about their job or employer names and duties, we calculate the likelihood of that student being in managerial or professional employment. We do this by apportioning the partial response between managerial or professional and non-managerial or professional employment in the same ratio as has been derived for the provider, mode, and level of study of the student in question.

Further information about use of the Graduate Outcomes survey to define progression measures

81. The Graduate Outcomes survey asks respondents which of 11 possible activities they had been doing during the census week, 15 months after they gained their higher education qualification. They could respond that they were undertaking multiple activities. Table 12 shows the 11 possible activities and whether they count as positive outcomes towards the progression measure. Where a student reported that they were undertaking multiple activities, the student's outcome counts as a positive outcome towards the progression measures if **any** of their activities are counted as positive.

Table 12: Activities reported by students in the Graduate Outcomes survey

Activity	Counted as a positive outcome towards progression measures
Paid work for an employer	Yes, if managerial or professional employment.
Self-employment/freelancing	Yes, if managerial or professional employment.
Running own business	Yes, if managerial or professional employment.
Developing a creative, artistic or professional portfolio	Yes, if managerial or professional employment.
Voluntary/unpaid work for an employer	Yes, if managerial or professional employment.
Engaged in a course of study, training or research	Yes.
Taking time out to travel – this does not include short-term holidays	Yes.
Caring for someone (unpaid)	Yes.
Retired	Yes.
Unemployed and looking for work	No, not unless the student reports this activity in combination with another one that does count as positive.
Doing something else	No, not unless the student reports this activity in combination with another one that does count as positive.

82. Whether a respondent to the Graduate Outcomes survey is identified as being in **managerial or professional employment** is based on the details they have provided about their job or

employer names and duties. Within the Graduate Outcomes survey, jobs are mapped to the 10 major groupings of the ONS Standard Occupational Classification (SOC 2020). Table 13 shows which of the major groups are classified as being managerial or professional employment.[39]

Table 13: SOC 2020 groupings counted as professional employment

SOC2020 major group code	SOC2020 major group label	Counted as managerial or professional employment
1	Managers, directors, and senior officials	Yes
2	Professional occupations	Yes
3	Associate professional and technical occupations	Yes
4	Administrative and secretarial occupations	No
5	Skilled trade occupations	No
6	Caring, leisure, and other service occupations	No
7	Sales and customer service occupations	No
8	Process, plant, and machinery operatives	No
9	Elementary occupations	No

83. In some cases, the respondent identified that they were in employment but not engaged in any of the other activities that count positively (as shown in Table 12) and did not provide details of their job or the information provided cannot be mapped to a SOC code. In these cases the response is apportioned between professional and non-professional employment in the same ratio between professional and non-professional employment that has been derived for that provider, mode of study, and broad level of study.

84. For example, suppose that a provider has 100 students from a given mode and broad level of study who responded to the Graduate Outcomes survey that they were in employment (with known SOC codes): 65 of these are in professional employment and the remaining 35 are in non-professional employment. They are undertaking no other activities that count as positive outcomes. For this example provider there are also 10 respondents from the same mode and broad level of study that report that they are in employment with no other activities that count as a positive outcome, and their associated SOC codes are not known. In this case each of the 10 responses are individually weighted so that each one contributes 0.65 towards the number in professional employment for that provider and 0.35 towards the number in non-professional employment.

85. The Graduate Outcomes survey includes a question as to whether the respondent has undertaken any employment in the interim 15-month period between qualifying and the census week. Similarly, the survey includes a question as to whether the respondent is due to start any employment in the next month. This information has not been used in the calculation of

[39] "The ONS recently introduced a new code to its Standard Occupational Classification, for Henna/Tattoo Artists. The code has been assigned by the ONS to a SOC major group that maps to our established definition of a professional or managerial occupation. We are proposing to align with ONS and include the new SOC code for Henna/Tattoo Artists as a professional or managerial occupation in the graduate outcomes data for 2021-22 qualifiers onwards. The classification for earlier years has not changed.

progression measures and therefore does not count towards the identification of employment activities.

86. Whether a respondent to the Graduate Outcomes survey is identified as being in **further study** is based only on the response to the question about which of 11 possible activities they had been doing during the census week. Because any further study counts towards a positive progression outcome (even if it is at the same or lower level than the qualification the student recently obtained) it is not necessary to use any other information from the survey.

87. The survey includes a question as to whether the respondent has undertaken any further study in the interim 15-month period between qualifying and census week. Similarly, the survey includes a question as to whether the respondent is due to start any study in the next month. This information has not been used in the calculation of progression measures and therefore does not count towards the identification of further study activities.

Structure and reporting

This section is aimed at readers who are seeking to better understand how students contribute to our student outcome and experience measures, and how the measures are presented.

The structure and reporting principles described in this section are relevant to the **access and participation data dashboard, student outcomes data dashboard,** and the **TEF data dashboard**.

88. Student outcome and experience measures are constructed and reported through a series of indicators and split indicators. The reporting structure includes the following components:

    a. An **indicator** is defined as the student outcome or experience measure in question being reported separately according to the combination of a students' **mode** and **level** of study. It means, for example, that we generate indicators which report:

        - Continuation outcomes for full-time students on first degree programmes,

        separately from:

        - Continuation outcomes for part-time students on first degree programmes,

        separately from:

        - Continuation outcomes for part-time students on postgraduate research degrees,

        and so on.

For further information about the definition of indicators, see Indicators.

    b. **Split indicators** are defined as the student outcome or experience measure being reported as a further breakdown of student groups within the mode and level of study to which the indicator refers. Split indicators report separately on subject studied, student characteristics, year of entry or qualification (as appropriate to the student outcome in question), specific course types and provider partnership arrangements.

For further information about the definition of split indicators, see Split indicators.

    c. For the access and participation data dashboards, indicators represent both an aggregate and the individual years of a time series of multiple years of data. This means that the reporting structure involves additional indicators and split indicators which accommodate intersections of year with each of the different student characteristics, as well as a limited selection of intersections between student characteristics, in order to support our regulatory objectives for access and participation.

d. Indicators and split indicators are reported within up to four separate **views of a provider's student population** (registered students, taught students, students taught or registered by the provider (TorR), and students associated with the provider through validation or subcontractual partnerships).

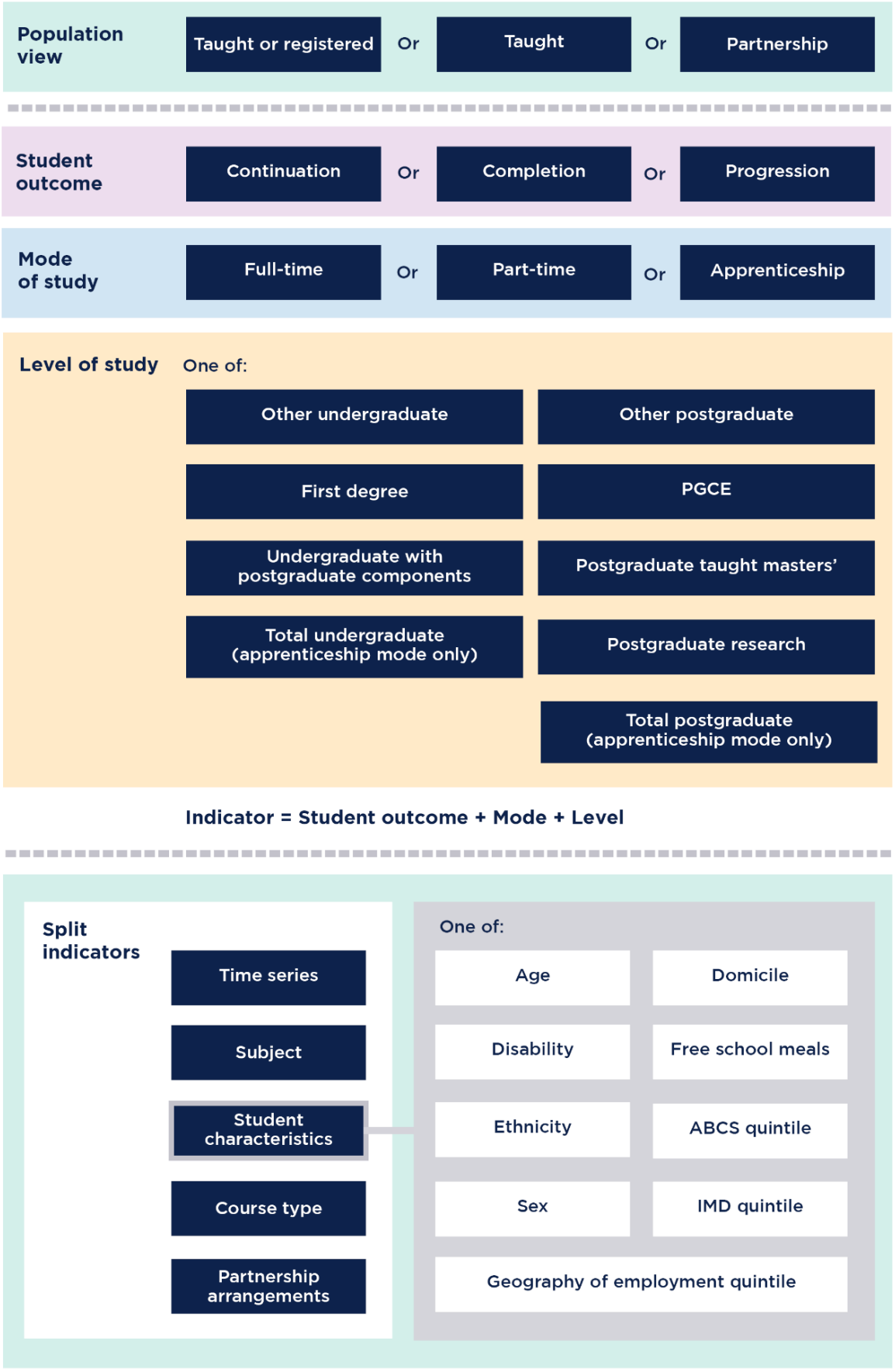
For further information about the definition of views of a provider's student population, see Views of a provider's student population.

89. We use a variety of different sections of the general reporting structure to construct the data resources that inform our assessments, depending on the function of the assessment. In doing so, we select the sections and student populations that represent the closest alignment with the scope of our regulation in respect of that function and focus on the data that best meets our regulatory objectives and user needs. This allows for tailored data resources, which benefit from consistency and transparency in their underlying definition. For example, indicator and split indicator values reported in relation to a given mode or level of study, or student characteristic, may differ across the outputs produced for the TEF data dashboard compared with the access and participation data dashboard on account of their different regulatory scope and objectives. All individual students are, however, subject to calculations based on the same definitions of positive outcomes, and mode and levels of study categories. This means that an individual student's contribution to our student outcome and experience measures remains unchanged, whether or not they fall into the relevant population for a given function.

90. The indicators and split indicators for use in the regulation of student outcomes, the TEF and regulation of access and participation are illustrated in Figures 5, 6 and 7.

91. The indicators and split indicators constructed to inform our regulation of student outcomes through condition B3 is represented in Figure 5.

Figure 5: Reporting structure for indicators and split indicators used in the regulation of student outcomes

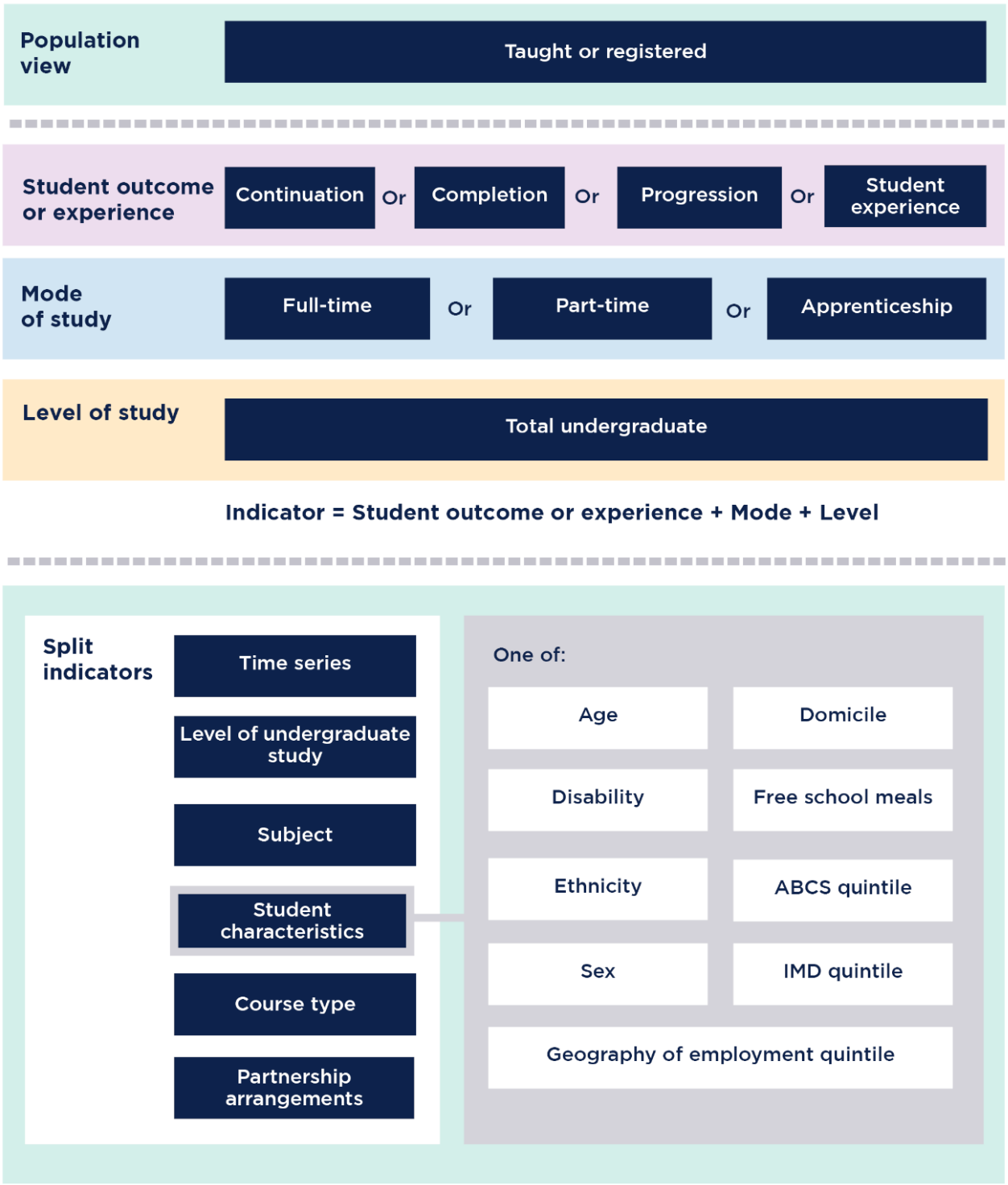


92. The indicators and split indicators shown in Figure 5 are reported for each combination of mode of study and individual level of study. They cover students at undergraduate and postgraduate levels of study to provide appropriate alignment with the scope of our regulation of student outcomes.

93. The indicators and split indicators shown in Figure 5 report data for students that are taught or registered by the provider (or both), students that are taught by the provider, and students associated with the provider through partnerships arrangements. In doing so, we report all the indicators and split indicators shown in Figure 5 when looking at either the taught or taught or registered (TorR) view of the student population. For the partnerships view the split indicators would only include the indicators and the split indicators showing subject studied, year of entry or qualification and type of partnership.

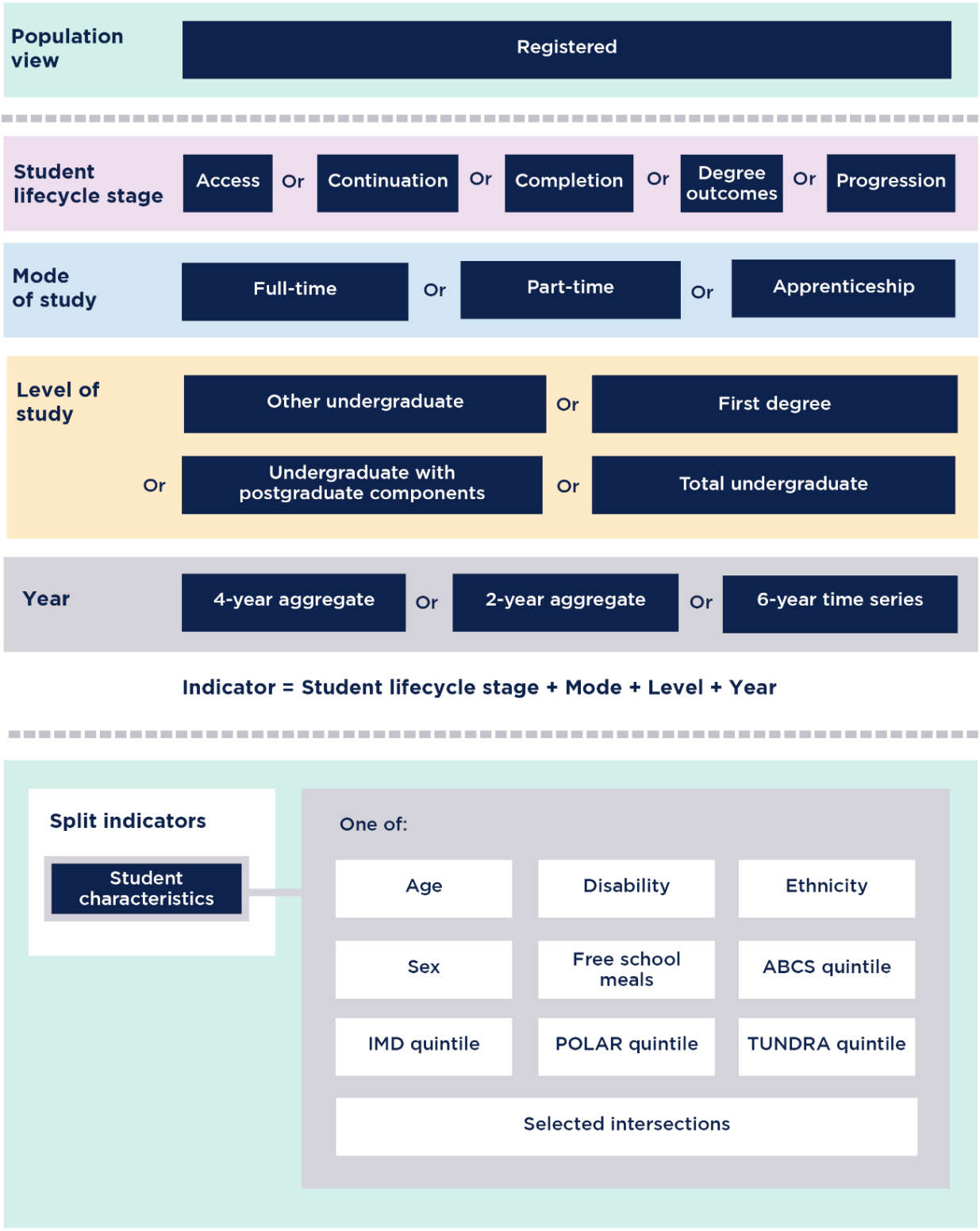
94. The reporting structure used for the purposes of constructing TEF indicators is shown in Figure 6. It covers students at undergraduate levels of study to provide appropriate alignment with the scope of TEF assessment for which we have data available. The indicators and split indicators shown in Figure 6 are reported for each mode of study based on the combination of students at all undergraduate levels of study. They report data for students that are taught or registered by the provider (or both).

Figure 6: Reporting structure for indicators and split indicators used in the TEF data dashboard



95. The reporting structure for measures constructed for the access and participation data dashboard is shown in Figure 7. Coverage is limited to UK-domiciled undergraduates throughout, to provide appropriate alignment with the scope of access and participation plans, as prescribed through regulations made under HERA. The indicators and split indicators shown in Figure 7 are reported for each year, for each combination of mode of study and level of undergraduate study. They report data for students that are registered by the provider.

Figure 7: Reporting structure for indicators and split indicators used in the access and participation data dashboard



Indicators

96. Calculation of an **indicator** involves each measure being reported separately according to the unique combinations of a student's mode and level of study. They represent overall performance across all types of courses, subjects, and student groups studying within the given combination of mode and level of study.

97. For the regulation of student outcomes and the TEF, indicators represent an aggregate of the **most recent four years** of available data that contribute to the construction of the relevant measure. The data dashboards we have published refer to these as the 'Overall indicator'.

98. For the regulation of access and participation, indicators currently represent both aggregates and the individual years of the **most recent six years** of available data for the relevant measure. The access and participation data dashboard reports indicators based on two aggregations of this time series, based on the most recent two and four years of available data for the relevant measure.

For further information about the years of data that contribute to the construction of indicators for each student outcome or experience measure, see:

Indicator definitions: Access to higher education measures

Indicator definitions: Continuation measures

Indicator definitions: Completion measures

Indicator definitions: Student experience measures

Indicator definitions: Degree outcomes measure

Indicator definitions: Progression measures

99. The definition of indicators involves reporting on the unique combinations of each category of modes and levels of study. There are three categories of mode of study (full-time, part-time and apprenticeship), which intersect with the levels of study as shown in Table 1 to create:

- 16 unique combinations (or indicators) for each student outcome measure used in our regulation of student outcomes
- 3 unique combinations for each student outcome and experience measure used in the TEF
- 9 unique combinations for each student outcome and experience measure for use in the access and participation data dashboard (noting that when these intersect with the six individual years of data to form indicators for this purpose, the number of indicators would increase to 54).

Table 14: Levels of study which result in the construction of an indicator when combined with mode of study categories

Level of study	Access and participation	Student outcomes	TEF
Other undergraduate	Full-time and part-time	Full-time and part-time	Not an indicator
First degree	Full-time and part-time	Full-time and part-time	Not an indicator
Undergraduate with postgraduate components	Full-time and part-time	Full-time and part-time	Not an indicator
All undergraduates	Full-time, part-time and apprenticeship	Apprenticeship	Full-time, part-time and apprenticeship
Other postgraduate	Not in scope	Full-time and part-time	Not in scope
PGCE	Not in scope	Full-time and part-time	Not in scope
Postgraduate taught masters'	Not in scope	Full-time and part-time	Not in scope
Postgraduate research	Not in scope	Full-time and part-time	Not in scope
All postgraduates	Not in scope	Apprenticeship	Not in scope

Mode and level of study definitions

100. Students are always attributed to the mode of study categories on the basis of the mode of study reported in the first year of their programme of study:

a. Student outcome and experience indicators which report on entrant cohorts – those measuring access to higher education, and continuation and completion rates – take the mode of study from the year in which we identify a student as an entrant who contributes to calculation of the measure.

b. Student outcome and experience indicators which report on cohorts other than entrants – those measuring student experience, degree outcomes and progression rates – track students back to the earliest student record submitted by their provider for the programme on which they are a final year student in the year that they contribute to calculation of the measure. The student will be categorised according to the mode of study reported by the provider in the earliest student record located for the student, even if later records for the same student identify that they subsequently changed to a different mode.

101. Students are attributed to levels of study according to level of the qualification aimed for in the relevant year for the student outcome or experience measure in question:

a. Student outcome indicators which report on entrant cohorts – those measuring access to higher education, and continuation and completion outcomes – are associated with the level of study of the student's qualification aim in the first year of study.

b. Student outcome indicators which report on qualifier cohorts other than entrants – those measuring degree outcomes and progression rates – are associated with the level of study of the student's qualification aim in the final year of study.

c. Student experience indicators are associated with the level of study of the student's qualification aim in the year in which the survey population is drawn, usually their penultimate year of study

102. Table 15 provides information about the levels of study used in the definition of indicators (and split indicators) and the types of courses that are included in each category.[40]

Table 15: Levels of study used in the definition of indicators

Level of study	Definition
First degree	First degree courses mostly consist of study for qualifications such as honours or ordinary degrees, including Bachelor of Arts (BA) and Bachelor of Science (BSc) degrees.
Other undergraduate	Other undergraduate includes courses such as foundation degrees, diplomas and certificates of higher education at Levels 4 and 5 (including those accredited by professional or statutory bodies, such as the Association of Accounting Technicians or the Chartered Institute of Building), Higher National Diploma (HND) and Higher National Certificate (HNC).
Undergraduate with postgraduate components	Examples of undergraduate courses with postgraduate elements include: integrated undergraduate-postgraduate taught masters' degrees on the enhanced or extended pattern; pre-registration medical degrees regulated by the General Medical Council; and pre-registration dentistry degrees regulated by the General Dental Council.
All undergraduates	The aggregation of first degree, other undergraduate and undergraduate with postgraduate components.
Other postgraduate	Examples of other postgraduate courses include: graduate or postgraduate diplomas, certificates or degrees at Levels 5 and 6 where a Level 5 or 6 qualification is a pre-requisite for course entry; postgraduate certificates and diplomas; diplomas in teaching in the lifelong learning sector at Level 7; post-registration health and social care qualifications at Level 7; and taught qualifications at Level 7

[40] For full technical detail on the data definitions of the levels of study used, see the 'Technical algorithms for student outcome and experience measures' document at https://www.officeforstudents.org.uk/publications/description-and-definition-of-student-outcome-and-experience-measures/

	leading towards obtaining eligibility to register to practice with a health or social care or veterinary statutory regulatory body.
PGCE	A PGCE is a higher education programme providing both professional training leading to qualified teacher status (QTS) and a course of academic study leading to an academic qualification. The title Postgraduate Certificate in Education is reserved for those PGCE qualifications that are set at Level 7. The title Professional Graduate Certificates in Education is used for those PGCE qualifications that are set Level 6.
Postgraduate taught masters'	Examples of postgraduate taught masters' courses include masters' degrees (such as MA, MSc).
Postgraduate research	Examples of postgraduate research courses include: doctoral degrees (such as PhD/DPhil, EdD); masters' degrees by research (such as MPhil, MRes).
All postgraduates	The aggregation of other postgraduate, PGCE, postgraduate taught masters' and postgraduate research.

Split indicators

103. Calculation of a **split indicator** involves a further break down of the student population that has been included in the calculation of an indicator, where these break downs relate to various categories of a provider's students and provision. As a further breakdown of an indicator, it follows that the split indicators are each reported separately by mode and level of study.

104. The split indicators we construct vary according to the regulatory function we are constructing them for.

Split indicators for the regulation of student outcomes

105. For the regulation of student outcomes, split indicators cover:

- a four-year time series based on year of entry or qualification (as appropriate to the student outcome in question)
- student characteristics
- age on entry; disability; ethnicity; sex
- domicile, ABCS quintile, deprivation quintile (IMD), eligibility for free school meals (FSM); geography of employment quintile (for progression measures only)
- these split indicators are not constructed for the partnerships view of the provider's student population
- subject studied

- subjects based on level 2 of the common aggregation hierarchy (except that Celtic studies is aggregated with the languages and area studies grouping)
- specific course types
- other undergraduate courses at Level 4, and at Level 5+
- first degree courses with integrated foundation years
- Higher Technical Qualifications will be included as the data becomes available
- these split indicators are not constructed for the partnerships view of the provider's student population
- provider partnership arrangements
- for further information see Table 16.

For further information about the student and course characteristics used in the construction of split indicators for the purposes of regulating student outcomes, see Annex B: Further information about the definition of split indicators.

Split indicators for use in the TEF

106. For TEF purposes, split indicators cover:

- a four-year time series based on year of entry or qualification (as appropriate to the student outcome or experience in question)
- level of undergraduate study
- student characteristics
- age on entry; disability; ethnicity; sex.
- domicile, ABCS quintile (for continuation, completion and progression measures), deprivation quintile (IMD), eligibility for free school meals (FSM); geography of employment quintile (for progression measures only)
- subject studied
- subjects based on level 2 of the common aggregation hierarchy (except that Celtic studies is aggregated with the languages and area studies grouping)
- specific course types
- other undergraduate courses at Level 4, and at Level 5+
- first degree courses with integrated foundation years
- Higher Technical Qualifications will be included as the data becomes available

- provider partnership arrangements
- for further information see Table 16.

For further information about the student and course characteristics used in the construction of split indicators for the purposes of the TEF, see Annex B: Further information about the definition of split indicators.

Split indicators for use in the access and participation dashboard

107. For the access and participation dashboard, indicators represent both an aggregate and the individual years of a time series of multiple years of data. Consequently, the split indicators we construct for the following student characteristics report on their intersection with individual years of a time series as well as at an aggregate level based on the whole time series:

- age on entry; disability; disability type; ethnicity; sex
- ABCS quintile (for access, continuation, completion and progression measures), deprivation quintile (IMD 2015 and IMD 2019), eligibility for free school meals (FSM); Participation of Local Areas quintile (POLAR4); tracking underrepresentation by area quintile (TUNDRA).

For further information about the student characteristics used in the construction of split indicators for the purposes of the access and participation data dashboard, see Annex B: Further information about the definition of split indicators.

Views of a provider's student population

108. There are four different views of a provider's student population which are defined as:

a. **Taught or registered (TorR) population:** This view is used in the student outcomes data dashboard, and in the TEF data dashboard. These are students who are either registered or taught at the provider in question, including those who are taught and registered by the same provider, subcontracted in to the provider for teaching, and subcontracted out to another provider for teaching.

b. **Registered population:** This view is used in the access and participation data dashboard. These are students who are registered at the provider in question. They may also be taught at that provider, or they may be taught elsewhere, at another provider, under a subcontractual or partnership arrangement (subcontracted out).

c. **Taught population**: This view is used in the student outcomes data dashboard. These are any students who are taught at the provider in question. This may be the same provider where they are registered (taught and registered) or it may be that the provider in question is teaching the student under a subcontractual partnership arrangement (subcontracted in).

d. **Partnership population:** This view is used in the student outcomes data dashboard. These are students who are either:

   - Registered by the provider in question and taught elsewhere, at another provider, under a subcontractual partnership arrangement (subcontracted out); or

   - Neither taught nor registered by the provider in question, but that provider acts as the awarding body for the qualification that a student is studying (validation-only).

109. The teaching provider is the provider where the student received the majority of their teaching in the relevant year for the student outcome or experience measure in question:

   a. Student outcome indicators which report on entrant cohorts – those measuring access to higher education, and continuation and completion outcomes – identify the teaching provider as the one which delivered the majority of the teaching in the student's first year of study.

   b. Student outcome indicators which report on qualifier cohorts other than entrants – those measuring degree outcomes and progression rates – identify the teaching provider as the one which delivered the majority of the teaching in the student's final year of study.

   c. Student experience indicators identify the teaching provider as the one which delivered the majority of the teaching in the penultimate year of study.

Split indicators for provider partnership arrangements for different views of a provider's student population

110. The provider partnership arrangements split indicators vary in their definition according to the view of a provider's student population they are being constructed for. The categories of the partnership arrangements split indicators are shown in Table 16, along with their relevance to the different provider views of student populations.

Table 16: Summary of partnership arrangement split indicator categories and their relevance to provider views of student populations

Split indicator category	Nature of the teaching arrangement	Provider views of student populations to which the category is relevant
Taught and registered	The provider registering the student is also teaching them directly	Taught view (used in the student outcomes data dashboard)
Taught only	The students are subcontracted in to the provider	Taught view (used in the student outcomes data dashboard)
Registered only	The students are subcontracted out from the provider	Taught or registered view (used in the student outcomes data dashboard and the TEF data dashboard)

Split indicator category	Nature of the teaching arrangement	Provider views of student populations to which the category is relevant
		Partnership view (used in the student outcomes data dashboard)
Validation only	The students are neither taught nor registered by the provider, but study for an award of that provider	Partnership view (used in the student outcomes data dashboard)
Taught	The provider is teaching the students and may or may not also be registering them	Taught or registered view (used in the student outcomes data dashboard and the TEF data dashboard)

Presentation

This section is aimed at readers who are seeking to better understand how the student outcome and experience measures are presented, and how they can be interpreted.

Most definitions and principles discussed in this section are relevant to the **access and participation data dashboard, the student outcomes data dashboard** and the **TEF data dashboard.**

The section relating to **rounding and suppression** is also relevant to the **size and shape of provision dashboard.**

The section relating to **benchmarking** is only relevant to the **student outcomes data dashboard** and to the **TEF data dashboard.**

Elements included in the presentation of student outcome and experience measures

111. When reporting student outcome and experience indicators and split indicators, the following information contributes to, or results from, our calculations. Most of these are included in our data outputs:

a. Denominator of the indicator: The total number of students in the population for which we are measuring outcomes or experiences.

b. Numerator of the indicator: The number of students who achieve the outcome or experience in question.

c. Indicator value (as a proportion): Calculated in percentage terms as the numerator divided by the denominator. This is the rate at which students have achieved the outcome or experience in question, expressed as a point estimate providing a factual representation of the actual population of students present at a particular provider at a particular time.

d. The distribution of statistical uncertainty around the indicator values that we have calculated, where this distribution relies on the calculation of a series of discrete confidence intervals.

e. Survey response rate (for progression outcomes and student experience measures): Calculated in percentage terms as the number of students who responded to the relevant survey, divided by the total number of students eligible to be surveyed.

f. Benchmark value (as a proportion): Calculated in percentage terms for each provider as a weighted sector average which takes account of that provider's particular mix of students. Benchmarks give information about the values that the sector overall might have achieved for the indicator if the characteristics included in the benchmarking factors are the only ones that are important. They are reported in the student outcomes data dashboard, and the TEF data dashboard.

g. Difference between indicator and benchmark values: This is a point estimate of the difference between the indicator and benchmark (expressed as indicator minus benchmark). This is reported in the student outcomes data dashboard, and the TEF data dashboard.

h. The distribution of statistical uncertainty around the difference between indicator and benchmark values that we have calculated, where this distribution relies on the calculation of a series of discrete confidence intervals. This is reported in the student outcomes data dashboard, and the TEF data dashboard.

i. Contribution to own benchmark: Calculated in percentage terms for each provider as the weighted average of the provider's own students contributing to the sector averages that are used to calculate their benchmark. This is reported in the student outcomes data dashboard, and the TEF data dashboard.

j. Proportions of the statistical uncertainty distribution that fall above and below the provider's benchmark value. This is reported in the student outcomes data dashboard, and the TEF data dashboard. For the TEF data dashboard, the proportion of the statistical uncertainty distribution that falls broadly in line with the benchmark is also reported.[41]

k. Proportions of the statistical uncertainty distribution that fall above and below the relevant numerical threshold used in regulation of student outcomes. This is reported in the student outcomes data dashboard.

l. Percentage point gap: the difference between two split indicators within the same split type. This is reported in the access and participation data dashboard.

m. The distribution of statistical uncertainty around the percentage point gaps we have calculated, where this distribution relies on the calculation of a series of discrete confidence intervals. This is reported in the access and participation data dashboard.

n. Ratio: the ratio between two indicators with the same split type. This is reported in the access and participation data dashboard.

o. The change from year 5 to year 6: the change in indicators and gaps in the latest year of the time series. This is reported in the access and participation data dashboard.

p. Two- and four-year aggregates across the time series: in addition to the six-year time series in the access and participation dashboard, denominators, numerators, indicators, gaps, ratios, response rates are reported as an aggregate of the latest two and four years of the time series. This is reported in the access and participation data dashboard.

[41] For TEF purposes, we report the proportions of the statistical uncertainty distribution which fall above, below or between a pair of 'guiding lines' which illustrate where the indicator value could be considered as materially above or below the benchmark value, compared with being broadly in line with the benchmark value. These guiding lines are positioned at +/- 2.5 percentage points difference between the indicator and benchmark values.

For further information about the distributions of statistical uncertainty, and the proportions of the statistical uncertainty distribution that fall above and below particular values, see Statistical uncertainty and visualisation of this.

For further information about the benchmark values, see Benchmarking.

Rounding and suppression

112. In this section, where we refer to rounding and suppression of indicators, we use the terms numerator and denominator. When talking about rounding and suppression of overall size and shape of provision data we use the term headcounts.

113. The data has been rounded as follows:

a. Denominators or headcounts have been rounded to the nearest 10.

b. Indicators, percentages, ratios, benchmarks, differences and their confidence intervals have been rounded to the nearest 0.1.

114. Any data point that is not reportable will be replaced with a symbol to indicate why, applied according to the following hierarchical order:

- [DQ]: Data has been suppressed due to a known issue with this provider's reported 2022-23 data.[42] For further information see https://www.officeforstudents.org.uk/data-and-analysis/student-outcome-and-experience-measures/data-quality/.
- [N/A]: Not applicable, where the data item is not applicable to that population or cannot be calculated.[43]
- [none]: where there are no students in the denominator (or two or fewer). Indicators in this category are displayed in the Excel workbooks with this suppression code, but they are not included in the Tableau dashboards at all.
- [low]: Low numbers of students, where there are more than two but fewer than 23 students in the denominator.
- [DP]: Data protection, where data has been suppressed for data protection reasons. The code [DP] has been used where further data protection has taken place for sensitive data. This includes scenarios as in paragraph 115.

[42] [DQ] will also be used in place of other suppression codes where indicators are suppressed for this reason.

[43] In a limited number of cases, the statistical tests associated with the difference between indicator and benchmark values that we have calculated are replaced with the symbol [N/A]. This occurs where the contribution to own benchmark is 100 per cent. This prohibits the calculation of the confidence intervals and the proportions of the statistical uncertainty distribution that fall above and below the provider's benchmark value.

- [RR]: Response rate, for the progression or student experience measures where the provider participated in the relevant survey (Graduate Outcomes survey or NSS respectively) but has not met response rate threshold required (50 per cent for the NSS, 30 per cent for the Graduate Outcomes survey).[44]

- [BK]: Benchmark suppressions, where the benchmarks are suppressed because at least 50 per cent of the provider's students have unknown information for one or more of the factors used for that benchmark calculation.[45]

- [DPL]: Data protection for low numerators, where data has been suppressed for data protection reasons. The code [DPL] has been used to indicate where the data has been suppressed due to a numerator or headcount that is less than or equal to two, meaning that the indicator will take on a value close to 0 per cent.[46]

- [DPH]: Data protection for high numerators, where data has been suppressed for data protection reasons. For the indicators data, the code [DPH] has been used to indicate where data has been suppressed due to a numerator that is greater than two but is within two of the denominator. For the overall shape and size of provision data, the code [DPH] has been used to indicate where data has been suppressed due to a headcount for a particular category of students being greater than two but within two of the total number of students who are taught or registered by the provider, meaning that the indicator will take on a value close to 100 per cent.[46]

115. When reporting split indicators which report on students who were or were not eligible for free school meals, the sensitivity of this information at an individual student level means that it is appropriate for us to take further steps for data protection reasons. If one of the data protection reasons we have already described causes us to suppress one of the free school meals split indicators reported for a given provider, we will also select one other free school meals split indicator calculated for that provider to be suppressed. In selecting another free school meals split for secondary suppression, we will normally select the indicator which refers to the smallest population, working across the different undergraduate levels of study and different views of a provider's student population to limit the impact of this suppression. We take the view that this is necessary to fully mitigate the risks of data disclosure.

116. In the reporting of gaps and ratios in the access and participation dashboard, where any of the constituent indicators have been suppressed (according to any of the reasons set out above), the gaps and ratios will also be suppressed.

[44] Where response rate suppression occurs, the corresponding denominator and response rate are still reported but all other elements included in the presentation of that indicator or split indicator are replaced with the [RR] symbol.

[45] Where benchmark suppression occurs ([BK] symbol), the benchmark value, the contribution to own benchmark, the difference and any statistical tests associated with the difference will be replaced with this symbol.

[46] Where data has been suppressed for data protection reasons ([DPL] or [DPH] symbols), the denominator, response rate (if applicable), benchmark value and the contribution to benchmark are not replaced with this symbol.

Statistical uncertainty and visualisation of this

117. As a producer of official statistics, we are committed to effectively communicating our statistics so that users can have confidence in their use and interpretation of them. This means we aim to use meaningful and effective ways to understand the potential extent of statistical uncertainty within the indicators and split indicators that inform our regulation of student outcomes and the TEF, as well as those that are reported through the access and participation data dashboard.

What is statistical uncertainty?

When calculating student outcome and experience measures, each indicator and split indicator that we calculate is a factual representation of the outcomes or experiences of students observed at a particular provider at a particular point in time. If one is interested only in the actual population of students present at a particular provider at a particular time, then it would be appropriate to rely solely on this value.

Within our regulatory uses of student outcome and experience indicators, we want to instead think about indicator values as representing the underlying performance of the provider in relation to a whole population of students who could have attended that provider, or may do so in the future. This whole population is known as a **superpopulation**.

It is not possible to say exactly what a provider's underlying performance looks like for the superpopulation, because students who could have attended the provider in question but did not do so, and students who may attend the provider in future, cannot be known to us.

The group of students which **did** attend are just one set of students from this superpopulation, and the value calculated from data about this group is used to infer what we would expect in the superpopulation.

However, the group of students which **did** attend is – in various respects – a random realisation of the whole population who **could** have attended. For example, if the observed population at the provider had included a few more 'morning people' and fewer 'night owls', would attendance at morning lectures have had a different influence over continuation or completion outcomes? If it happened to be raining on the day that students chose to complete a survey, how differently would student experiences be reported compared with the responses that would have been made if it happened to be sunny instead?

This randomness could give rise to a slight difference in the observed population which could give rise to slightly different indicator values being calculated, even though the underlying performance of the provider and their course delivery remained the same. This potential for random variation in the indicator values we calculate and interpret as the provider's performance, is known as **statistical uncertainty**.

Why is statistical uncertainty important?

Statistical uncertainty is unavoidable in the calculation of any statistic that is unable to identify and refer to its superpopulation: it cannot be rectified through adjustments to the underlying data or the calculations we are performing.

This means there will always be a question as to how exact any calculated indicator value is as an estimate for the superpopulation.

This question of exactness (or of statistical uncertainty) is important when indicator and split indicator values are being used to inform a judgement about the underlying performance of a provider in respect of the outcomes and experiences it delivers for students. This is because the judgement of underlying performance is intended to be a judgement about the superpopulation.

Statistical uncertainty, not measurement error

Statistical uncertainty should not be confused for measurement error (sometimes known as observational error).

Measurement error occurs when there are inaccuracies either in the underlying data on which we are performing our calculations (for example, a student is erroneously reported as studying full-time rather than part-time), or within the calculations that we are performing (for example, a formula that should include a 'greater than or equals to' condition mistakenly includes a 'strictly greater than' condition instead).

While neither example of measurement error can be entirely ruled out, we aim to identify and rectify any such errors through our sharing of the data and methods used with providers and other stakeholders. We are confident that the indicators we have calculated are an accurate factual representation of student outcomes and experiences as they have been reported to us through the student data returns that inform those indicators.

Communication of statistical uncertainty

118. As described in paragraph 111, when reporting student outcome and experience measures we show the value of each indicator and split indicator, and for some uses, the difference from the provider's benchmark or the gap between different split indicators. Shaded bars are used in our presentation of the student outcomes data dashboard, the TEF data dashboard and the access and participation data dashboard, to communicate the statistical uncertainty associated with each of those values.

119. Figure 1 and Figure 2 provided an illustration of the shaded bars we use in our presentation of the student outcomes and TEF data dashboards. Figure 3 and Figure 4 provided an illustration of the shaded bars we use in our presentation of the indicators shown in the access and participation data dashboard. The Summary of key features of our approach noted that these aim to represent the continuous spread (or distribution) of statistical uncertainty around the different values that we have calculated to understand a provider's performance.

120. The shading of the bars indicates the changing likelihood that underlying provider performance takes different values, with the darkest shading representing the range in which there is the greatest likelihood that true provider performance might lie. Much like the 'bell curve' of the normal distribution, as the shading lightens in both directions it represents a lower likelihood that true underlying performance falls at that point. Wider shaded bars mean we need to consider the potential for the provider's true performance falling within a wider range of values around the point estimate that has been observed.

121. The presentation of the shaded bars used throughout the presentation of student outcomes and experience measures is intentionally similar and each can be thought of as representing a series of discrete confidence intervals around the point estimate we have observed, where each confidence interval in the series corresponds to a different confidence (or significance) level.

For further technical detail about the statistical methods and calculations underpinning our presentation of statistical uncertainty, and the role of confidence intervals in doing this, see Annex C: Further information about how we calculate and present statistical uncertainty.

Interpreting the shaded bars used in our presentation of indicators and split indicators in the student outcomes and TEF data dashboards

122. To support consistency and transparency of interpretation about the shaded bars and the statistical uncertainty they represent, Figure 1 and Figure 2 showed that we also include summary figures in a table to the right of the shaded bars.

123. Our regulation of student outcomes and the TEF seek to interpret the shaded bars with reference to their statistical confidence that a provider's underlying performance provider is above or below a given numerical value (for example, one of the minimum numerical thresholds used in respect of condition B3 or a difference from benchmark of zero).[47]

124. The summary figures reported alongside the shaded bars describe the proportion of the distribution of statistical uncertainty, represented by the shaded bar, that falls above or below certain values (the minimum numerical thresholds for condition B3, or the provider's own benchmark value). It is intended that the summary figures are used together with the shaded bars to aid interpretation of users' statistical confidence related to student outcome and experience measures. They are highlighted where they show that at least 75 per cent of the distribution falls above or below those values, but users can use the shaded bars to make other interpretations of a provider's performance.

125. In designing the shaded bars, we have sought to avoid selecting a single confidence interval significance level. To do so would create a 'cliff edge' at a single significance level pre-determined by the OfS for our specific use, which would facilitate a binary interpretation of performance as definitively above or below a given threshold by most users. Instead, we illustrate the distribution of statistical uncertainty up to a maximum of a 99.7 per cent confidence interval and our own assessments of a provider's performance will take into account the statistical confidence we have in relation to the indicator, and difference between indicator and benchmark, values. We also anticipate that other users of the data will be empowered to better understand the confidence in which they can hold their own judgements of student outcomes and experience indicators.

[47] For further information about the indicative categories of statistical confidence, see Annex B of Regulatory Advice 20: Regulating student outcomes, at www.officeforstudents.org.uk/publications/regulatory-advice-20-regulating-student-outcomes/. Our analysis of responses to the TEF consultation (www.officeforstudents.org.uk/publications/student-outcomes-and-teaching-excellence-consultations/the-tef/) confirmed that the same categories will be relevant to TEF assessment.

For further information about the summary figures we report alongside the shaded bars, see Annex C: Further information about how we calculate and present statistical uncertainty .

Multiple comparisons adjustments

126. When multiple statistics are calculated on a given topic, it is often expected that users will wish to make comparisons between those statistics. In statistics, the issue of 'multiple comparisons' arises when a user considers multiple statistical tests at once. With more tests there is more opportunity for unlikely events to occur simply due to the influence of random chance. This means that when looking at multiple indicators or split indicators at once, for any that appear to be significantly above or below a numerical threshold, or benchmark, there is a greater chance of finding a result that appears significant but has occurred through random chance alone.

127. To account for this, when conducting multiple tests, it is sometimes appropriate to make formulaic adjustments to what we consider to be unlikely to have occurred by random chance alone. We do not consider that a formulaic adjustment for multiple comparisons is appropriate for the student outcome and experience measures we construct. However, we do suggest that users who wish to make multiple comparisons in order to identify outlying data points (or indicators that are significantly above or below a benchmark or numerical threshold) consider adjusting to a higher level of confidence when making their judgements. This is because of the higher risk of false discovery when using lower levels of statistical confidence. In this context, users may wish to be more conservative in their interpretation of statistical uncertainty the more comparisons they are making. Users can mitigate the risk of making a false discovery by adjusting to use higher levels of statistical confidence.

128. Similar adjustments can be made when interpreting the statistical uncertainty associated with gaps between split indicators in the access and participation data dashboard. This means that users of the access and participation data dashboards can consider the statistical uncertainty presented by the shaded bars and the potential impact of multiple comparisons. Users may wish to do this when interpreting gaps across multiple years of data, or levels and modes of study to identify the challenges a provider may be facing with respect to equality of opportunity. Users who choose to consider the impact of making multiple comparisons adjustments should balance the risk of acting on a 'false discovery' against the risk that statistical evidence regarding inequality of opportunity is overlooked.

For further information aimed at supporting users to understand when multiple comparison adjustments may be appropriate, see Annex D: Further information about making multiple comparison adjustments.

Benchmarking

129. We use benchmarking to inform our regulation of student outcomes and the TEF, to help interpret a provider's actual performance relative to that in the sector overall once we have taken into account the mix of students at the provider or the provision being offered. Each indicator that we calculate represents the outcomes that we have observed for the students at

a particular provider at a particular point in time. The calculation of a benchmark gives us a counterfactual for the observed outcomes, which we intend can be used in two ways:

- to understand how well a provider has performed compared with performance for similar types of students on similar types of courses in the higher education sector as a whole
- to assess similarities between individual providers.

130. In making these comparisons, we take account of factors which describe the profile of students and provision delivered by higher education providers and which are correlated with the outcomes we are measuring. The benchmarking methodology we use involves consideration of unique combinations of the student and course characteristics that we have selected to act as benchmarking factors: we refer to these unique combinations as benchmarking groups.

131. The methodology allows us to ask the question: 'What would the observed student outcome have been at this provider if its distribution of students across benchmarking factor groups had been what it was, but its outcomes across those same benchmarking groups were replaced by the sector-overall rates?'.

132. When there are known differences between the outcomes and experiences of some groups of students or providers, observed average values for the whole of the higher education sector are not necessarily helpful when forming this expectation. Instead, we calculate the benchmark as a weighted sector average reflecting the number of students in that group at the provider. As such, benchmarks give information about the values that the sector overall might have achieved for the indicator if the characteristics included in the benchmarking factors are the only ones that are important. Where differences exist between an indicator and its corresponding benchmark, these may be due to the provider's performance, or they may be due to some other characteristic which is not included in the weighting.

133. Benchmark values may change between years and between releases within a year as overall sector data changes.

General approach to benchmarking

134. To create benchmarks for a given measure, we calculate the observed rates for the higher education sector as a whole, for each benchmarking group. The benchmark for each provider is then calculated by taking a weighted average of the overall sector outcomes for each benchmarking group, taking account of the particular mix of students across those groups at the provider in question.

135. The benchmarking methodology we use means that a provider is not being compared with a pre-set group of providers, but rather the outcomes for a provider's students are compared with the outcomes of similar students across the entirety of the higher education sector. For the purpose of calculating these benchmarks for OfS registered providers, the higher education sector within which we are making comparisons of the outcomes for similar students is made up of all English higher education providers registered with the OfS at the time that we produce the indicators.

For a worked example of how we calculate benchmarks, see Annex F: Worked example of benchmarking calculations.

For details of the formulae used in the calculation of benchmarks, see Annex H: Technical detail about benchmarking calculations.

The benchmarking factors we use

136. The basis on which we select, define, and apply the factors used in benchmarking student outcome and experience indicators is key to the integrity and robustness of the benchmark values calculated and assessed.

137. The benchmarking factors we use is the result of us following the set of Annex E: Principles for the selection and application of benchmarking factors.[48]

138. The benchmarking factors used for each measure were summarised in Table 3. In Tables 17 to 19 we provide further information about the benchmarking factors and groupings we use for each student outcome measure.

For further information about the entry qualification and subject of study factors to which Tables 14 to 16 refer, see Annex G: Definitions of entry qualifications and subject areas of study groupings used in benchmarking.

Table 17: Benchmarking factors for continuation measures

Benchmarking factor	Continuation: full-time	Continuation: part-time	Continuation: apprenticeship
Level of study (First degree, other undergraduate, undergraduate with postgraduate components)	✓ (Other undergraduate separated into those at Level 4 and those at Level 5+)	✓	✓
Subject of study (CAH level 1 groups)	✓	✓	✓
Entry qualifications	✓ (11 groupings)	✓ (5 groupings)	✓ (5 groupings)
Expected course length (Expected course length of less than a year, or otherwise)	×	✓	×

[48] We have published our review of the selection and grouping of benchmarking factors at www.officeforstudents.org.uk/publications/review-of-selection-and-grouping-of-benchmarking-factors/.

Benchmarking factor	Continuation: full-time	Continuation: part-time	Continuation: apprenticeship
ABCS quintile (Continuation ABCS Quintiles 1 to 5 (including unmatched) for the relevant mode of study, non-UK domiciled)[49]	✓	✓	✓
Total distinct benchmarking groups	5,544	3,780	1,890

Table 18: Benchmarking factors for completion measures

Benchmarking factor	Completion: full-time	Completion: part-time	Completion: apprenticeship
Level of study (First degree, other undergraduate, undergraduate with postgraduate components)	✓ (Other undergraduate separated into that at Level 4 and that at Level 5+)	✓	✓
Subject of study (CAH level 1 groups)	✓	✓	✓
Entry qualifications	✓ (11 groupings)	✓ (5 groupings)	✓ (5 groupings)
Expected course length	✓ (Expected course length of less than two years, two years, or at least three years)	✓ (Expected course length of less than a year, or otherwise)	×
ABCS quintile (Completion ABCS Quintiles 1 to 5 (including unmatched) for the relevant mode of	✓	✓	✓

[49] The ABCS method constructs separate quintiles relevant to each student outcome measure, where necessary differentiating by mode of study. The ABCS analysis for continuation outcomes considers full- and part-time students separately at www.officeforstudents.org.uk/data-and-analysis/associations-between-characteristics-of-students/. Full-time continuation ABCS quintiles are used in respect of apprenticeship students.

Benchmarking factor	Completion: full-time	Completion: part-time	Completion: apprenticeship
study, non-UK domiciled)50			
Total distinct benchmarking groups	16,632	3,780	1,890

Table 19: Benchmarking factors for progression measures

Benchmarking factor	Progression: full-time	Progression: part-time	Progression: apprenticeship
Year qualification obtained	✓	✓	✓
Level of study (First degree, other undergraduate, undergraduate with postgraduate components)	✓	✓	✓
Subject of study	✓ (CAH level 2 groups51 for first degree level of study; Broadly defined subject groups otherwise)	✓ (Broadly defined subject groups)	✓ (Broadly defined subject groups)
Entry qualifications	✓ (5 groupings for other undergraduate level of study; 11 groupings otherwise)	✓ (3 groupings)	✓ (3 groupings)
ABCS quintile (Progression ABCS quintiles 1 to 5 (including unmatched) for the	✓	✓	✓

[50] The ABCS method constructs separate quintiles relevant to each student outcome measure, where necessary differentiating by mode of study. The ABCS analysis for completion outcomes considers full- and part-time students separately at www.officeforstudents.org.uk/data-and-analysis/associations-between-characteristics-of-students/. Full-time completion ABCS quintiles are used in respect of apprenticeship students.

[51] For benchmarking purposes, the CAH level 2 group for Celtic studies (CAH19-02) has been combined into the Languages and area studies group (CAH19-04).

Benchmarking factor	Progression: full-time	Progression: part-time	Progression: apprenticeship
relevant mode of study)[52]			
Geography of employment quintile (Quintile 1, Quintiles 2 and 3, Quintiles 4, 5 and unknown)	✓	✓	✓
Total distinct benchmarking groups	Varies by level of study: Other undergraduate: 3,000 First degree: 22,440 Undergraduate with postgraduate components: 6,600	5,400	5,400

Table 20: Benchmarking factors for student experience measures

Benchmarking factor	Student experience: full-time	Student experience: part-time	Student experience: apprenticeship
Year of survey	✓	✓	✓
Level of study (First degree, other undergraduate, undergraduate with postgraduate components)	✓	✓	✓
Subject of study	✓ (CAH level 2 groups)[53]	✓ (Broadly defined subject groups)	✓ (Broadly defined subject groups)
Age on entry (Under 21 or unknown, 21 to 30, 31 and over)	✓	✓	✓
Disability (Disability	✓	✓	✓

[52] The ABCS method constructs separate quintiles relevant to each student outcome measure, where necessary differentiating by mode of study. The ABCS analysis for progression outcomes considers full- and part-time students separately at www.officeforstudents.org.uk/data-and-analysis/associations-between-characteristics-of-students/. Full-time progression ABCS quintiles are used in respect of apprenticeship students.

[53] For benchmarking purposes, the CAH level 2 group for Celtic studies (CAH19-02) has been combined into the Languages and area studies group (CAH19-04).

Benchmarking factor	Student experience: full-time	Student experience: part-time	Student experience: apprenticeship
reported, no disability reported)			
Ethnicity (Asian, Black, Mixed, Other, Unknown or White, non-UK domiciled)	✓	✓	✓
Sex (Female or other, Male)	✓	×	×
Total distinct benchmarking groups[54]	29,376	4,320	4,320

Risks of self-benchmarking

139. When constructing the benchmark for an individual provider, the students at that provider contribute to the sector averages we calculate. We recognise that where the characteristics of students at the provider in question do not frequently occur among student populations in the wider sector, these sector averages may be heavily influenced by that provider. This is referred to as the risk of 'self-benchmarking'. In such a scenario, the provider's own students would be making a substantial contribution to the calculation of its benchmark, making the calculation less robust and the resulting benchmark value less meaningful. The benchmark value will become more similar to the indicator value as the provider's contribution increases. This is because there is less other sector data that can provide the information necessary to make the benchmark a reliable estimate of the values that might have been expected for the provider.

140. The risk of self-benchmarking becomes more acute when benchmarking groups are populated by only one or two students: the sector averages calculated for those groups will tend to a small range of values. If the sector average is calculated in reference to a single student, it can only result in an 'average' of either 0 per cent or 100 per cent. If it refers to only two students, the average can only be 0 per cent, 50 per cent or 100 per cent. Sector averages that include large numbers of 0 per cent and 100 per cent values can lead to an ineffectual weighting which will skew the resulting benchmark and increase the standard errors of the calculated difference between indicator and benchmark values.[55]

141. Our selection of benchmarking factors has sought to minimise the occasions on which we might encounter self-benchmarking, by selecting and grouping factors in such a way as to ensure as far as possible that reasonable numbers of students from multiple providers are contributing to each sector average that we calculate. However, we are aware that the diversity of the higher education sector means that we cannot mitigate this risk entirely and our

[54] The total number of benchmarking groups for student experience measures reflects the four years of NSS responses that will be used in the construction of student experience indicators in steady state.

[55] The standard errors of a statistic represent the amount by which one would expect that statistic to change, based solely on random sampling.

proposed benchmarking factors tolerate a risk of self-benchmarking on a small scale. To facilitate an understanding of where this situation may occur, and the resulting benchmark value may be of more limited use, we report information about the provider's own contribution to that benchmark within the datasets we construct.

For details about our calculation of a provider's contribution to its own benchmark, see Annex H: Technical detail about benchmarking calculations.

Benchmarking split indicators

142. The approach to benchmarking split indicators follows the general approach described in paragraphs 134 to 138. However, instead of creating a benchmark for the provider using data from the sector overall, we repeat that process per split indicator and subset both the provider and the sector to the population of the split indicator in question.

143. For example, to benchmark the 'male' split indicator we subset the provider and the sector to only male students, so that we can compare the student outcomes for male students at the provider to a benchmark created from male students across the sector. We then separately benchmark the 'female' split indicator by sub-setting the provider and the sector to only female students.

144. This approach can lead to cases where a provider's difference between the indicator and benchmark values could show performance below benchmark for indicator, but above benchmark for every split indicator.

Benchmarking indicators and split indicators for the 'taught or registered (TorR)' student population

145. While the general approach to benchmarking can be applied to each of the registered and taught populations that our reporting of student outcome and experience measures use as Views of a provider's student population, for the view of students who are who are either taught or registered at the provider in question (or both) we need to vary our approach.

146. The benchmarking methodology assumes that students per provider per unique combination of benchmarking factors are independent from any other combination. To construct benchmarks for the taught or registered (TorR) student population we therefore need to account for the potential that students are associated with more than one provider and contribute to more than one provider's indicators and split indicators. We make some minor adjustments to the benchmarking methodology to accommodate students contributing to the benchmarks of multiple providers.

For details about the adjustments we make to the benchmarking methodology for the taught or registered view of a provider's student population, see Annex H: Technical detail about benchmarking calculations.

Benchmarking suppression

147. Some of the benchmarking factors are known to include attributes identifying the characteristic or information as unknown, not required or not applicable. This occurs where student data has not been returned for us to be able to classify students appropriately, for example, because this information was not shared with the provider so it has been unable to include it in its DDB or ILR student data submissions.

148. A large number of students being reported with unknown attributes reported for a benchmarking factor can impact on the reliability of the benchmarking calculations. Our benchmarking method is effective in taking account of the mix of a provider's students and provision when the grouping of attributes within benchmarking factors forms coherent groups which share a consistency of student backgrounds, outcomes, or behaviours with respect to the indicator to which they refer. By virtue of the attribute being reported as unknown, we cannot know the extent to which students reported in this way do form coherent, homogeneous groups, nor the extent to which weighting the sector average for the size of this group becomes akin to comparing apples and pears. We therefore take the view that a large number of students being reported with unknown attributes dilutes the effect of that characteristic on the efficiency of the calculated benchmark.

149. Consequently, as shown in paragraph 114, benchmark values are not reported when a provider's student data reports at least 50 per cent of the students with unknown information for one or more of the factors used for that benchmark calculation. For example, where entry qualifications are proposed as a benchmarking factor, the benchmark value (and the calculated difference between the indicator and the benchmark) is suppressed if at least 50 per cent of the provider's students have unknown entry qualifications.

Data about the size and shape of higher education provision

This section is aimed at readers who are seeking to better understand the data available to support understanding of the student outcome and experience measures.

This section is relevant to the **size and shape of provision data dashboard**.

150. The student outcome and experience measures are supported by a set of data resources which shows, in broad terms, the size and shape of the provider's student population.[56] A number of charts and data tables are provided as part of these data resources, and these are described in more detail in paragraphs 151 to 159:

- student numbers
- partnership arrangements
- size and shape of provision by student and course characteristics.

Student numbers

151. This table provides data on full-time, part-time and apprenticeship student numbers broken down by mode and level of study, plus information on the numbers of those studying for credit or modules, or unspecified qualification aims. The student population is based on the taught or registered (TorR) view of the provider and considers students who are mainly studying in the UK and are actively studying on a qualification aim or aiming for credit or modular provision. This excludes students who are dormant, taking a sabbatical or writing up. Headcounts of entrants, qualifiers and total student numbers for the most recent four years (2019-20 to 2022-23) are provided. Apprenticeship student numbers are only broken down into broad levels of study (undergraduate and postgraduate).

152. The second section of the table in the data output provides data on offshore transnational education (Offshore TNE), which is sourced from the DDB's Aggregate Offshore Record (AOR). This includes students studying wholly outside the UK who are registered at the provider or who are studying for an award of the provider. Only total student numbers are shown for this population and only a breakdown by level of study is provided.

153. The final section of the table in the data output provides data on students who are mainly studying abroad. This student population is based on the taught or registered (TorR) view of the provider. Headcounts of entrants, qualifiers and total student numbers for the most recent four years (2019-20 to 2022-23) are provided and are broken down by level of study.

[56] See www.officeforstudents.org.uk/data-and-analysis/size-and-shape-of-provision-data-dashboard/.

Partnership arrangements

154. This table in the data output provides a summary of headcount student numbers shown for different types of partnership. Numbers are shown separately for each mode and broad level of study. Headcounts of entrants, qualifiers and total student numbers for the most recent four year (2019-20 to 2022-23) are provided.

155. The types of partnership arrangements presented are:

- all students registered or taught by this provider (TorR population)
- all students registered and taught by this provider
- only students registered by this provider (subcontracted out from this provider)
- only students taught by this provider (subcontracted in to this provider)
- only students validated by this provider.

Size and shape of provision

156. This table in the data output provides a summary of headcount student numbers shown separately for undergraduate and postgraduate population for each mode of study and then further broken down by student or course characteristics.

157. The student population is based on the taught or registered view of the provider and considers students who are mainly studying in the UK and are actively studying on a qualification aim. This excludes students who are dormant, sabbatical or writing up. Headcounts of entrants, qualifiers, and total student numbers for the most recent four year (2019-20 to 2022-23) are provided.

158. Some of the student or course characteristics are only available for particular cohorts within the student population. The characteristics considered and any restrictions on coverage are:

- Study characteristics
- course length – classification of student based on the number of years that the qualification they are studying for is expected to last
- course type – full-time, first-degree students with integrated foundation year
- higher education courses recognised and not recognised for OfS funding purposes
- subject of study – classification using Common Aggregation Hierarchy Level 2)
- year of study type – students on courses that have the option of a sandwich placement
- Student characteristics
- age on entry

- disability type
- ethnicity – only available for UK-domiciled students[57]
- sex
- sexual orientation – only available for providers that submit data to the DDB's student record[57]
- ABCS quintiles, reported separately for each of continuation, completion and progression – only available for UK-domiciled undergraduates who are registered with English providers[57]. ABCS continuation and completion quintiles available for entrants only, while ABCS progression quintiles are available for qualifiers only
- deprivation quintile (IMD) – only available for UK-domiciled students[57]
- domicile
- eligibility for free school meals – only available for undergraduate students aged under 21 years on entry who were found in the NPD KS4 data attending a state-funded mainstream school in England in 2009-10 or later
- entry qualifications
- geography of employment quintile – only available for qualifying undergraduate student cohorts who responded to the Graduate Outcomes survey[57]
- socio-economic background - only available for providers that submit data to the DDB's student record and for UK-domiciled undergraduate students[57]
- study location
- tracking underrepresentation by MSOA (TUNDRA MSOA) – only available for England domiciled undergraduate students.[57]

Presentation of the data on the overall size and shape of provision

159. Paragraphs 112 to 114 explain the rounding and suppression that has been applied in the presentation of data on the overall size and shape of provision.

[57] Students who do not meet this population restriction are reported in the 'unknown or not applicable' category.

Data about the reporting of interim study activities

This section is aimed at readers who are seeking to better understand the data available to support understanding of the student outcome and experience measures.

This section is relevant to the **access and participation data dashboard**, **student outcomes data dashboard** and the **TEF data dashboard**.

160. The definition of the progression measure does **not** count as a positive outcome those students who were unemployed at the census date, or not otherwise engaged in activities that count as a positive outcome, but reported in their response to the Graduate Outcomes survey that they had undertaken interim study since completing their higher education course.

161. We recognise that information about the number of students who count negatively towards the progression indicator but reported interim study may provide valuable context for students who have followed certain courses and could support users in understanding the potential influence of these interim activities on a provider's performance in relation to student outcomes. We therefore publish additional data (separately from the progression measure) to provide this information.

162. The information we provide about interim study activities focusses on the students who counted negatively towards the progression indicator. These students are defined through the Indicator definitions: Progression measures section and by the variable IPEMPIND within the 'Core algorithms' document.[58]

163. We report two figures in relation to interim study activities:

a. The number and proportion of students who counted negatively towards the progression indicator but reported in their response to the Graduate Outcomes survey that they had undertaken **any** interim study.[59]

b. The number and proportion of students who counted negatively towards the progression indicator but reported in their response to the Graduate Outcomes survey that they had undertaken **significant** interim study.[60]

164. The figures in paragraph 163 will both be reported for each breakdown of the student population represented by the indicators and split indicators. It is available within the student outcomes, TEF and access and participation data dashboards alongside the indicators and

[58] See www.officeforstudents.org.uk/data-and-analysis/institutional-performance-measures/technical-documentation/.

[59] We identify whether any of these students reported any interim study through their Graduate Outcomes survey response using the FURSTU variable. See www.hesa.ac.uk/collection/c19072/a/furstu or, equivalently, IPGOINTSTUDY = FT, PT or OTH within the OfS 'Core algorithms' document.

[60] We will identify whether any of these students reported significant interim study using definitions consistent with the HESA derived field XINTSTUDY as it was defined for the 2019-20 Graduate Outcomes survey: see www.hesa.ac.uk/collection/c19072/derived/xintstudy. The definition used by the OfS is provided by the variable IPGOSIGINTSTUDY within the 'Core algorithms' document.

split indicators to which it corresponds.[61] However, this information should be considered separately from the values of the progression indicators and split indicators because these interim study activities do **not** count as a positive progression outcome in our regulatory approaches, in particular for the purposes of measuring a provider's performance with reference to the minimum numerical thresholds for condition B3.

165. It should be noted that the figures described in paragraph 163.b are a subset of those described in paragraph 163.a, meaning that the two figures will need to be considered separately and users should not add the two together.

Presentation of the data about interim study activities

166. Paragraphs 112 to 114 explain the rounding and suppression that has been applied in the presentation of data about interim study activities. Note that suppression for data protection reasons (indicated by the codes [DPL] and [DPH]) is applied to interim study rates separately to the associated progression indicators.

[61] See the user guide information that we have published alongside each data dashboard for information about how to view this information.

Annex A: List of abbreviations and key terminology

Abbreviation	Meaning
ABCS	Associations between characteristics of students
DDB	Designated data body
ESFA	Education and Skills Funding Agency
FPE	Full-person equivalent
FSM	Free school meals
GO	Graduate Outcomes (survey)
HTQ	Higher Technical Qualification
ILR	Individualised Learner Record
IMD	Index of multiple deprivation
KS4	Key stage 4
LSOA	Lower layer super output area
MSOA	Middle layer super output area
NPD	National Pupil Database
NSS	National Student Survey
OfS	Office for Students
ONS	Office for National Statistics
PGCE	Postgraduate or Professional Graduate Certificate in Education
POLAR	Participation of Local Areas (measure)
SKE	Subject knowledge enhancement
SOA	Super output area
SOC	Standard Occupational Classification
TEF	Teaching Excellence Framework
TorR	Taught or registered (view of provider)
TTWA	Travel to work area
TUNDRA	Tracking underrepresentation (a measure by area)

Annex B: Further information about the definition of split indicators

This annex is aimed at readers who are seeking to better understand how students contribute to our student outcome and experience measures, and how the measures are presented.

This annex provides information about the student and course characteristics used in the definition of split indicators for the student outcome and experience measures reported through the **student outcomes data dashboard,** the **TEF data dashboard** and the **access and participation data dashboard**.

1. The student and course characteristics used in the construction of split indicators for the purposes of regulating student outcomes and the TEF are detailed in Table B1.

Table B1: Types of split indicators constructed for use in regulation of student outcomes and the TEF, and the split indicators considered

Split indicator type	Split indicators considered
Time series	Year 1 (earliest) Year 2 Year 3 Year 4 (most recent) The academic year that these relate to depends on the measure, as described in the definitions for each of the measures.
Level of study	Other undergraduate First degree Undergraduate with postgraduate components
Teaching arrangements	Taught by this provider Only registered by this provider (subcontracted out from this provider) And, for the student outcomes data dashboard only: • Registered and taught by this provider • Only taught by this provider (subcontracted in to this provider) • This provider is the degree awarding body only Some of these attributes are only available on some of the provider views
Course type: other undergraduate	Other undergraduate Level 4 Other undergraduate Level 5+ Higher technical qualifications (HTQs) (HTQs will be added when the data becomes available.)

Split indicator type	Split indicators considered	
Course type: foundation year	First degree with integrated foundation year	
Subject of study	Business and management	Business and management
	Design, and creative and performing arts	Creative arts and design Performing arts
	Education and teaching	Education and teaching
	Engineering, technology and computing	Computing Engineering Materials and technology
	Humanities and languages	Combined and general studies English studies History and archaeology Languages and area studies Media, journalism and communications Philosophy and religious studies
	Law and social sciences	Economics Health and social care Law Politics Sociology, social policy and anthropology
	Medicine, dentistry and veterinary sciences	Medicine and dentistry Veterinary sciences Nursing and midwifery
	Natural and built environment	Agriculture, food and related studies Architecture, building and planning Geography, earth and environmental studies
	Natural and mathematical sciences	Biosciences Chemistry General, applied and forensic sciences Mathematical sciences Physics and astronomy Sport and exercise sciences
	Nursing, allied health and psychology	Allied health subjects Medical sciences Pharmacology, toxicology and pharmacy

Split indicator type	Split indicators considered
	Psychology
Age on entry (on 31 August in the student's year of entry to higher education programme)	For undergraduate courses: Under 21 years 21 to 30 years 31 years and over For postgraduate courses: Under 25 years 25 to 30 years 31 years and over
Disability[62]	Disability reported No disability reported
Ethnicity (for UK-domiciled students only)	Asian Black Mixed Other White
Sex	Female Male
ABCS quintile[63] (for the continuation, completion and progression measures only, and only for UK-domiciled students on undergraduate courses)	Quintile 1 Quintile 2 or quintile 3 Quintile 4 or quintile 5 (Students in quintile 1 have the lowest modelled rates and those in quintile 5 have the highest modelled rates. For example, for the continuation indicator, students in quintile 1 have characteristics that tend to have the lowest rates of continuation across the sector, whilst quintile 5 tend to have the highest rates of continuation across the sector.)

[62] Disability information included has been recorded on the basis of the student's own self-assessment. Changes in the number of students in this category may occur as a result of changes in data reporting.

[63] ABCS is a set of analyses which aims to improve our understanding of the outcome of different groups of young people. It uses statistical modelling to calculate modelled access, continuation, completion and progression rates. These rates are then used to separate groups of students into quintiles defined for each stage of the student lifecycle. Further detail can be found at: www.officeforstudents.org.uk/publications/update-to-associations-between-characteristics-of-students/. The ABCS quintiles use data from the ABCS analyses published in Autumn 2022 but applied to more recent student cohorts.

Split indicator type	Split indicators considered
Deprivation quintile (IMD)[64]	Quintile 1 or quintile 2 Quintile 3, quintile 4 or quintile 5 (Quintile 1 areas have the highest level of deprivation and those in quintile 5 have the lowest.)
Domicile	UK Non-UK
Eligibility for free school meals (FSM)[65] (for undergraduate students aged under 21 years on entry who were found in the NPD KS4 data attending a state-funded mainstream school in England in 2009-10 or later)	Eligible Not eligible
Geography of employment quintile[66] (for the progression measure only)	Quintile 1 Quintile 2 or quintile 3 Quintile 4 or quintile 5 (Graduates in quintile 1 are in an area with the lowest rates of positive progression outcomes and those in quintile 5 are in an area with the highest rates of positive progression outcomes.)

2. The student and course characteristics used in the construction of split indicators for the purposes of the access and participation data dashboard are detailed in Table B2.

[64] For students domiciled in England at registering providers in England, this will be based on the English IMD 2019 quintile. For students domiciled in Wales at registering providers in Wales, this will be based on the Welsh IMD 2019 quintile. For students domiciled in Scotland at registering providers in Scotland, this will be based on the Scottish IMD 2020 quintile. For students domiciled in Northern Ireland at registering providers in Northern Ireland, this will be based on the Northern Ireland Multiple Deprivation Measure 2017 quintile.

[65] The population that can be matched to NPD KS4 data is often smaller when based on more historic entrant cohorts (such as the completion indicator) or qualifier cohorts due to the length of time required to track students from when students were in school and the earliest NPD KS4 data available. Care should be taken when comparing free school meals split indicators across a time series since the population sizes can be quite different.

[66] Geography of employment quintiles are assigned based on the graduate's travel to work area (TTWA) which is derived from information reported in a graduate's response to the Graduate Outcomes survey. Further information on the methodology can be found in the OfS report 'A geography of employment and earnings', www.officeforstudents.org.uk/publications/a-geography-of-employment-and-earnings/.

Table B2: Types of split indicators constructed for use in the access and participation data dashboard, and the split indicators considered

Split indicator type	Split indicators considered
ABCS quintile[67] (for the access, continuation, completion and progression measures only)	Individual quintiles 1, 2, 3, 4 and 5 Aggregation of quintiles 2 and 3 Aggregation of quintiles 4 and 5 (Students in quintile 1 have the lowest modelled rates and those in quintile 5 have the highest modelled rates. For example, for the continuation indicator, students in quintile 1 have characteristics that tend to have the lowest rates of continuation across the sector, while quintile 5 tend to have the highest rates of continuation across the sector.)
Age on entry (on 31 August in the student's year of entry to higher education programme)	Young (under 21 years) Mature (21 years and over) 21 to 25 years 26 to 30 years 31 to 40 years 41 to 50 years 51 years and over
Disability[68]	Disability reported No disability reported
Disability type[69]	Cognitive or learning difficulties Mental health conditions No disability reported or unknown disability type Multiple or other impairments Sensory, medical or physical impairments Social or communication impairments
Ethnicity (for UK-domiciled students only)	Asian Black Mixed Other

[67] ABCS is a set of analyses which aims to improve our understanding of the outcome of different groups of young people. It uses statistical modelling to calculate modelled access, continuation, completion and progression rates. These rates are then used to separate groups of students into quintiles defined for each stage of the student lifecycle. Further detail can be found at: www.officeforstudents.org.uk/publications/update-to-associations-between-characteristics-of-students/. The ABCS quintiles use data from the autumn 2022 ABCS analyses.

[68] Disability information included has been recorded on the basis of the student's own self-assessment. Changes in the number of students in this category may occur as a result of changes in data reporting.

[69] See footnote 68.

Split indicator type	Split indicators considered
	White Aggregation of Asian, black, mixed and other (ABMO)[70] Aggregation of Asian, mixed, other and white Aggregation of black, mixed, other and white Aggregation of Asian, black, other and white Aggregation of Asian, black, mixed and white
Sex	Female Male
English deprivation quintile (IMD) (2015) Based on English-domiciled students	Individual quintiles 1, 2, 3, 4 and 5 Aggregation of quintiles 1 and 2 Aggregation of quintiles 3, 4 and 5 Aggregation of quintiles 2, 3, 4 and 5 Aggregation of quintiles 1, 3, 4 and 5 Aggregation of quintiles 1, 2, 4 and 5 Aggregation of quintiles 1, 2, 3 and 5 Aggregation of quintiles 1, 2, 3 and 4 (Quintile 1 areas have the highest level of deprivation and those in quintile 5 have the lowest.)
English deprivation quintile (IMD) (2019) Based on English-domiciled students	Individual quintiles 1, 2, 3, 4 and 5 Aggregation of quintiles 1 and 2 Aggregation of quintiles 3, 4 and 5 Aggregation of quintiles 2, 3, 4 and 5 Aggregation of quintiles 1, 3, 4 and 5 Aggregation of quintiles 1, 2, 4 and 5 Aggregation of quintiles 1, 2, 3 and 5 Aggregation of quintiles 1, 2, 3 and 4 (Quintile 1 areas have the highest level of deprivation and those in quintile 5 have the lowest.)
Eligibility for free school meals (FSM)[71] (for undergraduate students aged under 21 years on entry who were found in the NPD KS4 data attending	Eligible Not eligible

[70] Also referred to as 'black, Asian and minority ethnic'.

[71] The population that can be matched to NPD KS4 data is often smaller when based on more historic entrant cohorts (such as the completion indicator) or qualifier cohorts due to the length of time required to track students from when students were in school and the earliest NPD KS4 data available. Care should be taken when comparing free school meals split indicators across a time series since the population sizes can be quite different.

Split indicator type	Split indicators considered
a stage-funded mainstream school in England in 2009-10 or later)	
Participation of Local Areas classification (POLAR4) Based on young students (aged under 21 in year of entry to higher education programme)	Individual quintiles 1, 2, 3, 4 and 5 Aggregation of quintiles 1 and 2 Aggregation of quintiles 3, 4 and 5 Aggregation of quintiles 2, 3, 4 and 5 Aggregation of quintiles 1, 3, 4 and 5 Aggregation of quintiles 1, 2, 4 and 5 Aggregation of quintiles 1, 2, 3 and 5 Aggregation of quintiles 1, 2, 3 and 4 (Quintile 1 areas have the lowest rate of participation and quintile 5 have the highest.)
Tracking underrepresentation by area (TUNDRA) Based on young students (aged under 21 in year of entry to higher education programme) who attended state-funded mainstream schools in England	Individual quintiles 1, 2, 3, 4 and 5 Aggregation of quintiles 1 and 2 Aggregation of quintiles 3, 4 and 5 Aggregation of quintiles 2, 3, 4 and 5 Aggregation of quintiles 1, 3, 4 and 5 Aggregation of quintiles 1, 2, 4 and 5 Aggregation of quintiles 1, 2, 3 and 5 Aggregation of quintiles 1, 2, 3 and 4 (Quintile 1 areas have the lowest rate of participation and quintile 5 have the highest.)
Interaction of ethnicity and English Index of Multiple Deprivation (2019 version only) Based on English-domiciled students	ABMO and IMD quintile 1 or 2 ABMO and IMD quintile 3, 4 or 5 White and IMD quintile 1 or 2 White and IMD quintile 3, 4 or 5
Interaction of sex and English Index of Multiple Deprivation (2019 version only) Based on English-domiciled students	Female and IMD quintile 1 or 2 Female and IMD quintile 3, 4 or 5 Male and IMD quintile 1 or 2 Male and IMD quintile 3, 4 or 5
Interaction of ethnicity and POLAR4 classification Based on young students (aged under 21 in year of entry to higher education programme)	ABMO and POLAR4 quintile 1 or 2 ABMO and POLAR4 quintile 3, 4 or 5 White and POLAR4 quintile 1 or 2 White and POLAR4 quintile 3, 4 or 5
Interaction of sex and POLAR4 classification	Female and POLAR4 quintile 1 or 2 Female and POLAR4 quintile 3, 4 or 5 Male and POLAR4 quintile 1 or 2 Male and POLAR4 quintile 3, 4 or 5

Split indicator type	Split indicators considered
Based on young students (aged under 21 in year of entry to higher education programme)	

Annex C: Further information about how we calculate and present statistical uncertainty

This annex is aimed at readers seeking a fuller explanation of the statistical methods used to present statistical uncertainty.

It provides information and technical detail about our approach to the calculation, presentation and interpretation of the statistical uncertainty associated with indicator values and difference between indicator and benchmark values within the **student outcomes data dashboard**, the **TEF data dashboard** and the **access and participation data dashboard**.

Some of the information in the annex is aimed at readers with an in-depth knowledge of advanced statistical methods and assumes a familiarity with statistical formulae and notation.

1. As described in our discussion of Communication of statistical uncertainty, our presentation of data to inform our regulation of student outcomes, the TEF and access and participation uses shaded bars to communicate the statistical uncertainty associated with indicators, split indicators, and for some uses, the difference between the indicator and benchmark values.

2. Figure 1, Figure 2, Figure 3 and Figure 4 illustrated how we communicate the distribution of statistical uncertainty through the shaded bars in the student outcomes, TEF and access and participation data dashboards that we show in respect of:

    a. The observed value of the indicator as a point estimate, reporting the proportion of students that we observe to have achieved a certain outcome or reported a certain experience.

    b. The observed value of the difference between the indicator and its associated benchmark, as a point estimate. This is relevant to the student outcomes and TEF data dashboards.

    c. The observed value of the gap between two indicators of the same split type, as a point estimate. This is relevant to the access and participation data dashboard.

3. The shaded bars aim to represent the distribution of statistical uncertainty around the different values that we have calculated to understand a provider's performance. They can be thought of as representing a series of discrete confidence intervals around the point estimate we have observed, where each confidence interval in the series corresponds to a different confidence (or significance) level.

What is a confidence interval?

4. One way in which statistics can help to describe the level of statistical uncertainty associated with an indicator or split indicator is to supply a range of reasonable values for a provider's true performance. This range of reasonable values is called a confidence interval.

5. A confidence interval has an associated confidence level, which represents the likelihood that the confidence interval contains the true value of performance in the superpopulation. In other words, on average, 95 per cent of confidence intervals computed at the 95 per cent confidence

level would contain the true value of performance in the superpopulation. Similarly, 90 per cent of confidence intervals computed at the 90 per cent confidence level would contain the true value, and likewise for other confidence levels.

6. The width of an indicator's confidence interval is influenced by the chosen confidence level and by the number of students informing the calculation of the observed indicator value (otherwise known as the sample size). The higher the confidence level, the wider the confidence interval, since a greater range of values is required to be more confident that it will contain the true value of performance. When the sample size increases, the width of the confidence interval at a given confidence level tends to decrease, since the potential influence of randomness is reduced. The variability of the observed outcomes or experiences across students can also influence the width of the interval, with more variation generating wider confidence intervals. Wider confidence intervals mean that we become less confident that the observed point estimate is close to the true value in the superpopulation.

7. We do not report limits of confidence intervals that are above 100 per cent or below zero per cent. In such cases, when it is clearly impossible for the proportion to fall below 0 per cent or above 100 per cent, confidence intervals can appear truncated at one end and not be symmetrical.

Using confidence intervals to construct the shaded bars

8. The shaded bars are constructed around the point estimates by calculating a set of confidence intervals, starting with the 75 per cent confidence interval with further intervals calculated at 2.5 percentage point increments up to a maximum of a 99.7 per cent confidence interval.

9. The bar is shaded between each of these intervals to represent the shape of the underlying distribution, with the darkest shading representing the range in which has the highest likelihood that true underlying provider performance might lie. Much like the 'bell curve' of the normal distribution, as the shading lightens in both directions it represents a lower likelihood that true underlying performance falls at that point.

10. By illustrating the distribution up to a maximum of a 99.7 per cent confidence interval, we maximise the chance that the shaded bars encapsulate the true underlying performance. This empowers users to better understand the confidence in which they can hold their own judgements of student outcomes and experiences, by making their own choice of confidence intervals.

11. Our construction of the shaded bars requires a set of assumptions to be made about the statistical distributions from which the statistics are drawn. These assumptions, and their resulting influence over the methods we have selected, vary in respect of our consideration of indicator values and the difference between indicator and benchmark values. The sections that follow explain these assumptions in greater detail, and provide the calculations which underpin the summary figures we provide to help with Interpreting the shaded bars used in our presentation of indicators and split indicators.

Constructing the shaded bars for indicator values

12. Shaded bars are used to represent the distribution of statistical uncertainty associated with the **indicator value** as a point estimate.

13. Typically for this type of observed outcome, you would create a binomial proportion confidence interval, where the probability of success and the number of trials is given by the observed indicator value and the number of students informing the indicator respectively (the denominator).

14. The confidence intervals which underpin the construction of the shaded bars are created using the Jeffreys interval.[72] We have used the Jeffreys interval method because it has been shown to perform well in a wide range of circumstances in the assessment of many and diverse providers, including where the denominator is small, or the observed proportion is close to 0 per cent or 100 per cent.[73] The Jeffreys interval is calculated using the Jeffreys prior[74] for the binomial proportion, p, given n trials. Confidence intervals are calculated from the posterior distribution for p which is a Beta distribution with parameters $(np + 0.5, n - np + 0.5)$. In our case, p is the observed proportion and n is the denominator for the indicator in question. As the standard deviation of the binomial distribution decreases as the probability of success approaches 1 (i.e. an observed rate near 100 per cent), this results in a clear asymmetry in some of the bars.

15. For gaps relating to the difference between student outcomes indicators in the access and participation data dashboard, and for population-referenced gaps in the access lifecycle stage, we have created these shaded bars by calculating confidence intervals using a normal approximation. The shaded bars represent a normal distribution with mean equal to the calculated gap, and variance estimated by summing the estimated variances for each of the two split indicators (assuming independence between the two populations). Variance for each split indicator is estimated as the variance of a binomial proportion by $p(1-p)/n$, where p is the observed split indicator value and n is the denominator of the split.

16. For gaps in the access lifecycle stage of the access and participation data dashboard which are not population-referenced, it would not be appropriate to assume independence between the two split indicators. Instead, we assume a multinomial distribution and estimate the variance of the gap between access proportions p_1 and p_2 from a population of n entrants as follows.[75]

$$Var(p_1 - p_2) = \frac{p_1(1-p_1)}{n} + \frac{p_2(1-p_2)}{n} - 2Cov(p_1, p_2), \text{ where } Cov(p_1, p_2) = -(p_1 p_2)/n$$

[72] Jeffreys, Harold (1946). An invariant form for the prior probability in estimation problems. Proc. Royal Society, London. A186453–461. http://doi.org/10.1098/rspa.1946.0056.

[73] Brown et al (2001). Interval estimation for a binomial proportion Statistical Science. Vol. 16, No. 2, pages 101-133. http://dx.doi.org/10.1214/ss/1009213286.

[74] Although the Jeffreys interval has a Bayesian derivation it can also be justified from a frequentist perspective. See Brown et al (2001) – details in footnote 55.

[75] As above, each proportion is assumed to be binomial, with variance estimated accordingly. The covariance term is estimated based on the covariance of two multinomial proportions. See https://statproofbook.github.io/P/mult-cov.

17. Otherwise, the approach is the same; the shaded bars represent normal distributions with mean equal to the observed gap.

18. In this release of the access and participation dashboard, statistical uncertainty has not been presented for ratios or changes from year 5 to year 6, but we will explore the possibility of doing this in future. When considering ratios, users are encouraged to consider the statistical uncertainty presented via shaded bars for the associated percentage point gap.

Interpreting the shaded bars for indicator values presented in the student outcomes data dashboard

19. The summary figures shown alongside the green shaded bars report the proportions of the statistical uncertainty distribution that fall above and below the minimum numerical thresholds used in respect of condition B3. It informs our regulatory approach to assessment of condition B3.[76]

20. To produce the figures in this supplementary table we have determined the proportion of the distribution represented by the bar that falls above and below the numerical threshold. To do this, the cumulative distribution function (CDF) for the Jeffreys posterior distribution is used. The calculation is as follows:

    a. Proportion of the statistical uncertainty distribution above the numerical threshold: one minus the CDF at the numerical threshold.

    b. Proportion of the statistical uncertainty distribution below the numerical threshold: the CDF at the numerical threshold.

Constructing the shaded bars for the difference between indicator and benchmark values presented in the student outcomes and TEF data dashboards

21. Blue shaded bars are used to represent the distribution of statistical uncertainty associated with the **difference** between a provider's indicator and its corresponding benchmark as a point estimate.

22. We use benchmarking to create a comparator to the indicator values. The method to determine the benchmark, and hence the difference between the indicator value and the benchmark, follows the methodology described by Draper and Gittoes (2004)[77] and the most relevant elements of this methodology are described in our discussion of Benchmarking. The method includes a derivation of the standard deviation[78] of the difference between the indicator value and the benchmark, which incorporates uncertainty in both components. They describe the

[76] For further information about the indicative categories of statistical confidence, see Annex B of Regulatory Advice 20: Regulating student outcomes, at www.officeforstudents.org.uk/publications/regulatory-advice-20-regulating-student-outcomes/.

[77] Draper, D and Gittoes, M (2004). Statistical analysis of performance indicators in UK higher education. Journal of the Royal Statistical Society. Series A (Statistics in Society), 167, Part 3, pages 449-474.

[78] Because these are standard deviations of a statistic (the difference), they are more usually called standard errors.

relationship between the indicator value and the benchmark and present evidence that the differences are normally distributed.

23. Each of the blue shaded bars represent a normal distribution with the distribution mean equal to the observed difference from benchmark and the distribution variance as the standard deviation squared. The distribution formula for the difference is:

    $N(\text{Difference}, (\text{Standard deviation})^2)$

24. Where the observed indicator value is near 0 per cent or 100 per cent, it is possible for the distribution of the difference from benchmark represented by the blue shaded bar to imply that the indicator value (i.e. if you centred this distribution around the observed indicator value) could extend below 0 per cent or above 100 per cent. In constructing these bars, we have explicitly not adjusted for this and have instead tried to mitigate this issue by also presenting the green shaded bar. This is because the green shaded bar does not have this issue due to its derivation. The use of both charts reduces the risk that a user will misinterpret the uncertainty on the difference from benchmark in these cases.

Interpreting the shaded bars for the difference between indicator and benchmark values

25. The summary figures shown alongside the blue shaded bars report the proportions of the statistical uncertainty distribution that fall above and below provider's benchmark. It informs our regulatory approach to assessment of condition B3 and the TEF.[79]

26. To produce the figures in the supplementary table alongside the blue shaded bar we have determined the proportion of the distribution represented by the bar that falls around the numerical thresholds. To do this, the cumulative distribution function (CDF) for the normal distribution is used. To the left of the boundary of the numerical threshold the proportion is given by the CDF, to the right of the boundary of the numerical threshold the proportion is given by one minus the CDF. The numerical thresholds used differ between our regulation of student outcomes and the TEF:

    - For our regulation of student outcomes:
    - Proportion of the statistical uncertainty distribution above the benchmark: one minus the CDF at 0
    - Proportion of the statistical uncertainty distribution below the benchmark: the CDF at 0.
    - For the TEF:
    - Proportion of the statistical uncertainty distribution materially above benchmark: one minus the CDF at 2.5

[79] For further information about the indicative categories of statistical confidence, see Annex B of Regulatory Advice 20: Regulating student outcomes, at www.officeforstudents.org.uk/publications/regulatory-advice-20-regulating-student-outcomes/. Our analysis of responses to the TEF consultation (www.officeforstudents.org.uk/publications/student-outcomes-and-teaching-excellence-consultations/the-tef/) confirmed that the same categories will be relevant to TEF assessment.

- Proportion of the statistical uncertainty distribution materially below benchmark: the CDF at -2.5
- Proportion of the statistical uncertainty distribution broadly in line with benchmark: one minus the sum of the results for materially above benchmark and materially below benchmark.

Annex D: Further information about making multiple comparison adjustments

This annex is aimed at readers seeking more information about whether and how adjustments should be made to their interpretation of indicators and split indicators to account for multiple comparisons.

It provides information about our approach which is relevant to the **student outcomes data dashboard**, the **TEF data dashboard** and the **access and participation data dashboard**.

Some of the information in the annex is aimed at readers with an in-depth knowledge of advanced statistical methods and assumes a familiarity with statistical formulae and notation.

Multiple comparisons adjustments

1. In statistics, the issue of 'multiple comparisons' arises when a user considers multiple statistical tests at once. With more tests there is more opportunity for unlikely events to occur simply due to the influence of random chance. To account for this, when conducting multiple tests, it may be appropriate to make adjustments to what we consider to be unlikely to have occurred by random chance alone compared to a single test (at equivalent levels of confidence).

2. This means that when looking at multiple indicators or split indicators at once, for any that appear to be significantly above or below a numerical threshold, or benchmark, there is a greater chance of finding a result that appears significant but has occurred through random chance alone.

3. We can also think of multiple comparisons as an issue of **selection bias**. If we consider many indicators simultaneously in search of outliers, then an indicator value that we identify as anomalous may be less likely to be representative of the true value due to the influence of random chance than a randomly selected indicator value. Similarly, the confidence intervals constructed around such an outlier may be less likely to contain the true value than their significance levels would suggest. The more indicators we look at in order to find such an outlier, the greater the potential for selection bias.

4. Adjustments for multiple comparisons rates typically limit the risk of making a 'false discovery' (in statistics, a type 1 error) across the statistical tests, but also increases the risk that statistical evidence may be overlooked (in statistics, a type 2 error).

5. We have not made any formulaic adjustments for multiple comparisons within our construction of student outcome and experience indicators to inform our regulation of student outcomes, access and participation and the TEF. This is because we do not consider an arbitrary adjustment based on an assumed number of comparisons to be proportionate:

    a. The shaded bars we have included in the presentation of the data indicates the distribution of statistical uncertainty around the observed value and we do not rely on, nor calculate the results of, a single confidence interval or significance test. We consider that the

presentation of uncertainty up to the 99.7 per cent confidence interval is sufficient to encapsulate the true underlying performance.

b. The number of comparisons that users might make within and across the full set of available data points could vary substantially depending on the use case and is difficult to predict.

c. While an adjustment based on an arbitrary number of comparisons may reduce the risk of users making incorrect assumptions due to statistical variation, we consider that it would simultaneously increase the risk that good statistical evidence is overlooked. We consider that showing artificially wider distributions of the statistical uncertainty associated with each indicator would be a particular issue for our regulatory uses and where users are considering an indicator in isolation or looking across a smaller number of indicators than are accounted for by the arbitrary adjustment.

6. We acknowledge that there are some circumstances in which it may be desirable to consider making adjustments for multiple comparisons. We suggest that when lower levels of statistical confidence are being used to help identify outlying data points, or indicators that are above or below a benchmark or numerical threshold, users should consider adjusting to a higher level of confidence when making their judgements. This is because of the higher risk of false discovery when using lower levels of statistical confidence. In this context, users may wish to be more conservative in their interpretation of statistical uncertainty the more comparisons they are making. Users can heavily mitigate the risk of making a false discovery by adjusting to use higher levels of statistical confidence. However, in doing so, they should note the consequence being an increased risk that sound statistical evidence may be overlooked. We provide information in this annex that is intended to support users to understand when such adjustments may be appropriate, and how an appropriate balance between the risks of type 1 and type 2 errors might be achieved.

7. To explore when it is appropriate to consider a multiple comparisons adjustment, and how to judge how many comparisons are being made, we will consider the following illustrative scenarios:

   a. A provider is looking at their own student outcomes measures and chooses to focus on the overall indicator values for each of the modes and levels of study that they deliver. They are looking for any obvious areas of strength or weakness in their performance relative to the benchmark values.

   b. A regulator is identifying which providers it considers should fall in scope for a regulatory assessment of their student outcomes. The regulator is interested in a specific set of indicators across many providers.

   c. A head of department at a given provider is considering the split indicators for full-time continuation for the same level of study as a particular course that is offered by their department.

8. The scenarios are elaborated in more detail in paragraphs 9 to 23. Once you have considered whether and how you need to consider an adjustment for multiple comparisons, our discussion of The potential impact of multiple comparison adjustments illustrates how adjustments could be made for a given number of comparisons.

Scenario A

9. In this scenario, a provider is considering their own student outcomes measures across each mode and level of study for which they deliver provision. They are predominantly interested in identifying indicators that appear to demonstrate notable areas of strength or weakness relative to benchmark.

10. They should be aware of how many indicators they are looking at in order to identify any that appear anomalous. In doing this they may want to consider:

    a. How many indicators are reportable; any that are not reportable can be ignored.

    b. The population sizes associated with each indicator. Larger populations should translate to less statistical uncertainty and narrower shaded bars and the provider may also be more concerned about indicators that relate to more students.

    c. The scale of the difference between the observed indicator values and the benchmarks.

11. If we suppose that one indicator appears to suggest weaker outcomes than any of the others, has an observed rate that is clearly below the benchmark value, and is labelled as 90 per cent below benchmark. The provider may be concerned about this apparent area of weaker outcomes but unsure how confident they should be that the observed indicator is not below benchmark as a product of random chance alone.

12. They could consider the impact of an adjustment for the number comparisons identified from the populated indicators that they have considered. They can understand the impact of the adjustment on the shaded bar and proportion of the uncertainty distribution below benchmark by considering the lookup tables and diagrams included in our discussion of The potential impact of multiple comparison adjustments.

13. The provider should consider the strength of the statistical evidence that the indicator is below benchmark and weigh up their relative risk appetite between the risk of reacting to cases that occurred purely by chance, compared with the risk of ignoring sound evidence that is representative of true performance. In doing this they might consider:

    a. The range of values within which they feel confident that the true indicator lies after considering adjustments for multiple comparisons.

    b. That when adjusting for multiple comparisons, the confidence intervals that relate to higher levels of confidence would shift less than those that relate to lower levels of confidence following a multiple comparisons adjustment. This is illustrated within the diagrams contained in our discussion of The potential impact of multiple comparison adjustments.

    c. Whether they have other evidence that either corroborates or contradicts the idea that underlying performance is relatively weak.

Scenario B

14. A regulator is looking at three student outcome measures for the purpose of identifying which providers it considers should fall in scope for a regulatory assessment of their student outcomes. The regulator is interested in providers with indicator values that fall below a certain numerical value.

15. If there are around 200 regulated providers, then at face value the regulator might be considering around 600 indicators here, but there are some considerations that mean an adjustment for that many comparisons would be overly cautious:

    a. Some of the indicators will have no students, or are otherwise not reportable (e.g. due to small populations or low survey response rates).

    b. Some of the indicators will relate to large populations and underlying performance that falls close to 100 per cent. In many cases, these will not be of interest to the regulator because the large population sizes often give rise to low levels of statistical uncertainty (and very limited potential for results to have occurred through random variation alone) associated with an indicator value that is very high in absolute terms.

16. The more cautious the regulator is to avoid making incorrect assumptions as a result of random variation, the greater the risk that pockets of poor performance go unidentified and unaddressed. If the regulator's approach includes other corroborating evidence, then it may not need to be so conservative in adjusting for multiple comparisons; if it does not include any other corroborating evidence, then a larger adjustment for multiple comparisons may be appropriate.

17. The adjustment may not need to account for the full 600 indicators (for the reasons outlined above in paragraph 15 above). It should, however, provide a reasonable estimate for the number of reportable indicators with populations large enough that they give rise to low levels of statistical uncertainty and that could conceivably have generated an indicator value below the regulator's numerical value in an alternative version of events (i.e. the observed indicator value is not so far away from the regulator's value that there is no realistic prospect it could have been observed below it).

18. In this scenario, the regulator will want to make judgements about the extent to which multiple comparison adjustments should be considered, and the consequences of poor performance going unidentified and unaddressed relative to the consequences of it acting on a 'false discovery'. The strength of statistical evidence required to prompt a regulatory assessment of their student outcomes should be made on a case-by-case basis, subject to the considerations outlined above. In doing so, they might consider seeking appropriate support for understanding the statistical concepts and approaches that influence their judgements.

Scenario C

19. Having considered the overall continuation indicator for the full-time, first-degree course that corresponds to a particular course that is offered by their department, the head of department at a given provider is looking across the full range of split indicators for full-time first degree continuation outcomes for their provider.

20. In this scenario, the head of department should:

    a. Use the shaded bars to understand the range of values likely to reflect underlying performance.

    b. Consider the strength of evidence across all the split indicators, recognising the statistical uncertainty associated with each.

c. Be aware of the risk of anomalous outcomes arising through random chance across the full range of split indicators under consideration.

21. If there are split indicators that appear to be outliers (because they have particularly strong or particularly weak outcomes when considering all the other split indicators), then the head of department may wish to consider an adjustment for multiple comparisons, to mitigate the risk that those observed split indicators arose through random chance alone.

22. In considering such an adjustment, the head of department should make a reasonable estimate of how many of the split indicators were considered concurrently before they started focusing on these apparent outliers. For full-time, first-degree continuation measures there are 67 split indicators, but it is likely that the effective number of comparisons was smaller than 67:

    a. Some split indicators may be suppressed (due to small populations or otherwise).

    b. Some split indicators may be reportable but based on relatively small populations and therefore given less weight by the student.

23. As discussed in previous scenarios, the head of department could take a less cautious approach to adjusting for multiple comparisons (and judging the strength of statistical evidence more generally) if the split indicators in question are corroborated by other evidence. In doing so, they might consider seeking appropriate support for understanding the statistical concepts and approaches that influence their interpretation.

Information for TEF panel members and B3 assessors regarding multiple comparison adjustments

24. This information builds on Scenario B and Scenario C provided above and represents the broad approach that we expect TEF panel members and B3 assessors to take when making judgements in our main regulatory contexts.[80] We take the view that the approach taken in individual cases will need to consider the merits of that individual case, statistical and otherwise, so our guidance to TEF panel members and B3 assessors is expressed in broad terms and should be considered illustrative rather than prescriptive. TEF panel and B3 assessors will need to make subjective judgements to inform their specific approach.

25. The illustrations we give in paragraphs 26 to 38 refer to two of the main regulatory contexts in which we consider it most likely that adjustments for multiple comparisons will require careful consideration.

Illustration 1: Multiple comparison adjustments in the context of prioritisation for assessment of compliance with condition B3

26. Our regulatory advice relating to our regulation of student outcomes describes the prioritisation approach that we will use to select providers for an assessment of compliance with condition B3. The specific criteria involved in the prioritisation approach vary over time, within the bounds of the approach described in the regulatory advice.

[80] 'B3 assessors' describes OfS staff that may be involved in the assessment of condition B3 in any circumstances.

27. This illustration supposes that in a particular year, we have decided that our prioritisation approach for that year involved three given indicators. We would therefore be looking at those three prioritised indicators across all OfS-registered providers, to select some providers for B3 assessment. Providers with indicators that appear below the relevant numerical threshold for condition B3 may be selected, particularly if a provider appears to demonstrate performance below the numerical threshold that affects large number of students, or the indicator appears to be a long way below the numerical threshold. In this respect, this illustration is similar to Scenario B above and the considerations described there.

28. If there are around 400 registered providers, then at face value we would be considering around 1,200 indicators here, but there are some considerations that mean an adjustment for that many comparisons would be overly cautious:

    a. Many of the indicators have no students, or are otherwise not reportable (e.g. due to small populations or low survey response rates).

    b. Even indicators that are reportable are less likely to lead to a provider being prioritised for assessment if they relate to small populations of students.

    c. Many of the indicators relate to large enough populations and underlying performance that is far enough above the numerical threshold that there is no realistic prospect that they would ever have seen their indicator value fall below the numerical threshold through random variation.

    d. For providers where one or more prioritised indicator appears below the threshold, B3 assessors may consider wider evidence available about the provider, including:

        - Other qualitative intelligence and contextual information.

        - Indicators and split indicators beyond the prioritised categories. If many indicators and split indicators appear to be below their numerical thresholds, then this is unlikely to be due to random chance alone (particularly if populations are large, performance is a long way below the threshold, and we have high level of statistical confidence).

29. The more cautious we are to avoid making incorrect assumptions as a result of statistical variation the greater the risk that pockets of poor performance go unidentified and unaddressed. The consequences of us allowing poor performance to go unaddressed will need to be weighed against the consequences of a provider being prioritised for an assessment of condition B3 on the basis of a false discovery.

30. For providers where the wider evidence discussed in paragraph 28 clearly highlights widespread weak outcomes or corroborates that performance is weak for the indicators among the prioritised categories, B3 assessors may not need to be so conservative in adjusting for multiple comparisons.

31. If a provider has an observed indicator within the prioritised categories below the numerical threshold but these weaker outcomes have not been corroborated by the other indicators for the provider or any other evidence, a larger adjustment for multiple comparisons may be appropriate. This adjustment would not need to account for the full 1,200 indicators (for reasons outlined above in paragraph 28), but should account for a reasonable estimate of the

number of prioritised indicators with large enough populations that either had an observed value below the numerical threshold or conceivably could have had an observed values below the numerical threshold in an alternative version of events.

32. Judgements about the extent to which multiple comparisons adjustments should be considered and the strength of statistical evidence required to prompt further assessment of compliance with condition B3 need to be made on a case-by-case basis, subject to the considerations outlined above. B3 assessors may seek support from OfS analysts in making these judgements.

Illustration 2: Multiple comparison adjustments in the context of TEF assessment

33. Our regulatory advice relating to the TEF describes the TEF assessment approach that the TEF panel members will use to award TEF ratings. TEF indicators and split indicators are considered in the assessment process alongside the TEF submission. The split indicators are considered secondarily to the indicators. However, the panel members consider the split indicators to:

    a. Consider how far very high quality and outstanding quality features might apply across all a provider's student groups and range of courses and subjects.

    b. Test the evidence in a provider's submission about its strengths and areas for improvement, including the provider's own analysis and use of the split indicators, alongside any other evidence it determines for itself.

34. It will therefore be necessary for TEF panel members to look across the full range of split indicators for a given indicator for a single provider. In this respect, this illustration is similar to Scenario C above and the considerations described there.

35. The TEF assessment approach describes that TEF panel members can use indicative categories of statistical confidence to help them to interpret the split indicators, but in most cases (where the uncertainty distribution spans one or more of the guiding lines) they should avoid making binary judgements about split indicator performance based on a single fixed significance level (i.e., they should not act as if a split indicator is definitely materially above or below benchmark just because there is significant evidence of that at a given level). Instead, they should:

    a. Use the shaded bars to understand the range of values likely to reflect underlying performance.

    b. Consider the strength of evidence across the split indicators, recognising the statistical uncertainty associated with each.

    c. Be aware of the risk of anomalous outcomes arising through random chance across the full range of split indicators under consideration.

36. If there are split indicators that appear to have particularly strong or particularly weak outcomes when considering all of the split indicators, and this is likely to impact judgements made through the TEF assessment process, then the TEF panel members may wish to consider an adjustment for multiple comparisons to mitigate the risk that the observed split indicators of interest arose through random chance alone.

37. In considering such an adjustment, panel members should make a reasonable estimate of how many of the split indicators were considered concurrently before focusing on these apparent outliers. For full-time continuation, for example, there are 67 split indicators, but it is likely that the effective number of comparisons was smaller than 67:

    a. Some split indicators may be not applicable or suppressed (due to small populations or otherwise).

    b. Some split indicators may be reportable but based on relatively small populations and therefore given less weight in assessment.

    c. Panel members may be focusing on particular split indicators that relate to evidence from the TEF submission.

38. As discussed in the previous scenarios, panel members may take a less cautious approach to adjusting for multiple comparisons (and judging the strength of statistical evidence more generally) if the split indicators in question are corroborated by evidence from the submission or evidence from other TEF indicators.

The potential impact of multiple comparison adjustments

39. A common approach to adjusting for multiple comparisons is the Bonferroni correction, which aims to limit the probability of making a false discovery (type 1 error) in any of the comparisons (limiting the family-wise error rate). This is a cautious approach, which, in aiming to limit the risk of making any false discoveries at all, can lead to very wide confidence intervals and substantially increase the risk of overlooking statistical evidence (type 2 errors), particularly when adjusting for many comparisons.

40. An alternative approach is to limit the rate of false rejections (false discovery rate, or FDR) of a hypothesis. That is, the rate at which false discoveries are made relative to the total number of discoveries. An approach like this has been taken for NSS confidence intervals.[81] In aiming to limit the proportion of false discoveries rather than the chance of making any across all comparisons, these methods are not as conservative as the Bonferroni correction and give narrower confidence intervals.

41. Two confidence interval adjustments were used within the NSS results but here we will opt for the more conservative of the two as it does not rely on any dependence assumptions. In some cases (when considering multiple independent indicators, for example) this adjustment may be overly cautious and lead to confidence intervals that are wider than necessary. To limit the FDR to a given type 1 error rate (α), for m comparisons, this approach uses an approximate adjusted error rate for each interval of:

$$\alpha^* = \frac{\alpha(m+1)}{2m(\log_e(m) + 0.6)}$$

[81] See the paper by Professor Harvey Goldstein to HEFCE, at: www.officeforstudents.org.uk/media/61c14141-5d4d-49ec-a7d3-6626a7ea0f90/confidence-interval-calculations.pdf

42. Table D1 shows the impact of such an adjustment. The first row gives the type 1 error rate (α) for individual confidence intervals in isolation. The subsequent rows show, for given numbers of comparisons, the individual confidence intervals required to limit the FDR to the same value of α. For example, Table D1 shows that if an adjustment equivalent to making 50 comparisons was considered appropriate, a 99.4% confidence interval could be said to give an equivalent level of confidence as a 95% confidence interval would for one indicator in isolation.

43. The values in Table D1 would also apply to the calculation of proportions of the uncertainty distribution which are displayed alongside the shaded bars. For example, if an adjustment equivalent to making 50 comparisons was considered appropriate then a finding that 98.9% of the uncertainty distribution falls below a numerical threshold could be said to give an equivalent level of confidence as finding that 90% of the uncertainty distribution falls below a numerical threshold would when considering one indicator in isolation.

Table D20: Confidence intervals with adjustments for multiple comparisons

	Confidence interval					
	99.0%	**97.0%**	**95.0%**	**90.0%**	**80.0%**	**70.0%**
Type 1 error rate (α)	0.01	0.03	0.05	0.1	0.2	0.3
Comparisons	**Adjusted confidence interval (1- α*)**					
5	99.7%	99.2%	98.6%	97.3%	94.6%	91.9%
10	99.8%	99.4%	99.1%	98.1%	96.2%	94.3%
50	99.9%	99.7%	99.4%	98.9%	97.7%	96.6%
100	99.9%	99.7%	99.5%	99.0%	98.1%	97.1%
300	99.9%	99.8%	99.6%	99.2%	98.4%	97.6%
600	99.9%	99.8%	99.6%	99.3%	98.6%	97.9%
1000	99.9%	99.8%	99.7%	99.3%	98.7%	98.0%
2000	99.9%	99.8%	99.7%	99.4%	98.8%	98.2%
5000	99.9%	99.8%	99.7%	99.5%	98.9%	98.4%

44. The following figures indicate how the presentation of our indicators might change if a multiple comparisons adjustment was applied for a range of different numbers of comparisons (10, 100, 600, or 5000). These figures are only intended to be illustrative and are intended to give users a broad understanding of the scale of change that might result from a multiple comparisons adjustment for varying numbers of comparisons.

45. Figure D1 shows an example of the green shaded bar for an example indicator value and how the presentation may change following a multiple comparisons adjustment. Figure D2 shows the same presentation for an example of the difference between that same indicator value and its benchmark. We can see the shaded bars (and the individual confidence intervals within them) become wider the more comparisons we adjust for. Notably, under these adjustments, the confidence intervals with lower levels of statistical confidence shift by more than the confidence intervals with greater levels of statistical confidence. This indicates that the more confident we are to begin with, the less important it is that multiple comparisons are accurately adjusted for. While this has been illustrated using horizontal shaded bars this would apply in a

similar way for the vertical shaded bars that are used to communicate statistical uncertainty in the access and participation data dashboard.

Figure D1: Presentation of an indicator, adjusted to account for multiple comparisons

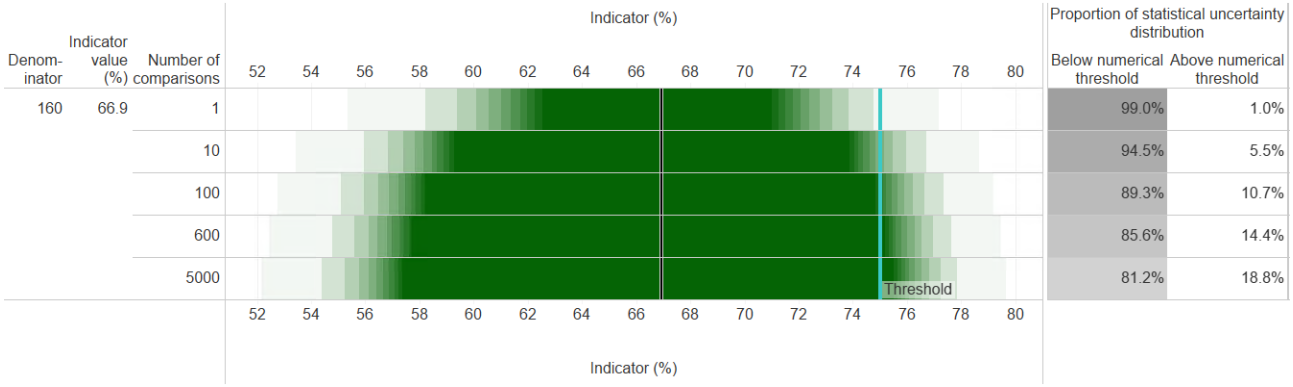
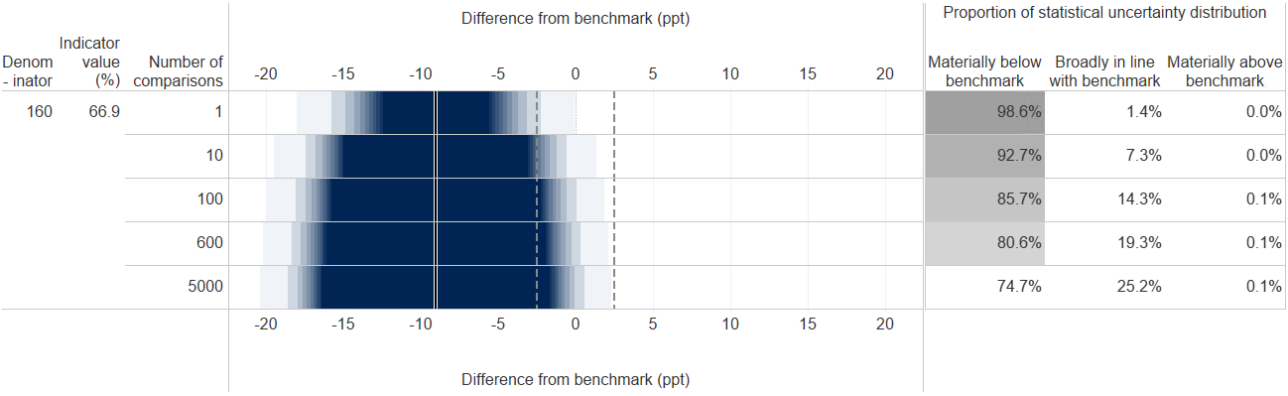


Figure D2: Presentation of the difference from benchmark for an indicator, adjusted to account for multiple comparisons



46. We do not expect users to re-calculate adjusted confidence intervals or statistical uncertainty distributions, nor to replicate these adjusted presentations. We anticipate that users will use the broad understanding this annex aims to support – of the considerations when establishing how many comparisons are being made, the visual impact of different numbers of comparisons on the shaded bars used in our data dashboards, and the potential for interpretation of a provider's performance to move between different levels of statistical confidence – to gauge the weight they may wish to place on particular pieces of statistical evidence.

Annex E: Principles for the selection and application of benchmarking factors

This annex is aimed at readers seeking more information about how we have selected benchmarking factors. The benchmarking principles outlined in this annex have guided the selection and application of benchmarking factors for benchmarking indicators and split indicators, for undergraduate levels of study, within the **student outcomes data dashboard** and the **TEF data dashboard**.

1. The principles outlined in this annex have been used to inform the approach that we take in selecting and applying the factors used in benchmarking calculations.

2. These principles are guiding rather than binding, but they are intended to provide an effective mechanism to build public trust and confidence in the benchmarks that we create and use in our student outcome and experience indicators.

3. When selecting benchmarking factors, our intention is that each principle is considered in turn and, where appropriate, evidence of its applicability would be sought from statistical analysis or modelling. We are aware that the principles may sometimes sit in tension with one another, and that in most cases a judgement will be required to confirm fit or applicability with the principle.

4. The **core principles** relating to the factors being considered for benchmarking are:

    a. The selection of benchmarking factors should be fit for purpose, evidence-based and robust, conforming to recognised best practice in the production of statistical information. In particular:

    b. Details of the selection process should be published for the benefit of providers and other users or interested parties.

    c. The selection of benchmarking factors should vary across different student outcome and experience indicators only when there is a clear and valid rationale.

    d. The number and definition of benchmarking factors selected should not compromise the statistical integrity of the broader benchmarking approach.

    e. Benchmarking factors should be applicable to, and available for, all types of providers across England that are delivering the higher education provision for which the indicator is measuring students' outcomes or experience.

    f. Benchmarking factors should contribute to an overall benchmarking approach which supports fair comparison of indicators across the higher education sector. A candidate benchmarking factor should therefore have relevance to help explain the context or differing characteristics of a provider's students or provision.

    g. The benchmarking approach should neutralise the effect of characteristics on a provider's performance where this is consistent with policy objectives. This approach guards against

inadvertently creating incentives for providers to change their behaviour in terms of the students they recruit or the range of provision they offer in ways that could undermine our ability to meet our duties around access and participation, and competition. It does not imply that it is acceptable for some student groups to receive lower quality provision, but recognises that this is currently the case, and the risks of not controlling for it. The benchmarking approach should only neutralise the effect of characteristics where there is such a risk of negative unintended consequences, as otherwise it risks creating perverse incentives.

h. Benchmarking factors should primarily reflect structural factors that contribute to variations in student outcomes or experience which are outside of a provider's control, or undesirable for it to control for. This means that characteristics of the provider will not normally act as benchmarking factors.

i. In selecting the range of benchmarking factors to apply for a given indicator, the need to preserve the statistical integrity of the broader benchmarking approach requires that consideration should be given to limit the number of factors on the basis of:

   - The size of the population for which the effect occurs: it is unlikely that a factor where the effect is limited to a small population will be selected where there are other factors with similar effects that have broad applicability.

   - The distribution of the population for which the effect occurs: it is unlikely that a factor where the effect is limited to a population concentrated in a small subsection of providers will be selected where there are other factors with similar effects that have applicability to a wider cross-section of provision.

   - The nature of the other candidate factors: where there are a number of similar candidate factors (for example, measures of disadvantage), it will normally be the case that only the one that has the greatest effect should be selected so that a balance of factors is achieved.

j. The factors used in benchmarking should be reviewed at regular intervals, to check that the evidence for, and applicability of, the approach remains current and fit for purpose, and to consider the impact achieved by previous benchmarking exercises.

5. The **availability and data quality** of candidate benchmarking factors should be considered in relation to the principles as follows:

   a. The quality of data items considered as candidate benchmarking factors should be understood and judged to be of sufficiently high quality for use in a benchmarking exercise. The data items should normally be collected in a consistent and fair way across the sector; it should have a good sample base and use transparent definitions.

   b. Where possible, benchmarking factors should be drawn from existing data sources. Any proposal to collect further data for the purpose of a benchmarking factor should be carefully considered against the principles for data burden included within the OfS data strategy.

6. The principles for the **statistical properties** that candidate benchmarking factors should demonstrate are:

a. Statistical models that seek to account for a range of characteristics should identify a remaining correlation between the benchmarking factor and the student outcome or experience that is being measured.

b. Once other factors have been accounted for, statistical modelling should identify that the performance being measured is not uniformly distributed across the attributes within a benchmarking factor, and that differences between these attributes are non-trivial.

c. A benchmarking factor should not be uniformly distributed across providers or performance units; rather, the factor should differentially affect the benchmarks that are calculated, meaning that factors which are distributed unevenly across providers or performance units should be considered as stronger candidates to be used as benchmarking factors.

d. Where possible, a benchmarking factor should be a direct measure, rather than a proxy.

e. As far as possible, the selection of benchmarking factors should limit the extent to which a benchmark value can be determined by a single provider. The selection of a benchmarking factor (and the subsequent grouping of attributes within it) should not compromise the statistical integrity of the broader benchmarking approach.

f. Benchmarking factors (and the data sources from which they are derived) should normally have longevity, with these statistical properties observed to continue over time.

7. Once benchmarking factors have been selected, the principles for **defining groupings of the attributes within the benchmarking factor** are:

   a. The grouping of attributes within benchmarking factors should be fit for purpose and determined through consideration of sound evidence.

   b. The number of categories formed when grouping attributes within benchmarking factors should be the minimum for the benchmarking factor to be effective. The number and definition of the groupings should not compromise the statistical integrity of the broader benchmarking approach.

   c. The grouping of attributes within benchmarking factors should avoid creating groups in which numbers of students possessing those attributes are either very small or very large in the sector overall. The effect of creating groups that are known to be very small or very large at individual provider level should be acknowledged where they cannot be avoided.

   d. The attributes that form a grouping should share a consistency of student backgrounds, outcomes or behaviours with respect to the indicator to which they refer. The consistency of attributes should be understood from the evidence of statistical analysis.

   e. The grouping of attributes within benchmarking factors should make practical sense, to form coherent groups which share a qualitative similarity.

   f. The grouping of attributes within benchmarking factors should vary across indicators only when there is a clear and valid rationale. Where variations are necessary, those deviations should use other groupings that exist elsewhere in a sector-wide hierarchical view of the

benchmarking factor in question, at a more aggregated or disaggregated level according to need.

g. The grouping of attributes within benchmarking factors should be reviewed periodically to ensure that it continues to comply with these principles.

Annex F: Worked example of benchmarking calculations

This annex is aimed at readers seeking more information about how we have calculated the benchmarks that are reported within the **student outcomes data dashboard** and the **TEF data dashboard**.

This annex includes a fictional, simplified example to demonstrate how we calculate benchmarks for continuation measures. This example demonstrates the method that applies to the calculation of benchmarks for continuation, completion, progression and student experience indicators for which the outcome is observed at an individual level before being aggregated to report on a provider.

1. In this fictional, simplified example, assume that we are seeking to calculate benchmarks for continuation measures using only two benchmarking factors which affect the outcomes we are measuring. Specifically, we want to take account of students' age on entry to higher education, and the subject that they are studying. Suppose that students' age is defined as either 'young' or 'not young' and that the higher education sector delivers provision in only three subject areas (agriculture, maths and history). The figures given in this annex are for illustrative purposes only and are not reflective of provider or sector behaviour.

2. That means that for this measure there are six possible distinct benchmarking groups, set out in Table F1.

Step one: the provider

3. The provider for which we are calculating a benchmark has 1,090 students studying agriculture and maths. Table F1 shows the provider's students, split across the six benchmarking groups, and the continuation rate that we observe for each of these groups.

4. Overall, the provider has a continuation rate of 94.3 per cent. This is effectively a weighted average of the rates for each group.

5. Note that the provider's observed continuation rate for young maths students is particularly low (92.0 per cent) in comparison to the observed rate for other groups at the provider. This low continuation rate is outweighed by the larger number of students in groups with higher observed continuation rates, such as young agriculture students.

Table F1: Distribution of the provider's observed continuation rates across benchmarking groups

Age group	Subject group	Number of students	Students in the benchmarking group as a proportion of total students	Observed continuation rate
Young	Agriculture	500	45.9%	95.0%

Age group	Subject group	Number of students	Students in the benchmarking group as a proportion of total students	Observed continuation rate
Young	History	0	0.0%	N/A
Young	Maths	150	13.8%	92.0%
Not young	Agriculture	400	36.7%	94.0%
Not young	History	0	0.0%	N/A
Not young	Maths	40	3.7%	98.0%
				Provider indicator
Total		1,090	100%	94.3%

Step two: the sector

6. There are 210,500 full-time students across the whole sector, studying agriculture, maths and history. Table F2 shows the sector's students, split across the six benchmarking groups, and the continuation rate that we observe for each of these groups across the sector as a whole.

7. Overall, the sector has a continuation rate of 96.6 per cent.

8. Note that the sector's overall continuation rate is driven by high continuation rates observed for young history students (99.0 per cent), and the small student numbers for agriculture subjects, for which we observe relatively low rates for both young (95.0 per cent) and not young (94.0 per cent) students.

Table F2: Distribution of the sector's observed continuation rates across benchmarking groups

Age group	Subject group	Number of students	Observed continuation rate
Young	Agriculture	20,000	95.0%
Young	History	80,000	99.0%
Young	Maths	95,000	95.0%
Not young	Agriculture	5,000	94.0%
Not young	History	6,500	98.0%
Not young	Maths	4,000	98.0%
			Sector indicator
Total		210,500	96.6%

Step three: calculating the provider specific benchmark

9. So far, in Table F2, the sector's continuation rates are weighted against the numbers of students in the **sector** in each of the six distinct benchmarking groups. In Table F3 below, the sector's continuation rates are instead weighted to reflect the students in the **provider**.

10. Table F3 shows that weighting the sector's continuation rates by the proportion of students in each benchmarking group at the provider results in a weighted sector benchmark of 94.7 per cent for this provider.

11. This weighted sector rate is lower than the original sector rate shown in Table F2 since it no longer reflects the (relatively high) rates for history students (because the provider has no history students), and because the agriculture groups have a much higher weighting, reflecting that the provider has a higher proportion of agriculture students than the sector as a whole.

12. The provider's indicator (94.3 per cent) can now be compared with the weighted sector benchmark (94.7 per cent). The provider's rate is still lower than the rate observed for students with similar characteristics across the sector.

Table F3: Calculation of the provider benchmark using the sector's observed continuation rates across benchmarking groups

Age group	Subject group	Students in the benchmarking group as a proportion of total students at the provider (a)	Sector observed continuation rate (b)	Weighted sector continuation numbers (= a x b)
Young	Agriculture	45.9%	95.0%	43.6%
Young	History	0.0%	99.0%	0.0%
Young	Maths	13.8%	95.0%	13.1%
Not young	Agriculture	36.7%	94.0%	34.5%
Not young	History	0.0%	98.0%	0.0%
Not young	Maths	3.7%	98.0%	3.6%
Total		100%	Sector indicator	**Provider benchmark**
			96.6%	**94.7%** (= 43.6% + 0.0% + 13.1% + 34.5% + 0.0% + 3.6%)

Annex G: Definitions of entry qualifications and subject areas of study groupings used in benchmarking

This annex is aimed at readers seeking more information about how we have calculated the benchmarks that are reported within the **student outcomes data dashboard** and the **TEF data dashboard**.

This annex includes definitions of the benchmarking factors of entry qualifications and subject areas of study, which are used to benchmark various student outcome and experience measures.

1. Table G1 shows the groupings of subject areas of study that we have decided to use as benchmarking factors. We have decided to use these groupings as follows:

    a. Broadly defined subject groups as benchmarking factors for the full-time other undergraduate and full-time undergraduate with postgraduate components, part-time and apprenticeship progression, and student experience indicators.

    b. CAH level 1 groups as benchmarking factors for the full-time, part-time and apprenticeship continuation and completion measures.

    c. CAH level 2 groups as benchmarking factors for the full-time first-degree progression, and student experience indicators.

Table G1: Groupings of subject areas used as benchmarking factors

Broadly defined subject group	CAH level 1 group	CAH level 2 group
Medicine, dentistry and veterinary sciences	CAH01: Medicine and dentistry	CAH01-01: Medicine and dentistry
	CAH05: Veterinary sciences	CAH05-01: Veterinary sciences
Nursing, allied health and psychology	CAH02: Subjects allied to medicine	CAH02-02: Pharmacology, toxicology and pharmacy
		CAH02-04: Nursing and midwifery
		CAH02-05: Medical sciences
		CAH02-06: Allied health
	CAH04: Psychology	CAH04-01: Psychology
		CAH03-01: Biosciences

Broadly defined subject group	CAH level 1 group	CAH level 2 group
Natural and mathematical sciences	CAH03: Biological and sport sciences	CAH03-02: Sport and exercise sciences
	CAH07: Physical sciences	CAH07-01: Physics and astronomy
		CAH07-02: Chemistry
		CAH07-04: General, applied and forensic sciences
	CAH09: Mathematical sciences	CAH09-01: Mathematical sciences
Engineering, technology and computing	CAH10: Engineering and technology	CAH10-01: Engineering
		CAH10-03: Materials and technology
	CAH11: Computing	CAH11-01: Computing
Law and social sciences	CAH15: Social sciences	CAH15-01: Sociology, social policy and anthropology
		CAH15-02: Economics
		CAH15-03: Politics
		CAH15-04: Health and social care
	CAH16: Law	CAH16-01: Law
Business and management	CAH17: Business and management	CAH17-01: Business and management
Humanities and languages	CAH19: Language and area studies	CAH19-01: English studies
		CAH19-04, CAH19-02: Languages and area studies
	CAH20: Historical, philosophical and religious studies	CAH20-01: History and archaeology
		CAH20-02: Philosophy and religious studies
	CAH23: Combined and general studies	CAH23-01: Combined and general studies
	CAH24: Media, journalism and communications	CAH24-01: Media, journalism and communications

Broadly defined subject group	CAH level 1 group	CAH level 2 group
Education and teaching	CAH22: Education and teaching	CAH22-01: Education and teaching
Design, and creative and performing arts	CAH25: Design, and creative and performing arts	CAH25-01: Creative arts and design
		CAH25-02: Performing arts
Natural and built environment	CAH06: Agriculture, food and related studies	CAH06-01: Agriculture, food and related studies
	CAH13: Architecture, building and planning	CAH13-01: Architecture, building and planning
	CAH26: Geography, earth and environmental studies	CAH26-01: Geography, earth and environmental studies

2. Table G2 shows the groupings of entry qualifications that we have decided to use as benchmarking factors. We have decided to use these groupings as follows:

   a. 11 entry qualification groups as benchmarking factors for the full-time continuation, completion and progression measures.

   b. 5 entry qualification groups as benchmarking factors for the part-time and apprenticeship continuation and completion measures.

   c. 3 entry qualification groups as benchmarking factors for the part-time and apprenticeship progression measures.

Table G2: Groupings of entry qualifications used as benchmarking factors

3 groups of entry qualifications	5 groups of entry qualifications	11 groups of entry qualifications	Detailed entry qualification group
Higher education qualifications, and other qualifications reported by non-UK domiciled students	Higher education qualifications, and other qualifications reported by non-UK domiciled students	Higher education level qualifications on entry	Higher education qualification: first degree
			Higher education qualification: other undergraduate
			Higher education qualification: postgraduate
		Other qualifications reported by non-UK domiciled students	Other qualifications reported by non-UK domiciled students
			A-level: A*A*A*A*

3 groups of entry qualifications	5 groups of entry qualifications	11 groups of entry qualifications	Detailed entry qualification group
A-levels, international baccalaureate, BTECs (DDM or higher) and other Level 3 qualifications at 105 tariff points or higher	A-levels, international baccalaureate, BTECs (DDM or higher) and other Level 3 qualifications at 105 tariff points or higher	A-levels (AAA or higher)	A-level: A*A*A*A
			A-level: A*A*AA
			A-level: A*AAA
			A-level: AAAA
			A-level: A*A*A*
			A-level: A*A*A
			A-level: A*AA
			A-level: AAA
		A-levels (ABB or higher)	A-level: AAB
			A-level: AAC
			A-level: ABB
		A-levels (BCC or higher) or international baccalaureate	A-level: ABC
			A-level: ACC
			A-level: BBB
			A-level: BBC
			A-level: BCC
			International baccalaureate
		A-levels (CDD or higher)	A-level: CCC
			A-level: CCD
			A-level: CDD
		A-levels (DDD or lower, other Level 3 at 105 tariff points or higher, or 2 A-levels and 1 BTEC	A-level: DDD
			A-level: Below DDD
			2 A-levels and 1 BTEC
			>115 tariff points
			>105 tariff points
		BTECs (at least DDM), or 1 A-level and 2 BTECs	1 A-level and 2 BTECs
			BTEC: D*D*D*
			BTEC: D*D*D
			BTEC: D*DD
			BTEC: DDD

3 groups of entry qualifications	5 groups of entry qualifications	11 groups of entry qualifications	Detailed entry qualification group
			BTEC: DDM
BTECs (lower than DDM), access and foundation courses, or other Level 3 at 65 tariff points or higher, none, unknown or other entry qualifications	BTECs (lower than DDM)	BTECs (lower than DDM)	BTEC: DMM
			BTEC: MMM and below
			BTEC: unknown grades
	Access and foundation courses, or other Level 3 at 65 tariff points or higher	Access and foundation courses, or other Level 3 at 65 tariff points or higher	Access to higher education course
			Foundation course
			>90 tariff points
			>80 tariff points
			>65 tariff points
			Other Level 3 qualifications
	None, unknown or other entry qualifications	None, unknown or other entry qualifications	>40 tariff points
			>0 tariff points
			Other qualifications
			No qualifications on entry
			Unknown qualifications on entry

Annex H: Technical detail about benchmarking calculations

This annex is aimed at readers seeking to understand the calculation of benchmark values from individualised student data. It provides information about the calculation of benchmarks which are relevant to the **student outcomes data dashboard** and the **TEF data dashboard**.

The information in the annex is aimed at readers with an in-depth knowledge of advanced statistical methods and assumes a familiarity with statistical formulae and notation.

1. The general approach to benchmarking follows the design-based adjustment method described in 'Statistical analysis of performance indicators in UK higher education' by Draper and Gittoes (2004).[82] This annex summarises the key information from that methodology.

General approach

2. In this method, for each unique combination of benchmarking factors (described as potential confounding factors (PCFs) in the literature), an observed rate for the measure, and the number of students that inform it, is calculated for both the sector and each provider.

3. The presentation of these rates and number of students for each unique combination of benchmarking factors can be visualised as two large grids as shown in Figure H1 below (the rates shown in the top table, with the number of students in the bottom table). In this figure, M represents the number of unique combinations of benchmarking factors. The method is based on a further cross-tabulation of the N providers by these M categories. The '.' and '+' notations in subscripts indicate averaging and summing over the relevant columns or rows of the table respectively. Within each table, each cell ij contains n_{ij} students from provider i with unique combination of benchmarking factors j. The observed rate of success of these students is $\hat{p}_{ij}$. Each weighted row mean, $\hat{p}_{i.}$ is the observed indicator value for provider i and $\hat{p}_{.j}$ is the observed indicator value for students with unique combination of benchmarking factor j across all students in the sector.

[82] Draper, D and Gittoes, M (2004). Statistical analysis of performance indicators in UK higher education. Journal of the Royal Statistical Society. Series A (Statistics in Society), 167, Part 3, pages 449-474.

Figure H1: A tabular presentation of the rates and number of students for each unique combination of benchmarking factors per provider

Provider	Unique combination of benchmarking factors 1	2	...	M	Weighted row mean
1	$\hat{p}_{11}$	$\hat{p}_{12}$	...	$\hat{p}_{1M}$	$\hat{p}_{1\cdot}$
2	$\hat{p}_{21}$	$\hat{p}_{22}$	...	$\hat{p}_{2M}$	$\hat{p}_{2\cdot}$
⋮	⋮	⋮	⋱	⋮	⋮
N	$\hat{p}_{N1}$	$\hat{p}_{N2}$	...	$\hat{p}_{NM}$	$\hat{p}_{N\cdot}$
Weighted column mean	$\hat{p}_{\cdot 1}$	$\hat{p}_{\cdot 2}$	...	$\hat{p}_{\cdot M}$	$\hat{p}_{\cdot\cdot}$

					Row sum
1	n_{11}	n_{12}	...	n_{1M}	n_{1+}
2	n_{21}	n_{22}	...	n_{2M}	n_{2+}
⋮	⋮	⋮	⋱	⋮	⋮
N	n_{N1}	n_{N2}	...	n_{NM}	n_{N+}
Column sum	n_{+1}	n_{+2}	...	n_{+M}	n_{++}

4. The observed indicator value, $\hat{p}_{i\cdot}$, for the provider can be directly read from the tables in Figure H1. The structure of the table allows us to consider the question: 'What would the observed indicator value have been at provider i, if its distribution of students across the unique combination of benchmarking factors had been what is was, but its rates were replaced by the sector rates, $\hat{p}_{\cdot j}$?'. These can be summarised as follows:

   The observed indicator value, $\hat{O}_i$, at provider i is:

$$\hat{O}_i = \hat{p}_{i\cdot} = \frac{1}{n_{i+}} \sum_{j=1}^{M} n_{ij} \hat{p}_{ij}$$

   The benchmark, $\hat{E}_i$, at provider i is:

$$\hat{E}_i = \frac{1}{n_{i+}} \sum_{j=1}^{M} n_{ij} \hat{p}_{\cdot j}$$

   The difference between the observed indicator value and benchmark, $\hat{D}_i$, at provider i is:

$$\hat{D}_i = \hat{O}_i - \hat{E}_i$$

5. To aid interpretation of the observed difference, the standard deviations of the differences between the indicator value and benchmark have been calculated. A standard deviation measures the amount by which one would expect a statistic to change, based solely on random sampling. Because these are standard deviations of a statistic (the difference), they are more usually called standard errors.

6. To calculate the standard deviation, the formula for the difference is adjusted using algebraic manipulation (the full manipulation can be found in the literature) to be written as a weighted sum of all cells in the tables shown in Figure H1:

$$\widehat{D}_i = \sum_{j=1}^{M}\sum_{k=1}^{N} \lambda_{ikj}\hat{p}_{kj}$$

where $\lambda_{ikj} = \frac{n_{ij}}{n_{i+}}(\delta_{ik} - \frac{n_{kj}}{n_{+j}})$

and $\delta_{ik} = 1 \quad if\ i = k,$
$\delta_{ik} = 0 \quad if\ i \neq k$

Assuming the $\hat{p}_{kj}$ terms are independent, the variance is given by:

$$Var(\widehat{D}_i) = \sum_{j=1}^{M}\sum_{k=1}^{N} \lambda_{ikj}^2 Var(\hat{p}_{kj})$$

The literature shows that a reasonable estimate for the variance of $\hat{p}_{kj}$ can be made by using a shrinkage estimation procedure:

$$Var(\hat{p}_{kj}) = \frac{\hat{p}^*_{kj}(1 - \hat{p}^*_{kj})}{n_{kj}}$$

where $\hat{p}^*_{kj} = 0.5\hat{p}_{..} + 0.5\hat{p}_{kj}$

and $\hat{p}_{..}$ is the overall rate of the sector.

The square root of the variance of $\widehat{D}_i$ gives the standard deviation

7. We calculate the average contribution to benchmark for provider, i, using a similar weighted average calculation. This statistic calculates the contribution of the provider's own students on the sector averages that informs the calculation of the provider's benchmark of the form:

$$average\ contribution\ to\ the\ benchmark_i = \sum_{j=1}^{M} \frac{n_{ij}^2}{n_{+j}n_{i+}}$$

Benchmarking split indicators

8. In the calculation of the standard deviation for the purposes of benchmarking split indicators a small adjustment is made within the formulae described in the general approach above. The approach to create an estimate for the variance of $\hat{p}_{kj}$ by using a shrinkage estimation is the same, but the value for $\hat{p}_{..}$ used in the derivation of $\hat{p}^{*}_{kj}$ remains the overall rate of the sector calculated at provider level. This is instead of using $\hat{p}_{..}$ created based on the subset of the provider and sector to the split indicator. This adjustment is made to ensure that the shrinkage estimation is applied consistently between the overall provider split indicator and other split indicators. For example, in a case where a provider delivers only a single subject, the standard deviation could appear different for the provider-level indicator and the split for the subject only because of the shrinkage estimation.

9. These differences in the approach to calculating benchmarks for split indicators is presented in the same tabular presentation as in Figure H1 in Figure H2, which assumes the split indicator being calculated is for 'Male' students. The $\hat{p}_{..}$ has been relabelled as $\widehat{Overallp}_{..}$. Otherwise, the notation is the same as described in paragraph 2.

Figure H2: A tabular presentation of the rates and number of students for each unique combination of benchmarking factors per provider for male students

Male students at provider...	Unique combination of benchmarking factors 1	2	...	M	Weighted row mean
1	$\hat{p}_{11}$	$\hat{p}_{12}$	...	$\hat{p}_{1M}$	$\hat{p}_{1\cdot}$
2	$\hat{p}_{21}$	$\hat{p}_{22}$	...	$\hat{p}_{2M}$	$\hat{p}_{2\cdot}$
⋮	⋮	⋮	⋱	⋮	⋮
N	$\hat{p}_{N1}$	$\hat{p}_{N2}$	...	$\hat{p}_{NM}$	$\hat{p}_{N\cdot}$
Weighted column mean	$\hat{p}_{\cdot 1}$	$\hat{p}_{\cdot 2}$	...	$\hat{p}_{\cdot M}$	$\hat{Overallp}_{..}$

					Row sum
1	n_{11}	n_{12}	...	n_{1M}	n_{1+}
2	n_{21}	n_{22}	...	n_{2M}	n_{2+}
⋮	⋮	⋮	⋱	⋮	⋮
N	n_{N1}	n_{N2}	...	n_{NM}	n_{N+}
Column sum	n_{+1}	n_{+2}	...	n_{+M}	n_{++}

Adjustments to the general approach to benchmarking for the ‘taught or registered (TorR)’ population

10. While the general approach to benchmarking can be applied to each of the registered and taught populations that our reporting of student outcome and experience measures use as Views of a provider’s student population, for the view of students who are who are either taught or registered at the provider in question (or both) we need to vary our approach.

11. This is because the taught or registered view of student populations allows for students being associated with more than one provider and contributing to more than one provider’s indicators and split indicators. However, the benchmarking methodology assumes that students per provider per unique combination of benchmarking factors are independent from another combination.

12. The design-based adjustment methodology by Draper and Gittoes (2004) is adjusted as follows. We are placing any students that would be allowed to contribute to more than one provider in its own ‘dummy’ provider. These are students that contribute to the provider's indicator who registers them, but also to another provider’s indicator who teaches them. To visualise this, the approach is presented in the same tabular presentation as in Figure H1 in Figure H3. In this figure, providers 1 and 2 share some duplicated students, *Y*, and their overall student population including these students is presented by *X*. The ‘dummy provider’ has been included as a separate row, shown as $1{:}\,2_Y$. Otherwise, the notation is the same as described in paragraph 2.

Figure H3: A tabular presentation of the rates and number of students for each unique combination of benchmarking factors per provider for the taught or registered population

Provider	Unique combination of benchmarking factors: 1	2	…	M	Weighted row mean
1_{X-Y}	$\hat{p}_{1_{X-Y}1}$	$\hat{p}_{1_{X-Y}2}$	$\cdots$	$\hat{p}_{1_{X-Y}M}$	$\hat{p}_{1_{X-Y}\cdot}$
2_{X-Y}	$\hat{p}_{2_{X-Y}1}$	$\hat{p}_{2_{X-Y}2}$	$\cdots$	$\hat{p}_{2_{X-Y}M}$	$\hat{p}_{2_{X-Y}\cdot}$
$1{:}2_Y$	$\hat{p}_{1:2_Y1}$	$\hat{p}_{1:2_Y2}$	$\cdots$	$\hat{p}_{1:2_YM}$	$\hat{p}_{1:2_Y\cdot}$
⋮	⋮	⋮	⋱	⋮	⋮
N	$\hat{p}_{N1}$	$\hat{p}_{N2}$	$\cdots$	$\hat{p}_{NM}$	$\hat{p}_{N\cdot}$
Weighted column mean	$\hat{p}_{\cdot 1}$	$\hat{p}_{\cdot 2}$	$\cdots$	$\hat{p}_{\cdot M}$	$\hat{p}_{\cdot\cdot}$

					Row sum
1_{X-Y}	$n_{1_{X-Y}1}$	$n_{1_{X-Y}2}$	$\cdots$	$n_{1_{X-Y}M}$	$n_{1_{X-Y}+}$
2_{X-Y}	$n_{2_{X-Y}1}$	$n_{2_{X-Y}2}$	$\cdots$	$n_{2_{X-Y}M}$	$n_{2_{X-Y}+}$
$1{:}2_Y$	$n_{1:2_Y1}$	$n_{1:2_Y2}$	$\cdots$	$n_{1:2_YM}$	$n_{1:2_Y+}$
⋮	⋮	⋮	⋱	⋮	⋮
N	n_{N1}	n_{N2}	$\cdots$	n_{NM}	n_{N+}
Column sum	n_{+1}	n_{+2}	$\cdots$	n_{+M}	n_{++}

13. This manipulation to create a 'dummy' provider means:

    a. The $\hat{p}_{kj}$ terms are independent across the whole grid because no students are duplicated within the grid.

    b. There is no effect on the calculation of the sector average, $\hat{p}_{.M}$ because no students are duplicated within the grid.

    c. The approach to estimating the variance of the difference used in the general approach for benchmarking can be used. In this example given in Figure H3, the variance of the difference is calculated for each provider, 1_{X-Y}, 2_{X-Y} and $1{:}\,2_{Y}$.

14. To calculate the difference and to estimate the variance per provider (including students that are duplicated across providers – in this example the variance for provider 1, rather than provider 1 without any students that are duplicated across providers), it is then necessary to combine the information calculated across the provider and any associated 'dummy' providers. Our derivation is as follows:

    Subscript Z represents the number of associated 'dummy' providers from provider i.

    n_{i_Z} represents the number of students from provider i, per 'dummy' provider Z.

    D_i represents the difference (indicator – benchmark) from provider i.

    D_{i_Z} represents the difference (indicator – benchmark) from provider i, per 'dummy' provider Z.

    The difference can be written as a weighted sum of the difference across multiple 'dummy' providers:

$$D_i = \frac{n_{i_1} D_{i_1} + n_{i_2} D_{i_2} + \cdots + n_{i_Z} D_{i_Z}}{n_{i_1} + n_{i_2} + \cdots + n_{i_Z}}$$

    Therefore, the variance of this weighted sum of difference is:

$$Var(D_i) = Var\left(\frac{n_{i_1} D_{i_1} + n_{i_2} D_{i_2} + \cdots + n_{i_Z} D_{i_Z}}{n_{i_1} + n_{i_2} + \cdots + n_{i_Z}}\right)$$

    This is equivalent to:

$$Var(D_i) = \left(\frac{1}{n_{i_1} + n_{i_2} + \cdots + n_{i_Z}}\right)^2 \times Var(n_{i_1} D_{i_1} + n_{i_2} D_{i_2} + \cdots + n_{i_Z} D_{i_Z})$$

$$Var(D_i) = \left(\frac{1}{n_{i_1} + n_{i_2} + \cdots + n_{i_Z}}\right)^2 \times \{Var(n_{i_1} D_{i_1}) + Var(n_{i_2} D_{i_2}) + \cdots + Var(n_{i_Z} D_{i_Z}) + covariance\}$$

15. As students do not appear more than once across Z 'dummy' providers, we can keep the assumption that the $\hat{p}_{kj}$ terms are independent. By combining 'dummy' providers we minimise the covariance between our differences, but inevitably there will a small amount of shared

data[83], and hence covariance between them. In these calculations we are assuming that the covariance term is near zero. By also bringing out the n_i terms:

$$Var(D_i) = \left(\frac{1}{n_{i_1} + n_{i_2} + \cdots + n_{i_Z}}\right)^2 \times \left\{{n_{i_1}}^2 Var(D_{i_1}) + {n_{i_2}}^2 Var(D_{i_2}) + \cdots + {n_{i_Z}}^2 Var(D_{i_Z})\right\}$$

16. This derivation shows that we can estimate the variance for the entire provider by taking a weighted sum of the estimated variances for each of its 'dummy providers'. The square root of this variance gives the standard deviation.

17. We have tested our assumption that the covariance term is near zero by comparing the standard deviations to the taught provider view (which does not need this adjustment because students are not duplicated across providers).

18. We also adjust the calculation of the average contribution to benchmark for provider, i, using a similar weighted average calculation across dummy providers. This can be written as a weighted sum of the difference across multiple 'dummy' providers, where:

    C_i is the contribution to the benchmark from provider i.

    C_{i_Z} is the contribution to the benchmark from provider i, per 'dummy' provider Z.

    $$C_i = \frac{n_{i_1} C_{i_1} + n_{i_2} C_{i_2} + \cdots + n_{i_Z} C_{i_Z}}{n_{i_1} + n_{i_2} + \cdots + n_{i_Z}}$$

[83] This only impacts the calculations of the uncertainty for the difference between indicator and benchmark value, and not the indicator value.