Access and participation data resources

Supporting data user guide

Enquiries to providermetrics@officeforstudents.org.uk

Publication date 24 March 2022

Contents

Introduction

1. This user guide is a basic introduction to using the Office for Students' (OfS's) access and participation data resources. It is intended for anyone new to the data, but it should help higher education providers in particular to navigate the data resources, understand their contents, and use the information to develop their access and participation plans. The data resources have been published as an analytical resource to support further interrogation of the information included in the access and participation data dashboard.[1]

2. The OfS has published the access and participation dashboard to help users build a good and consistent understanding of some of the key data that relates to a provider's performance in access and participation. The supporting access and participation data resources are intended to remove some of the barriers that we know can prevent effective use of data, and arise from the complexity of the access and participation data landscape.

3. We are aware that the information in these resources is not exhaustive in terms of the different types and sources of data, indicators and other measures being used in this area. However, the data allows comparisons to be made between different student groups and their peers, to identify some of the gaps in access and key student outcomes. Providers can use these resources to help them set appropriate targets and support their strategies for reducing gaps between student groups. The OfS will also use these data resources to review submitted plans. This document aims to help users understand what the data resources contain, and how to identify which measures are most relevant to each provider.

4. For a detailed understanding of all the data resources, and the measures and outcomes reported in them, this document should be read alongside the following documents:

    - 'Technical algorithms for institutional performance measures: Access and participation indicators, methodology and rebuild descriptions', which describes the indicators, including their coverage and presentation, and how they can be rebuilt from individualised student data[2]

    - 'Technical algorithms for institutional performance measures: Core algorithms', which includes further technical detail of the categorisations applied to individualised student data.[3]

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

[2] Available at www.officeforstudents.org.uk/data-and-analysis/institutional-performance-measures/technical-documentation/.

[3] Available at www.officeforstudents.org.uk/data-and-analysis/institutional-performance-measures/technical-documentation/.

Key terms

5. The following key terms should be borne in mind when reading this guidance:

- **Access and participation data resources:** A package of information including CSV and Excel files that contain the data underpinning the interactive data dashboard in a tabular format. It also contains the technical documentation defining the measures and outcomes reported (including this document and those detailed in paragraph 4).
- **Attribute:** A subset of a characteristic (for example, 'black' is an attribute within the 'ethnicity' characteristic).
- **Characteristic:** A categorisation of students (for example, 'ethnicity'). This also applies to combined categories, such as 'white and POLAR4 quintile 1 or 2').
- **Confidence interval:** A statistical term describing a range of values within which we are confident that the true value lies.[4]
- **English IMD:** Index of Multiple Deprivation, as calculated in England only. It is calculated from a set of measures which classifies areas in England by their level of deprivation. It is presented here as five quintiles, where quintile 1 contains the most deprived 20 per cent of the English population, and quintile 5 the least deprived 20 per cent.
- **Gap:** The difference between two indicators, in percentage points (for example, the gap between a continuation rate of 98 per cent for non-disabled students and 95 per cent for disabled students would be 3 percentage points).
- **KPM:** Key performance measure. The OfS has set various KPMs relating to access and participation measures:
  - **KPMs 1 and 2:** Gap in participation between most and least represented groups (participation rates of students from POLAR4 quintile 1 areas compared with quintile 5 areas)
  - **KPM 3:** Gap in continuation rates between most and least represented groups (students from POLAR4 quintile 1 areas compared with quintile 5 areas)
  - **KPM 4:** Gap in attainment between white students and black students
  - **KPM 5:** Gap in attainment between disabled students and non-disabled students.

[4] These figures are included to give an idea of the sensitivity of the indicator to small populations and the effects of random variation in the outcome being measured by the indicator. A large confidence interval typically indicates small numbers of students informing an indicator calculation, and it is advisable to be cautious when interpreting that indicator.

For example, if there are only 10 students in the population, and eight of them continued their studies, then just one student changing continuation outcome would make a 10 percentage point difference to the indicator. The confidence interval would reflect the magnitude of this effect.

For more details about KPMs see the OfS website.[5]

- **POLAR4:** Participation of Local Areas, an area-based measure, classifying areas according to the likelihood of young people living within them of participating in higher education. Quintile 1 contains the areas with the lowest 20 per cent of participation rates for the young population; quintile 5 the highest participating 20 per cent. For more details see the OfS website.[6]

- **Higher education sector:** The 'sector' includes all English higher education providers that submit data to the Higher Education Statistics Agency or to the Education and Skills Funding Agency, including both those that are registered with the Office for Students and those that are not.

- **Statistical significance:** If a gap is statistically significant, then it is unlikely to be a result of chance or random variation alone.

- **Student lifecycle:** The four stages of students' progression through higher education, as described in paragraph 6.

- **Numerator (of the indicator):** For the access indicators, this is the number of students with the attribute in question. For the other indicators, it is the number of students with the attribute in question who achieve the outcome being measured.

- **Denominator (of the indicator):** For the access indicators, this is the total number of entrants. For the other indicators, it is the total number of students in the population with the attribute in question.

- **Indicator:** Calculated (in percentage terms) as the numerator divided by the denominator. For the access indicators, this shows the proportion of the entrant cohort with the attribute in question. For the other indicators, it is the continuation, attainment or progression rate as appropriate.

- **Year:** Each indicator is reported across a five-year time series from years 1 to 5.

    - **Access and attainment indicators:** Year 1 corresponds to 2016-17, and Year 5 to 2020-21

    - **Full-time continuation indicators:** Year 1 corresponds to 2015-16, and Year 5 corresponds to 2019-20

    - **Part-time continuation indicators:** Year 1 corresponds to 2014-15, and Year 5 corresponds to 2018-19

[5] See 'Participation performance measures' (www.officeforstudents.org.uk/about/measures-of-our-success/participation-performance-measures/).

[6] See 'Young participation by area' (www.officeforstudents.org.uk/data-and-analysis/polar-participation-of-local-areas/).

    - **Progression indicators:** Year 1 corresponds to 2012-13, and Year 5 corresponds to 2016-17.

- **Aggregate**: In addition to the yearly time series, each indicator is also reported combining multiple years in the time series into one data point. This may allow data to be reported in aggregate where it cannot be reported as individual years due to small population sizes.

    - **Three-year aggregate**: reports the latest three years of the time series combined as one data point

    - **Five-year aggregate**: reports the latest five years of the time series combined as one data point.

- **Suppression rules:** In some cases, the following suppression rules have been applied:

    - **'N'** where there are fewer than 25 students in the population

    - **'N/A'** where the provider did not report any students in the population, or did not participate in the survey, or is otherwise not applicable

    - **'R'** where the provider participated in the Destination of Leavers from Higher Education (DLHE) survey but has not met the response rate threshold required (only applicable to progression lifecycle stage)

    - **'DP'** where the information has otherwise been suppressed for data protection reasons.

Overview of data resources

6. The dashboard provides measures of performance at each stage of the student lifecycle, for individual providers registered with the Office for Students:

    a. **Access indicators**, detailing the makeup of students entering higher education.

    b. **Continuation indicators**, reporting on the percentage of first year students who complete their course or remain in UK higher education 12 months (full-time students) or 24 months (part-time students) after starting their course.

    c. **Attainment indicators**, expressing the number of Level 6+ undergraduate leavers who were awarded 1st or 2:1class degrees, as a proportion of all those leavers from Level 6+ undergraduate courses who were awarded classified degrees.

    d. **Progression indicators**, reporting on the activity of students six months after leaving higher education, in terms of their progression to highly skilled employment and higher-level study. [7]

7. The data resources allow users to explore and understand patterns identified by these indicators for a range of student characteristics. They contain breakdowns of the various indicators described in paragraph 6, for a variety of demographic characteristics. Information is available for each stage of the student lifecycle, for individual providers and the whole sector, for each mode and level of study, and across a five-year time series, and for three- and five-year aggregates of the time series. We report numbers and proportions of students, and have included confidence intervals for the indicator values, to help users judge the potential sensitivity of an indicator to small numbers of students.

8. The data resources also compare students with different attributes within each characteristic (for example, within the characteristic of ethnicity, the continuation rate of black students compared with the continuation rate of white students). Resulting gaps between different groups are included, along with confidence intervals and statistical significance markers calculated for each year of the time series. Where a gap is marked as statistically significant, this indicates that the difference observed is unlikely to have been due to random variation alone.

9. The data resources include ‘Overall’ rates for the continuation, attainment and progression lifecycle stages. These detail their overall rate for each lifecycle stage across all student characteristics for each provider in the data. As a new addition, they now include ‘Overall’ denominators in the access lifecycle stage. This details the overall number of student entrants across all characteristics for each provider in the data.

[7] Also expressed as progression to ‘professional or managerial jobs’ and higher-level study.

Accessing the data resources

10. The data resources are available to download from the OfS website in several different formats.[8] The files and formats available are summarised in Table 1 below. These have been compressed in a zip file for download. The OfS provides a data definition file as part of the data resources. This contains descriptions of the variables along with the values that they can take and is included in each downloadable zip file.

Table 1: Summary of available data resource files

Filename	File format	Description
APP2022Mar_[*UKPRN*]_[*NAME*].csv APP2022Mar_SECTOR.csv	Comma delimited (CSV)	Individual files for each provider as well as one representing the aggregation to the English sector as a whole.
APP2022Mar_[*UKPRN*]_[*NAME*]_SIG.xlsx APP2022Mar_SECTOR_SIG.xlsx	Excel workbook	Individual files for each provider as well as one representing the aggregation to the English sector as a whole, detailing the statistically significant data.
APP2022Mar_ALL.csv	Comma delimited (CSV)	Data resources for all individual providers in the sector supplied in a single CSV file (for users of software packages capable of processing files containing large volumes of data).
APP2022Mar_Access.xlsx APP2022Mar_Continuation.xlsx APP2022Mar_Attainment.xlsx APP2022Mar_Progression.xlsx	Excel workbook	Data resource files for each lifecycle stage containing data resources for all individual providers in the sector. Within each workbook there are separate sheets containing data for each characteristic (these files can be useful if you want to look at data from more than one provider).

11. All data resources have the same variables (columns) and measures (rows) for each provider that has data to report.

[8] Available at www.officeforstudents.org.uk/data-and-analysis/access-and-participation-data-dashboard/guide-to-the-data-and-its-findings/get-the-dashboard-data/.

Navigating the data resources: Worked examples

12. The following worked examples have been provided to illustrate some techniques that you may wish to use when navigating a provider's data, as published in CSV format, and finding relevant measures. They are not intended as a guide to the measures that should be included in access and participation plans.

13. These examples are based on the sector file. The methods and concepts are equally applicable to individual provider files. For the purposes of this user guide, we assume that the data for a single provider at a time is being viewed in Microsoft Excel (version 2007 or later). In each case, begin by downloading the data resources for the sector and opening APP2022Mar_SECTOR.csv in Microsoft Excel. Highlight the header row, and choose **filter** from the **data** menu.

Worked example 1: Finding continuation rates for IMD quintile 1 students

14. Filter **LIFECYCLESTAGE** to 'Continuation', **SPLITTYPE** to 'EnglishIMDQuintile_2019' and **SPLIT1** to 'IMDQ1'.

15. This shows all the data around IMD quintile 1 continuation rates. To find only the continuation rates filter **MEASURETYPE** to 'INDICATOR'.

16. There are multiple continuation rates for different modes and levels of study. We can see that full-time undergraduate students had a continuation rate of 88.1 per cent during Year 5, compared with 82.0 per cent for part-time undergraduates with postgraduate components during the same year.

Worked example 2: Finding access measure for POLAR4 quintile 5 students

17. Filter **LIFECYCLESTAGE** to 'Access', **SPLITTYPE** to 'POLAR4Quintile' and **SPLIT1** to 'POLAR4Q5'.

18. Filter **MODE_OF_STUDY** to 'Full-time or apprenticeship' and **LEVEL_OF_STUDY** to 'First degree' to look at only full-time first-degree data.

19. Filtering **MEASURETYPE** to 'INDICATOR' will show the percentage of full-time first-degree students who come from POLAR4 quintile 5 areas.

20. Remove the filter on **MEASURETYPE** by selecting '(Select all)'. Filter **SPLIT2** to 'POLAR4Q1' and filter **MEASURETYPE** to 'GAP' to show the difference between the proportions of full-time first degree POLAR4 quintile 5 students and POLAR4 quintile 1 students in the sector in percentage points. We can see that the gap has decreased between Year 1 and Year 5.

21. Remove the filter on **MEASURETYPE** by selecting '(Select all)'. Remove the filter on **SPLIT2** by selecting '(Select all)'. Filter **SPLIT1** to 'POLAR4Q5_Population18'.

22. Filtering **MEASURETYPE** to ‘GAP’ will show the difference between the proportion of 18-year-olds from POLAR4 quintile 5 areas in the population and the proportion of 18-year-old full-time first degree students from POLAR4 quintile 5 areas in the sector, expressed in percentage points. Filtering **MEASURETYPE** to ‘RATIO’ will show the ratio between the proportion of 18-year-olds from POLAR4 quintile 5 areas in the population and the proportion of 18-year-old full-time first degree students from POLAR4 quintile 5 areas in the sector.

Worked example 3: Finding gaps in progression for young and mature students

23. Filter **LIFECYCLESTAGE** to ‘Progression’ and filter **SPLITTYPE** to ‘AgeOnCommencement’.

24. To look at all levels of full-time study filter **MODE_OF_STUDY** to ‘Full-time or apprenticeship’ and **LEVEL_OF_STUDY** to ‘All undergraduates’.

25. To find the gap between young and mature students, **SPLIT1** to ‘Young_Under21’, **SPLIT2** to ‘Mature_Age21andOver’ and filter **MEASURETYPE** to ‘GAP’. We can see that the gap for Year 1 is -7.3 percentage points and the gap for Year 5 is -3.4 percentage points. We can also see that the change between Year 1 and Year 5 is statistically significant.

Worked example 4: Finding significant gaps in a single year

26. Filter **LIFECYCLESTAGE** to ‘Continuation’. Filter **SPLITTYPE** to ‘POLAR4Quintile’. Filter **YEAR5** to ‘Yes’. Observe that **MEASURETYPE** shows ‘STATISTICAL_SIGNIFICANCE_OF_GAP’, as this is the only measure with a ‘Yes’/‘No’ flag.

27. There are significant gaps for a large number of POLAR4 splits. Filter **SPLIT1** to ‘POLAR4Q5’ and **SPLIT2** to ‘POLAR4Q1’.

28. This shows that the gaps are significant across most modes and levels of study. To look at full-time first degree only, filter **MODE_OF_STUDY** to ‘Full-time or apprenticeship’ and **LEVEL_OF_STUDY** to ‘First degree’.

29. Remove the filter on **YEAR5** by selecting ‘(Select all)’. You should then see five rows, each with a different **MEASURETYPE**. ‘GAP’ shows the values for the percentage point gap (3.2 per cent in **YEAR5**). Observe that the confidence interval is small (2.5 per cent to 4 per cent), as the population under consideration here (UK-domiciled young entrants across the whole sector with known POLAR4 quintiles) is relatively large.

30. If you wish to chart a time series, highlight the ‘GAP’ values from **YEAR1** to **YEAR5** and create a new chart in Excel. This illustrates that the participation gap between the most and least represented groups is widening over the period in question.

31. You can view all of the statistically significant data for the sector as a whole in a single table by opening the APP2022Mar_SECTOR_SIG.xlsx file (similar files are available for each provider individually).

Worked example 5: Finding significant changes between year 4 and year 5

32. Filter **LIFECYCLESTAGE** to ‘Progression’. Filter **SPLITTYPE** to ‘Sex’. Filter **SIGNIFICANCE_Y4TOY5CHANGE** to ‘Yes’. Observe that **MEASURETYPE** shows ‘GAP’, as this is the only measure that can have statistically significant changes from one year to another.

33. In this case, there are only significant changes in the gap between male and female progression rates for part-time study for other undergraduate level and all undergraduates.

34. The **CHANGE_YEAR4_TO_YEAR5** column shows -3.4 percentage points for all undergraduates; this can be confirmed by inspecting the values for the gap in the **YEAR4** and **YEAR5** columns. This shows that the gap between men and women progressing to highly skilled employment or higher-level study after graduation has reduced for all part-time undergraduates in the most recent year.

35. You can view all of the statistically significant data for the sector as a whole in a single table by opening the APP2022Mar_SECTOR_SIG.xlsx file (similar files are available for each provider individually).

Worked example 6: Finding data relevant to the OfS KPMs

36. Filter **LIFECYCLESTAGE** to ‘Attainment’. Filter **OFSKPM** to ‘Yes_KPM5’. Filter **SPLIT1** to ‘NoKnownDisability’. Filter **SPLIT2** to ‘Disabled’.

37. It is not immediately apparent which measures are significant or not. Hence, filter **MEASURETYPE** to ‘STATISTICAL_SIGNIFICANCE_OF_GAP’. The distribution of ‘Yes’ and ‘No’ in columns **YEAR1** to **YEAR5** shows which years and measures to focus on. In this case, almost all gaps are significant.

38. Filter **MODE_OF_STUDY** to ‘Full-time or apprenticeship’ and **LEVEL_OF_STUDY** to ‘All undergraduates’, and clear the filter from **MEASURETYPE**. We can now see the gaps, as well as the confidence intervals. We can see that the gap between disabled and non-disabled graduates achieving 1st or 2:1 degrees is statistically significant in each year, and seems to be reducing over the time series. In this case, the **SIGNIFICANCE_Y1TOY5CHANGE** column shows ‘Yes’. This means that it is unlikely that the change over this time is due to chance.