Associations between characteristics of students (ABCS) 2021

How do outcomes differ when accounting for multiple student characteristics?

Reference OfS 2021.46

Enquiries to Paula Duffin: official.statistics@officeforstudents.org.uk

Publication date 13 October 2021

Contents

Summary

1. Association between characteristics of students (ABCS) is a set of analyses that seeks to better understand how outcomes vary for groups of students holding different sets of characteristics. We define groups of students by looking at a set of characteristics so that we can determine the effect of not just one characteristic on an outcome, but the effect of multiple characteristics. This report is an update to the November 2020 publication Update to associations between characteristics of students.[1]

2. In this update, we have rolled forward the five years of data used in our statistical modelling. The modelling approach employed is the same as last time, but there have been some refinements to the methodology.

3. After consideration and improvements made since the previous two publications of ABCS, this latest update is no longer marked as experimental.

[1] See www.officeforstudents.org.uk/publications/update-to-associations-between-characteristics-of-students/

Introduction

4. Associations between characteristics of students (ABCS) is a set of analyses which aims to improve our understanding of the outcome different groups of people are likely to experience across the student lifecycle. As in the previous publication, it focuses on access to higher education for young people and continuation in higher education for students of all ages.

5. We have used statistical modelling to calculate modelled access and continuation rates for different student groups.[2] These student groups are defined by a combination of all the characteristics included in the model. We have then used these modelled access and continuation rates to split the various student groups into quintiles.

6. Those student groups with the lowest modelled rates will be in the lowest access or continuation quintiles and those with the highest will be in the highest access and continuation quintiles. As well as looking at the quintile for each complete student group (defined by combining all of the characteristics for each measure), the interactive dashboards allow users to examine how a particular characteristic or combination of characteristics is distributed across the quintiles.[3] For example, it is possible to see the proportion of those in student groups that include only females in each of the quintiles, or who are both Chinese and were eligible for free school meals. In this way, we clearly see how those who share a single, or even multiple, characteristics might not have the same experience when it comes to access to, or continuation in, higher education.

7. Since the ABCS quintiles are defined using modelled access and continuation rates, they group students by their likelihood of accessing or continuing in higher education based on the factors that are included in the statistical models (see the individual access, continuation for full-time students and continuation for part-time student sections for details of the characteristics included). In selecting the factors for use in these models we are looking for characteristics that should not influence the outcome in question, but where there is evidence that the outcomes for groups within these characteristics differ. For example, there is no reason why a student's ethnicity should have an impact on the likelihood of them continuing into the second year of their course. However, our analysis of continuation rates shows that black students have lower continuation rates than students from any other ethnic background.[4] Conversely, we know that the level 3 qualifications obtained by young people will have an impact on the likelihood of a young person entering higher education (for example, a student with three A-levels at A* is more likely to go into higher education than a student with three Es). However, prior attainment will not be included in the model because this is a justifiable – or valid – relationship.

8. As in the last update, there are three ABCS measures:

[2] These models are binary logistic regression models with two-level interactions. The interactions to include have been determined using stepwise regression. See Annex A for details.

[3] See www.officeforstudents.org.uk/data-and-analysis/associations-between-characteristics-of-students/

[4] See the continuation tab of the Access and Participation dashboard here: www.officeforstudents.org.uk/data-and-analysis/access-and-participation-data-dashboard/

a. ABCS access – this measure looks at the access rates of young people in England.

b. ABCS FT continuation – this measure looks at the continuation rates for full-time students.

c. ABCS PT continuation – this measure looks at the continuation rates for part-time students.

9. In this report, we describe these three ABCS measures and give examples of the findings for each of them. However, we encourage users to examine our interactive dashboards to explore the different combinations of characteristics for these three measures. We also set out the minor changes that we have made to the methodology.

ABCS access

10. For ABCS access, ‘access’ is defined as participation in higher education at age 18 or 19. The measure uses data from the Department for Education’s National Pupil Database (NPD) from the year in which pupils obtain their key stage four (KS4) qualifications – most commonly GCSEs.[5] The data is restricted to only those attending English state-funded mainstream schools. We then look for these individuals in higher education records two or three years later – those with a matching record have accessed higher education. This can be participation at any level of higher education and any mode of study. The only restriction placed on the higher education instance is that it must be two or three years after the completion of KS4.

11. Aside from the differences described in the ‘Changes to the methodology’ section, the other change that we have made to the ABCS access measure is that the five-year time series has been rolled forward a year. Therefore, the data included now is for pupils who received their KS4 qualifications from summer 2012 to the summer of 2016 and commenced higher education from 2015-16 to 2019-20.

12. As in the previous update, the access measure still looks at six characteristics, although the participation of local areas (POLAR4) characteristic has now been replaced by TUNDRA (tracking underrepresentation by area).[6] This is an area-based measure that uses tracking of state-funded mainstream school pupils in England to calculate young participation. For this ABCS measure, TUNDRA, like IMD and IDACI, is based on the individual’s home postcode recorded at KS4. The six characteristics are ethnicity, free school meal (FSM) eligibility, gender, income deprivation affecting children index (IDACI), index of multiple deprivation (IMD) and TUNDRA.[7]

13. We have used a statistical model to calculate predicted access rates for each of the student groups. We have included all the characteristics as main effects and used a statistical approach (stepwise) to determine which of the two-way interactions should be included. Details of the model, the interactions included, and the model results can be found in Annex C.

14. We have then used the predicted access rates to split the students included in the modelling into five quintiles. We have calculated the quintiles in such a way that no student groups will be split across quintiles, and if there are student groups with the same predicted access rates, these will also not be split across quintiles. This means that each quintile does not contain exactly 20 per cent of the population. Table 1 shows the number and proportion of students in each quintile, as well as the mean, minimum and maximum predicted access rate.

[5] The DfE does not accept responsibility for any inferences or conclusions derived from the NPD data by third parties

[6] See www.officeforstudents.org.uk/data-and-analysis/young-participation-by-area/about-tundra/

[7] See Annex B for detailed definitions.

Table 1: description of ABCS access quintiles 2021

Access quintile	Number of students	Proportion of students	Mean modelled access rate	Minimum modelled access rate	Maximum modelled access rate
Quintile 1	551,590	20.1%	17.0%	0.1%	24.9%
Quintile 2	538,865	19.7%	31.3%	24.9%	36.8%
Quintile 3	549,830	20.1%	41.6%	36.8%	46.5%
Quintile 4	549,590	20.1%	51.8%	46.5%	58.4%
Quintile 5	549,075	20.1%	66.3%	58.4%	95.9%

15. These quintiles allow us to explore the groups of young people who are least likely (access quintile 1) and most likely (access quintile 5) to access higher education. Each student group, made up of a particular combination of categories of the six characteristics used in the statistical model, is assigned to a quintile. However, the interactive dashboard allows users to look not only at which quintile is assigned to each student group, but what the distribution is across the quintiles when you select only one characteristic, or two – all the way up to the complete set of six.[8] There are only two groups of students whose access quintile is defined entirely by a single characteristic. The first is those whose ethnicity is Gypsy, Roma or Traveller – 100 per cent of whom are in access quintile 1 (those least likely to access higher education). This means that their ethnicity is having a very strong impact on the likelihood of them accessing higher education, regardless of their other characteristics. In contrast, 100 per cent of those whose ethnicity is Chinese are in access quintile 5.

16. For all other student groups, combining more than one characteristic shows how considering multiple characteristics tells us more about the likelihood of a student group accessing higher education.

17. Here, we highlight three examples:

    a. When looking at all student groups where the ethnicity is black Caribbean, 20 per cent are found in access quintile 5. When you restrict this to black Caribbean ethnicity and only those who were eligible for FSM, none are found in access quintile 5. This shows that while some black Caribbean groups are highly likely to enter higher education, this is not the case for those who were also eligible for FSM.

    b. When looking at student groups where the ethnicity is black Caribbean and gender is female, 39 per cent are in access quintile 5; however, for black Caribbean male groups, none are found in access quintile 5. This indicates that there is a difference in the likelihood of accessing higher education for black Caribbean males and females.

    c. When looking at student groups where IDACI is restricted to just quintile 1 areas, 42 per cent are found in access quintile 1. If we then look at those student groups where IDACI is restricted to quintile 1 and gender is restricted to female, this decreases to 32 per cent.

[8] See www.officeforstudents.org.uk/data-and-analysis/associations-between-characteristics-of-students/access-to-higher-education/

However, if we instead restrict to males, then this increases to 52 per cent. That is, student groups from IDACI quintile 1 areas are not likely to access higher education, and they are even less likely to do so if they are male.

Continuation for full-time students

18. ABCS FT continuation is a measure that assesses the likelihood of different student groups continuing in their full-time higher education studies. For full-time students, continuation measures the proportion of students who qualify, transfer or continue into a second year of study.[9] The measure looks at continuation for UK domiciled, full-time undergraduate students studying at English higher education providers.

19. Aside from the differences described in the 'Changes to the methodology' section, the other change that we have made to the measure is that the five-year time series has been rolled forward a year. Therefore, the data included now is for students starting their courses from 2014-15 to 2018-19.

20. As in the previous update, there are 12 characteristics in the measure: age, care experience, disability, ethnicity, FSM eligibility, IDACI, IMD, local or distance learner, national statistic – socio-economic classification (NS-SEC), parental higher education, TUNDRA and sex. TUNDRA has replaced POLAR4, in the same way as it has for the ABCS access measure. The three area-based characteristics: IDACI, IMD and TUNDRA, are all based on the student's home postcode recorded on entry to higher education. TUNDRA is most applicable to young students and so is included in the FT measure, where the majority of students starting their course under the age of 21 are found.

21. We have used a statistical model to calculate predicted continuation rates for each of the student groups. We have included all the characteristics as main effects and used a statistical approach (stepwise) to determine which of the two-way interactions should be included. Details of the model, the interactions included, and the model results can be found in Annex D.

22. We have then used these predicted continuation rates to split the students into quintiles. The creation of the quintiles ensures that no student groups are split across quintiles, and no student groups with the same predicted continuation rates are split across quintiles. This means that each quintile does not contain exactly the same number of students, although it is the case that each quintile does round to having 20.0 per cent of students due to the very small differences in the numbers in each quintile. Table 2 shows the number and proportion of students in each quintile, as well as the mean, minimum and maximum predicted continuation rate in each quintile.

Table 2: description of ABCS FT continuation quintiles 2021

FT continuation quintile	Number of students	Proportion of students	Mean modelled FT continuation rate	Minimum modelled FT continuation rate	Maximum modelled FT continuation rate
Quintile 1	415,890	20.0%	80.1%	1.1%	85.8%
Quintile 2	415,885	20.0%	88.1%	85.8%	90.1%
Quintile 3	415,905	20.0%	91.7%	90.1%	93.1%

[9] For details of how we calculate full-time continuation rates, see the OfS access and participation data methodology and rebuild instructions available from: www.officeforstudents.org.uk/data-and-analysis/institutional-performance-measures/technical-documentation/

Quintile 4	415,875	20.0%	94.4%	93.1%	95.5%
Quintile 5	415,890	20.0%	96.6%	95.5%	100.0%

23. Similarly to ABCS access, these quintiles allow us to explore the groups of students who are least likely (FT continuation quintile 1) and most likely (FT continuation quintile 5) to continue into the second year of their full-time undergraduate courses. Each student group, made up of a particular combination of the categories within the 12 characteristics used in the statistical model, is assigned to a quintile. However, the interactive dashboard allows users to look at which quintile is assigned to each student group, and the distribution across the quintiles when you select only one characteristic, or two – all the way up to the complete set of 12.[10] Unlike ABCS access, there are no groups of students whose FT continuation quintile is defined entirely by a single characteristic.

24. Combining more than one characteristic shows how considering multiple characteristics tells us more about the likelihood of a student group continuing in higher education, as demonstrated in the following examples:

    a. When looking at student groups where care status is restricted to only those who are care experienced, 46 per cent are in FT continuation quintile 1 (least likely to continue). However, for student groups restricted to those who are both care experienced and ethnicity is restricted to Chinese, only 1 per cent are in FT continuation quintile 1 and 66 per cent are in FT continuation quintile 5 (most likely to continue). This shows that the likelihood of continuing is higher for those who are care experienced and Chinese than the average for all those who are care experienced.

    b. When looking at student groups where age is restricted to 21-25 at the start of their course, 45 per cent are in FT continuation quintile 1. If we then look at student groups where age is restricted to 21-25 and local or distance learner is restricted to those who were neither distance nor local learners (those who are studying at a provider and away from home), this decreases to 31 per cent. This reduces again to only 21 per cent once we restrict the student groups further to include only female students. This shows that the likelihood of 21-25-year-olds continuing their studies is higher for those who were neither local or distance learners, and is higher again for 21-25-year-olds who are neither local or distance learners and are female.

[10] See www.officeforstudents.org.uk/data-and-analysis/associations-between-characteristics-of-students/continuing-in-full-time-higher-education/

Continuation for part-time students

25. ABCS PT continuation is a measure that assesses the likelihood of different student groups continuing in their part-time higher education studies. Continuation is measured differently for students studying part-time than for those studying full-time: those who have qualified, transferred or continue into the third year of their studies are said to have continued.[11]

26. Because of this difference in definition, the most recent available year of data for part-time continuation is for students who started their course in 2017-18, and so the five-year time series we use is from 2013-14 to 2017-18. The measure looks at UK-domiciled students on part-time undergraduate courses at higher education providers in England.

27. As in the previous update, the measure looks at seven characteristics: age, disability, ethnicity, IDACI, IMD, local or distance learners and sex.

28. As for the other measures, we used a statistical model to calculate predicted continuation rates for each of the student groups. We included all the characteristics as main effects and used an automated approach (stepwise) to determine which of the two-way interactions should be included. Details of the model, the interactions included and the model results can be found in Annex E.

29. We have then used these predicted continuation rates to split the students into quintiles. The creation of the quintiles ensures that no student groups are split across quintiles, and no student groups with the same predicted continuation rates are split across quintiles. This means that each quintile does not contain exactly 20 per cent of the population. Table 3 shows the number and proportion of students in each quintile, as well as the mean, minimum and maximum predicted continuation rate in each quintile.

Table 3: description of ABCS PT continuation quintiles 2021

PT continuation quintile	Number of students	Proportion of students	Mean modelled PT continuation rate	Minimum modelled PT continuation rate	Maximum modelled PT continuation rate
Quintile 1	87,365	20.0%	51.6%	0.0%	55.6%
Quintile 2	86,805	19.8%	58.8%	55.6%	61.8%
Quintile 3	89,095	20.4%	64.5%	61.8%	68.5%
Quintile 4	87,175	19.9%	70.6%	68.5%	73.9%
Quintile 5	86,985	20.0%	80.1%	73.9%	100.0%

30. As for the other two ABCS measures, these quintiles allow us to explore the groups of students who are least likely (PT continuation quintile 1) and most likely (PT continuation quintile 5) to continue into the third year of their part-time undergraduate courses. Each student group,

[11] For details of how we calculate part-time continuation rates, see the OfS access and participation data methodology and rebuild instructions available from: www.officeforstudents.org.uk/data-and-analysis/institutional-performance-measures/technical-documentation/

made up of a particular combination of the seven characteristics used in the statistical model, is assigned to a quintile. However, the interactive dashboard allows users to look at which quintile is assigned to each student group, and the distribution across the quintiles when you select only one characteristic, or two – all the way up to the complete set of seven.[12] Like ABCS FT continuation, there are no groups of students whose PT continuation quintile is defined entirely by a single characteristic.

31. Combining more than one characteristic shows how considering multiple characteristics tells us more about the likelihood of a student group continuing in higher education.

32. For example, when looking:

    a. At student groups where sex is restricted to female, only 18 per cent are found in PT continuation quintile 1, but when we look at student groups where the sex is restricted to female and disability is restricted to mental health condition, this increases to 63 per cent – indicating that female students who have reported a mental health condition are much less likely to continue their studies. The same is true when we restrict to female students of mixed – white and black Caribbean ethnicity: the percentage in PT continuation quintile 1 is 48 per cent.

    b. At student groups where IMD is restricted to quintile 1, 31 per cent are in PT continuation quintile 1, while 16 per cent are in PT continuation quintile 5. When we look at only males from IMD quintile 1 areas, the proportion in PT continuation quintile 1 increases to 44 per cent, but the proportion in PT continuation quintile 5 also increases to 28 per cent. This highlights the complexity of understanding how different student groups' continuation rates vary for some in IMD quintile 1 depending on other characteristics. Being male increases their risk of not continuing for some student groups who are also from IMD quintile 1 areas, but for other males from IMD quintile 1 areas, it decreases the risk. We can add other characteristics to explore which ones are having an impact on likelihood of continuing. For example, if we further narrow down the group to only those from IMD quintile 1 areas who are male and are distance learners, 99 per cent are in PT continuation quintile 1.

[12] See www.officeforstudents.org.uk/data-and-analysis/associations-between-characteristics-of-students/continuing-in-part-time-higher-education/

Changes to the methodology

Common changes

33. In the ABCS access and FT continuation models where previously POLAR4 was one of the characteristics used, this has now been replaced by TUNDRA.[13] TUNDRA is a different measure to POLAR4 because it focuses on the participation rate of state-funded mainstream school pupils and only applies to England. It uses data-linking to track cohorts of 16-year-olds who completed their KS4 in the summer of 2012 to 2016 and matches them to higher education records. These cohorts therefore align with the five-year time series used for the access measure.

34. We have changed how we handle instances where data for a characteristic is missing, unknown or not applicable, such as a TUNDRA quintile for a student over 21 years of age, or an IMD or TUNDRA quintile for one domiciled outside of England. Previously, we examined the access and continuation rates of these groups and put them in the groups with the most similar access or continuation rate. Now, we include them as their own 'Unknown or N/A' group in the modelling. As such, students with unknown IMD, IDACI or TUNDRA quintiles are now grouped this way.[14]

35. The same approach has been taken with some groups that previously were too small to be treated as their own group in the modelling: where possible, we have included them as their own distinct group.

36. All three ABCS measures look at the characteristics IDACI and IMD. In the previous publication these were based on the 2015 versions. These have now been updated and are based on the 2019 versions.[15]

37. Definitions of all the characteristics used in the three ABCS measures can be found in Annex B.

Changes specific to access

38. The base population used for the access model has been refined to exclude entrants who left their higher education course within 14 days of starting.

39. Students who attended independent schools are not included in the access measure because not all of the same characteristic data is collected for them as for those who attended state-funded mainstream schools. However, analysis of the original ABCS access measure showed that the access rate for these pupils is so high that they would automatically be placed in access quintile 5 had they been included. Now we have rolled forward our five-year time series

[13] See www.officeforstudents.org.uk/data-and-analysis/young-participation-by-area/about-tundra/

[14] Note these groups are omitted from the interactive dashboards, since understanding the behaviour of these groups is not very informative without knowing anything about them. However, they are included in the data downloads for transparency.

[15] See www.gov.uk/government/statistics/english-indices-of-deprivation-2019

we have checked this is still correct: across the five most recent cohorts, access rates for these students range from 71.8 per cent to 72.2 per cent, which according to table 1 puts them in access quintile 5.

Changes specific to FT continuation

40. As described in paragraph 35, there are now sufficient students with sex 'other' to be kept as their own group in the modelling. Conversely, students with unknown age have been placed into the '51+' age group, as their FT continuation rate is closest to that of this group and there are not enough of these students to warrant their own group.

41. In the previous update it was shown that for continuation of full-time students, there was sufficient data, and evidence of difference in continuation, for the following four additional characteristics to be included in the model: FSM eligibility, parental higher education, national statistic – socio-economic classification (NS-SEC) and care experience. These characteristics have again been used in the FT continuation model.

42. The application of our Data Quality Framework as part of the Differences in student outcomes: further characteristics work means that FSM eligibility is now restricted to those students 21 years and under on entry to higher education.[16] Although this data is not complete for all ages across the five years in our time series, we have chosen to maximise the amount of data used in the model and therefore use the full five years. This results in FSM eligibility being unknown for 37.8 per cent of students in our population in 2014-15, reducing slightly to 33.6 per cent in 2018-19.

43. The Data Quality Framework has also shown that the data for care experience is only of a sufficient quality for use from 2014-15 onwards and NS-SEC for 2015-16 onwards. As we have now rolled forward one year and our starting year is 2014-15, this will have affected the number of unknowns. Parental higher education data is shown to be useable for entrants from 2013-14 onwards, therefore there are no limitations on years of use for FT continuation.

Changes specific to PT continuation

44. As described in paragraph 34, students with unknown age are now placed in their own 'Unknown or N/A' group, as there are sufficient numbers of these students to warrant this.

45. In the previous report, we explored the possible use of the four additional characteristics in PT continuation but concluded that for reasons of insufficient data they could not be used. Now that we have rolled forward a year, we have again carried out analysis to determine if these characteristics could be used for PT continuation in this update. The proportions of missing data across the five-year time series for each of the characteristics are shown in Table 4, where they are compared with the proportions of missing data for FT continuation.

[16] See Annex A: Data quality framework – a method for assessing the quality of student characteristic data. Available from www.officeforstudents.org.uk/publications/differences-in-student-outcomes-further-characteristics/

Table 4: proportions of missing data in the PT and FT continuation models

Characteristic	Proportion of students with missing data for PT continuation	Proportion of students with missing data for FT continuation
Care experience	100%	30.3%
FSM eligibility	91.2%	33.5%
NS-SEC	100%	44.8%
Parental higher education	100%	29.4%

46. As described in paragraph 42, the application of our Data Quality Framework means that FSM eligibility is now restricted to those students 21 years and under on entry to higher education.[17] As part-time study tends to attract older students, this explains the high percentage of missing data and therefore rules out the inclusion of FSM eligibility data for PT continuation.

47. The use of the three remaining characteristics is also not possible since it is generally not collected for part-time students.

[17] See Annex A: Data quality framework – a method for assessing the quality of student characteristic data. Available from www.officeforstudents.org.uk/publications/differences-in-student-outcomes-further-characteristics/