Medr

Y Comisiwn Addysg Drydyddol ac Ymchwil
Commission for Tertiary Education and Research

Mae'r ddogfen hon hefyd ar gael yn y Gymraeg | This document is also available in Welsh

Medr statistics

Welsh Higher Education Initial Participation measure: 2016/17 to 2022/23

Statistics reference: Sta/Medr/05/2025

Date: 27 February 2025

Designation: Official statistics in development

Email: hestats@medr.cymru

Summary: This publication presents the methodology and results for a Higher Education Initial Participation (HEIP) measure for Wales. This measure estimates the probability that a Welsh domiciled person will participate in higher education by the time they are 30 years old**.** This includes the breakdown of initial participation by age and the differences between males and females.

As this is the first time Medr are publishing the HEIP measure these statistics have been labelled as Official Statistics in Development while we develop the measure further to meet users' needs. To help with this, any feedback on the methodology or contents of this output would be welcomed. To provide any feedback please contact us at hestats@medr.cymru.

Contents

Introduction

1. This report presents an estimate of initial participation in higher education (HE) for the 17 to 30 year old Welsh population for the academic years from 2016/17 until 2022/23.

2. The Higher Education Initial Participation (HEIP) measure is an estimate of the probability that a Welsh domiciled person will participate in HE by the age of 30. The analysis also looks at the difference in the HEIP between males and females. A full explanation of the methodology and data sources is in the methodology section.

3. The statistics in this publication are Official Statistics in Development as we are developing this measure and recognise that there are limitations to the methodology used. By publishing this information as Official Statistics in Development users can be involved in the development of these statistics and to contribute in making them as useful and relevant as possible.

4. We would welcome any feedback on the content of this publication whether relating to the methodology or what information could be included to make this useful for you. To provide feedback please email us at hestats@medr.cymru.

Why we are publishing these statistics

5. One of the strategic duties of Medr as set out in the Tertiary Education and Research (Wales) Act 2022 is to "encourage individuals who are ordinarily resident in Wales, particularly those who have additional learning needs, to participate in tertiary education." The intention in publishing this measure is to provide some evidence about participation in HE, feeding into the overall information for participation in the wider tertiary education sector.

6. Another strategic duty of Medr is to "promote increased participation in Welsh tertiary education by persons who are members of under-represented groups". As well as an overall HEIP measure for Wales, this publication also includes a split by sex to compare the initial participation in HE of males and females. While this is the only personal characteristic included here, part of the development of this measure will be to investigate if other characteristics could be included to provide a greater insight into the differences in participation in HE from different groups of the Welsh population.

7. Participation in tertiary education has been an area of increasing focus within the wider Welsh policy environment. In 2022, the Welsh Government commissioned the Welsh Centre for Public Policy to undertake a review evidence and best practice on inequity in tertiary education in Wales. The subsequent report was published in October 2024. In November 2024, the Senedd's Children, Young People and Education Committee commenced an inquiry into routes into post-16 education and training with a particular focus on participation. The inquiry is ongoing.

8. A measure of initial participation in HE for Wales has not been published since 2016 when the Higher Education Funding Council for Wales (HEFCW) published statistics for the 2012/13 academic year. During this time participation measures for England, Scotland and Northern Ireland have still been produced meaning that

there was a gap in evidence for Wales. However methodological differences limit how comparable the different measures across the UK are. Information on what is published in the rest of the UK is included in a later section.

Methodology

9. The HEIP measure is the sum of the initial participation rates for each age from 17 to 30 inclusive. The initial participation rates are the proportion of each age group that is participating in HE for the first time and to calculate this we need two pieces of information. The first piece of information is the number of students of each age who are participating in HE for the first time and the second is the overall population of that age in Wales.

Step 1: Estimating the number of students initially participating in HE

10. We use three data sources to estimate the number of students of each age who are participating in HE for the first time. These sources are the Higher Education Statistics Agency (HESA) Student Record, the HESA Student Alternative record (for the years 2014/15 to 2021/22) and the Lifelong Learning Wales Record (LLWR) from 2016/17 onwards collected by the Welsh Government.

11. For the HESA records we have linked data from 2004/05 to 2022/23 to identify when a person appears multiple times in the data. Details of this linking are provided in Annex A. With the records linked we find the earliest record for a student where they studied, or were still expected to study, for at least 6 months to ensure they have a considerable engagement in HE. We also check if they have previously obtained a HE level qualification and exclude those that have as they will have previously participated in HE.

12. For the LLWR data, we identify the first academic year where a student has either a learning programme or learning activity that is at the equivalent of Level 4 or above in the Credit and Qualifications Framework for Wales (CQFW). As with the HESA data we require that the relevant programme or activity lasts, or is expected to last, at least 6 months. Students who had a HE level qualification on entry are also excluded again.

13. The identification of students' initial participation is performed separately for the HESA and LLWR data so the number of initial participants in each are combined to give the total number of initial participants in each academic year. The initial participants are divided into their age as at 31 August at the beginning of the academic year, e.g. for the 2022/23 academic year the students ages are calculated as at 31 August 2022.

Step 2: Estimating the overall population

14. Two data sources are used to estimate the population of Wales. These are the 2021 Census out-of-term population estimate and the mid-year population estimates for Wales from 2016 to 2022, both produced by the Office for National Statistics (ONS).

15. The base for the population estimate is the 2021 Census out-of-term population estimates. The out-of-term population is produced by the ONS as part of their census outputs, and is the usual resident population but with full-time schoolchildren

and students counted at their out-of-term address. This population has been used as the base rather than using the mid-year estimates directly as we want to count students where they usually live rather than where they are studying.

16. The out-of-term population is based on the 2021 Census day of 21 March 2021 so an adjustment is made to age the population to 31 August 2021 to match the date used in the student data. For example, the number of 18 year olds is estimated as being a proportion of 17 year olds who have turned 18 since 21 March and the proportion of 18 year olds who have not yet turned 19 since 21 March.

17. A similar adjustment is made to the mid-year population estimates to age these populations from 30 June, the date of the mid-year estimates, to 31 August. We then calculate the percentage change between each adjusted mid-year estimate and the adjusted 2021 mid-year estimate for each age. These percentage changes are applied to the adjusted out-of-term population to produce at out of term-time population estimate as at 31 August for each year.

Step 3: Calculation of the initial participation rates and the HEIP measure

18. For each age from 17 to 30 years old we calculate the initial participation rate for that age by dividing the number of initial participants of that age from step 1 by the estimated population of that age from step 2.

19. The HEIP measure is calculated by summing the initial participation rates for each age. The idea behind this is that each individual participation rate represents the probability that someone of that age will participate in HE for the first time and by summing these you are building the probability that someone will participate in HE between the ages of 17 and 30 years old if these probabilities remain the same.

20. For clarity, the HEIP is not the same as dividing the total number of initial participants aged 17 to 30 in an academic year by the overall Welsh population of those ages. This would produce a much lower figure and would assume that someone is equally likely to be an initial participant in HE at any age, which is not the case.

Limitations

21. There are several limitations to note regarding the calculation of the HEIP used in this report.

    a) Initial participation in HE through studies that are not collected in the HESA records or LLWR are not be included in this measure. This would include any Welsh domiciled students studying at HE level at further education colleges in England, Scotland and Northern Ireland, some independent HE institutions in the UK or at higher education providers outside of the UK.

    If someone obtained a HE level qualification via the above routes then any further HE studies that were recorded in the HESA or LLWR data would also not be included in the measure as they would be excluded due having a HE level qualification on entry.

    This issue could be reduced by obtaining additional data sources that cover these other options for HE level studies.

b) As the HESA and LLWR data are not linked together, it would be possible for someone to appear as an initial participant in both if they participated but did not obtain a HE level qualification. For example someone could appear in the HESA data but drop out after one year without obtaining any qualifications. They could then appear in the LLWR data and still be considered an initial participant. This issue would be reduced by linking the datasets prior to looking for initial participants.

c) The measure assumes that the initial participation rates for each age will continue, however it doesn't account for differences in participation between cohorts. For example, the levels of participation of the cohort of 18 year olds in 2022/23 when they reach the age of 30 may differ from what those who are 30 in 2022/23 for a variety of reasons including policy changes and the wider economic landscape.

d) While the population estimates used for the overall populations are all accredited official statistics, there have been a number of assumptions made to adjust these to fit the purposes of this measure.

   The first adjustment is to age the estimates to 31 August so the age is comparable to the age used from the student data and the age is relevant to the academic years. However, this adjustment uses the assumption that birth dates are equally distributed which is not the case.

   The second adjustment is to 'grow' the adjusted out of term-time population to create a time series based on the percentage changes seen in the adjusted mid-year estimates. This assumes that the mid-year population and the out-of-term population change at the same rate.

e) Domicile is not static. This means we are not following a specific group of people and estimating how many of them participate in HE. Instead the population that is being considered is always changing and the population is affected by inward and outward migration.

   For example, someone could live in Wales until they are 24 before moving to England, if this person then participated in HE for the first time at the age of 25 they would not be included in the calculation as they would be English domiciled at that point. Conversely, someone living in England before moving to Wales and then participating in HE for the first time would be included.

f) As this methodology does not follow specific cohorts of people, it is difficult to produce reliable figures on more detailed characteristics. This is particularly difficult if characteristics change over time, for example whether someone lives in a more deprived area, or it is difficult to get accurate population estimates.

g) In the absence of a universal identifier to link records, algorithms are used to link the HESA student records and this will mean some incorrect links are made, or real links may be missed. In the case of incorrect links being made, then an individual's initial participation could be discounted as we will believe they have participated in HE previously. In the case of a real link being missed then an individual could be counted as an initial participant twice, although this should be minimised by excluding those who have a prior HE level qualification recorded in the data.

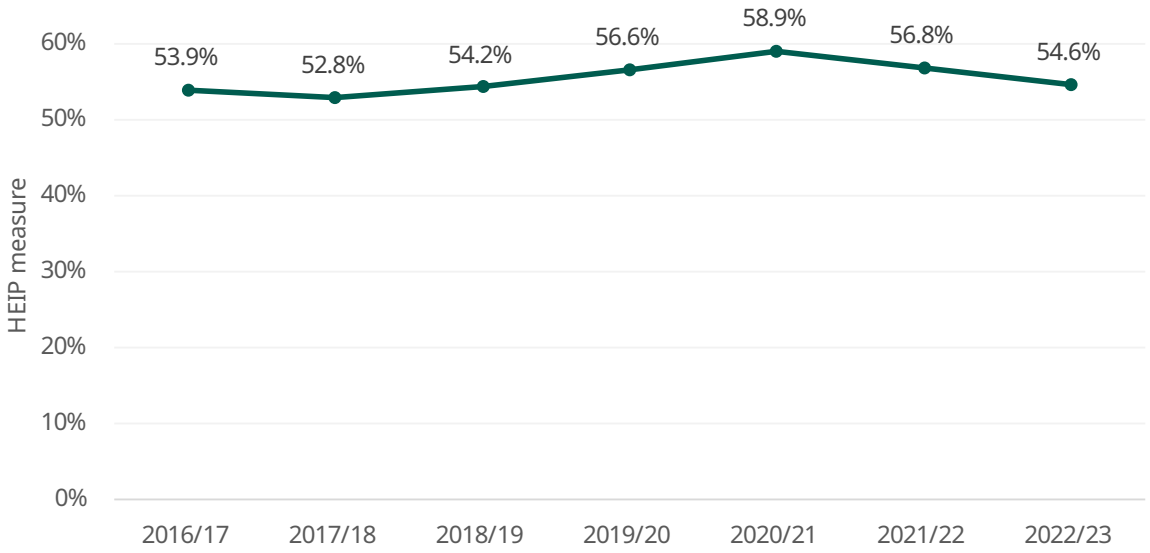
   These incorrect or missed links can occur due to data quality issues, such as incorrect information being recorded or digits being swapped in dates of birth. They can also occur when someone's data is correct but varies over time, for

example using different variations of their name or if someone changes their name.

Results

22. The HEIP measure is the sum of the initial participation rates for each age from 17 to 30 years old in a given academic year. The HEIP measure is not the percentage of 17 to 30 year olds who are participating in HE in that particular year. Instead the HEIP measure is an estimate of the probability that a Welsh domiciled person will participate in HE by the time they are 30 based on the initial participation rates in that year.

Figure 1: Higher Education Initial Participation measure – 2016/17 to 2022/23



Description of Figure 1: A line chart showing the HEIP measure increasing from 2017/18 to 2020/21 before decreasing in the following two years.

23. The HEIP measure in 2022/23 was 54.6%. This means that the estimated probability of a Welsh domiciled person participating in HE by the age of 30 is 54.6% based on the initial participation rates for each age from 17 to 30 in 2022/23.

24. After a drop between 2016/17 and 2017/18, the HEIP measure increased every year from 2017/18 to 2020/21 reaching a high of 58.9%. From this peak in 2020/21 there has been a decline in the following two years down to the figure of 54.6% in 2022/23. The Covid-19 pandemic will have been a factor on the levels of participation in the most recent years.

By Age

Table 1: Initial entry percentages by age – 2016/17 to 2022/23

Age	2016/17	2017/18	2018/19	2019/20	2020/21	2021/22	2022/23
17	0.8%	0.4%	0.5%	0.6%	0.3%	0.2%	0.2%
18	27.5%	27.4%	27.0%	28.1%	28.4%	29.5%	29.6%
19	10.4%	10.1%	10.3%	10.7%	11.0%	10.2%	9.4%
20	3.4%	3.5%	3.5%	3.5%	3.8%	3.4%	3.3%
21	1.9%	1.9%	2.2%	2.1%	2.4%	2.2%	1.8%
22	1.4%	1.5%	1.6%	1.6%	2.0%	1.7%	1.6%
23	1.2%	1.1%	1.3%	1.5%	1.7%	1.5%	1.3%
24	1.2%	1.2%	1.3%	1.4%	1.6%	1.3%	1.2%
25	1.1%	1.1%	1.2%	1.3%	1.5%	1.3%	1.1%
26	1.0%	1.0%	1.2%	1.3%	1.4%	1.2%	1.0%
27	1.0%	1.0%	1.1%	1.3%	1.3%	1.2%	1.2%
28	1.0%	0.9%	1.1%	1.1%	1.3%	1.1%	1.0%
29	1.1%	0.9%	0.9%	1.1%	1.2%	1.1%	0.9%
30	0.9%	0.8%	1.0%	1.1%	1.2%	1.0%	1.0%
HEIP measure	**53.9%**	**52.8%**	**54.2%**	**56.6%**	**58.9%**	**56.8%**	**54.6%**

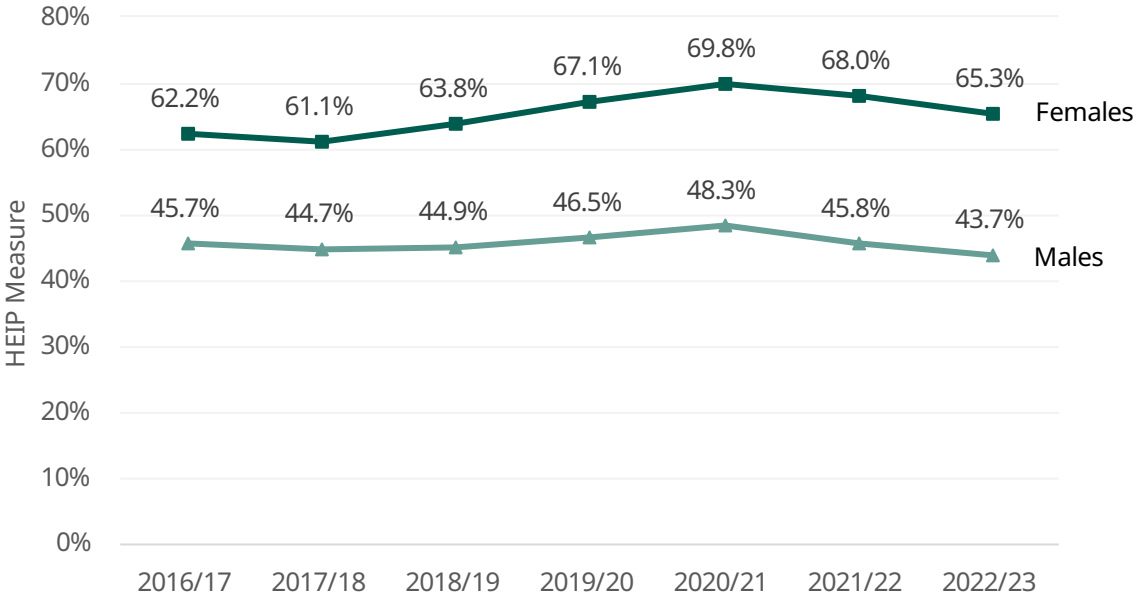
25. The largest contribution to the HEIP measure comes from 18 and 19 year olds. In 2022/23 the initial participation rates for these two ages contribute 38.9 percentage points to the overall HEIP measure of 54.6%.

26. The initial participation rate of 18 year olds has increased in every year since 2018/19.

27. For other ages the initial participation rates generally increased between 2017/18 to 2020/21, before falling in the following two years.

By Sex

Figure 2: Higher Education Initial Participation measure by sex – 2016/17 to 2022/23



Description of Figure 2: A line chart showing the HEIP measure for males and females. Both follow the same pattern as the overall HEIP measure but the figures are considerably higher for females than males.

28. As with the overall HEIP measure, the HEIP measure for males and females decreased between 2016/17 and 2017/18 before increasing every year until 2020/21. There was then a drop in each of the following two years.

29. The HEIP measure is considerably higher for females than for males, with the gap widening across the period. In 2016/17 there was a difference of 16.5 percentage points while in 2022/23 there was a 21.6 percentage point difference.

30. The HEIP measure for females peaked at 69.8% in 2020/21 compared to 48.3% for males in the same year. The HEIP has since dropped to 65.3% and 43.7% for females and males respectively in 2022/23.

By Age and Sex

Table 2: Initial entry percentages by age and sex – 2021/22 and 2022/23

Age	Females 2021/22	Females 2022/23	Males 2021/22	Males 2022/23
17	0.2%	0.3%	0.1%	0.2%
18	35.9%	35.1%	23.5%	24.0%
19	11.8%	11.4%	8.5%	7.4%
20	3.9%	3.7%	2.9%	2.8%
21	2.4%	2.1%	2.0%	1.5%
22	2.0%	1.8%	1.4%	1.3%
23	1.9%	1.6%	1.1%	1.0%
24	1.7%	1.5%	1.0%	0.9%
25	1.6%	1.4%	1.0%	0.8%
26	1.4%	1.4%	0.9%	0.7%
27	1.4%	1.4%	0.9%	1.1%
28	1.3%	1.3%	0.8%	0.7%
29	1.4%	1.1%	0.9%	0.7%
30	1.2%	1.2%	0.8%	0.8%
HEIP measure	**68.0%**	**65.3%**	**45.8%**	**43.7%**

31. Table 2 shows that the initial participation rates are higher for females than males at every age from 17 to 30 years old for 2021/22 and 2022/23. This is also the case when looking back to 2016/17, with the exception of 17 year olds between 2017/18 and 2019/20 where the rates were level.

32. The overall initial participation rate for 18 year olds increased between 2021/22 and 2022/23, however when looking at the measure by sex this was only the case for males. The initial participation rate for female 18 year olds dropped by 0.8 percentage points between 2021/22 and 2022/23 while there was a 0.5 percentage point increase for male 18 year olds.

Participation measures in the rest of the UK

33. There is no single measure of participation across the UK making difficult to make comparisons. This section covers the differences and similarities in other participation measures across the UK.

England

34. The Department for Education (DfE) have a statistical series called 'Participation measures in higher education'. The methodology for this series was similar to what has been used here up to the academic year 2019/20 release of the DfE statistics.

35. A new methodology called the Cohort-based Higher Education Participation (CHEP) measure was introduced for the 2020/21 academic year. Instead of estimating future participation by age 30 using current participation levels as the HEIP methodology does, the CHEP tracks cohorts of school pupils to measure participation.

36. While CHEP is quite different from the HEIP methodology, the 2021/22 release does have a section 'Projecting future HE participation' that uses the cohort data to produce a projection that is more similar to how the HEIP measure is constructed.

37. The rationale for changing methodology was that while the HEIP produced a timely measure there were some known limitations such as:

    - estimating a higher participation rate than the real rate for a particular entry cohort when there is steady growth in entry rates for younger age groups.
    - not being able to create reliable figures by region and key demographics.

38. DfE felt the CHEP methodology lessened the impact of inward and outward migration flows over time and that it would also not be affected by revisions to the ONS population estimates that occur following each Census.

39. The other benefit was that the CHEP approach would allow them to analyse participation by pupil characteristics taken from the school census such as breakdowns by gender and region of school attended.

40. One drawback of the new methodology is that it is less timely than the HEIP as it requires each 15-year-old school cohort to reach a particular age before reporting on it. In other words you would only report on the percentage participating in HE by the age of 25 for those aged 15 in the 2024/25 academic year, once the 2034/35 academic year data are available.

Scotland

41. The Scottish Funding Council (SFC) include a Higher Education Initial Participation Rate (HEIPR) in the background tables of their 'HE Students and Qualifiers at Scottish Institutions' statistical publication.

42. This is produced using a similar methodology to what has been presented for Wales in this publication. Although there will be differences in the exact methodology for how initial participation is identified. One difference is that it covers those aged 16 to 30 rather than 17 to 30.

43. One similarity to note is that the HEIPR for Scotland also reaches a peak in 2020/21. However unlike the HEIP for Wales, after falling in 2021/22 it then increased again in 2022/23.

Northern Ireland

44. The Northern Ireland Statistics and Research Agency (NISRA) have produced an 'Age Participation Index for Northern Ireland' for 1998/99 to 2021/22. This is the number of Northern Irish domiciled young entrants (aged under 21) to full-time Higher Education in the UK or Republic of Ireland as a percentage of the 18-year-old population in Northern Ireland.

Future developments

45. Any feedback received will help direct how the HEIP could be improved. Developments will be informed by the discussions we have with those with an interest in this area, but possible developments include:

- Extending the coverage of HE activity by obtaining data on Welsh domiciled initial participants studying at HE level in Further Education Providers in England, Scotland and Northern Ireland.
- Investigating whether it would be possible to robustly report on a wider range of characteristics. For example ethnicity, disability and living in more deprived areas.
- Investigate whether there are possibilities to produce initial participation rates using a cohort based methodology as the DfE do for England. When the work on this HEIP was started in the Higher Education Funding Council (HEFCW) a cohort methodology was not feasible due to the lack of availability of longitudinal data. However the establishment of Medr may provide new opportunities.
- Look at how the measure could be adapted for the wider tertiary education sector instead of only focusing on HE.

Quality information

Official Statistics in Development

The HEIP measure in this report is an official statistic in development. Official statistics in development may be new or existing statistics, and will be tested with users, in line with the standards of trustworthiness, quality, and value in the Code of Practice for Statistics. The methodology and data sources used to create this measure are still evolving and further developments are planned to improve the coverage of the measure and how useful it is.

We welcome and value any feedback on the methodology or contents of this output. Please contact us at hestats@medr.cymru.

Our statistical practice is regulated by the Office for Statistics Regulation (OSR).

OSR sets the standards of trustworthiness, quality and value in the Code of Practice for Statistics that all producers of official statistics should adhere to.

You are welcome to contact us directly with any comments about how we meet these standards.

Alternatively, you can contact OSR by emailing regulation@statistics.gov.uk or via the OSR website.

Compliance with the Code of Practice for Statistics

This section provides a summary of how these statistics comply with the three pillars of the Code of Practice for Statistics – trustworthiness, quality and value.

Trustworthiness

This has been produced by analysts complying with the Code of Practice for Statistics and following Medr's Statement of Compliance. Release dates are pre-announced and protocols around data confidentiality are followed.

Quality

The HESA data and LLWR data used to identify the number of initial participants both have rigorous quality assurance processes in place when collecting data to ensure it is of sufficient quality. These data sources are also used regularly for Official Statistics and Accredited Official Statistics produced by other organisations.

The linking methodology used to link different years of the HESA data was based on a methodology produced by the Office for Students, with the output being used in the HEIP for England produced by DfE until the change to a cohort-based method. A sample of the links were assessed by Medr to check that the links made were accurate.

The 2021 Census out of term-time population and the mid-year estimates used for the overall populations used in the calculation of initial participation rates are accredited

official statistics. However, adjustments are made to these and as acknowledged in the ‘Limitations’ section.

While the overall methodology has limitations it is similar to what is currently used by the Scottish Funding Council, and what was previously used by the DfE. Given the data availability at the time of production it was felt that this was the most suitable method, although there is recognition that it may need further development.

The similarity in methodology allows some comparability between this HEIP and what is produced by the SFC, and while the DfE have moved to a cohort-based method, their most recent publication still includes a limited comparison to their old method which provides a low level of comparability to this measure.

Value

The HEIP provides a measure of HE initial participation for Wales. This is something that has been missing for several years and could help with evidencing how participation in HE in Wales changes over time, and how it compares with the rest of the UK.

Annex A – Linking algorithm rules

This annex contains a summary of the conditions used within the algorithm to link HESA student records. Information about the variables mentioned can be found on the HESA website at www.hesa.ac.uk/collection/c21051 for the HESA Student record for 2021/22 and earlier and www.hesa.ac.uk/collection/22056 for data collected for 2022/23 onwards. Information for the HESA Student Alternative record can be found at www.hesa.ac.uk/collection/c21054/.

Within the code used to link records together there is also some standardisation of first names to account for nicknames and common variations.

Two student records are considered to be the same person if any of the following conditions are met:

Conditions A

Date of birth matches exactly and one of the following groups of conditions holds:

1 a) HUSID/SID matches – HUSID was used prior to 2022/23, and SID used since then. These are unique student identifier variables in the data.

b) UK Provider Reference Number (UKPRN) matches.

2 a) One of HUSID/SID, UCASPERID and ULN matches. UCASPERID and ULN are two more identifiers although not provided for all students, these are the UCAS personal identifier and the Unique Learner Number.

b) First name is sufficiently similar. Either:

- Same names but can be reordered.
- One is a substring of the other.
- SPEDIS score of under 40, where SPEDIS is a function that determines the likelihood of two words matching.
- When combined with surname it matches the data in surname at 16.
- First name and surname are in reverse order.

3 a) There is a postcode match, either:

- The term-time postcode in an earlier year matches the entry postcode of a later year.
- Term-time postcodes match exactly.
- Entry postcodes match exactly.

b) First name is sufficiently similar. See 2b) but SPEDIS score is reduced to 21 and first name and surname being reversed is not allowed.

c) Surname is sufficiently similar. Either:

- One is a substring of the other.
- SPEDIS score of under 21.
- SPEDIS score compared to surname at 16 is under 21.

4 a) One first name is a substring of the other.

b) Surnames match exactly, or match the surname at 16 exactly.

c) Entry postcodes are not blank and differ by one character at most.

Conditions B

HUSID/SID matches and one of the following groups of conditions hold:

1) Date of birth matches exactly.

2) First name matches exactly.

3) Surname matches exactly.

4 a) First name is similar – in this case either:

- One is a substring of the other.
- the same names re-ordered.
- a SPEDIS score of under 40.
- matching sname16 when combined with surname.

b) Surname is either a substring of the other, or a spedis score of under 40.

c) Entry postcode matches exactly.

Conditions C

A link is made if all of the following conditions are true:

- One first name is a substring of the other.
- Surname matches exactly.
- Entry postcode matches exactly.
- Date of birth only differs by one character.

Other links

Links are also made if they can be deduced logically from other links that have been made. For example if record X links to record Y, and Y links to Z. Then X and Z are linked even if they do not directly meet any of the above conditions.

Medr

Y Comisiwn Addysg Drydyddol ac Ymchwil
Commission for Tertiary Education and Research

www.medr.cymru

2 Cwr y Ddinas
Stryd Tyndall
Caerdydd
CF10 4BZ

2 Capital Quarter
Tyndall Street
Cardiff
CF10 4BZ

Noddir gan
Lywodraeth Cymru
Sponsored by
Welsh Government