Office for Students

2019-20 Transparency attainment data summary:

HESA Student, HESA Student Alternative and ILR data checking tool technical document and rebuild instructions

Enquiries to transparency@officeforstudents.org.uk

Publication date 1 September 2020

Contents

Summary

1. This document details the algorithms applied to Higher Education Statistics Agency (HESA) student record, HESA student alternative record (SAR) and Individualised Learner Record (ILR) data to determine the number (headcount) of UK-domiciled undergraduate qualifiers in 2019-20 at a provider by qualification classification achieved, mode of study, ethnicity, gender and English Index of Multiple Deprivation (EIMD) 2019. This data (the Transparency attainment data summary) forms part of a provider's transparency information.[1]

2. This document is aimed at readers with in-depth knowledge of the data. Readers are advised to have a copy of the HESA student record coding manual, HESA student alternative record coding manual and ILR specification for the relevant year as well as the Office for Students' 'Technical algorithms for institutional performance measures: 2020 core algorithms'[2] to hand when using this document.

3. The Transparency attainment data summary is formed of two files. These are:

    a. Transparency attainment data summary workbook
        'TR19_Attainment_DCT_XXXXXXXX.xlsx'

    b. Transparency attainment data summary individualised file
        'TR19_Attainment_DCT_XXXXXXXX_IND.csv'

    where XXXXXXXX is your provider's eight digit UK Provider Reference Number (UKPRN).

4. The data in the Transparency attainment data summary data checking tool outputs are provided as a data quality tool, to aid providers in getting their data correct and must not be used for any other purpose. The final HESA student, HESA student alternative and ILR R14 data will be used to generate a provider's final Transparency attainment data summary which forms part of a provider's transparency information.

Note on terminology

5. The algorithms below apply to three different data collections, each of which has a different structure. In the HESA student and HESA student alternative return, the fields that we use are mainly associated with a student instance. In the ILR return, the fields that we use are associated with a learner on a course; a learner may study multiple courses.

6. Below, we use the generic term 'record' to refer to either an instance, or a learner-course combination, depending on the data source.

---

[1] For details of the transparency information, see www.officeforstudents.org.uk/publications/regulatory-advice-8-guidance-for-providers-about-condition-of-registration-f1-transparency-information/
[2] Available at www.officeforstudents.org.uk/data-and-analysis/institutional-performance-measures/technical-documentation/

7. Fields referenced in this document prefixed with 'TR', 'OFS' or 'IP' are derived by the OfS, all others are taken directly from the respective student data collection.[3]

8. Where derived fields prefixed with 'IP' are referenced in this document, readers should consult the Office for Students' 'Technical algorithms for institutional performance measures: 2020 core algorithms'[4] for their definitions. Details are given where the definitions of these fields differ from the published algorithms.

Fields used in the Transparency attainment algorithms

Table 2: HESA student fields used in the Transparency attainment calculations

Entity	Field	Description
Institution	UKPRN*	UK Provider Reference Number
Student	HUSID*	HESA unique student identifier
Instance	NUMHUS*	Student instance identifier
Student	SEXID	Sex of student
Student	ETHNIC	Ethnicity of student
Entry profile	POSTCODE	Postcode of student's permanent address prior to entry
Course subject(s)	SBJCA (1-5)	Subject(s) appropriate to course
Instance	INITIATIVES (1-3)	Specific scheme identifier
Instance	PROGTYPE	Type of program student is undertaking
Course	REGBODY (1-2)	Regulatory body for medical, dental, health and social care, and veterinary students.
HESA derived field	XQMODE01	Qualification obtained mode of study
HESA derived field	XDOMHM01	Domicile of student
HESA derived field	XPQUAL01	Qualifications obtained population identifier
HESA derived field	XQOBTN01	Highest qualification obtained
HESA derived field	XCLASSF01	Classification of qualification

*These fields are not used directly in the algorithms but are included in the individualised file to allow easy identification of students

---

[3] For definitions of HESA fields labelled with 'HESA derived field' in table 2 and table 3, users should consult the 'HESA derived fields specification' available on the HESA website. For the HESA student record see www.hesa.ac.uk/collection/c19051/derived/contents. For the HESA student alternative record see www.hesa.ac.uk/collection/c19054/derived/contents.

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

Table 3: HESA student alternative fields used in the Transparency attainment calculations

Entity	Field	Description
Institution	UKPRN*	UK Provider Reference Number
Student	HUSID*	HESA unique student identifier
Instance	NUMHUS*	Student instance identifier
Student	SEXID	Sex of student
Student	ETHNIC	Ethnicity of student
Entry profile	POSTCODE	Postcode of student's permanent address prior to entry
Course subject(s)	SBJCA (1-5)	Subject(s) appropriate to course
Instance period	INITIATIVES (1-3)	Specific scheme identifier
Course	REGBODY (1-2)	Regulatory body for medical, dental, health and social care, and veterinary students.
HESA derived field	XQMODE01	Qualification obtained mode of study
HESA derived field	XDOMHM01	Domicile of student
HESA derived field	XPQUAL01	Qualifications obtained population identifier
HESA derived field	XQOBTN01	Highest qualification obtained
HESA derived field	XCLASSF01	Classification of qualification

*These fields are not used directly in the algorithms but are included in the individualised file to allow easy identification of students.

Table 4: ILR fields used in the Transparency attainment calculations

Field	Description
UKPRN*	UK Provider Reference Number
LEARNREFNUMBER*	Learner reference number
LEARNAIMREF*	Learning aim reference code
AIMSEQNUMBER*	Learning aim sequence number
POSTCODEPRIOR	Postcode of learner prior to enrolment
DOMICILE	Domicile of learner
ETHNICITY	Ethnicity of learner
SEX	Sex of learner
LEARNSTARTDATE	Learning start date
LEARNPLANENDDATE	Planned learning end date
MODESTUD	Mode of study
PROGTYPE	Programme type
OUTGRADE	Grade awarded to learner
OUTCOME	Indicates whether learner has achieved learning aim

*These fields are not used directly in the algorithms but are included in the individualised file to allow easy identification of students.

Description of derived fields

Table 5: Description of OfS derived fields for recreating Transparency attainment tables

Derived field name	Description	Paragraph
TRSOURCE	Identifies the source of the record	13
OFSQAIM	Recognised higher education qualification aim (ILR only)	16
TRCRSELGTH	Expected duration of study (ILR only)	17
TRMODE	Mode of study linked to qualification	18
TRDOM	Broad level domicile of student	19
TRETHNIC	Ethnicity of student	20
TRSEX	Sex of student	21
TREIMD	English Index of Multiple Deprivation (EIMD) 2019 quintile of student's address prior to entry	22
TRAWARDLEVELNUM	Numeric level of award according to the higher education qualification framework	23
TRDENT	Student has at least one instance of a 'Pre-clinical dentistry' or 'Clinical dentistry' associated with programme of study	25
TRAWARD_DETAIL	Detailed grouping of qualification awarded to student	26
TRAWARDLEVEL	Broad grouping of qualification awarded to student	29
TRAWARD	Broad grouping of qualification awarded to student relevant to the Transparency attainment population	31
TRDEGCLASS	Degree classification of award	32
TRATTAINPOP	Indicator for inclusion of student in the Transparency attainment population	33

Derived fields for recreating Transparency attainment tables

TRSOURCE

9. Identifies the source of the record.

Value	Definition
HESASTU	Record is taken from the HESA student record
HESASAR	Record is taken from the HESA alternative student record (previously known as the HESA alternative provider (AP) student record)
ILR	Record is taken from the ILR

10. Where an algorithm cannot be applied in the same way to each TRSOURCE, this will be indicated in the description of the algorithm.

OFSQAIM (ILR only)

11. This field classifies qualification aims into broad levels of higher education. For a full definition of this field please refer to '2019-20 ILR data checking tool: Classifying learning aims technical document'.[5]

TRCRSELGTH (ILR only)

12. This field contains the number of years that the qualification aim is expected to last. If a student's planned end date falls within the first 14 days of a second year of study, the number of years is taken as 1. Otherwise, it is rounded up to the nearest whole number. TRCRSELGTH is the difference in years between LEARNSTARTDATE and LEARNPLANENDDATE. For example, a course with LEARNSTARTDATE = 1 October 2014 and LEARNPLANENDDATE = 5 January 2016 will have TRCRSELGTH = 2. This field has the same definition as IPCRSELGTH.

TRMODE

13. This field classifies the qualification mode of study. This filed has the same definition as IPMODE.

TRSOURCE = HESASTU

Value	Description	Definition
APP	Apprenticeship	INITIATIVES1 = K or INITIATIVES2 = K or INITIATIVES3 = K or PROGTYPE in (20, 21, 22, 23, 25)
FT	Full-time	XQMODE01 = 1 and not above
PT	Part-time	XQMODE01 = 2 and not above
OTH	Otherwise	Otherwise

TRSOURCE = HESASAR

Value	Description	Definition
APP	Apprenticeship	INITIATIVES1 = K or INITIATIVES2 = K or INITIATIVES3 = K
FT	Full-time	XQMODE01 = 1 and not above
PT	Part-time	XQMODE01 = 2 and not above
OTH	Otherwise	Otherwise

---

[5] Available at www.officeforstudents.org.uk/data-and-analysis/data-checking-tools/2019-20-ilr-data-checking-tool/

TRSOURCE = ILR

Value	Description	Definition
APP	Apprenticeship	PROGTYPE in (20, 21, 22, 23, 25)
FT	Full-time	MODESTUD in (1, 2) or (MODESTUD in (99, BLANK) and (TRCRSELGTH = 1 or (TRCRSELGTH ≤ 2 and OFSQAIM in (HIGHER, FIRST, FOUDEG, DIPHE, HND)) or (TRCRSELGTH ≤ 3 and OFSQAIM in (HIGHER, FIRST)) or (TRCRSELGTH ≤ 4 and OFSQAIM = ENHANCED))) and not above
PT	Part-time	Otherwise

TRDOM

14. This field classifies if the student is domiciled in England, Scotland, Wales, Northern Ireland or elsewhere or is not known. This field has the same definition as IPDOM for values of E, S, W and N.

TRSOURCE = HESASTU, HESASAR

Value	Description	Definition
E	England	XDOMHM01 = 1
S	Scotland	XDOMHM01 = 2
W	Wales	XDOMHM01 = 3
N	Northern Ireland	XDOMHM01 = 4
NON-UK	Not domiciled in UK	Otherwise

TRSOURCE = ILR

Value	Description	Definition
E	England	DOMICILE = XF or (DOMICILE in (XJ, XK, GB) and (POSTCODEPRIOR is in England or ((POSTCODEPRIOR = BLANK or POSTCODEPRIOR begins ZZ) and

		Provider in England)))
S	Scotland	DOMICILE = XH or
		(DOMICILE in (XJ, XK, GB) and
		(POSTCODEPRIOR is in Scotland or
		((POSTCODEPRIOR = BLANK or
		POSTCODEPRIOR begins ZZ) and
		Provider in Scotland)))
W	Wales	DOMICILE = XI or
		(DOMICILE in (XJ, XK, GB) and
		(POSTCODEPRIOR is in Wales or
		((POSTCODEPRIOR = BLANK or
		POSTCODEPRIOR begins ZZ) and
		Provider in Wales)))
N	Northern Ireland	DOMICILE = XG or
		(DOMICILE in (XJ, XK, GB) and
		(POSTCODEPRIOR is in Northern Ireland or
		((POSTCODEPRIOR = BLANK or
		POSTCODEPRIOR begins ZZ) and
		Provider in Northern Ireland)))
UNKNOWN	Domicile not known	DOMICILE in (ZZ, BLANK)
NON-UK	Not domiciled in UK	Otherwise

TRETHNIC

15. This field classifies the ethnicity of the student. This field has the same definition as IPETHNIC.

TRSOURCE = HESASTU, HESASAR

Value	Description	Definition
A	Asian or Asian British	ETHNIC in (31, 32, 33, 34, 39)
B	Black, African, Caribbean or Black British	ETHNIC in (21, 22, 29)
M	Mixed or multiple ethnic background	ETHNIC in (41, 42, 43, 49)
O	Other ethnic background	ETHNIC in (14, 15, 50, 80)
W	White	ETHNIC in (10, 13, 19)
U	Refused or unknown	Otherwise

TRSOURCE = ILR

Value	Description	Definition
A	Asian or Asian British	ETHNICITY in (39, 40, 41, 42, 43)
B	Black, African, Caribbean or Black British	ETHNICITY in (44, 45, 46)
M	Mixed or multiple ethnic background	ETHNICITY in (35, 36, 37, 38)
O	Other ethnic background	ETHNICITY in (33, 47, 98)
W	White	ETHNICITY in (31, 32, 34)
U	Refused or unknown	Otherwise

TRSEX

16. This field classifies the sex of the student. This field has the same definition as IPSEX.

TRSOURCE = HESASTU, HESASAR

Value	Description	Definition
M	Male	SEXID = 1
F	Female	SEXID = 2
O	Other	Otherwise

TRSOURCE = ILR

Value	Description	Definition
M	Male	SEX = M
F	Female	SEX = F
O	Other	Otherwise

TREIMD

17. This field shows the English Index of Multiple Deprivation (EIMD) 2019[6] quintile of the postcode of the student's address prior to entry. In order to assign an EIMD 2019 quintile to a student's address prior to entry, the postcode is matched to the National Statistics Postcode Lookup (NSPL) for the United Kingdom (May 2020 release) to determine the Lower Layer Super Output Area (LSOA) in which the postcode exists. The EIMD 2019 quintile of the LSOA is then assigned.

TRSOURCE = HESASTU, HESASAR

Value	Description	Definition
1, 2, 3, 4, 5	Assigned EIMD 2019 quintile 1 (most deprived) to 5 (least deprived)	TRDOM = E and POSTCODE matched to postcode and LSOA in NSPL
U	EIMD 2019 quintile unknown	TRDOM = E and (POSTCODE = BLANK or POSTCODE not matched to postcode and LSOA in NSPL)
NA	Student domicile outside of England	Otherwise

TRSOURCE = ILR

Value	Description	Definition
1, 2, 3, 4, 5	Assigned EIMD 2019 quintile 1 (most deprived) to 5 (least deprived)	TRDOM = E and POSTCODEPRIOR matched to postcode and LSOA in NSPL
U	EIMD 2019 quintile unknown	TRDOM = E and (POSTCODEPRIOR = BLANK or POSTCODEPRIOR not matched to postcode and LSOA in NSPL)
NA	Student domicile outside of England or student domicile unknown	Otherwise

TRAWARDLEVELNUM (HESASTU and HESASAR only)

18. This field gives the numeric level of study according to the higher education qualifications framework. This also aligns with NVQ levels. This field has the same definition as IPAWARDLEVELNUM.

TRDENT (HESA and HESASAR only)

19. This field indicates whether a student has at least one instance of 'Pre-clinical dentistry' or 'Clinical dentistry' associated with their programme of study. This is calculated in the same way as IPDENT but uses the equivalent Higher Education Classification of Subjects (HECoS) codes in place of Joint Academic Coding System (JACS) codes in accordance with the updated HESA specification for 2019-20.[7]

---

[6] For information on the English Index of Multiple Deprivation 2019 see
https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019
[7] For mappings of JACS codes to HECoS codes see www.hesa.ac.uk/innovation/hecos

Value	Description	Definition
1	The student has at least one instance of a 'Pre-clinical dentistry' or 'Clinical dentistry' programme of study	SBJCA(1-5) in (100275, 100266)
0	The student does not have an instance of a 'Pre-clinical dentistry' or 'Clinical dentistry' programme of study	Otherwise

TRAWARD_DETAIL

20. This field allocates the qualification awarded to the student during the reporting year to a level of qualification awarded.

TRSOURCE = HESASTU, HESASAR

21. This field has the same definition as IPAWARD_DETAIL but uses the revised specification for TRDENT shown above in place of IPDENT and TRAWARDLEVELNUM in place of IPLEVELNUM.

TRSOURCE = ILR

22. This field has the same definition as IPAWARD_DETAIL but uses OFSQAIM (presented earlier in this document) in place of IPOFSQAIM.

TRAWARDLEVEL

23. This field allocates the qualification awarded to the student during the reporting year to a broad grouping. This field has the same definition as IPAWARDLEVEL but uses TRAWARD_DETAIL in place of IPAWARD_DETAIL.

TRAWARD

24. This field simplifies the broad grouping of the qualification awarded to the student during the reporting year allocated by TRAWARDLEVEL to those levels relevant to the Transparency attainment population.

Value	Description	Definition
FDEG	First degree	TRAWARDLEVEL in (DEG, PUGD)
OUG	Other undergraduate	TRAWARDLEVEL in (OUG, PUGO)
OTH	Other including no award	Otherwise

TRDEGCLASS

25. This field indicates the degree classification awarded.

TRSOURCE = HESASTU, HESASAR

Value	Description	Definition
1	First class honours degree	TRAWARD = FDEG and XCLASSF01 = 01
2_1	Upper second class honours degree	TRAWARD = FDEG and

		XCLASSF01 = 02
2_2	Lower second class honours degree	TRAWARD = FDEG and XCLASSF01 = 03
3	Third class honours/pass	TRAWARD = FDEG and XCLASSF01 in (05, 09)
UN	Degree awarded without classification	TRAWARD = FDEG and not above
NA	Otherwise	Otherwise

TRSOURCE = ILR

Value	Description	Definition
1	First class honours degree	TRAWARD = FDEG and OUTGRADE = FI
2_1	Upper second class honours degree	TRAWARD = FDEG and OUTGRADE = SU
2_2	Lower second class honours degree	TRAWARD = FDEG and OUTGRADE in (SL, SE)
3	Third class honours/pass	TRAWARD = FDEG and OUTGRADE in (TH, FO)
UN	Degree awarded without classification	TRAWARD = FDEG and not above
NA	Otherwise	Otherwise

TRATTAINPOP

26. This field indicates whether the student is included in the Transparency attainment population.

TRSOURCE = HESASTU, HESASAR

Value	Description	Definition
1	Student is counted in the Transparency attainment population	TRDOM in (E, S, W, N) and TRAWARD in (FDEG, OUG) and XPQUAL01 = 1
0	Otherwise	Otherwise

TRSOURCE = ILR

Value	Description	Definition
1	Student is counted in the Transparency attainment population	TRDOM in (E, S, W, N) and TRAWARD in (FDEG, OUG) and OUTCOME = 1
0	Otherwise	Otherwise

Rebuild instructions

27. The individualised file, TR19_Attainment_DCT_XXXXXXXX_IND.csv, can be used to interrogate the data and our calculation of derived fields by filtering to certain groups of records. Full details of how to access and use an individualised file are given on the OfS website (www.officeforstudents.org.uk/data-and-analysis/supplying-data/working-with-individualised-files). Specific instructions to rebuild the Transparency attainment data summary figures in the Transparency attainment data summary workbook, TR19_Attainment_DCT_XXXXXXXX.xlsx, are provided here.

28. A number of additional fields not described in the field tables or algorithms above have been included in the individualised file for reference. These additional fields are listed in the tables below.

TRSOURCE = HESASTU

Entity	Field name	Description
Instance	COMDATE	Start date of instance
Instance	ENDDATE	End date of instance
Course	COURSEID	Course identifier
Course	CTITLE	Course title
Course	COURSEAIM	General qualification aim of course
Course subject(s)	SBJPCNT(1-5)	Proportion of time allocated for each subject studied on a course

TRSOURCE = HESASAR

Entity	Field name	Description
Instance	COMDATE	Start date of instance
Instance	ENDDATE	End date of instance
Instance period	PERIODSTART	Instance period start date
Instance period	PERIODEND	Instance period end date
Course	COURSEID	Course identifier
Course	CTITLE	Course title
Course	COURSEAIM	General qualification aim of course
Course subject(s)	SBJPCNT(1-5)	Proportion of time allocated for each subject studied on a course

TRSOURCE = ILR

Field name	Description	Data source
QUAL_TIT	Learning aim title	ILR
NOTIONALNVQLEVELV2	Level of the learning aim	LARS

Note: 'LARS' = 'Learning Aim Reference Service'.

29. The following instructions assume the user has the individualised file open in Microsoft Excel.

30. To apply filters to the column headings, highlight the headings. From the 'DATA' tab in the toolbar select 'Filter'.

31. The headcount present in the summary workbook is the number of records present in the individualised data file once the relevant filters have been applied.

32. To filter to the Transparency attainment population, TRATTAINPOP = 1 should be selected.

33. Filters can then be applied to TRMODE, TRDEGCCLASS, TREIMD, TRETHNIC and TRSEX to reproduce the headcount of qualifiers (number of records) in the Transparency attainment summary workbook.

34. For example, to reproduce the headcount for upper second class honours degrees awarded to female qualifiers who studied full-time you would apply the following filters: TRATTAINPOP = 1 and TRDEGCLASS = 2_1 and TRMODE = FT and TRSEX = F. The number of records in the filtered dataset represents the headcount in TR19_Attainment_DCT_XXXXXXXX.xlsx.

35. To reproduce the headcount of classified first degrees awarded for a given mode of study and characteristic, filter to the required mode and characteristic split and select TRAWARD = 1, 2_1, 2_2, 3.

36. For example, to reproduce the headcount of classified first degrees awarded to qualifiers from EIMD quintile 1 areas who studied full-time you would apply the following filters: TRATTAINPOP = 1, TRDEGCLASS = 1, 2_1, 2_2, 3 and TRMODE = FT and TREIMD = 1. The number of records in the filtered dataset represents the headcount in TR19_Attainment_DCT_XXXXXXXX.xlsx.