Calculating student numbers

Technical document

Summary

1. This document details the algorithms applied to Higher Education Statistics Agency (HESA) student record, HESA alternative provider student record and Individualised Learner Record (ILR) data in order to determine a provider's higher education and further education student numbers. We will use this approach to assess applications for degree awarding powers and university title, calculate Office for Students registration fees, and determine whether a provider is required to participate in the Teaching Excellence and Student Outcomes Framework.

2. This will apply to degree awarding powers applications received on or after 1 April 2018; and for university title applications and applications to transfer out of the Further Education Sector and become a Higher Education Corporation received on or after 1 April 2019.

3. 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 alternative provider student record coding manual and ILR specification for the relevant year, and the Office of Qualifications and Examinations Regulation (Ofqual) Register Data Extract Specification to hand when using this document.

4. Some providers submit both a HESA data return and an ILR return. Student instances should not be returned in both collections, except in the case of higher apprenticeships. If a provider submits a HESA return it should return any records associated with higher apprenticeships to both HESA and the ILR. Therefore to avoid double counting these records, we have excluded from the population all records related to higher apprenticeships from the ILR for providers that submit HESA data.

Note on terminology

5. The algorithms below apply to three different data collections, each of which has a different structure. In the HESA student return, the fields that we use are associated with a student instance. In the HESA alternative provider student return, the fields that we are use are (mainly) associated with a student instance period; there may be many instance periods per student instance. In the ILR return, the fields that we use are associated with a course; a learner may study multiple courses.

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

7. The year '20Y1-Y2' below refers to the academic year starting on 1 August 20Y1 and ending on 31 July 20Y2. These algorithms apply to data from the 2017-18 academic year and onwards.

20Y1-Y2 fields used in the student number algorithms

Table 1: HESA student fields used in the student number calculations

Entity	Field	Description
Course	COURSEAIM	General qualification aim of course
Instance	COMDATE	Start date of instance

Instance	ENDDATE	End date of instance
Instance	EXCHANGE	Identifies whether the student is an exchange student
Instance	LOCSDY	Location of study
Instance	MODE	Mode of study
Instance	NUMHUS	Student instance identifier
Instance	STULOAD	Student instance full-time equivalence (FTE)
Institution	UKPRN	UK Provider Reference Number
Student	HUSID	HESA unique student identifier

Table 2: HESA alternative provider student fields used in the student number calculations

Entity	Field	Description
Course	COURSEAIM	General qualification aim of course
Instance	NUMHUS	Student instance identifier
Instance period	EXCHIND	Identifies whether the student is an exchange student
Instance period	LOCSDY	Location of study
Instance period	PERIODEND	Instance period end date
Instance period	PERIODSTART	Instance period start date
Instance period	STULOAD	Student Instance FTE
Provider	UKPRN	UK Provider Reference Number
Student	HUSID	HESA unique student identifier

Table 3: ILR, Learning Aim Reference Service and Ofqual fields used in the student number calculations

Data source	Field	Description
ILR	COMPSTATUS	Degree of completion of the learning activities leading to the learning aim
ILR	LEARNACTENDDATE	Learning actual end date
ILR	LEARNPLANENDDATE	Learning planned end date
ILR	LEARNSTARTDATE	Learning start date
ILR	STULOAD	Learner FTE
ILR	PLANLEARNHOURS	Planned Learning Hours
ILR	PRIORLEARNFUNDADJ	The proportion of this learning aim still to be delivered
ILR	PROGTYPE	The type of programme which the learner is undertaking
ILR	UKPRN	UK Provider Reference Number
ILR	WITHDRAWREASON	The reason why the learner has withdrawn from the learning aim

LARS	ENGLANDFEHESTATUS	Further or higher education status of a course ('England FE/HE status')
LARS	LEARNAIMREF	Learning aim reference
LARS	LEARNAIMREFTITLE	Learning aim title
LARS	LEARNAIMREFTYPE	Learning aim type
LARS	NOTIONALNVQLEVELV2	The detailed learning aim level on the NVQ scale
Ofqual	TOTAL QUALIFICATION TIME	The total time a learner may typically be expected to expend in achieving a qualification
Ofqual	MAXIMUM GUIDED LEARNING HOURS	The maximum number of hours a learner can be expected to be taught or instructed by their tutor or teacher to complete the qualification
Ofqual	TOTAL CREDITS	The total credits required to obtain the qualification

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

Derived fields for calculating total higher education FTE

SNSOURCE

8. Identifies the source of the record.

Value	Definition
HESA	Record taken from the HESA student record
HESAAP	Record taken from the HESA alternative provider student record
ILR	Record taken from the ILR

SNHESAILRFLAG

9. Identifies providers that returned both HESA and ILR data in the academic year 20Y1-Y2.

Value	Definition
1	Provider returned both HESA and ILR data for the academic year 20Y1-Y2
0	Provider did not return both HESA and ILR data for the academic year 20Y1-Y2

SNEXCL1

10. Identifies records that start after the academic year of interest.

Value	Description	Definition
1	Excluded	**SNSOURCE= HESA** COMDATE >31/07/Y2 **SNSOURCE = HESAAP** PERIODSTART > 31/07/Y2 **SNSOURCE= ILR**

		LEARNSTARTDATE > 31/07/Y2
0	Included	Otherwise

SNEXCL2

11. Identifies records that finish before the academic year of interest.

Value	Description	Definition
1	Excluded	**SNSOURCE= HESA** ENDDATE <01/08/Y1 and ENDDATE is not BLANK) **SNSOURCE = HESAAP** PERIODEND < 01/08/Y1 and PERIODEND is not BLANK **SNSOURCE= ILR** (LEARNACTENDDATE < 01/08/Y1 and LEARNACTENDDATE is not BLANK
0	Included	Otherwise

SNEXCL4

12. Identifies records in the HESA data are dormant or on sabbatical.

Value	Description	Definition
1	Excluded	**SNSOURCE= HESA** MODE = 51, 63, 64, 98
0	Included	Otherwise

SNEXCL8

13. Identifies incoming exchange students.

Value	Description	Definition
1	Excluded	**SNSOURCE= HESA** EXCHANGE = 4,G or LOCSDY = S **SNSOURCE = HESAAP** EXCHIND = 1 or LOCSDY = S
0	Included	Otherwise

SNEXCL16

14. Identifies apprenticeship standard 'wrapper' programme aims.

Value	Description	Definition
1	Excluded	**SNSOURCE= HESA** COURSEAIM = Z99 **SNSOURCE = ILR** LEARNAIMREF = ZPROG001
0	Included	Otherwise

SNEXCL32

15. Identifies higher apprentices in the ILR that have been returned by providers that also submit HESA data.

Value	Description	Definition
1	Excluded	**SNSOURCE= ILR** SNHESAILRFLAG = 1 and PROGTYPE = 20, 21, 22, 23, 25
0	Included	Otherwise

SNEXCL64

16. Identifies an ILR record which has been closed to correct an incorrect LEARNPLANENDDATE.

Value	Description	Definition
1	Excluded	**SNSOURCE= ILR** COMPSTATUS = 4[1] and WITHDRAWREASON = 40 and LEARNACTENDDATE = 1 August 2017
0	Included	Otherwise

SNEXCL128

17. Identifies records in the ILR which we believe relate to teaching provision rather than registered provision. This will be identified if two records returned by two providers for the same student are found to have the same LEARNAIMREF, the courses start within 31 days of each other and the UKPRN for one record matches the PARTNERUKPRN of the other record. The record with missing PARTNERUKPRN will then be flagged.

Value	Description	Definition
1	Excluded	**SNSOURCE= ILR** Student appears to exist in another provider's ILR and this record will be excluded from the student numbers calculation

---

[1] ILR records returned to the ESFA with COMPSTATUS=3 and WITHDRAWREASON=40 will have had the value of COMPSTATUS changed to 4 by the ESFA.

0	Included	Otherwise

SNEXCL

18. This field indicates whether the record will be included in the calculation of student numbers. SNEXCL is the sum of all applicable values from the table below.

Value	Description	Definition
1	Record starts after academic year of interest	SNEXCL1 = 1
2	Record ends before academic year of interest	SNEXCL2 = 1
4	Student is dormant or on sabbatical	SNEXCL4 = 1
8	Incoming exchange student	SNEXCL8 = 1
16	Apprenticeship standard 'wrapper' programme aim	SNEXCL16 = 1
32	Higher apprentices returned in ILR by providers that also submit HESA data	SNEXCL32 = 1
64	ILR records closed to correct LEARNPLANENDDATE	SNEXCL64 = 1
128	ILR record returned by the teaching provider also returned by the registering provider	SNEXCL128 = 1
0	Otherwise	None of the above

19. This field contain(s) the exclusion reason(s) for the instance. It is calculated as $(1 \times EXCL1) + (2 \times EXCL2) + \ldots + (128 * EXCL128)$.

SNLEVEL

20. Identifies whether the record is of higher education or further education.

Value	Description	Definition
HE	Record of higher education	COURSEAIM = D, E, L, M, H, I, J, C or ENGLANDFEHESTATUS = H or NOTIONALNVQLEVELV2 = 4, 5, 6, 7, 8, H
FE	Record of further education	COURSEAIM = P, Q, R, S, X or (ENGLANDFEHESTATUS = F and LEARNAIMREF ≠ ZPROG001 and not above)

SNHESTUDYLEVEL

21. Indicates the level of study for higher education records, using either the Quality Assurance Agency for Higher Education's Frameworks for higher education qualifications (for qualifications awarded using degree awarding powers) or else the Regulated qualifications framework[2].

---

[2] For the Frameworks for higher education qualifications see www.qaa.ac.uk/en/quality-code/the-existing-uk-quality-code/part-a-setting-and-maintaining-academic-standards.

Value	Description	Definition
6	Records at Level 6 or above	SNLEVEL = HE and (COURSEAIM = D, E, L, M, H, I00, I11, I12, I16 or NOTIONALNVQLEVELV2 = 6, 7, 8)
5	Records at Level 5	SNLEVEL = HE and (COURSEAIM=I60, I61, I70, I71, I72, I73, I74, I76, I78, I79, I80, I81, I90, I91, I99, J or NOTIONALNVQLEVELV2 = 5)
4	Records at Level 4	SNLEVEL=HE and not above

SNTITLEHRS

22. The maximum number of hours per record as extracted from the course title on the Learning Aim Reference Service. For example, the value of SNTITLEHRS for the course listed as 'Non regulated SFA formula funded provision, Entry level, Medicine and Dentistry, 389 to 580 hrs, PW B' would be 580.

SNTITLECREDITS

23. The maximum number of credits per record as extracted from the course title on the LARS. For example, the value of SNTITLECREDITS for the course listed as 'Innovation code, Entry level, Medicine and Dentistry, 3-5 credits' would be 5.

SNPRIORLEARNADJ

24. The funding adjustment for prior learning expressed as a decimal value. For ILR records only.

Value	Description	Definition
PRIORLEARNFUNDADJ/100		SNSOURCE = ILR and PRIORLEARNFUNDADJ > 0
0		SNSOURCE = ILR and PRIORLEARNFUNDADJ = 0

SNQUALHOURS

25. The total number of hours the course associated with the record is expected to take

Value	Description	Definition
TOTAL QUALIFICATION TIME × SNPRIORLEARNADJ	Source of hours is TOTAL QUALIFICATION TIME	SNSOURCE = ILR and STULOAD = MISSING and TOTAL QUALIFICATION TIME ≠ MISSING
MAXIMUM GUIDED LEARNING HOURS × SNPRIORLEARNADJ	Source of hours is MAXIMUM GUIDED LEARNING HOURS	SNSOURCE = ILR and STULOAD = MISSING and MAXIMUM GUIDED LEARNING HOURS ≠ MISSING and not above

(TOTAL CREDITS × 10) × SNPRIORLEARNADJ	Source of hours is TOTAL CREDITS	SNSOURCE = ILR and STULOAD = MISSING and TOTAL CREDITS ≠ MISSING and not above
(SNTITLEHRS) × SNPRIORLEARNADJ	Source of hours is SNTITLEHRS	SNSOURCE = ILR and STULOAD = MISSING and SNTITLEHRS ≠ MISSING and not above
(SNTITLECREDITS × 10) × SNPRIORLEARNADJ	Source of hours is SNTITLECREDITS	SNSOURCE = ILR and STULOAD = MISSING and SNTITLECREDITS ≠ MISSING and not above

SNENDDATE

26. Determines the end date to be used when calculating the number of days studied (SNAYDAYSSTUDIED) in the academic year 20Y1-Y2. Used only for ILR data where SNQUALHOURS will be used to calculate the FTE of the record.

Value	Description	Definition
LEARNACTENDDATE	Learner completed learning activities in 20Y1-Y2	SNSOURCE=ILR and LEARNACTENDDATE ≠ MISSING and LEARNACTENDDATE<=31/07/Y2
31/07/Y2	Learner is still studying at the end of 20Y1-Y2	SNSOURCE=ILR and not above

SNAYDAYSSTUDIED

27. The number of days studied in the 20Y1-Y2 academic year, calculated as the numbers of days between SNENDDATE and either 1 August 20Y1 or LEARNSTARTDATE, whichever is later. Used only for ILR data where SNQUALHOURS will be used to calculate the FTE of the record.

SNAVHOURSPERDAY

28. Calculates the average number of hours that would be studied each day for the duration of the course, calculated as SNQUALHOURS divided by the number of days between the LEARNSTARTDATE and LEARNPLANENDDATE. Used only for ILR data where SNQUALHOURS will be used to calculate the FTE of the record.

SNHOURSPERAYR

29. Calculates the total number of hours studied in the academic year for each record. Used only for ILR data where SNQUALHOURS will be used to calculate the FTE of the record. We cap SNHOURSPERAYR at the value of SNQUALHOURS for the course.

- SNHOURSPERAYR= Min(SNAVHOURSPERDAY × SNAYDAYSSTUDIED, SNQUALHOURS).

SNHESTULOADCASE

30. Defines the fields we will use in calculating the FTE for a given higher education record.

Value	Description	Definition
0	STULOAD will be used in calculating FTE for this record	SNEXCL = 0 and SNLEVEL = HE and STULOAD ≠ MISSING
1	STULOAD will be set to 0 for this record	SNEXCL = 0 and SNLEVEL = HE and SNSOURCE = HESA, HESAAP and STULOAD = MISSING
2	SNQUALHOURS will be used in calculating FTE for this record	SNEXCL = 0 and SNLEVEL = HE and SNSOURCE = ILR and STULOAD = MISSING and SNQUALHOURS ≠ MISSING
3	A value of 25 STULOAD will be assigned to this record	SNEXCL = 0 and SNLEVEL = HE and SNSOURCE = ILR and STULOAD = MISSING and not above

SNHESTULOAD

31. The STULOAD for each higher education record.

Value	Definition
STULOAD	SNHESTULOADCASE = 0
0	SNHESTULOADCASE = 1
(SNHOURSPERAYR ÷ 540)×100	SNHESTULOADCASE = 2
25	SNHESTULOADCASE = 3

SNHEINSTLEVSTULOAD

32. The total STULOAD at a given SNHESTUDYLEVEL for each higher education instance[3]

Value	Definition
SNHESTULOAD	SNSOURCE=HESA
Sum of SNHESTULOAD across all instance periods at a given SNHESTUDYLEVEL per instance	SNSOURCE=HESAAP and SNEXCL=0
Sum of SNHESTULOAD across all records for each SNHESTUDYLEVEL per student	SNSOURCE=ILR and SNEXCL=0

SNHEINSTLEVSTULOADADJ

33. The capped STULOAD at a given SNHESTUDYLEVEL for each higher education instance.

- SNHEINSTLEVSTULOADADJ= Min(SNHEINSTLEVSTULOAD,150).

SNTOTHESTULOAD

34. SNTOTHESTULOAD is the total STULOAD per student including any capping, calculated as the sum of SNHEINSTLEVSTULOADADJ across each unique combination of instance and level per student where SNEXCL=0.

---

[3] The ILR data does not include the concept of an 'instance'. Solely for the purpose of calculating SNHEINSTLEVSTULOAD we treat a level of study in the ILR data as analogous to an instance.

SNTOTHEFTE

35. The total higher education FTE, calculated as SNTOTHESTULOAD÷100.

Additional fields required to calculate total further education FTE

SNFEPLANLEARNSTULOAD

36. Converts PLANLEARNHOURS to STULOAD for further education records.

- SNFEPLANLEARNSTULOAD = max (((PLANLEARNHOURS÷540)×100)-SNTOTHESTULOAD,0).

SNFESTULOADCASE

37. Determines the fields that will be used to calculate the FTE for further education records.

Value	Description	Definition
0	STULOAD will be used when calculating FTE for this record	SNEXCL = 0 and SNLEVEL = FE and SNSOURCE = HESA, HESAAP and STULOAD ≠ MISSING
1	STULOAD will be set to 0 for this record	SNEXCL = 0 and SNLEVEL = FE and SNSOURCE = HESA, HESAAP and STULOAD = MISSING
2	PLANLEARNHOURS will be used when calculating FTE for this record	SNEXCL = 0 and SNLEVEL = FE and SNSOURCE = ILR and PLANLEARNHOURS ≠ MISSING
3	SNQUALHOURS will be used when calculating FTE for this record	SNEXCL = 0 and SNLEVEL = FE and SNSOURCE = ILR and SNQUALHOURS ≠ MISSING and not above
4	A value of 10 STULOAD will be assigned to this record	SNEXCL = 0 and SNLEVEL = FE and SNSOURCE = ILR and not above

SNNUMFERECORDS

38. The number of SNEXCL = 0, SNLEVEL = FE records per student. Used only for records from the ILR.

SNFESTULOAD

39. The STULOAD for each further education record.

Value	Definition
STULOAD	SNFESTULOADCASE = 0
0	SNFESTULOADCASE = 1
SNFEPLANLEARNSTULOAD ÷ SNNUMFERECORDS	SNFESTULOADCASE = 2
(SNHOURSPERAY÷ 540)×100	SNFESTULOADCASE = 3
10	SNFESTULOADCASE = 4

SNTOTFESTULOAD

40. The sum of SNFESTULOAD across all SNEXCL = 0 records for the student. This is capped at a maximum of 150 per student.

SNTOTFEFTE

41. The total further education FTE, calculated as SNTOTFESTULOAD÷100.