2018-19 HESA AP Student data checking tool:

Student loans company (SLC) metrics technical document

Contents

2018-19 SLC comparison algorithms

Purpose

1. This document details the algorithms applied to the 2018-19 HESA AP Student record which allow for a comparison with SLC data.

2. This document is aimed at readers with in-depth knowledge of the data. Readers are advised to have a copy of the HESA AP Student Record Coding Manual 2018-19 to hand when using this document.

3. The purpose of this comparison is to ensure that the 2018-19 HESA data is consistent with SLC records. For each provider, we check:

    – whether all students who received student support from the SLC during the 2018-19 academic year are included in the 2018-19 HESA data

    – whether all students who are recorded as receiving SLC tuition fee loans in the 2018-19 HESA data appear in the SLC data

    – whether all students recorded in the SLC data with a withdrawal date have a corresponding end date in the HESA data.

4. We compare the HESA data with SLC data for all students who are active in the 2018-19 academic year. This will include students whose year of study beings in the 2017-18 academic year and continues into the 2018-19 academic year.

2018-19 HESA fields used in the SLC comparison algorithms

5. Only certain fields, detailed in Table 1, are used to generate the SLC comparison derived fields.

6. Throughout this document, fields taken or derived from the HESA AP student record are shown in capitals using the names given in Tables 1 and 2 respectively.

Using the individualised file

7. When working through this document it is necessary to use the individualised file SLCCOM18_DCT_XXXXXXXX_IND.csv, where XXXXXXXX is the UKPRN. This will show the allocation of students to cells within the output and, where relevant, details of why they were excluded from the output.

Table 1 HESA fields used in the SLC comparison

Entity	Field	Description	Data set
Course[*]	COURSEID	Course identifier	HESA
Instance	ENDDATE	End date of instance	HESA
Instance	NUMHUS	Student instance identifier	HESA
Instance[*]	OWNINST	Institution's own instance identifier	HESA
Instance	XINACT01	Inactive for entire reporting year indicator	HESA

Entity	Field	Description	Data set
Instance Period	MSTUFEE	Major source of tuition fees	HESA
Instance Period	PERIODSTART	Instance period start date	HESA
Provider	UKPRN	UK Provider Reference number	HESA
Student	HUSID	HESA unique student identifier	HESA
Student*	OWNSTU	Institution's own internal identifier for the student	HESA
Student	SSN	Student Support Number	HESA
	SLC1819TERMSTART_DATE	The first term start date within the 2018-19 academic year from the SLC data	SLC
	SLCATTENDANCE_DATE	Earliest attendance confirmation date for a term relating to the 2018-19 academic year from the SLC data	SLC
	SLCSSN	Student Support Number from the SLC data	SLC
	SLCWITHDRAWN_DATE	The latest withdrawal date within the 2018-19 academic year from the SLC data	SLC
	SLCYEARSTART_DATE	The term 1 start date from SLC data. Used to determine whether the student record relates to a year of study which began in the 2017-18 or 2018-19 academic year	SLC

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

Description of derived fields for re-creating tables

8. Here we give details of the derived fields in the individualised file. These fields are used in the SLC comparison.

Table 2 SLC comparison derived fields

Derived field name	Description	Paragraph
SLCCOMENDCHECK	Identifies students in the SLC data with a withdrawal date where no ENDDATE or a substantially later or earlier ENDDATE is recorded in the HESA data	15
SLCCOMHESAPOP	Identifies students in the HESA data for the provider	12
SLCCOMHESASLCPOP	Identifies students in the HESA data whose main source of tuition fees is the SLC	13
SLCCOMMSTUFEE1-3[†]	Major source of tuition fees	10
SLCCOMPERIODSTART1-3[†]	Instance period start date	11
SLCCOMSLCPOP	Identifies students who have had their registration or attendance confirmed in the SLC data	14

[†] Due to file size restrictions the number of occurrences of this field is restricted to three.

Instance period data

9. The number of instance periods that can be returned for a given instance is unbounded; however, due to file size considerations we restrict the number of separate instance periods displayed in the individualised file to three. In the unusual case where an instance has more than three instance periods, all will be considered in our algorithms, although only three will be displayed in the individualised file. In selecting instance period data to show we will first select those with the major source of tuition fees recorded as the SLC (MSTUFEE = 02, 03, 04, 52, 53, 54) and then those with the earliest period start date (PERIODSTART).

SLCCOMMSTUFEE1-3

10. These fields contain the major source of tuition fees (MSTUFEEs) that are associated with the instance. For providers making a HESA Student return, only SLCCOMMSTUFEE1 will be populated.

SLCCOMPERIODSTART1-3

11. These fields contain the instance period start dates (PERIODSTARTs) for the student. For example SLCCOMPERIODSTART1 will contain the instance period start date for SLCCOMMSTUFEE1.

SLCCOMHESAPOP

12. This field identifies active students who are in the HESA data for the provider.

Value	Description	Definition
1	Student is in the HESA data for the provider with an active record	A record was found in the HESA AP Student data for the provider in the 2018-19 academic year and XINACT01 = 0
0	Otherwise	Otherwise

SLCCOMHESASLCPOP

13. This field identifies active students in the HESA data whose source of tuition fees is the SLC, in whole or in part.

Value	Description	Definition
1	Students whose source of tuition fees is the SLC	SLCCOMHESAPOP = 1 and (SLCCOMMSTUFEE1 = 02, 03, 04, 52, 53, 54 or SLCCOMMSTUFEE2 = 02, 03, 04, 52, 53, 54 or SLCCOMMSTUFEE3 = 02, 03, 04, 52, 53, 54)
0	Otherwise	Otherwise

SLCCOMSLCPOP

14. This field identifies students in the SLC data who were paid some student support in the 2018-19 academic year, and either did not withdraw within 14 days of the first term start date in that year, or else had their attendance confirmed during 2018-19.

Value	Description	Definition
1	Students who have had their registration or attendance confirmed in the SLC data	SLCSSN ≠ BLANK and (ENDDATE = BLANK or ENDATE > 31 July 2018) and (SLCWITHDRAWN_DATE > SLC1819TERMSTART_DATE + 14 or SLCWITHDRAWN_DATE = BLANK or SLC1819TERMSTART_DATE = BLANK or SLCATTENDANCE_DATE ≠ BLANK)
0	Otherwise	Otherwise

SLCCOMENDCHECK

15. This field identifies students in the SLC data with a withdrawal date where no ENDDATE or a later ENDDATE or an earlier ENDDATE is recorded in the HESA data.

Value	Description	Definition
1	Students who have a withdrawal date in the SLC data but no ENDDATE is recorded in the HESA data	A record was found in the HESA data and SLCSSN ≠ BLANK and SLCWITHDRAWN_DATE ≠ BLANK and SLCWITHDRAWN_DATE < 01

Value	Description	Definition
		August 2019 and ENDDATE = BLANK
2	Students who have a substantially later ENDDATE than the withdrawal date in the SLC data	A record was found in the HESA data and SLCSSN ≠ BLANK and SLCWITHDRAWN_DATE ≠ BLANK and ENDDATE ≠ BLANK and ENDDATE > SLCWITHDRAWN_DATE + 14
3	Students who have a substantially earlier ENDDATE than the withdrawal date in the SLC data	A record was found in the HESA data and SLCSSN ≠ BLANK and SLCWITHDRAWN_DATE ≠ BLANK and ENDDATE ≠ BLANK and ENDDATE < SLCWITHDRAWN_DATE - 14
0	Otherwise	Otherwise

Re-building the 2018-19 SLC comparison using the individualised file

16. The summary figures derived from HESA data shown in the SLCCOM18_DCT_XXXXXXXX.xlsx workbooks are listed below. For each of item, we describe how to identify the records in the individualised file SLCCOM18_DCT_XXXXXXXX_IND.csv that contribute to the figure. In each case we specify the values of the derived fields needed to rebuild the output.

Item in workbook	Derived field selection in individualised file
Students receiving student support and appearing in the HESA data	SLCCOMHESAPOP = 1 and SLCCOMSLCPOP = 1
Students receiving student support and not appearing in the HESA data	SLCCOMHESAPOP = 0 and SLCCOMSLCPOP = 1
Students recorded in the HESA data as receiving a tuition fee loan and not appearing in the SLC data	SLCCOMHESASLCPOP = 1 and SLCCOMSLCPOP = 0
Students appearing in the SLC data with a withdrawal date but either no ENDDATE or a substantially later ENDDATE or a substantially earlier ENDDATE is recorded in the HESA data	SLCCOMENDCHECK = 1, 2, 3