HESA AP data checking tool 2017-18

Pearson qualification comparison technical document

Contents

2017-18 Pearson qualification comparison algorithms

Purpose

1. This document details the algorithms applied to the 2017-18 HESA AP Student record which allow a comparison with Pearson 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 2017-18 to hand when using this document.

3. The purpose of this comparison is to ensure that the 2017-18 HESA data is consistent with the Pearson qualification records. For each provider we check:

   a. whether all students who received a qualification from Pearson during the 2017-18 academic year are included in the 2017-18 HESA data as having this qualification

   b. whether all students who are recorded as having achieved a Pearson qualification in the 2017-18 HESA data appear in the Pearson data

   c. whether the type of qualification recorded in the 2017-18 HESA data matches the type of qualification in the Pearson data.

4. We compare the HESA data with the Pearson qualification data for all students who are active in the 2017-18 academic year. This will include students whose instance period begins in the 2016-17 academic year and continues into the 2017-18 academic year.

2017-18 HESA fields used in the Pearson qualification comparison algorithms

5. Only certain fields, detailed in Table 1, were used to generate the Pearson 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.

Table 1 HESA fields used in the Pearson qualification comparison

Entity	Field	Description	Data set
†	OWNSTU / OWNINST		
Course	CTITLE	Course title	HESA
Course	COURSEAIM	Course aim	HESA
Student	HUSID	HESA unique student identifier	HESA
Provider	UKPRN	UK Provider Reference number	HESA
Student	NUMHUS	HESA Student Instance Identifier	HESA
Instance	QUAL	Qualification Awarded	HESA
Instance	PERIODSTART	Instance period start date	HESA

Entity	Field	Description	Data set
	PEAR1718REGNO	Learner's Registration Number (Unique Identifier in the Pearson data)	Pearson
	PEAR1718CTITLE	Course title from Pearson data	Pearson
	PEAR1718CMPLTDATE	Learner's date of completion from the Pearson data	Pearson
	PEAR1718QUAL	Learner's qualification aim from the Pearson data	Pearson
	PEAR1718STATUS	Learner's outcome status	Pearson
	PEAR1718GRADE	Learner's outcome grade where a qualification has been awarded	Pearson

* The individualised file PEARCOM17_DCT_AXXXXXXXX_IND.csv, downloadable as an IRIS output from the HESA data collection system.

† 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

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

Table 2 Pearson qualification comparison derived fields

Derived field name	Description	Paragraph
PEARCOMHESAPOP	Identifies students in the HESA data for the provider	13
PEARCOMPEARPOP	Identifies students in the Pearson data	15
PEARCOMPERIODSTART1-3†‡	Instance period start date	11
PEARCOMQUAL1_1-3_2†	Identifies the qualification awarded to the student in the HESA data.	12
PEARCOMPEARQUAL1-2	Identifies the qualification awarded to the student in the Pearson data	14
PEARCOMQUALMATCH	Identifies students with matching qualification types in the HESA and Pearson data	16

* The individualised file PEARCOM17_DCT_AXXXXXXXX_IND.csv, downloadable as an IRIS output from the HESA data collection system.

† Due to file size restrictions the number of occurrences of this field is restricted to 3.

‡ This field is only relevant to providers making a HESA AP student return.

Instance period data

8. 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 3. In the unusual case where an instance has more than 3 instance periods, all will be considered in our algorithms, although only 3 will be displayed in the individualised file. In selecting instance period data to show we will select those with a qualification ordered by:

   a) Those with a HND

   b) Those with a HNC

   c) Period start date (PERIODSTART) ordered by the earliest first.

Linking

9. The HESA and Pearson records are linked using the UKPRN, First Name(s), Last Name(s) and Date of Birth. A link is defined as an identical match of UKPRN, First Name(s), Last Name(s) and Date of Birth. Allowances are made for missing or incorrectly ordered names.

10. In the case that no link can be made between the 2017-18 HESA data and the Pearson data, we will attempt to link the Pearson data to the 2016-17 HESA data. If a link can be made, these students will be excluded from the output.

PEARCOMPERIODSTART1-3

11. These fields contain the instance period start dates (PERIODSTARTs) for the student. They are only relevant to providers making a HESA AP student return.

PEARCOMQUAL1_1-3_2

12. These fields contain the qualifications awarded (QUALs) for the student as present in the HESA data. For example, PEARCOMQUAL1_1 will contain the first qualification awarded associated with PEARCOMPERIODSTART1.

PEARCOMHESAPOP

13. This field identifies students who are in the HESA data for the provider and whether they have been awarded a Pearson qualification.

Value	Description	Definition
2	Student is in the HESA data for the provider and has been awarded a Pearson qualification	A record was found in the HESA data for the provider and one of PEARCOMQUAL1_3-3_2 = J30, C30
1	Student is in the HESA data for the provider but has not been awarded a Pearson qualification	A record was found in the HESA data for the provider and not above.
0	Otherwise	Otherwise

PEARCOMPEARQUAL1-2

14. This field identifies students who have received a valid qualification in the Pearson data.

Value	Description	Definition
J30	Student has received a HND qualification in the Pearson data	A record was found in the Pearson data where PEAR1718QUAL = J30 and PEAR1718STATUS = Complete and PEAR1718GRADE = D, M, P
C30	Student has received a HNC qualification in the Pearson data	A record was found in the Pearson data where PEAR1718QUAL = C30 and PEAR1718STATUS = Complete and PEAR1718GRADE = D, M, P
0	Otherwise	Otherwise

PEARCOMPEARPOP

15. This field identifies students who are in the Pearson data for the provider and whether they have a qualification recorded.

Value	Description	Definition
2	Student is in the Pearson data for the provider and they have a qualification recorded	A record was found in the Pearson data for the provider and PEARCOMPEARQUAL1-2 = J30, C30
1	Student is in the Pearson data for the provider and they do not have a qualification recorded	A record was found in the Pearson data for the provider and not above.
0	Otherwise	Otherwise

PEARCOMQUALMATCH

16. This field identifies students who have received the same qualification in the Pearson and the HESA data.

Value	Description	Definition
1	Student's recorded qualification type in the HESA data is the same as in the Pearson data	PEARCOMQUAL1_1 = PEARCOMPEARQUAL1
0	Otherwise	Otherwise

Using the individualised file

17. When working through this document it is necessary to use the individualised file PEARCOM17_DCT_AXXXXXXXX_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.

Re-building the 2017-18 Pearson qualification comparison using the individualised file

18. The summary figures derived from HESA data shown in the PEARCOM17_DCT_AXXXXXXXX.xlsx workbooks are listed below. For each item we describe how to identify the records in the individualised file PEARCOM17_DCT_AXXXXXXXX_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 recorded in both HESA and Pearson data as receiving the same qualification	PEARCOMHESAPOP = 2 and PEARCOMPEARPOP = 2 PEARCOMQUALMATCH = 1
Students recorded with a qualification in the Pearson data but not appearing in the HESA data	PEARCOMHESAPOP = 0 and PEARCOMPEARPOP = 2
Students recorded with a qualification in the HESA data but that qualification does not appear in the Pearson Data	PEARCOMHESAPOP = 2 and PEARCOMPEARPOP = 1
Student recorded with a qualification in the Pearson data but that qualification does not appear in the HESA data	PEARCOMHESAPOP = 1 and PEARCOMPEARPOP = 2
Students recorded in both HESA and Pearson data as receiving a different qualification.	PEARCOMHESAPOP = 2 and PEARCOMPEARPOP = 2 PEARCOMQUALMATCH = 0