

Office for
Students



2018-19 HESA data checking tool:

Pearson qualification comparison technical document

Enquiries to apdata@officeforstudents.org.uk

Publication date August 2019

Contents

2018-19 Pearson qualification comparison algorithms	3
Purpose	3
2018-19 HESA fields used in the Pearson qualification comparison algorithms	3
Table 1 HESA fields used in the Pearson qualification comparison	3
Description of derived fields for re-creating tables	5
Instance period data	5
Linking	5
Table 2 Pearson qualification comparison derived fields	5
PEARCOMPERIODSTART1-3	5
PEARCOMQUAL1_1-3_2	6
PEARCOMHESAPOP	6
PEARCOMPEARQUAL1-2	6
PEARCOMPEARPOP	6
PEARCOMQUALMATCH	7
Re-building the 2018-19 Pearson qualification comparison using the individualised file	8

2018-19 Pearson qualification comparison algorithms

Purpose

1. This document details the algorithms applied to the 2018-19 HESA AP Student record which allow for 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 2018-19 to hand when using this document.
3. The purpose of this comparison is to ensure that the 2018-19 HESA AP Student data is consistent with the Pearson qualification records. For each provider we check:
 - whether all students who received a qualification from Pearson during the 2018-19 academic year are included in the 2018-19 HESA AP Student data as having a qualification
 - whether all students who are recorded as having achieved a Pearson qualification in the 2018-19 HESA AP Student data appear in the Pearson data
 - whether the type of qualification recorded in the 2018-19 HESA AP Student data matches the type of qualification in the Pearson data.
4. We compare the HESA AP Student data with the Pearson qualification data for all students who are active in the 2018-19 academic year. This will include students whose instance period begins in the 2017-18 academic year and continues into the 2018-19 academic year.

2018-19 HESA fields used in the Pearson qualification comparison algorithms

5. Only certain fields, detailed in Table 1, are 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.

Using the individualised file

7. When working through this document it is necessary to use the individualised file PEARCOM18_DCT_AXXXXXXXXX_IND.csv*, where XXXXXXXXX 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 Pearson qualification comparison

Entity	Field	Description	Dataset
Provider	UKPRN	UK Provider Reference number	HESA
Student	HUSID	HESA unique student identifier	HESA
Student	OWNSTU†	Provider's own student identifier	HESA
Instance	NUMHUS	HESA Student Instance identifier	HESA

Entity	Field	Description	Dataset
Instance	OWNINST†	Provider's own instance identifier	HESA
Instance	QUAL	Qualification Awarded	HESA
Instance Period	PERIODSTART	Instance period start date	HESA
Course	CTITLE	Course title	HESA
Course	COURSEAIM	Course aim	HESA
	PEAR1819REGNO	Learner's Registration Number (Unique identifier in the Pearson data)	Pearson
	PEAR1819CTITLE	Course title from Pearson Data	Pearson
	PEAR1819CMPLTDATE	Learner's date of completion from the Pearson data	Pearson
	PEAR1819QUAL	Learner's qualification aim from the Pearson data	Pearson
	PEAR1819STATUS	Learner's outcome status	Pearson
	PEAR1819GRADE	Learner's outcome grade where a qualification has been awarded	Pearson

* The individualised file PEARCOM18_DCT_AXXXXXXXXX_IND.csv, is 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 for 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 Pearson comparison.

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 select those with a qualification ordered by:
1. Those with a HND
 2. Those with a HNC
 3. Those whose period start date (PERIODSTART) is the earliest

Linking

10. The HESA and Pearson records are linked using UKPRN, First Name(s), Last Names(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.
11. In the case that no link can be made between the 2018-19 HESA AP Student data and the Pearson data, we will attempt to link the Pearson data to the 2017-18 HESA AP Student data. If a link can be made, these students will be excluded from the output.

Table 2 Pearson qualification comparison derived fields

Derived field name	Description	Paragraph
PEARCOMHESAPOP	Identifies students in the HESA data for the provider	14
PEARCOMPEARPOP	Identifies students in the Pearson data	16
PEARCOMPERIODSTART1-3†	Instance period start date in the HESA data	12
PEARCOMQUAL1_1-3_2	Identifies the qualification awarded to the student in the HESA data	13
PEARCOMPEARQUAL1-2	Identifies the qualification awarded to the student in the Pearson data	15
PEARCOMQUALMATCH	Identifies students with matching qualification types in the HESA and Pearson data	17

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

PEARCOMPERIODSTART1-3

12. These fields contain the first three instance period start dates (PERIODSTARTs) for the student. See paragraph 8 for details on how these PERIODSTARTs are selected.

PEARCOMQUAL1_1-3_2

13. These fields contain the qualification awarded (QUALs) for the student as present in the HESA AP Student data. A maximum of two QUALs can be recorded per instance period, each qualification will be recorded in a separate column in the individualised file. For example, PEARCOMQUAL1_1 will contain the first qualification awarded associated with PEARCOMPERIODSTART1.

PEARCOMHESAPOP

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

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

PEARCOMPEARQUAL1-2

15. 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 PEAR1819QUAL = J30 and PEAR1819STATUS = Complete and PEAR1819GRADE = D, M, P
C30	Student has received a HNC qualification in the Pearson data	A record was found in the Pearson data where PEAR1819QUAL = C30 and PEAR1819STATUS = Complete PEAR1819GRADE = D, M, P
0	Otherwise	Otherwise

PEARCOMPEARPOP

16. 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 One of PEARCOMPEARQUAL1-2 = J30, C30

Value	Description	Definition
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

17. This field identifies students who have received the same qualification in the Pearson and the HESA AP Student 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

Re-building the 2018-19 Pearson qualification comparison using the individualised file

18. The summary figures derived from the HESA AP Student data shown in the PEARCOM18_DCT_XXXXXXXXX.xlsx workbooks are listed below. For each item we describe how to identify the records in the individualised file PEARCOM18_DCT_XXXXXXXXX_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 the 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 and 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
Students 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



© The Office for Students copyright 2019

This publication is available under the Open Government Licence 3.0 except where it indicates that the copyright for images or text is owned elsewhere.

www.nationalarchives.gov.uk/doc/open-government-licence/version/3/