# 2019-20 HESA Student Alternative data checking tool

## Pearson comparison 2019-20

# Contents

# 2019-20 Pearson qualification comparison algorithms

## Summary

1. This document describes the algorithms applied to the 2019-20 Higher Education Statistics Agency (HESA) Student Alternative 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 Student Alternative Record Coding manual 2019-20[1] to hand when using this document.

3. The purpose of this comparison is to ensure that the 2019-20 HESA Student Alternative 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 2019-20 academic year are included in the 2019-20 HESA Student Alternative data as having a qualification.

   b. Whether all students who are recorded as having achieved a Pearson qualification in the 2019-20 HESA Student Alternative data appear in the Pearson data.

   c. Whether the type of qualification recorded in the 2019-20 HESA Student Alternative data matches the type of qualification in the Pearson data.

4. We compare the HESA Student Alternative data with the Pearson qualification data for all students who are active in the 2019-20 academic year. This will include students whose instance period begins in the 2018-19 academic year and continues into the 2019-20 academic year.

5. The Pearson comparison outputs are formed of two files. These are:

   a. Pearson comparison workbook 'PEARCOM19_AAA_XXXXXXXX.xlsx'

   b. Pearson comparison individualised file 'PEARCOM19_AAA_XXXXXXXX_IND.csv'

   where XXXXXXXX is your provider's eight digit UK Provider Reference Number (UKPRN), and AAA identifies the output type as listed in the table below.

| File abbreviation | Source of outputs |
| --- | --- |
| DCT | Data checking tool |
| AST | Amendments submission tool |
| PCO | Post-collection outputs |

---

[1] Available at https://www.hesa.ac.uk/collection/c19054

# 2019-20 HESA fields used in the Pearson qualification comparison algorithms

6. Only certain fields, described in Table 1, are used to generate the Pearson comparison derived fields.

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

## Using the individualised file

8. When working through this document it is necessary to use the individualised file PEARCOM19_AAA_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 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 |
| 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 |
| | PEAR1920REGNO | Learner's Registration Number (Unique identifier in the Pearson data) | Pearson |
| | PEAR1920CTITLE | Course title from Pearson Data | Pearson |
| | PEAR1920CMPLTDATE | Learner's date of completion from the Pearson data | Pearson |
| | PEAR1920QUAL | Learner's qualification aim from the Pearson data | Pearson |
| | PEAR1920STATUS | Learner's outcome status | Pearson |
| | PEAR1920GRADE | Learner's outcome grade where a qualification has been awarded | Pearson |

\* The individualised file PEARCOM19_DCT_XXXXXXXX_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

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

## Instance period data

10. 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:

    a.  Those with a HND.

    b.  Those with a HNC.

    c.  Those whose period start date (PERIODSTART) is the earliest.

## Linking

11. 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.

12. In the case that no link can be made between the 2019-20 HESA Student Alternative data and the Pearson data, we will attempt to link the Pearson data to the 2018-19 HESA alternative provider (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 |
| --- | --- | --- |
| 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 |
| PEARCOMHESAPOP | Identifies students in the HESA data for the provider | 14 |
| PEARCOMPEARQUAL1-2 | Identifies the qualification awarded to the student in the Pearson data | 15 |
| PEARCOMPEARPOP | Identifies students in the Pearson data for the provider | 16 |
| 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

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

## PEARCOMQUAL1_1-3_2

14. These fields contain the qualification awarded (QUALs) for the student as present in the HESA Student Alternative 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

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

| Value | Description | Definition |
|-------|-------------|------------|
| 2 | Student is in the HESA Student Alternative data for the provider and has been awarded a Pearson qualification | A record was found in the HESA Student Alternative data for the provider and one of PEARCOMQUAL1_1-3_2 = J30, C30 |
| 1 | Student is in the HESA Student Alternative data for the provider but has not been awarded a Pearson qualification | A record was found in the HESA Student Alternative data for the provider and not above |
| 0 | Otherwise | Otherwise |

## PEARCOMPEARQUAL1-2

16. These fields identify 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 PEAR1920QUAL = J30 and PEAR1920STATUS = 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 PEAR1920QUAL = C30 and PEAR1920STATUS = Complete PEAR1920GRADE = D, M, P |
| 0 | Otherwise | Otherwise |

# PEARCOMPEARPOP

17. 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 |
| 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

18. This field identifies students who have received the same qualification in the Pearson and the HESA Student Alternative 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 2019-20 Pearson qualification comparison using the individualised file

19. The summary figures derived from the HESA Student Alternative data shown in the PEARCOM19_AAA_XXXXXXXX.xlsx workbooks are listed below. For each item we describe how to identify the records in the individualised file PEARCOM19_AAA_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 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 and PEARCOMQUALMATCH = 0 |