Office for Students

2020-21 HESA Student Alternative data checking tool

PEARCOM20 Pearson comparison technical document and rebuild instructions

Contents

Summary

1. This document describes the 2020-21 Pearson comparison files generated by the 2020-21 HESA Student Alternative data checking tool and rebuild instructions. Throughout the document, fields taken or derived from the Higher Education Statistics Agency (HESA) Student Alternative record are shown in capitals.

2. The purpose of this comparison is to ensure that the 2020-21 HESA data is consistent with Pearson qualification records. For each provider, we check:

   - whether all students who received a qualification from Pearson during the 2020-21 academic year are included in the 2020-21 HESA Student Alternative data as having a qualification

   - whether all students who are recorded as having achieved a Pearson qualification in the 2020-21 HESA Student Alternative data appear in the Pearson data

   - whether the type of qualification recorded in the 2020-21 HESA Student Alternative data matches the type of qualification in the Pearson data.

3. The 2020-21 Pearson comparison can be used by higher education providers to verify and correct their 2020-21 individualised student data before signing off the final data with HESA.

4. The 2020-21 Pearson comparison outputs are formed of two files. These are:

   - 2020-21 Pearson comparison workbook 'PEARCOM20_AAA_XXXXXXXX.xlsx'

   - 2020-21 Pearson comparison individualised file 'PEARCOM20_AAA_XXXXXXXX_IND.csv'

   where XXXXXXXX is the 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

   Details of how to use individualised files are given on the Office for Students (OfS) website (www.officeforstudents.org.uk/data-and-analysis/supplying-data/working-with-individualised-files/).

5. These files can be accessed via the Information Reporting Interface Service (IRIS) report which is available on HESA's data collection site (https://datacollection.hesa.ac.uk/). Access details will be sent to the appropriate contacts at providers when the data checking tool is available.

6. This document is aimed at readers with in-depth knowledge of the data. Readers are advised to refer to the HESA Student Alternative Record Coding Manual 2020-21 (https://www.hesa.ac.uk/collection/c20054).

7. We will document any changes or corrections to the algorithms in this document on our website (www.officeforstudents.org.uk/data-and-analysis/data-checking-tool/2020-21-hesa-student-alternative-data-checking-tool/ under 'Updates').

Changes from the 2019-20 HESA Student Alternative data checking tool

8. This section describes any changes to the algorithms since the equivalent was published for the 2019-20 algorithms in 'Pearson comparison 2019-20' available from the data checking tool pages (www.officeforstudents.org.uk/data-and-analysis/data-checking-tool/).

9. Any deletions are shown with a ~~strikethrough~~. Any other changes are shown as **bold and underlined.**.

10. There are currently no changes to the algorithms since 'Pearson comparison 2019-20'.

2020-21 Pearson comparison workbook

11. The 2020-21 Pearson comparison workbook is the Excel workbook PEARCOM20_AAA_XXXXXXXX.xlsx which contains the following worksheets:

Table 1: Worksheets in the 2020-21 Pearson comparison workbook

Worksheet*	Title
Coversheet	Pearson comparison summary figures

* This worksheet reference corresponds to the spreadsheet tabs.

2020-21 Pearson comparison individualised file

12. When working through this document it is necessary to use the individualised file, PEARCOM20_AAA_XXXXXXXX_IND.csv. The individualised file contains the values of all fields used to generate the PEARCOM20_AAA_XXXXXXXX.xlsx workbook 2020-21 Pearson comparison figures.

13. Fields taken from the HESA Student Alternative return or derived as part of the 2020-21 Pearson comparison are shown in capitals using the names given in tables 2 and 3 respectively. Fields prefixed with 'PEARCOM' are derived, all others are taken directly from the HESA Student Alternative record, or Pearson data, as indicated in Table 2.

14. Specific instructions for rebuilding the 2020-21 Pearson comparison workbook from the individualised file are provided in this document.

15. The individualised file contains one record per instance (each unique combination of UKPRN, HUSID and NUMHUS) in the latest academic year. Additionally, the individualised file will contain one record per student as recorded in the Pearson data, where that record has not been matched to an instance in the latest academic year.

HESA Student Alternative fields used to create the 2020-21 Pearson comparison

16. The HESA Student Alternative fields used to generate the 2020-21 Pearson comparison figures are listed in Table 2.

Table 2: List of HESA Student Alternative fields used to create the 2020-21 Pearson comparison

Name	Description	Data source	Column in individualised file
UKPRN	UK Provider Reference Number	HESA	A
HUSID	HESA unique student identifier	HESA	B
OWNSTU†	Provider's own student identifier	HESA	D
NUMHUS	HESA Student Instance identifier	HESA	C
OWNINST†	Provider's own instance identifier	HESA	E
QUAL	Qualification Awarded	HESA	N/A
PERIODSTART	Instance period start date	HESA	N/A
CTITLE	Course title	HESA	G
COURSEAIM	Course aim	HESA	F
PEAR2021REGNO	Learner's Registration Number (Unique identifier in the Pearson data)	Pearson	AA-AB
PEAR2021CTITLE	Course title from the Pearson data	Pearson	W-X
PEAR2021CMPLTDATE	Learner's date of completion from the Pearson data	Pearson	U-V
PEAR2021QUAL	Learner's qualification aim from the Pearson data	Pearson	AC-AD
PEAR2021STATUS	Learner's outcome status	Pearson	Y-Z
PEAR2021GRADE	Learner's outcome grade where a qualification has been awarded	Pearson	AE-AF

† 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

17. The derived fields used to generate the 2020-21 Pearson comparison are listed in Table 3.

Table 3: List of derived fields used in the 2020-21 Pearson comparison

Name	Description	Paragraph	Column in individualised file
PEARCOMPERIODSTART1-3[†]	Instance period start date in the HESA data	20	K-M
PEARCOMQUAL1_1-3_2	Identifies the qualification awarded to the student in the HESA data	21	N-S
PEARCOMHESAPOP	Identifies students in the HESA data for the provider	22	H
PEARCOMPEARQUAL1-2	Identifies the qualification awarded to the student in the Pearson data	23	AG-AH
PEARCOMPEARPOP	Identifies students in the Pearson data for the provider	24	I
PEARCOMQUALMATCH	Identifies students with matching qualification types in the HESA and Pearson data	25	J

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

Instance period data

18. 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:

    i.   Those with an HND

    ii.   Those with an HNC

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

Linking

19. 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.

20. In the case that no link can be made between the 2020-21 HESA Student Alternative data and the Pearson data, we will attempt to link the Pearson data to the 2019-20 HESA Student Alternative data. If a link can be made, and a qualification can be matched, these students will be excluded from the output.

PEARCOMPERIODSTART1-3

21. These fields contain the first three instance period start dates (PERIODSTARTs) for the student. See paragraph 17 for details on how these PERIODSTARTs are selected.

PEARCOMQUAL1_1-3_2

22. 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

23. 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

24. 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 PEAR2021QUAL = J30 and PEAR2021STATUS = Complete and PEAR2021GRADE = D, M, P
C30	Student has received a HNC qualification in the Pearson data	A record was found in the Pearson data where PEAR2021QUAL = C30 and PEAR2021STATUS = Complete PEAR2021GRADE = D, M, P
0	Otherwise	Otherwise

PEARCOMPEARPOP

25. 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

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

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

Rebuild instructions

27. The individualised file, PEARCOM20_AAA_XXXXXXXX_IND.csv, can be used to interrogate the data and our calculation of derived fields by filtering to certain groups of records. Full details of how to access and use individualised files are given on our website (www.officeforstudents.org.uk/data-and-analysis/supplying-data/working-with-individualisedfiles).

28. Specific instructions to rebuild the 2020-21 Pearson comparison summary figures shown in the output workbook, PEARCOM20_AAA_XXXXXXXX.xlsx, are provided here.

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 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