

**Office for
Students**



2022-23 Student data data checking tool

**2022-23 Pearson comparison technical
document and rebuild instructions**

Enquiries to hesastudentdata@officeforstudents.org.uk

Publication date 7 September 2023

Contents

| | |
|---|-----------|
| Summary | 3 |
| Changes from the 2021-22 HESA Student Alternative data checking tool | 5 |
| 2022-23 Pearson comparison data files | 6 |
| Workbook | 6 |
| Individualised file | 6 |
| Fields used to create the 2022-23 Pearson comparison | 7 |
| Description of derived fields | 8 |
| Rebuild instructions | 10 |

Summary

1. This document describes the Pearson comparison files generated by the Student (22056) record, the algorithms used to generate the data, and rebuild instructions. Throughout the document, fields taken or derived from the Student (22056) record or the Pearson data are shown in capitals.
2. The Pearson comparison can be used by higher education providers to verify and correct their 2022-23 individualised student data before signing off the final data with the Designated Data Body (DDB).
3. The purpose of this comparison is to ensure that the Student (22056) record data is consistent with the Pearson qualification records. For each provider we check:
 - whether all students who received a qualification from Pearson during the 2022-23 academic year are included in the Student (22056) record as having a qualification.
 - whether all students who are recorded as having achieved a Pearson qualification in the Student (22056) record also have a qualification in the Pearson data.
 - whether the type of qualification recorded in the Student (22056) record matches the type of qualification in the Pearson data.
4. The Student (22056) and Pearson records are linked using UKPRN, First Name(s), Last Name(s), Date of Birth, and if possible qualification aim and start date. A link is defined as a match of UKPRN, First Name(s), Last Name(s), Date of Birth, and if possible, qualification aim and start date. Allowances are made for missing or incorrectly ordered names and start dates not matching by less than 40 days.
5. The 2022-23 Pearson comparison is formed of two files. These are:
 - Pearson comparison workbook 'PEARCOM22_DCT_XXXXXXXX.xlsx',
 - Pearson comparison individualised file 'PEARCOM22_DCT_XXXXXXXX_IND.csv'

where XXXXXXXX is your provider's eight-digit UK Provider Reference Number (UKPRN).

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

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

7. This document is aimed at readers with in-depth knowledge of the data. Readers are advised to refer to a copy of the Student (22056) record coding manual¹ when using this document.
8. We will document any changes or corrections to the algorithms in this document in the data checking tool section of the OfS website.²

¹ See [Student record – 22056 coding manual](#).

² See [The OfS website – Data checking tool](#).

Changes from the 2021-22 HESA Student Alternative data checking tool

9. Derivations for some derived fields have been updated to take in to account the change in data model from the previous HESA Student Alternative data collection to the new Student (22056) model. More information on our general approach to converting algorithms to the Student (22056) model can be found on the 'Data checking tool' pages of the OfS website under 'Data futures'.³
10. The change in data source means this output is now created for all providers that submit student data to the DDB, rather than just those that submitted data to the Student Alternative Record.
11. Unlike previous Pearson Comparison outputs, if no link between the Student (22056) record and the Pearson data we won't attempt to link the Pearson data to the previous year's student data (in this case the 2021-22 Student / Student Alternative Record).
12. PEARCOMPERIODSTART1-3 is no longer included in the individualised file due to differences in the structure between the Student Alternative record and the Student (22056) record.
13. The algorithms for PEARCOMPEARPOP, PEARCOMPEARQUAL1-2 and PEARCOMDDBPOP have been updated to account for the fact that HND is coded as J0003 and HNC is coded as C0001.
14. CENTREREF is now included in the individualised file to allow for easier identification of students.

³ See [The OfS website - Data checking tool](#).

2022-23 Pearson comparison data files

Workbook

15. The 2022-23 Pearson comparison workbook is the Excel workbook PEARCOM22_DCT_XXXXXXXX.xlsx which contains the following worksheets:

Table 1: Worksheets in the 2022-23 Pearson comparison workbook

| Worksheet* | Title |
|------------|------------------------------------|
| Coversheet | Pearson comparison summary figures |

* This worksheet reference corresponds to the spreadsheet tabs.

Individualised file

16. When working through this document it is necessary to use the individualised file, PEARCOM22_DCT_XXXXXXXX_IND.csv. The individualised file contains the values of all fields used to generate the Pearson comparison.

17. Fields taken from the Student (22056) record, taken from the Pearson data or derived as part of the 2022-23 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 Student (22056) record or Pearson data as indicated in Table 2.

18. Specific instructions for rebuilding the 2022-23 Pearson comparison workbook from the individualised file are provided in this document.

19. The individualised file contains one row per engagement (each unique combination of UKPRN, SID 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 engagement in the latest academic year.

20. The number of qualifications awarded that can be returned for a given student engagement is unbounded; however, due to file size considerations we restrict the number of separate qualifications awarded per engagement displayed in the individualised file to three. In the unusual case where an engagement has more than three qualifications awarded, all will be considered in our algorithms, although only three will be displayed in the individualised file. In selecting qualifications awarded data to show, we will select those ordered by:

- a. Those with a HND
- b. Those with a HNC
- c. Those with the highest level of study (excluding HNDs and HNCs).

Fields used to create the 2022-23 Pearson comparison

21. The Student (22056) and Pearson fields used to generate the data summary are listed in Table 2.

Table 2: List of Student (22056) and Pearson fields used in the 2022-23 Pearson comparison

| Source | Entity | Name | Description |
|---------|---------------|--------------------------------|---|
| Student | Engagement | UKPRN | UK Provider Reference Number |
| Student | Engagement | SID | Student identifier |
| Student | Engagement | OWNSTU [†] | Provider's own student identifier |
| Student | Engagement | NUMHUS | Engagement number |
| Student | Engagement | OWNENGID [†] | Provider's own engagement identifier |
| Student | Engagement | ENGSTARTDATE | Engagement start date |
| Student | Qualification | QUALCAT | Records the qualification category |
| Student | Course* | COURSETITLE [†] | Course title |
| Student | Course* | COURSEID [†] | Course identifier |
| Pearson | | PEAR2223CENTREREF [†] | Student identifier from the Pearson data |
| Pearson | | PEAR2223REGNO | Learner's registration number (unique identifier in the Pearson data) |
| Pearson | | PEAR2223COMMENCEMENT DATE | Learner's enrolment date |
| Pearson | | PEAR2223CTITLE | Course title from Pearson data |
| Pearson | | PEAR2223QUAL | Learner's qualification aim from the Pearson data |
| Pearson | | PEAR2223STATUS | Learner's outcome status |
| Pearson | | PEAR2223GRADE | Learner's outcome grade where a qualification has been awarded |

[†] These fields are not used in the comparison but are included in the individualised file to allow for easy identification of students.

*The values of these fields are taken from the latest student course session associated with the student's engagement.

Description of derived fields

22. The derived fields used to generate the 2022-23 Pearson comparison are listed in Table 3.

Table 3: List of derived fields used in the 2022-23 Pearson comparison

| Name | Description |
|--------------------|--|
| PEARCOMQUAL1-3 | Identifies the qualifications awarded to the student in the Student (22056) record |
| PEARCOMDDBPOP | Identifies students in the Student (22056) record for the provider |
| PEARCOMPEARQUAL1-2 | Identifies the qualification awarded to the student in the Pearson data |
| PEARCOMPEARPOP | Identifies students in the Pearson data for the provider |
| PEARCOMQUALMATCH | Identifies students with matching qualification types in the Student (22056) record and Pearson data |

PEARCOMQUAL1-3

23. These fields contain the qualifications awarded (QUALCATs) for the student as present in the Student (22056) record. See paragraph 20 for details on how these QUALCATs are selected.

PEARCOMDDBPOP

24. This field identifies students who are in the Student (22056) record for the provider and whether they have been awarded a Pearson qualification.

| Value | Description | Definition |
|-------|--|--|
| 2 | Student is in the Student (22056) record for the provider and has been awarded a Pearson qualification | A record was found in the Student (22056) record for the provider and one of PEARCOMQUAL1-3 = J0003, C0001 |
| 1 | Student is in the Student (22056) record for the provider but has not been awarded a Pearson qualification | A record was found in the Student (22056) record for the provider and not above |
| 0 | Student is not in the Student (22056) record for the provider. | Otherwise |

PEARCOMPEARQUAL1-2

25. These fields identify students who have received a valid qualification in the Pearson data.

| Value | Description | Definition |
|-------|--|---|
| J0003 | Student has received a HND qualification in the Pearson data | A record was found in the Pearson data where PEAR2223QUAL = J0003 and PEAR2223STATUS = Complete and PEAR2223GRADE = D, M, P |

| Value | Description | Definition |
|-------|---|--|
| C0001 | Student has received a HNC qualification in the Pearson data | A record was found in the Pearson data where PEAR2223QUAL = C0001 and PEAR2223STATUS = Complete PEAR2223GRADE = D, M, P |
| 0 | Student has not received a HND or HNC qualification in the Pearson data, or is not included in the Pearson data | Otherwise |

PEARCOMPEARPOP

26. 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= J0003, C0001 |
| 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 | Student is not in the Pearson data for the provider | Otherwise |

PEARCOMQUALMATCH

25. This field identifies students who have received the same qualification in the Pearson data and the Student (22056) record.

| Value | Description | Definition |
|-------|--|---|
| 1 | Student's recorded qualification type in the Student (22056) record is the same as in the Pearson data | One of PEARCOMQUAL1-3 = one of PEARCOMPEARQUAL1-2 |
| 0 | Student's recorded qualification type in the Student (22056) record if not the same as in the Pearson data | Otherwise |

Rebuild instructions

27. The individualised file, PEARCOM22_DCT_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 the OfS website (www.officeforstudents.org.uk/data-and-analysis/supplying-data/working-with-individualised-files/).
28. Specific instructions to rebuild the 2022-23 Pearson comparison summary figures shown in the Pearson comparison workbook, PEARCOM22_DCT_XXXXXXXX.xlsx, are provided here.

| Item in the workbook | Derived field selection in individualised file |
|---|---|
| Student engagement recorded in both the Student (22056) record and Pearson data as receiving the same qualification | PEARCOMDDBPOP = 2 and PEARCOMPEARPOP = 2 and PEARCOMQUALMATCH = 1 |
| Student engagement recorded with a qualification in the Pearson data but not appearing in the Student (22056) record | PEARCOMDDBPOP = 0 and PEARCOMPEARPOP = 2 |
| Student engagement recorded with a qualification in the Student (22056) record but that qualification does not appear in the Pearson data | PEARCOMDDBPOP = 2 and PEARCOMPEARPOP = 1 |
| Student engagement recorded with a qualification in the Pearson data but that qualification does not appear in the Student (22056) record | PEARCOMDDBPOP = 1 and PEARCOMPEARPOP = 2 |
| Student engagement recorded in both Student (22056) record and Pearson data as receiving a different qualification | PEARCOMDDBPOP = 2 and PEARCOMPEARPOP = 2 PEARCOMQUALMATCH = 0 |



© The Office for Students copyright 2023

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/