



2017-18 data checking tools

**Higher Education Statistics Agency
Student data, Higher Education Statistics
Agency Alternative Provider Student
data and Individualised Learner Record data**

Reference OfS 2018.30

Enquiries to

HESA Student data: HESAStudentData@officeforstudents.org.uk (for Office for Students outputs) and Data_Queries@re.ukri.org (for Research England outputs)

HESA Alternative Provider Student data: APData@officeforstudents.org.uk

ILR data: ILRData@officeforstudents.org.uk

Publication date 25 July 2018

Contents

Summary	3
Section 1: 2017-18 HESA Student data checking tool	4
Action required	4
Data checking tool outputs and release schedule	4
Data verification questions	5
Key assumptions	6
Data confidentiality	6
Further information	6
Section 2: 2017-18 HESA Alternative Provider Student data checking tool	8
Action required	8
Data checking tool outputs and release schedule	8
Data verification questions	9
Data confidentiality	10
Further information	10
Section 3: 2017-18 ILR data checking tool	11
Action required	11
Data checking tool outputs and release schedule	11
Data verification questions	13
Key assumptions	13
Data confidentiality	13
Further information	14
Explanation of abbreviations and terms	15

Summary

1. This document notifies providers of the forthcoming release of the Office for Students (OfS) data checking tools for 2017-18 Higher Education Statistics Agency (HESA) Student data, 2017-18 HESA Alternative Provider Student data and the Education and Skills Funding Agency's (ESFA's) 2017-18 Individualised Learner Record (ILR).
2. The data checking tools will produce a number of outputs that can be used by higher education providers to verify, and where appropriate correct, their 2017-18 individualised student data before signing the final data off with HESA or the ESFA. In some cases, these outputs use data that was submitted to the Higher Education Funding Council for England (HEFCE) as a comparison to help providers verify their individualised data. The OfS HESA Student data checking tool will also include an output which has been produced by Research England.
3. The OfS and Research England strongly encourage the use of the data checking tools, which we regard as an essential element of all providers' data quality processes. They will help providers to:
 - return accurate data to HESA and the ESFA
 - verify the accuracy of the derived fields that may be used for regulatory purposes
 - respond to data verification questions raised by OfS and Research England staff
 - reduce the likelihood of selection for data assurance activity for 2017-18
 - identify errors in or discrepancies with in-year OfS data returns.
4. Data submitted to HESA or the ESFA and signed off as correct by the accountable officer will be used by the OfS for a number of regulatory purposes and by Research England for funding purposes. The OfS, and Research England where relevant, will only accept amendments to the data after this point in exceptional circumstances, where any data errors identified are widespread and significant and make a material difference to the OfS's or Research England's use of the data. For further information regarding data amendments, please see the 'Amendments to data' page of the OfS website (www.officeforstudents.org.uk/data-and-analysis/amendments-to-data/).
5. Further information regarding each specific data collection and the corresponding data checking tool, as well as release schedules and contact information, are listed in each section of this publication. We will notify the relevant contacts at providers when the tools are initially released, and when the outputs are made available.

Section 1: 2017-18 HESA Student data checking tool

Action required

6. Higher education institutions will be able to access the 2017-18 data checking tool from the Information Reporting Interface Service (IRIS) report which is available on HESA's data collection site.
7. We strongly encourage providers to analyse the data checking tool outputs as part of their HESA Student data quality processes. OfS staff may query, via HESA's Minerva data quality system, potential errors in the HESA data that could affect its use. If, during the quality review process, errors are found to have occurred in previous data returns, we expect providers to take these into account when preparing future returns.
8. We would be grateful for feedback on the algorithms used to create the data checking tool outputs, in particular where they have changed from previous years. The algorithms used for each output will be made available on the OfS or Research England websites once released. Feedback for OfS outputs should be emailed to HESAStudentData@officeforstudents.org.uk and for Research England to Data_Queries@re.ukri.org.

Data checking tool outputs and release schedule

9. The 2017-18 HESA Student data checking tool will produce the following outputs:
 - Higher Education Students: Early Statistics survey 2017 (HESES17) comparison
 - Quality-related Research (QR) Research Degree Programme (RDP) supervision fund data summary (provided by Research England).
 - 'HEFCE/OfS-fundable'¹ full-time equivalences for the Transparent Approach to Costing for Teaching (TRAC(T)) data summary
 - student premium allocations data comparison

Additional outputs not listed above may be released as part of the 2017-18 HESA Student data checking tool. Information regarding such outputs will be sent to contacts at providers.
10. The data checking tool will not provide outputs that explicitly derive institutional performance metrics, such as those used in the Teaching Excellence and Student Outcomes Framework (TEF). HESA and the OfS encourage institutions to pay particular attention to the credibility checks and frequency counts that HESA's 2017-18 Student data collection provides in relation to key data items. These include credibility checks on those data items which act as benchmarking factors, or that define 'splits' and populations of the TEF metrics.

¹ The Office for Students was established in April 2018. The TRAC(T) data summary shows the full-time equivalences of students who were HEFCE-fundable. From April 2018 these students became OfS-fundable.

HESES17 comparison

11. The HESES17 comparison compares 2017-18 HESA Student data with the provider's HESES17 data returned to HEFCE at the start of the 2017-18 academic year. There is also a comparison between the provider's latest 2018-19 OfS teaching grant and the grant as it would have been if calculated using the 2017-18 HESA Student data.
12. The algorithms used to generate the HESES17 comparison output are intended to be those used for the post-collection outputs next spring. However, we may make changes where we believe these will improve our algorithms.
13. The HESES17 comparison will be released in the **week commencing 13 August 2018**.

QR RDP supervision fund data summary

14. This output is provided by Research England and applies the 2018-19 QR RDP supervision fund method to the 2017-18 HESA student data, to illustrate the effect that changes to HESA student data may have on the QR RDP supervision fund.
15. The QR RDP supervision fund data summary will be released in the **week commencing 13 August 2018**.
16. For queries relating to the QR RDP supervision fund data summary please contact Research England by email at Data_Queries@re.ukri.org.

HEFCE/OfS-fundable full-time equivalences for the TRAC(T) data summary

17. The TRAC(T) data summary provides the HEFCE/OfS-fundable full-time equivalent student numbers by cost centre and price group using the 2017-18 HESA student data. It allows providers to check the data that will be used in calculating the subject-related full average costs of teaching a HEFCE/OfS-fundable student in their TRAC(T) return.
18. The TRAC(T) data summary will be released in the **week commencing 20 August 2018**.

Student premium data comparison

19. The student premium data comparison applies the 2018-19 student premium funding methods to the 2017-18 HESA Student data, to illustrate the effect that changes to HESA Student data may have on student premium funding allocations. It also compares headcounts of students with particular characteristics derived from the 2016-17 and 2017-18 HESA Student data to allow providers to check the accuracy of the data we may use to calculate the 2019-20 student premium allocations.
20. The student premium data comparison will be released in the **week commencing 3 September 2018**.

Data verification questions

21. In addition to the data checks that providers undertake using the reports and outputs, data quality analysts and specialists at HESA, the OfS and Research England will also examine the data, although responsibility for providing data that is fit for purpose remains entirely the responsibility of the provider. If any anomalies are found, HESA, the OfS and Research England will raise queries through the HESA Minerva data quality database. Providers are

required either to resubmit data to amend these anomalies, or to provide an explanation, through Minerva, of why these apparent anomalies are genuine. This is an iterative process during which providers may need to submit, commit, review and decommit their data several times to ensure the final submission is credible. We expect all providers to respond to these queries before the sign-off deadline of **2 November 2018**. We will not allow providers' data to be set to 'credible' by HESA until all Minerva queries have been responded to or are cleared by submission of amended data.

22. We recognise that at the early stages of our verification activity the data uploaded to HESA is unlikely to be the final version of the record. However, dealing with issues at an early stage is particularly important because of the potential impact of incorrect data being submitted to HESA may have on a provider.

Key assumptions

23. The data checking tool calculations will incorporate any HESA Student data amendments for 2016-17 and earlier years that were approved by the OfS and Research England data amendment panel, and signed off by providers by 29 June 2018. We will not update the data checking tool outputs to reflect any amendments or overrides that may be signed off later.

Data confidentiality

24. OfS and Research England staff will access data processed via the data checking tool to support providers in returning high-quality data to HESA. Such support may include answering queries from providers about their outputs and asking data verification questions as outlined in paragraphs 21 and 22. We do not intend to use data processed by the data checking tool to make decisions about individual providers, but may use responses to data verification questions in our assessments of data quality. Further information on how we handle personal data can be provided on request.

25. OfS and Research England staff may examine outputs from the data checking tool in any of the following circumstances:

- to ask data verification questions and consider providers' responses to them
- to assist providers with queries about their outputs and data
- to test the functionality of forthcoming outputs and the application of overrides
- to produce early modelling
- to verify future data returns
- where a provider explicitly gives permission.

Further information

26. Further information about how to use the 2017-18 HESA Student data checking tool and the outputs will be made available on the OfS and Research England websites in due course.

27. For any further information or guidance please email HESASStudentData@officeforstudents.org.uk (for Office for Students outputs) and Data_Queries@re.ukri.org (for the QR RDP supervision fund data summary output provided by Research England).

Section 2: 2017-18 HESA Alternative Provider Student data checking tool

Action required

28. Alternative providers of higher education will be able to access the 2017-18 data checking tool from the IRIS report which is available on HESA's data collection site.
29. We strongly encourage providers to analyse the data checking tool outputs as part of their HESA Alternative Provider Student data quality processes.

Data checking tool outputs and release schedule

30. The 2017-18 HESA Alternative Provider Student data checking tool generates the following outputs:

- Higher Education in Alternative Providers Early Statistics survey 2017 (HEAPES17) comparison
- 2017-18 final year outcome metrics
- 2017-18 Student Loans Company (SLC) comparison
- 2017-18 Pearson qualification comparison.

Additional outputs not listed above may be released as part of the 2017-18 HESA Alternative Provider Student data checking tool. Information regarding such outputs will be sent to contacts at providers.

31. The data checking tool will not provide outputs that explicitly derive institutional performance metrics, such as those used in the TEF. HESA and the OfS encourage institutions to pay particular attention to the credibility checks and frequency counts that HESA's 2017-18 Alternative Provider Student data collection provides in relation to key data items. These include credibility checks on those data items which act as benchmarking factors, or that define 'splits' and populations of the TEF metrics.

HEAPES17 comparison

32. The HEAPES17 comparison compares the original HEAPES17 data submitted with a recreation of the survey from the 2017-18 HESA Alternative Provider Student data.

33. The output will contain illustrative recruitment against the 2017-18 student number controls. Once the data has been signed off we will produce a final comparison against the provider's HEAPES17 return using this data. This will be used to determine whether the provider has recruited above its 2017-18 student number controls.

34. The algorithms used to generate the outputs are generally expected to be those used in the final HEAPES17 comparison. We may, however, make changes where we believe these will improve the algorithms.

35. The HEAPES17 comparison will be released in the **week commencing 13 August 2018**.

2017-18 final year outcome metrics

36. The Department for Education (DfE) has asked the OfS to provide metrics which analyse the outcomes of full-time students who reach their final year of study. The DfE has indicated that it will use these metrics, together with the UK performance indicators published by HESA, to assess providers' academic performance.
37. The algorithms used to generate the outputs are generally expected to be the same as the final metrics, which will be shared with the DfE once the 2017-18 HESA Alternative Provider Student data is signed off. We may, however, make changes where we believe these will improve the algorithms.
38. The 2017-18 final year outcome metrics will be released in the **week commencing 13 August 2018**.

2017-18 SLC comparison

39. This output compares a provider's 2017-18 HESA Alternative Provider Student data with records the SLC holds to establish whether there are inconsistencies between the two, with regard to students receiving student support.
40. The 2017-18 SLC comparison will be released in the **week commencing 15 October 2018**.

2017-18 Pearson qualification comparison

41. This output compares a provider's 2017-18 HESA Alternative Provider Student data with records Pearson holds to establish where there are inconsistencies between the two, with regard to students receiving Pearson qualifications.
42. The 2017-18 Pearson qualification comparison will be released in the week commencing **15 October 2018**.

Data verification questions

43. In addition to the data checks that providers undertake using the reports and outputs, data quality analysts and specialists at HESA and the OfS will also examine the data, although responsibility for providing data that is fit for purpose remains entirely the responsibility of the provider. If any anomalies are found, HESA and the OfS will raise queries through the HESA Minerva data quality database. Providers are required either to resubmit data to amend these anomalies, or to provide an explanation, through Minerva, of why these apparent anomalies are genuine. This is an iterative process during which providers may need to submit, commit, review and decommit their data several times to ensure the final submission is credible. We expect all providers to respond to these queries before the sign-off deadline of **21 November 2018**. We will not allow providers data to be set to credible by HESA until all Minerva queries have been responded to or are cleared by submission of amended data.
44. We recognise that at the early stages of our verification activity the data uploaded to HESA is unlikely to be the final version of the record. However, dealing with issues at an early stage is particularly important because of the potential impact of incorrect data being submitted to HESA may have on an organisation.

Key assumptions

45. The data checking tool calculations will incorporate any HESA AP Student data amendments for 2016-17 and earlier years that were approved by the OfS and Research England data amendment panel, and signed off by providers by 29 June 2018. We will not update the data checking tool outputs to reflect any amendments or overrides that may be signed off later.

Data confidentiality

46. OfS staff will access data processed via the data checking tool to support providers in returning high quality data to HESA. Such support may include answering queries from providers about their outputs and asking data verification questions as outlined in paragraphs 43 and 44. We do not intend to use data processed by the data checking tool to make decisions about individual providers, but may use responses to data verification questions in our assessments of data quality. Further information on how we handle personal data can be provided on request.

47. OfS staff may examine outputs from the data checking tool in any of the following circumstances:

- to ask data verification questions and consider providers' responses to them
- to assist providers with queries about their outputs and data
- to test the functionality of forthcoming outputs and the application of overrides
- to produce early modelling
- to verify future data returns
- where a provider explicitly gives permission.

48. The outputs generated from a provider's SLC and Pearson return should be kept in confidence. In viewing this output, through HESA's Minerva Data Quality System, a provider accepts this commitment of confidentiality.

Further information

49. Further information about how to use the 2017-18 HESA Alternative Provider Student data checking tool and the outputs will be made available on the OfS website in due course.

50. For any further information or guidance please email APData@officeforstudents.org.uk.

Section 3: 2017-18 ILR data checking tool

Action required

51. Further education and sixth form colleges will be able to access the 2017-18 ILR data checking tool through the OfS portal (<https://extranet.officeforstudents.org.uk/Data>). Access to the data checking tool will be given via an access key which will be posted to the principal at each college shortly before release.
52. We strongly encourage colleges to analyse the data checking tool outputs as part of their ILR data quality processes. OfS staff may query potential errors in the ILR data that could affect its use. If, during the quality review process, errors are found to have occurred in previous data returns, we expect providers to take these into account when preparing future returns.
53. We would be grateful for feedback on the algorithms used to create the data checking tool outputs, in particular where they have changed from previous years. The algorithms used for each output will be made available on the OfS website once released. Feedback should be emailed to ILRData@officeforstudents.org.uk.

Data checking tool outputs and release schedule

54. The 2017-18 ILR data checking tool will produce the following outputs:
 - Higher Education in Further Education: Students survey 2017 (HEIFES17) comparison
 - student characteristics data summary
 - Unistats continuation summary
 - National Student Survey (NSS) target list
 - Graduate Outcomes survey target list

Additional outputs not listed above may be released as part of the 2017-18 ILR data checking tool. Information regarding such outputs will be sent to contacts at colleges.

HEIFES17 comparison

55. The HEIFES17 comparison compares 2017-18 ILR data with the college's HEIFES17 data returned to HEFCE at the start of the 2017-18 academic year. The output includes a comparison between the college's latest 2018-19 OfS teaching grant and grant modelled using the 2017-18 ILR data.
56. The algorithms used to generate the HEIFES17 comparison output are intended to be those used for the post-collection outputs next spring. However, we may make changes where we believe these will improve our algorithms.
57. The HEIFES17 comparison will be released in the **week commencing 30 July 2018**.

Student characteristics data summary

58. The student characteristics data summary identifies fields taken from the ILR that are used in the construction of institutional performance metrics and their associated benchmarks. These are fields that classify a student's background or other characteristics for use in the benchmarking process for quality and institutional performance measurements, such as the TEF.
59. The data summary is provided to help identify potential errors and reduce the numbers of unknown or unpopulated student characteristics in ILR data that will affect our uses of the data.
60. We are not providing a separate student premium allocations output for the 2017-18 ILR data checking tool. The student characteristics calculations that would have been in the student premium allocations output are contained in this output. These student characteristics may inform the basis of the 2019-20 student premium allocations.
61. The student characteristics data summary will be released in the **week commencing 30 July 2018**.

Unistats continuation summary

62. We intend to publish a summary of the continuation statistic for Unistats from the 2017-18 ILR R14 data in 2018. The data checking tool provides an opportunity for colleges to check, before submission, that their ILR R14 data is fit for publication on the Unistats website.
63. This output uses data submitted to the data checking tool and the ILR data for the 2016-17 academic year submitted to ESFA, to calculate whether entrants in 2016-17 continued to study, qualified, are taking a break from study, or left a course before completion. This allows providers to see how submitted data would affect the Unistats website and to ensure consistency with data collected in the previous year.
64. Information on entry qualifications and classification of degree outcomes for Unistats can now be found in the student characteristics output.
65. The Unistats continuation summary will be released in the **week commencing 3 September 2018**.

NSS target list

66. The 2017-18 ILR data submitted to the data checking tool will be used to produce a provisional target list of students to be included in the 2019 NSS. Colleges are encouraged to use these lists to start preparing the contact details for these students.
67. These will be passed to the agency running the survey. These contact details will need to be returned to the agency in mid-November 2018. Further details about the arrangements for the 2019 NSS will be published in due course.
68. On **Friday 19 October 2018** the OfS will extract final target lists for the 2019 NSS from data submitted to the data checking tool. Colleges should therefore ensure that the most recent data submitted to the data checking tool on this date generates a complete target list. Once we have

received the 2016-17 ILR R14 data from the ESFA, we will use it to validate colleges' target lists. OfS staff will routinely access NSS target lists generated by the data checking tool.

69. The NSS target list output will be released in the **week commencing 3 September 2018**.

Graduate Outcomes survey target list

70. The 2017-18 ILR data submitted to the data checking tool will be used to produce provisional target lists of students to be included in the 2017-18 Graduate Outcomes survey. Colleges are encouraged to use these lists to start preparing the contact details for these students and to check that the data submitted to ESFA will generate a valid target list.

71. Once we have received the 2017-18 ILR R14 data from the ESFA, we will use it to validate colleges' target lists.

72. The Graduate Outcomes survey target list output will be released in the **week commencing 17 September 2018**.

Data verification questions

73. In addition to the data checks that colleges undertake using the reports and outputs, data specialists at the OfS will also examine the data, although responsibility for providing data that is fit for purpose remains entirely the responsibility of the college. If any anomalies are found, the OfS will raise queries via email. Colleges are required either to resubmit data to amend these anomalies, or to provide an explanation of why these apparent anomalies are genuine. This is an iterative process during which colleges may need to submit and review their data several times to ensure the final submission is credible. We expect all providers to respond to these queries before the ESFA submission deadline of **18 October 2018**.

74. We recognise that at the early stages of our verification activity the data uploaded to the OfS portal is unlikely to be the final version of the record. However, dealing with issues at an early stage is particularly important because of the potential impact of incorrect data being submitted to ESFA may have on an organisation.

Key assumptions

75. The data checking tool calculations will incorporate any ILR amendments for 2016-17 and earlier years that were approved by the OfS and Research England data amendment panel and signed off by providers by 29 June 2018. We do not intend to update the data checking tool outputs to reflect any amendments or overrides that may be signed off later.

76. The version of the Learning Aims Search database that we refer to is taken from 20 July 2018. Although we are not using the total qualification time field from this database in any of the data checking tool outputs, we may use this field in our algorithms in future, including in calculations for registration and ongoing conditions of registration.

Data confidentiality

77. OfS staff will access data processed via the data checking tool to support providers in returning high-quality data to the ESFA. Such support may include answering queries from providers

about their outputs and asking data verification questions as outlined in paragraphs 73 and 74. We do not intend to use data processed by the data checking tool to make decisions about individual providers, but may use responses to data verification questions in our assessments of data quality. Further information on how we handle personal data can be provided on request.

78. OfS staff may examine outputs from the data checking tool in any of the following circumstances:

- to ask data verification questions and consider providers' responses to them
- to assist providers with queries about their outputs and data
- to test the functionality of forthcoming outputs and the application of overrides
- to produce early modelling
- to verify future data returns
- where a provider explicitly gives permission.

Further information

79. Further information about how to use the 2017-18 ILR data checking tool and the outputs will be made available on the OfS website in due course.

80. For any further information or guidance please email ILRData@officeforstudents.org.uk.

Explanation of abbreviations and terms

Abbreviations and terms	Explanation
DfE	Department for Education
ESFA	Education and Skills Funding Agency
HEAPES	Higher Education in Alternative Providers Early Statistics
HEFCE	Higher Education Funding Council for England
HEIFES	Higher Education in Further Education: Students
HESA	Higher Education Statistics Agency
HESES	Higher Education Students: Early Statistics
ILR	Individualised Learner Record
IRIS	Information Reporting Interface Service
Minerva	The online database containing information and queries relating to data submissions to HESA
NSS	National Student Survey
OfS	Office for Students
QR	Quality-related Research
RDP	Research Degree Programme
SLC	Student Loans Company
TEF	Teaching Excellence and Student Outcomes Framework
TRAC(T)	Transparent Approach to Costing for Teaching



© The Office for Students copyright 2018

This publication is available under the Open Government Licence 3.0.

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