

2019-20 data checking tools

- **Higher Education Statistics Agency Student data**
- **Higher Education Statistics Agency Student Alternative data**
- **Individualised Learner Record data**

Reference OfS 2020.36

Enquiries to

- HESA Student data: HESAStudentData@officeforstudents.org.uk
- HESA Student Alternative data: APData@officeforstudents.org.uk
- ILR data: ILRData@officeforstudents.org.uk

Publication date 23 July 2020

Contents

Summary	2
Section 1: 2019-20 HESA Student data checking tool	3
Action required	3
Data checking tool outputs and release schedule	3
Data verification questions	4
Key assumptions	5
Data confidentiality	5
Further information	5
Section 2: 2019-20 HESA Student Alternative data checking tool	6
Action required	6
Data checking tool outputs and release schedule	6
Data verification questions	7
Key assumptions	8
Data confidentiality	8
Further information	8
Section 3: 2019-20 ILR data checking tool	9
Action required	9
Data checking tool outputs and release schedule	9
Data verification questions	11
Key assumptions	11
Data confidentiality	12
Further information	12
Explanation of abbreviations and terms	13

Summary

1. This document notifies providers of the forthcoming release of the Office for Students (OfS) data checking tools for 2019-20 Higher Education Statistics Agency (HESA) Student data, 2019-20 HESA Student Alternative data and the Education and Skills Funding Agency's (ESFA's) 2019-20 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 2019-20 individualised student data before signing off the final data with HESA or the ESFA.
3. The OfS strongly encourages 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, supporting compliance with conditions of registration F4 and F3
 - verify the accuracy of the derived fields that may be used for regulatory purposes
 - respond to data verification questions raised by OfS staff
 - reduce the likelihood of selection for data assurance activity for 2019-20
 - 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. The OfS 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 use of the data. For further information regarding data amendments, please see the 'Amendments to data' page of the OfS website.¹
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.

¹ See www.officeforstudents.org.uk/data-and-analysis/amendments-to-data/.

Section 1: 2019-20 HESA Student data checking tool

Action required

6. Providers will be able to access the 2019-20 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 Issue Management 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 website once released. Feedback for OfS outputs should be emailed to HESAStudentData@officeforstudents.org.uk.

Data checking tool outputs and release schedule

9. The 2019-20 HESA Student data checking tool will produce the following outputs:
 - Higher Education Students: Early Statistics survey 2019 (HESES19) comparison
 - 2019-20 Student numbers data summary
 - 2019-20 Student Loans Company (SLC) comparison
 - Transparency 2019-20 attainment summary.
10. The data checking tool will not provide outputs that explicitly derive institutional performance measures, such as those used in the Teaching Excellence and Student Outcomes Framework (TEF) and the access and participation data dashboard. Nor will it provide a student premiums allocation output, following the introduction of linking to other data sources at other providers in the allocation method. HESA and the OfS encourage institutions to pay particular attention to the credibility checks and frequency counts that HESA's 2019-20 Student data collection provides in relation to key data items. These include credibility checks on those data items that act as benchmarking factors, or that define 'splits' and populations of the TEF metrics and other institutional performance measures.

HESES19 comparison

11. The HESES19 comparison compares 2019-20 HESA Student data with the provider's HESES19 data returned to the OfS at the start of the 2019-20 academic year. There is also a comparison between the provider's latest 2020-21 OfS teaching grant and the grant as it would have been if calculated using the 2019-20 HESA Student data.

12. The algorithms used to generate the HESES19 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 HESES19 comparison will be released in the **week commencing 10 August 2020**.

2019-20 Student numbers data summary

14. The Student numbers data summary provides the full-time equivalent student numbers by level of study using the 2019-20 HESA student data. These are the student numbers that the OfS intends to use for various regulatory purposes, including for setting registration fees, assessing applications for degree awarding powers and assessing applications for university title and university college title.

15. The 2019-20 Student numbers data summary will be released in the **week commencing 10 August 2020**.

2019-20 Student Loans Company (SLC) comparison

16. The SLC comparison compares 2019-20 HESA Student data with data held by the SLC for the provider as of 20 September 2020. It allows providers to establish whether there are inconsistencies between the two sets of data, with regard to students receiving student support.

17. The SLC comparison will be released in the **week commencing 5 October 2020** reflecting the availability of data from the SLC.

Transparency 2019-20 attainment summary

18. The transparency 2019-20 attainment summary provides information on the attainment of UK-domiciled students who qualified in the 2019-20 academic year. For each mode of study (full-time, part-time and study as part of an apprenticeship), the attainment summaries are split by ethnicity, gender and English Index of Multiple Deprivation (IMD) of qualifiers.

19. The transparency 2019-20 attainment summary will be released in the **week commencing 31 August 2020**.

Data verification questions

20. 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 with the provider. If any anomalies are found, HESA and the OfS will raise queries through the HESA Issue Management System. Providers are required either to resubmit data to amend these anomalies, or to provide an explanation, through the Issue Management System, 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 **6 November 2020**. We will not allow providers' data to be set to 'credible' by HESA until all queries have been responded to or are cleared by submission of amended data. Providers should allow time in their processes for the OfS to review submissions and raise queries. This

will be particularly important where there have previously been a significant number of queries or there have been significant changes to the data since the last submission.

21. 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 that submitting incorrect data to HESA may have on a provider.

Key assumptions

22. The data checking tool calculations will incorporate any HESA Student data amendments for 2018-19 and earlier years that were approved by the OfS data amendment panel, and signed off by providers by 17 June 2020. We will not update the data checking tool outputs to reflect any amendments or overrides that may be signed off later.

Data confidentiality

23. 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 20 and 21. 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.

24. 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.

25. The outputs generated from a provider's SLC data should be kept in confidence. In viewing this output, through IRIS, a provider accepts this commitment of confidentiality.

Further information

26. Further information about how to use the 2019-20 HESA Student data checking tool and the outputs will be made available on the OfS website in due course.

Section 2: 2019-20 HESA Student Alternative data checking tool

Action required

27. Providers will be able to access the 2019-20 data checking tool from the Information Reporting Interface Service (IRIS) report which is available on HESA's data collection site.
28. We strongly encourage providers to analyse the data checking tool outputs as part of their HESA Student Alternative data quality processes.

Data checking tool outputs and release schedule

29. The 2019-20 HESA Student Alternative data checking tool generates the following outputs:
 - Higher Education Students: Early Statistics survey 2019 (HESES19) comparison
 - 2019-20 Student numbers data summary
 - 2019-20 Student Loans Company (SLC) comparison
 - 2019-20 Pearson qualification comparison
 - Transparency 2019-20 attainment summary.
30. The data checking tool will not provide outputs that explicitly derive institutional performance measures, such as those used in the TEF and the access and participation data dashboard. Nor will it provide a student premiums allocation output, following the introduction of linking to other data sources at other providers in the allocation method. HESA and the OfS encourage institutions to pay particular attention to the credibility checks and frequency counts that HESA's 2019-20 Student Alternative data collection provides in relation to key data items. These include credibility checks on those data items that act as benchmarking factors, or that define 'splits' and populations of the TEF metrics and other institutional performance measures.

HESES19 comparison

31. The HESES19 comparison compares 2019-20 HESA Student Alternative data with the provider's HESES19 data returned to the OfS at the start of the 2019-20 academic year. For providers registered as approved (fee cap), this output includes a comparison between the provider's latest 2020-21 OfS teaching grant and grant modelled using the 2019-20 HESA Student Alternative data. This output will only be produced for providers that submitted a HESES19 return.
32. The algorithms used to generate the HESES19 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.
33. The HESES19 comparison will be released in the **week commencing 10 August 2020**.

2019-20 Student numbers data summary

34. The Student numbers data summary provides the full-time equivalent student numbers by level of study using the 2019-20 HESA Student Alternative data. These are the student numbers that the OfS intends to use for various regulatory purposes, including for setting registration fees, assessing applications for degree awarding powers and assessing applications for university title and university college title.
35. The 2019-20 Student numbers data summary will be released in the **week commencing 10 August 2020**.

2019-20 Student Loans Company (SLC) comparison

36. The SLC comparison compares 2019-20 HESA Student Alternative data with data held by the SLC for the provider as of 20 September 2020. It allows providers to establish whether there are inconsistencies between the two sets of data, with regard to students receiving student support.
37. The SLC comparison will be released in the **week commencing 5 October 2020**, reflecting the availability of data from the SLC.

2019-20 Pearson qualification comparison

38. This output compares a provider's 2019-20 HESA Student Alternative data with records Pearson holds to establish where there are inconsistencies between the two, with regard to students receiving Pearson qualifications.
39. The 2019-20 Pearson qualification comparison will be released in the week commencing **5 October 2020**.

Transparency 2019-20 attainment summary

40. The transparency 2019-20 attainment summary provides information on the attainment of UK-domiciled students who qualified in the 2019-20 academic year. For each mode of study (full-time, part-time and study as part of an apprenticeship), the attainment summaries are split by ethnicity, gender and English Index of Multiple Deprivation (IMD) of qualifiers.
41. The transparency 2019-20 attainment summary will be released in the **week commencing 31 August 2020**.

Data verification questions

42. 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 Issue Management System. Providers are required either to resubmit data to amend these anomalies, or to provide an explanation, through the Issue Management System, 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 **16 November 2020**. We will not allow providers' data to be set to 'credible' by HESA until all

queries have been responded to or are cleared by submission of amended data. Providers should allow time in their processes for the OfS to review submissions and raise queries. This will be particularly important where there have previously been a significant number of queries or there have been significant changes to the data since the last submission.

43. 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, it is important to begin the verification process at an early stage to allow as much time as possible to resolve potential data error.

Key assumptions

44. The data checking tool calculations will incorporate any HESA Student Alternative data amendments for 2018-19 and earlier years that were approved by the OfS data amendment panel, and signed off by providers by 17 June 2020. We will not update the data checking tool outputs to reflect any amendments or overrides that may be signed off later.

Data confidentiality

45. 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 42 and 43. 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.

46. 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.

47. The outputs generated from a provider's SLC and Pearson data should be kept in confidence. In viewing this output, through IRIS, a provider accepts this commitment of confidentiality.

Further information

48. Further information about how to use the 2019-20 HESA Student Alternative data checking tool and the outputs will be made available on the OfS website in due course.

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

Section 3: 2019-20 ILR data checking tool

Action required

50. Further education and sixth form colleges will be able to access the 2019-20 ILR data checking tool through the OfS portal (<https://extranet.officeforstudents.org.uk/Data>). To access the data checking tool, a provider's user administrator will be able to add users to the '2019-20 ILR data checking tool' area via the user administrator facility on the OfS portal once the tool has been released. Guidance for user administrators is available to download from the homepage (<https://extranet.officeforstudents.org.uk/Users>).
51. 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.
52. We would be grateful for feedback on the algorithms used to create the data checking tool outputs, particularly 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

53. The 2019-20 ILR data checking tool will produce the following outputs:
- Higher Education Students: Early Statistics survey 2019 (HESES19) comparison
 - Higher education-level apprenticeship data summary
 - National Student Survey (NSS) target list
 - Graduate Outcomes survey target list
 - 2019-20 student numbers data summary
 - 2019-20 ILR quality control summary
 - Transparency 2019-20 attainment summary

HESES19 comparison

54. The HESES19 comparison compares 2019-20 ILR data with the provider's HESES19 data returned to the OfS at the start of the 2019-20 academic year. For providers registered as approved (fee cap), this output includes a comparison between the provider's latest 2020-21 OfS teaching grant and grant modelled using the 2019-20 ILR data. This output will only be produced for providers that submitted a HESES19 return.
55. The algorithms used to generate the HESES19 comparison output are intended to be those used for the post-collection outputs next spring. However, we may make changes where we believe they will improve our algorithms.

56. The HESES19 comparison will be released in the **week commencing 10 August 2020**.

Higher education-level apprenticeship data summary

57. The higher education-level apprenticeship data summary uses fields taken from the ILR to categorise students that started a higher education-level apprenticeship in 2019-20. The summary includes the number of these apprenticeship starts categorised by student characteristic, subcontractual partner, subject and outcome, for the purpose of identifying potential errors in ILR data that may affect the OfS's uses of the data. It provides separate information on student apprenticeship programme aim and component aim starts.

58. The apprenticeship data summary will be released in the **week commencing 10 August 2020**.

NSS target list

59. The 2019-20 ILR data submitted to the data checking tool will be used to produce a provisional target list of students to be included in the 2021 NSS. Providers are encouraged to use these lists to start preparing the contact details for these students.

60. These will be passed to the agency running the survey. These contact details will need to be returned to the agency in mid-November 2020. Further details about the arrangements for the 2021 NSS will be published in due course.

61. After the ILR submission date, the OfS will extract final target lists for the 2021 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 2019-20 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.

62. The NSS target list output will be released by the **week commencing 31 August 2020**.

Graduate Outcomes survey target list

63. The 2019-20 ILR data submitted to the data checking tool will be used to produce provisional target lists of students to be included in the 2019-20 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.

64. Once we have received the 2019-20 ILR R14 data from the ESFA, we will use it to validate colleges' target lists.

65. The Graduate Outcomes survey target list output will be released by the **week commencing 31 August 2020**.

2019-20 Student numbers data summary

66. The Student numbers data summary provides the full-time equivalent student numbers by level of study using the 2019-20 ILR data. These are the student numbers that the OfS intends to use for various regulatory purposes, including for setting registration fees, assessing applications for degree awarding powers and assessing applications for university title and university college title.

67. The 2019-20 Student numbers data summary will be released in the **week commencing 10 August 2020**.

2019-20 ILR quality control summary

68. The ILR quality control output provides a summary of potential errors in the ILR return where records of students or courses may cause issues for OfS processes. These checks include searches for unknown values and failed continuity, as well as where combinations of information do not appear to be credible. The summary also includes breakdowns of student headcounts by demographic, highlighting where the change between years is greater than expected.

69. The 2019-20 ILR quality control summary will be released in the **week commencing 10 August 2020**.

Transparency 2019-20 attainment summary

70. The transparency 2019-20 attainment summary provides information on the attainment of UK-domiciled students who qualified in the 2019-20 academic year. For each mode of study (full-time, part-time and study as part of an apprenticeship), the attainment summaries are split by ethnicity, gender and English Index of Multiple Deprivation (IMD) of qualifiers.

71. The transparency 2019-20 attainment summary will be released in the **week commencing 31 August 2020**.

Data verification questions

72. 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 the OfS portal. 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. Providers should allow time in their processes for the OfS to review submissions and raise queries. This will be particularly important where there have previously been a significant number of queries or there have been significant changes to the data since the last submission.

73. 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, it is important to begin the verification process at an early stage to allow as much time as possible to resolve potential data error.

Key assumptions

74. The data checking tool calculations will incorporate any ILR amendments for 2018-19 and earlier years that were approved by the OfS data amendment panel and signed off by providers by 17 June 2020. We do not intend to update the data checking tool outputs to reflect any amendments or overrides that may be signed off later.

75. The version of the Learning Aims Search database that we refer to is taken from July 2020. 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

76. 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 72 and 73. 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.

77. 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

78. Further information about how to use the 2019-20 ILR data checking tool and the outputs will be made available on the OfS website in due course.

79. 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
HESA	Higher Education Statistics Agency
HESES	Higher Education Students: Early Statistics survey
ILR	Individualised Learner Record
IRIS	Information Reporting Interface Service
Issue Management System	The online database containing information and queries relating your data submissions to HESA
NSS	National Student Survey
OfS	Office for Students
Pearson	An awarding body for HNC and HND qualifications
SLC	Student Loans Company
TEF	Teaching Excellence and Student Outcomes Framework



© The Office for Students copyright 2020

This publication is available under the Open Government Licence 3.0.

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