Higher Education Statistics Agency data: quality assurance of administrative data used in population statistics, Feb 2017

Report detailing the quality assurance undertaken on administrative data for Higher Education Statistics Agency data used within Population Statistics Division (PSD) publications.

Table of contents

1 . Executive summary

This report describes the use of HESA (Higher Education Statistics Agency) data on students when used in the production of migration and population statistics.

The data provides basic demographic data (age and sex); term-time and domicile (parental home) postcode; information on the course (course length and type) and institutional information.

As the data provide term and home postcode they can be used to help track student migration, as well as support other administrative sources where the information may not be as up to date. This addresses a significant known issue with other sources where young adults, particularly students, can be slow to update details and may not do so at all for their university stays. The data covers all publicly funded higher education institutions and one private institution across the UK.

2 . Introduction

Higher Education Statistics Agency (HESA) collects, processes and publishes data about higher education (HE) in the UK. HESA is a charitable company operating under a statutory framework on behalf of the funding councils and UK government departments. They support HE providers in fulfilling HE data reporting requirements and produce Official Statistics, regulated by the UK Statistics Authority. They are principally funded by subscriptions from HE providers.

We use a range of data from HESA including both aggregate data and microdata.

In 2014 to 2015 HESA data covered 162 HE providers (131 in England, 9 in Wales, 18 in Scotland and 4 in Northern Ireland) with around 2,266,000 students.

This report covers the use of HESA data when used in the production of migration and population statistics by Population Statistics Division (PSD).

2.1. Uses in PSD

Internal migration

Record-level HESA data are used to supplement Patient Register data to help estimate internal migration; that is migration within England and Wales[1].

Internal migration is a component of mid-year population estimates for England and Wales and a National Statistics publication, which we publish.

Local Authority distribution of international migrants

The HESA data are used as one of the sources to distribute estimates of long-term international migration into the UK to geographic areas based on international students. Data on students in their first year of study whose domicile is outside the UK are used (together with other sources) to help allocate estimates of international migration into the UK to local authority areas.

International migration is a component of mid-year population estimates for England and Wales and a National Statistics publication at regional level based on survey data, which we publish; for more information please see the Long-term International Migration Estimates: Methodology Document 1991 onwards.

Short-term international migration

The data are not used to calculate the level of short-term international migration to the UK, but they are used as one of the sources to distribute estimates of short-term international migration into the UK to smaller geographic areas. Data on students' course of study of one academic year or less, whose domicile is outside the UK are used (together with other sources) to help allocate estimates of short-term international migration into the UK to local authority areas.

2.2. Assessment of Quality Assurance Level

The QAAD Toolkit sets out 4 levels for the quality assurance that will be required of a dataset:

- A0 – no assurance

- A1 – basic assurance

- A2 – enhanced assurance

- A3 – comprehensive assurance

The UK Statistics Authority states that the A0 level is not compliant with the Code of Practice for Official Statistics . The assessment of the assurance level is in turn based on a combination of assessments of data quality risk and public interest.

The toolkit sets out the level of assurances required as follows.

QAAD toolkit levels of assurances

A0: No assurance	Not compliant with the Code of Practice for Official Statistics
A1: Basic assurance	Statistical producer has reviewed and published a summary of the administrative data QA arrangements
A2: Enhanced assurance	Statistical producer has evaluated the administrative data QA arrangements and published a fuller description of the assurance
A3: Comprehensive assurance	Statistical producer has investigated the administrative data QA arrangements, identified the results of independent audit, and published detailed documentation about the assurance and audit

Quality Assurance Level for HESA Data: A3 – comprehensive assurance

Where we use the same data source for a number of different purposes it is normally more efficient to undertake some of the quality assurance only once. Each separate use may have a different level of assurance required. As a consequence, the highest level of quality assurance, from the different uses, will be applied to common quality assurance.

Summary of Quality Assurance Level for HESA data

Use	Risk of quality concern	Public interest	Assurance Level
Overall Assessment			A3 – comprehensive
Internal Migration	High	Medium	A3 – comprehensive
International Migration – allocation of data to Local Authority area	Low	High	A1 – Basic
Short-term International Migration	Low	Medium	A1 – Basic

Internal migration

Quality Assurance level for internal migration A3 – comprehensive assurance

Level of risk of data quality concerns: High

- There are a range of quality issues that apply to the use of these data for internal migration statistics.

- There are multiple collection bodies (higher education institutions), with collection practices differing between and within collection bodies (as collection may take place at the academic school level).

Public interest profile of the statistics: Medium

- Internal migration estimates are published National Statistics.

- Internal migration estimates are a key component of mid-year population estimates, which:
  - are very high profile National Statistics

  - are politically sensitive

  - are used in resource allocation

  - have wide media interest

  - are required under European legislation

  - are used as the basis of population projections which are used to allocate local authority funding and for planning

  - have a wide variety of other onward uses

Additional information

The quality issues can have a material effect on the quality of internal migration statistics, while not necessarily being significant for the administrative requirement for these data. Quality issues include:

- that over time there has been an increase in the number of ways a university has to contact students and, as such, there is now less importance on universities maintaining up-to-date address information than there used to be

- the snapshot nature of these data, combined with a tendency for students to move on rapidly from their first address, means that address information provided may be out of date

- due to the timing of data extracts some university drop outs will be included within the data

These data are also used in population estimates and projections. Because of these concerns, and the onwards use of the data, an A3 – comprehensive assurance level has been chosen.

Local Authority distribution of international migrants

Quality Assurance level for Local Authority distribution of international migrants : A1 - basic assurance

Level of risk of data quality concerns: Low

- There are a range of quality issues that apply to the use of these data for the local authority distribution of international migrants.

- There are multiple collection bodies (higher education institutions), with collection practices differing between and within collection bodies (as collection may take place at the academic school level).

- The data are one of several sources used in the methodology.

- The data are used for distributional purposes only (for international migration) and the size of the likely error is relatively small in comparison with potential errors from other sources.

Public interest profile of the statistics: High

- Local authority distribution of international migrants are part of a very high profile National Statistic, however this source contributes only to the geographical estimation of flows to local areas and does not affect the total estimates.

- International migration estimates are a key component of mid-year population estimates, which:
  - are very high-profile National Statistics

  - are politically sensitive

  - have wide media interest

  - are required under European legislation

  - are used as the basis of population projections which are used to allocate local authority funding

  - have a wide variety of other onward uses

Additional information

Because the local authority distribution of international migrants are a distribution of people to a previously determined total and the data only contributes toward the distribution, the likely impact of any error is low. As a result of these 2 factors an assurance level of A1 – Basic assurance has been chosen.

Short-term international migration

Quality Assurance level for short-term international migration: A1 – basic assurance

Level of risk of data quality concerns: Low

- There are a range of quality issues that apply to the use of these data for short-term international migration statistics.

- There are multiple collection bodies (higher education institutions), with collection practices differing between and within collection bodies (as collection may take place at the academic school level).

- The data are one of several sources used in the methodology.

- The data are used for distributional purposes only (for short-term international migration) and the size of the likely error is relatively very small in comparison with potential errors from other sources.

Public interest profile of the statistics: Medium

- Short-term international migration estimates are a National Statistic, however this source contributes only to the geographical estimation of flows to local areas.

- It is the Long-Term International Migration that receives greater political and media attention as the headline figures of international migration.

Additional information

Although short-term international migration statistics are a National Statistic, the geographical apportionment of the estimates is of lower profile. The likely impact of any error is very low in proportion to other possible error sources. As a result of these 2 factors an assurance level of A1 – Basic assurance has been chosen.

2.3. Selection of dataset

Higher Education Statistics Agency (HESA) provides the only comprehensive nationally consistent source of data on higher education students.

The main source for estimating internal migration is the NHS patient register, however evidence has shown that young healthy adults, particularly males, are slow to update their details. The most significant reason for internal migration for this age group is for moves both into and out of higher education. Therefore, data on students can help address a significant shortcoming when using Patient Register data in internal migration. HESA provide the best quality data on residential location of this section of the population.

The data collected from the International Passenger Survey (IPS), the main source for both short-term and long-term international migration, can be split into students and non-students. Since the migration patterns of students differ from those of non-students, data from HESA can be used to provide a more accurate geographical distribution of the estimates to local areas. The IPS sample size is not large enough to enable a reliable direct estimate of the local authority distribution of international migration estimates.

2.4. Practice areas associated with data quality

The QAAD Toolkit describes 4 practice areas for quality assurance. The assurance undertaken for each practice area is described in the following sections.

Notes for Introduction:

1. Migration within Scotland and within Northern Ireland are calculated separately by, respectively, National Records of Scotland and the Northern Ireland Statistics and Research Agency. A separate adjustment for cross border flows between the countries is made.

3 . Operational context and administrative data collection

3.1. Detailed description of the administrative system and operational context

Organisations collecting the data

The data are collated by Higher Education Statistics Agency (HESA) from returns from all 161 public higher education establishments in the UK plus the University of Buckingham, a private institution. These are mainly universities, but also include higher education colleges and other specialist providers of higher education. A full list is provided in Annex 1. The universities in turn collect the data from students.

Original purpose for data collection

The data are collected to record the number of students at each institution for higher education establishment funding purposes and for performance indicators. The data are also collected to enable statistical reporting and analysis.

Legal basis for collection

The Higher and Further Education Act (1992) requires universities and other higher education providers to submit data about their activities to the UK funding councils.

HESA was established in 1993 in order to support higher education providers fulfil their statutory reporting requirements. HESA is a sector-owned shared service, set up by agreement between the relevant government departments, the higher education funding councils and the universities and colleges.

The sharing of record-level data from HESA to ONS is enabled by an Information Sharing Order under the Statistics and Registration Service Act (2007).

Detailed description of how the data are collected

Potential students fill in an application form which is sent to the relevant higher education institution(s). The higher education institution(s) create a record of the application within the student record system. A decision is made whether or not to make the applicant an offer. Successful applicants will then choose whether or not to accept the offer as multiple applications may be made simultaneously. Should an offer be accepted, then the student will confirm their information upon the commencement of their course. The data then undergoes quality checks to ensure that the information entered meets the required standards of the institution. Once checks have been successfully completed at both the record and the dataset level, the data are ready for the institution's HESA return. The data are subject to further checks at HESA and once a satisfactory return has been received from all institutions a combined dataset is created. The data is then provided to us under an Information Sharing Order under the Statistics and Registration Service Act 2007. We carry out some pre-processing checks on the HESA dataset before it is loaded into the secure research environment and the secure processing environment.

Collection of data from students

Collection of data at application from students

Higher education institutions obtain some of the data used to report to HESA through application forms. The most common route for undergraduates, and for PGCE (Post Graduate Certificate in Education) applicant, is via the Universities and Colleges Admissions Service (UCAS). UCAS also administers some post graduate applications. Application to UCAS is via an online form, with some in-built validation, validation is also undertaken by UCAS during the application process. Data from UCAS are then passed on to the relevant institutions.

Direct application to the university for undergraduates is also possible. Direct application is the more common approach for postgraduate courses. Direct entry may be via online form or via paper (postal) application. Universities will differ in the amount of validation that is undertaken for direct applications.

Data collection at and post registration

Higher education institutes will also use information collected from students from registration when starting at the institute. Notably, this is often where a student's term-time postcode is recorded. This process will differ from institution to institution, and even within institution. Typically an online form is completed, or a paper-based form is completed and later entered into a database by staff. Alternatively a member of staff may complete the information in a one-to-one session. It is also possible that some of the information is captured by the student interacting with the accommodation office, who in turn provide information to the central institution.

Updates normally rely on the student informing the institution of any changes. However, in some cases it is possible that when a student changes address with the assistance the accommodation office, or into or out of accommodation office managed property, then the accommodation office will update the records held by the central institution.

Supply of data from higher education institution to HESA

Student data are collected by higher education institutions (HEIs) and sent to HESA through a password protected website. Validation checks are then carried out on the data; example checks include ensuring the data provided are in the correct format, such as date of birth being recorded in dd/mm/yyyy format. If the data fails at this point, then the HEI will have to review and resubmit their data. Once validated, the data undergoes an internal quality assurance review by the HEI. Again, if data does not pass this process, it will have to be reviewed and resubmitted. Once all data successfully passes internal QA and the HEI is happy with their submission, then the data are committed to HESA to review. HESA undertake data quality assurance using their Minerva tool to determine whether the data are credible. If they uncover any issues then the data are returned to the HEI to make amendments and restart the process. If data are credible then the HEI will sign off the data to say that it comprises a complete and accurate submission of their student information. Following sign off from the institutions HESA compiles the data and produces a range of statistical outputs. The data are produced for us after the published outputs and are sent to us using a secure transmission.

Differences across areas in the collection and recording of the data

Whilst the information is provided to HESA in a consistent way, and has to pass the same checks, approaches within institution will vary. Key differences can include:

- initial student contact details are determined from applications either made directly, through UCAS or through another third party

- how and when institutions validate and check or cross check information

- how institutions track address changes through time

- how institutions capture and update changes of details at the start of second and subsequent academic years

Data item issues

Timeliness of data

Address data on students is collected at the start of a student's period of study and may not get collected again. Practice varies from institution to institution, and even within institution, depending on the type of study and whether a student is in institution sponsored accommodation or private accommodation. Some institutions will record address information at the start of each academic year. Some institutions will track student addresses via the institution's accommodation office (where the student is in university related accommodation). Data quality on sandwich student address information is also variable in this regard.

However, anecdotally, students who move address during a year often move address a short period into an academic year. This may mean that the address recorded on HESA does not reflect the student's address on 30 June, which is the standard reference date used for population statistics.

Transcription errors

These are likely to be rare. Although some data entry is based on handwritten forms, and validation of data will vary, between and within institutions, most data entry is from electronic forms completed by the student. This reduces the scope for transcription errors, however there is still the possibility of typo errors (students may input data without checking and miss spelling mistakes) and incorrect formatting of dates.

One issue that may affect postcodes in particular is that the address will be new to the student. Most home [1] address postcode information is very reliable – people generally know their postcode. However, for a new address the student may not know, or have committed to memory, the correct postcode: this could lead to some error in postcode reporting. The size of this effect is unknown, and may vary between UK and foreign students.

Duplication and longitudinal matching

The data recorded for a student is held and collected by course of study. A student may enrol on several courses of study – either simultaneously, or sequentially within an academic year, or sequentially from year to year.

As a result students may be duplicated in the data. This may be entirely legitimate, but is undesirable for our statistical usages within population applications where the need is to capture the student once.

We wish to match students longitudinally (over time), as length of study is important to the statistical usage. Where a student changes course, especially between institutions, this can become problematic. Matching is often reliant on name and date of birth, and the name form used may vary (for example the use of diminutives, such as Bob as opposed to Robert, or alternative spellings – particularly for foreign students). This can lead to either false positive or false negative matches.

3.2. Issues in design and definition of targets

Most funding for higher education institutions is based around the number of students undertaking study and there are strict rules that help ensure that the information collected is consistent. The data supplied to the funding bodies, via HESA, are scrutinised both by HESA and the funding bodies. The approach (and the need to tie funding to individual students) means that there is an incentive for institutions to ensure their data are as accurate as possible and not miss students. As a result there are thought to be no significant issues with distortion of the data due to targets or other similar incentives.

However, there is an incentive for institutions to count their students as soon as possible after the start of the term, so that students who later drop out are included within funding.

3.3. Potential sources of bias and error in the administrative system

The likely sources of error are covered above and there are no other known likely causes of bias in the data.

3.4. Safeguards used to minimise the risks to data quality

HESA is an independent body that is responsible for collecting data from higher education institutions. It applies a detailed range of technical and credibility checks to the data to independently assure the quality of the data. It provides detailed instructions on how to supply accurate information.

There is not a formal independent audit of the data; however, as the independent body applies checks HESA itself performs many of the activities that would be undertaken in an audit as part of its standard processing. The respective funding councils also have an interest in reliable data, and as such provide an additional safeguard that the data are reliable.

HESA maintains a code of practice that institutions are required to follow.

Changes over time

The collection process and the data collected have remained largely stable over the period that we have been using HESA student data. We use data back to 2011 in research and statistics production.

Over this time period, although HESA collection has changed little, individual institutions have moved towards greater use of electronic data collection. This is likely to improve the quality of data by reducing the scope for errors, such as data transcription errors.

3.5. Implications for accuracy and quality of the data

Overall the supplied data are thought to be accurate, however the nature and purpose of the collection can give timeliness of data and over-coverage errors when used in the production of our statistics. In addition there is a residual risk of error for a number of individual data items. The main implications for the accuracy of the student data we use are:

- timeliness of data: Students may move on quickly to a new term-time address after registering their details with their institution, and their new address will often not feed into the data held by HESA

- over-coverage: Students who drop out will normally still be included in the HESA data, provided they have registered and attended the first 2 weeks of the course

- data item error – postcode: either through transcription or an incorrectly supplied postcode from the student, the term-time address or (less likely) the home address postcodes for the student may be incorrect

- data item error – date of birth: either through transcription or incorrectly supplied information (including reversal of date and month for the latter) the recorded date of birth for the student may be incorrect

- data item error – unique identifier: students may be incorrectly matched to a unique student identifier (either false positive, or false negative) meaning an incorrect student identifier and impacting on subsequent matching when used in population statistics

Notes for Operational context and administrative data collection:

1. HESA uses the term domicile for a student's normal address before starting on a period of study, with the concept that the student will often return to this address outside of term-time. This report uses the terms home and domicile interchangeably, normally in relation to address.

4 . Communication with data supply partners

4.1. Interaction with supply partners

Communication with data supply partners is managed by the Data Sharing and Supplier Management (DSSM) team within the Data as a Service Division of ONS.

Liaison with the supplier starts in November for the delivery in January or February. Communication is initiated by email, and will include phone calls, and conference calls if required. Over the lead-up to the data delivery date the following issues are discussed, in approximate time order:

- whether there are any changes to the data or supply

- ensuring that metadata is up to date

- confirmation that the Data Supply Template is up to date

- confirmation of the logistics of the data exchange

Written supply agreement

There is a formal contract in place with Higher Education Statistics Agency (HESA). The document is in line with our policy on data supply agreements. This contract is managed by the DSSM team within Data as a Service Division.

Roles and responsibilities

The parties to the agreement, as set out in the agreement, are:

- HESA Services Limited

- Higher Education Funding Council for England

- Higher Education Funding Council for Wales

- Department for Employment and Learning, Northern Ireland

- Office for National Statistics

- Northern Ireland Statistics and Research Agency

- Higher Education Statistics Agency Limited

The parties have the following roles set out in the agreement:

- HESA Services is the Data Processor

- the Higher Education Funding Councils for England and Wales, and the Department for Education and Learning Northern Ireland are the Data Controllers

The agreement gives nominated officers for each party.

Date of the agreement

The date of the agreement is June 2014, and the date of the variation is September 2014.

The agreement remains in force until terminated by a party.

Legal basis for data supply

The agreement lists the key relevant legislation, including the legislation under which the data are supplied:

- The Statistics and Registration Service Act 2007 (Disclosure of Higher Education Student Information) Regulations 2009
- The Statistics and Registration Service Act 2007 (Disclosure of Higher Education Student Information) Regulations (Northern Ireland) 2013

Data supply and transfer process

The agreement sets out that the data will be supplied in an electronic format using encryption and password protection. The agreement says that further details will be agreed by the appropriate parties.

A secure transmission method is used.

Security and confidentiality protection

The agreement acknowledges that the data include Sensitive Personal Information.

We are required to ensure that any data released or published are in an anonymised and aggregated form. This includes our application of HESA's Standard Rounding Methodology.

We are required to use appropriate technological and organisational security controls, including restricting access to a limited number of authorised personnel with appropriate training, and who are under obligation to maintain security, and for authorised purposes.

The data are held securely in the Statistical Research Environment, as set out in Safeguarding Data for Research: Our Policy, and in the secure processing environment. Data are subject to a detailed disclosure controlled process before they are transferred or published.

Schedule for data provision

The agreement sets out that the data will be delivered to a mutually agreed schedule. The agreement says that this should not be expected within 20 working days of the release of the student record data.

HESA publish their statistics on students in January, in relation to the previous academic year. The data for us are drawn from the same underlying data. The understanding between HESA and ourselves is that the data for us will normally be shared as soon as possible after this publication, which is 20 working days after the publication or release. In practice the delivery is often possible much sooner than the 20 working days, often only a few days after publication. Thus the data are normally received in January or February.

Content specification

The agreement sets out the variables for data content. These are:

- student Record Key (primary key)

- HESA unique student identifier (HUSID)

- forename(s)

- surname

- date of birth (BIRTHDTE)

- sex (GENDER)

- domicile (DOMICILE)

- mode of study (MODE)

- UK Provider Reference Number (UKPRN)

- campus identifier (CAMPID)

- campus name

- campus local authority

- location of study (LOCSDY)

- postcode (POSTCODE)

- last institution attended (PREVINST)

- term-time postcode (TTPCODE)

- term-time accommodation (TTACCOM)

- start date of instance (COMDATE)

- year of student in this instance (YEARSTU)

- expected length of study

- end date of instance (ENDDATE)

- reason for ending instance (RSNEND)

- reduced record indicator

- course aim

Data management arrangements

The agreement sets out that the data can be held indefinitely by us. There are no arrangements specified for data archive or deletion.

The agreement specifies that only a limited number of expressly authorised staff should be able to access the data and that these staff should have had appropriate security and data protection training and who are under a contractual obligation to uphold confidentiality and abide by the Data Protection Act and other security policies.

The agreement specifies the data should be held in a secure technical infrastructure that is appropriate for the level of sensitivity of personal data and in a manner that is consistent with the Data Protection Act (1998).

The agreement specifies arrangements in the event of a security breach. The agreement specifies arrangement in the event of a complaint.

Data usage

Our permitted purposes are defined in the agreement as those permitted by the Statistics and Registration Service Act 2007 (Disclosure of Higher Education Student Information) Regulations 2009. These are the:

- production of population statistics under section 20 (production of statistics) of the Act

- making of arrangements for a census under section 2 of the Census Act 1920(1) (duty of Registrar-General to carry out census, and provision for expenses)

- assessment of the census returns

Supplementary information

The agreement sets out the arrangements for us to onwardly share data with the Northern Ireland Statistics and Research Agency.

4.2. Change management process

HESA are required by the agreement to notify us of any changes to the data, on receipt of the annual request to supply the data.

Other variations from the agreement can be agreed only by a variation of the agreement signed by each party.

4.3. Engagement with users

Population Statistics Division (PSD) continually engages with users, through a variety of means, to understand how our outputs are meeting their requirements. Feedback provided tends to relate to the overall statistical methodology and the impact on the final statistics, rather than to any individual data source. To date no specific feedback on the use of this data source has been provided.

5 . Quality assurance principles, standards and checks applied by data suppliers

5.1. Data suppliers' principles, standards (quality indicators) and quality checks

Principles and standards

The Higher Education Statistics Agency (HESA) code of practice for data-suppliers sets out principles for data supply of honesty, impartiality and rigour.

The rigour principle sets out the standard that HESA expects from data supply:

"Data should be collected, prepared and submitted using repeatable and documented processes that can withstand scrutiny. When processes change, records should be kept of previous versions. Estimates and assumptions should be defensible, evidence-based and documented, and the effect on the data tested. Assumptions and estimates should be reviewed regularly."

Quality checks by the supplying institution.

The degree of quality checks applied will vary from institution to institution. However, the emergence of 2 main software providers [SITS by Tribal](#) and [Banner® by Ellucian](#) will tend to apply a degree of standardisation. In addition HESA encourages the use of its validation kit.

Validation kit

HESA provide a validation kit for higher education institutions to download. HESA strongly recommend that institutions use this kit to validate their data before transmission, however, this is not compulsory. This kit applies many of the validation checks that HESA will, in any-case, undertake following transmission. The kit allows institutions to check before transmission rather than submit and receive error reports.

Quality checks by HESA

Validation

Validation consists of 2 separate stages.

Schema checks check that the submitted file is structurally correct, for example:

- it is in a structurally correct .XML format and complies with .XML rules

- that the elements of the file are structurally correct

- that the data are in the correct form (string, dates, etc)

The submission has to pass the schema checks before it can go through to the second validation stage.

Business stage quality rules will follow a successful pass of the schema check. These checks include:

- internal consistency

- that compulsory fields are present

- range checks

Continuity validation checks that entries for individual students over time are consistent. It checks that the details of ongoing students can be matched back to previous records with consistent details. It also checks that every student has either a match from one year to the next or an explicit end (or suspension) of study has been recorded.

Review and commit

When an institution's submission has passed the validation, the institution is given the chance and is encouraged to quality assure the data. The HESA system provides a range of reports to facilitate this checking.

When the institution is content with the submission they commit the data, locking down the data so that it cannot be edited further, except by applying for them to be released (de-committed) by HESA.

Credibility checking

When data are committed HESA will undertake a range of credibility checks. HESA apply quality rules that require comparisons with data across an entire return and/or against reference data. This stage uses checks like frequency counts, and focuses on the entire dataset not the individual student. HESA also refers to this as exception stage validation, and uses its Minerva tool to manage the process.

To address issues raised at this stage the institution can request de-committing to correct data, confirm that the data are genuine, or state that the issue cannot be resolved in the current year. In reaching theses outcomes the institution can submit questions to HESA, or commit to resolving the issue in a re-submission.

When HESA is content all the issues at the credibility stage have been successfully resolved, it will mark the submission as credible and allow it to proceed to the sign-off stage.

Quality reports

No quality reports are supplied with the data.

Audit arrangements

There are no known formal audit arrangements, however, the checks applied on the data as described above serve some of the purposes of an audit. In addition there is a sign-off process for the data.

Sign-off

After HESA has determined the data as credible, a sign-off form is automatically generated. This requires the head of the institution to sign-off the data provided.

Sign-off marks the end of the data collection.

Implications for official statistics

HESA apply a wide range of checks to ensure the accuracy of the data. The data are therefore suitable for a range of uses and provide a good quality source. However, there are a number of issues that will have an impact on the statistics produced from the data.

The following sections describe the impacts of each of the main input data quality issues on the official statistics that use HESA data. Each issue is taken in turn, with each relevant use addressed within the issue.

Timeliness of data

Students can move quickly after registering their term-time address with their institution, meaning that the term-time address can become out of date. Measures of population use a usual residence definition. The application of this definition for students is that students should be placed at their term-time address; including international students who are usually resident in the UK for more than 12 months. Thus where an address becomes out of date, the information may not provide a balanced picture of the usual residence of students.

However this has some mitigation. Assuming a student stays at the same institution, most in-term student moves will be of a short distance and may be within the same local authority. Since the data are used at local authority level any such moves will not affect the statistics produced from the data. On the assumption that such moves are not patterned, for example from one student area to another, then although there may be some random error the moves will tend to cancel out. Finally, for the application of the data in the internal migration methodology information on activity and timeliness of data from the Patient Register can be drawn on to apply another address where more recent information is available through data-linkage.

These mitigations will, however, not work in all circumstances. There are some institutions that are near local authority boundaries, where even a short-move can affect the statistics. This will apply particularly to major cities, most notably London, that comprise of a number of local authorities. Another notable example is the University of Leicester, where a hall of residence is located outside of Leicester, in the local authority of Oadby and Wigston. There is also the potential in some areas for there to be movement patterns, for example from expensive to cheaper locations. Student address data are considered particularly valuable as they help resolve the issue in other datasets of not having the timely address information for young adults. As such, there will be a degree of error in the statistics, particularly for London and certain other areas in the short-term. However, people will update their information on this or other sources and this will lead to the statistics being corrected over time.

Over-coverage

Students who drop out of the course are still included on HESA data, provided they have completed two weeks of attendance on the course. (There are other limited data on students who don't complete the two weeks that are not required by our population-related uses of the data.) The data will accurately meet the suppliers' definitions, however for our statistical usage, these students who have dropped out are considered to contribute to the name of over-coverage. This may give rise to a small error in the statistics as those who have dropped out are unlikely to be resident at the recorded term-time address.

The timeliness of data issues previously discussed will also give rise to issues of over- and under-coverage within a geographical area as students are recorded in a previous location.

Data-item error and failure to match

Errors in the student identifier or date of birth recorded for a student may result in the failure to match a student to other data-sources, or longitudinally from one year's HESA data to another.

For internal migration this may mean that a student is not correctly allocated to their term-time address within the data. Thus a migration may be missed, and the population estimates at the home local authority being high by 1 and the term-time area being low by 1 for each student that fails to match.

When estimating long-term international migration as part of the mid-year estimates, the data are used to help apportion those migrating for study to local authority areas. The definition for a long-term international migrant is someone who has the intention to change their place of usual residence for at least a year. The assumption is therefore made that international students who are on a course of study that is less than one year are excluded.

Data-item error – date of birth

As well as the impact of an incorrect date of birth on failure to match (see Data-item error and failure to match), an incorrect date of birth can lead to an incorrect age being calculated for a student. Any errors transposing month and day will either have no effect or lead to the age as at 30 June being plus or minus one year. The volume of errors is likely to be small in any given year, with some tendency for errors to cancel each other out. No mitigation is applied to this error and over time the issue is likely to be corrected due to updates in the datasets used for statistical production.

6 . Producer's quality assurance investigations and documentation

We are moving increasingly to the collect once, use many times approach to data. As part of this ethos production of statistics is distributed across different teams within ONS, with common processing undertaken once centrally before there are internal handovers of data to teams who take on the production of statistics. This section therefore reflects the common processing that is undertaken, with a section for common checks undertaken, and then a section covering the specific relating to particular outputs or output areas.

6.1. Common processing

Process flow

Checks are made at various stages through the processing of the data. We receive a source file from Higher Education Statistics Agency (HESA); this is held in the secure research environment where it is loaded into SAS (a statistical processing software package). Initial processing takes place and the data are subject to geo-referencing. Further checks take place before 3 versions are created. The first version has a number of variables removed. These are identifier variables such as name and full address. These are removed to reduce the sensitivity of the data so it can be used in the secure production environment. This version is used for the production of population statistics. For the second version, identifier variables are hashed (a 1-way encryption process). When the variables have been hashed the data are made available to researchers. This second version is used for research into future population statistics methodologies. These versions are stored in separate secure environments and security arrangements mean that our staff does not have access to more than one environment at any one time. A third version relates to Northern Ireland data and is exported so that it can be handed over to the Northern Ireland Statistics and Research Agency.

Validation checks

On loading of the data onto SRE (Statistical Research Environment) systems a sample of records is examined (eyeballed), this is a non-random sample of approximately ten records from each of the start, middle, and end of the datafile. A manual check of these records is undertaken; this includes a check to ensure that the records supplied are in the expected format. This check will also identify, after loading the data into SAS, if columns and records do not correctly align, for example checking that forenames are consistently read into the forenames slot and do not drop into the surname slot.

Throughout processing SAS produces a log. The log is checked for the words "Error", "Warning", or "Note". If any of these words are found the cause is investigated. SAS log files are examined (after each stage of code is run). Examination of the log-file will[1] detect the following types of error:

- character (text) data when numeric data are expected

- fields that are truncated (for example 10 characters in a field where 8 are expected)

- data that are in the incorrect format (for example dates in not in the expected format)

- unexpected characters (unexpected non standard characters within text, beyond normal punctuation for example control – ctrl codes)

A manual check that matchkeys[2] have been correctly generated is undertaken. This checks that the matchkeys have the correct components, in the correct format for each matchkey. The resultant matchkeys are also checked to ensure that they are composed of the correct data from within the specific records. This check ensures that the matchkeys correctly concatenated the specific data values, and that values such as date of birth have been (re) formatted to the correct format. For example in matchkeys (unlike other standard output) dates are all included in SAS's Date9 format. This check is undertaken on a small non-random selection of records. (Further information on the detailed use of matchkeys has been published by ONS.) If matchkeys are not correctly generated then matches cannot be successfully made. Matchkeys are hashed[3] before supply to internal customers so it is not possible to retrospectively correct after onwards supply, without repeating the processing, starting again with the input data we received.

A manual check using a text viewer is made of output (CSV) files to check that the correct standardised date format has been applied (this is the SAS DDMMYY10 format). This check is done on final files as well as files for hashing, including within matchkeys. If an alternative format is used then matches using matchkeys will not be possible.

Sense checks

The manual sample check is mainly used as a sense check. The sense check looks to see that the data look as expected, examples include:

- names look to be normal names, and not other text

- surnames look like surnames and not first names (and vice-versa)

- addresses look to be in right and consistent order

- that fields are appropriately populated

- that data do not conform to odd patterns (for example, each date is one day more recent than the previous).

Sense (eyeball) checks are also undertaken whenever files are split or merged (such as during geo-referencing or hashing processes). These checks are mainly there to check that the correct variables are in each split and merged file. They are also used to check that the local unique identifier is on each file, so that a merge can be successfully made to join the data back-up.

Derived variables are calculated for person's age at 4 dates within the year, based on date of birth. Spot checks are applied to a small (non-random) number of records to check these have been correctly calculated. This check is undertaken before hashing as date of birth is one of the variables that are hashed.

Consistency checks

When the data are first loaded into SAS a note of the number of records is undertaken. This number is then used at stages throughout the processing to ensure that the number of records continues to add up. That is the total number of records across file splits and after merges sums back to this total number of records. This check is undertaken after every merge or split.

Acceptance testing

Following pre-processing of the data, they are passed to internal customers for acceptance testing. For the HESA data the acceptance testing has been based on frequency comparisons. Thus for every variable a spreadsheet is produced that identifies the most frequently occurring values. These frequencies are compared to previous years, and large changes flagged for further investigation. This approach allows an assessment of the shape of the data compared to previous years, which can help identify certain types of errors and also if there has been a large change in the source data from year to year.

6.2. Processing in secure production environment

Data are exported to a secure production environment before further processing for internal migration, the local authority distribution of long-term international migration and short-term international migration. These data include date of birth but are otherwise de-identified.

Population Statistics Division (PSD) makes use of date of birth: used to determine age at 30 June; for internal migration, the international migration component of population change and short-term international migration estimates.

6.3. Corrections to data

Postcodes are cleaned to correct for common errors (for example changing a zero to a letter O where a zero is invalid for example P015 to PO15).

The data have a standard set of cleaning applied. Cleaning includes the removal from names of any punctuation, including the symbol found in some HESA data, but excluding hyphens; and converting name fields to uppercase. Names are only used in the statistical matching of data. As the treatment of punctuation in names can differ within system practice and between systems, the consistent removal of punctuation improves the matching reliability, so that for example Darcy will match with D'arcy within automated matching.

HESA data include a single field for all forenames. There is a separate surname field. Therefore to achieve a consistent standard with other data, the forenames field is split into separate first-name and middle-names fields. This is done after data cleaning. The text before any space in the forenames field becomes the first-name. The text after the first space, if any, becomes the middle-names field. For example:

Corrections to data example

Forenames	First-name	Middle-names
John	John	
John Andrew	John	Andrew
John Andrew Kevin	John	Andrew Kevin
John-Andrew Kevin	John-Andrew	Kevin

A student may be registered on more than one course (either at the same or different institutions). Our current uses of the record level student data are concerned with student level data and not the individual courses they are enrolled on. Consequently where a student is registered on more than one course this duplication is identified.

6.4. Metrics

Missingness

A missingness report is generated as part of the standard processing. This details the proportion of each variable where data are missing from within a record. This is checked to see that it has correctly run. The report is passed to internal users.

Geo-referencing

Geo-referencing is the process where address information is used to determine higher levels of geography, such as local authority. Geo-referencing rates are compared from one year to the next and investigative action undertaken if there is a substantive difference in any of the rates. For home address the rate was 75% in 2014, and 74% in 2015; for term-time address the rate was 87% in both years. The rates are lower than for some other sources for a number of reasons for example, only a postcode is available, not a full address and home address includes international students (where no postcode is present).

6.5. PSD specific checks

Internal migration

The principal variables of interest are date of birth, sex, home address, term-time address, domicile (country of residence), campus identifier, mode of study and student identifier. These variables give rise to age and geo-referenced local authorities.

Checks are performed to determine: the level of missingness of these variables; that records have valid entries; that non-UK student levels and levels of duplication seem plausible and that the level of these are consistent over time. Duplicates are expected as students can register for more than one course at the same time and these are resolved to a single record during processing.

Local authority distribution of international migrants

Checks are undertaken to ensure that counts (total and local authority) are similar to the previous year for relevant data items and that proportions are of a similar magnitude. Where checks reveal changes to historical patterns, possible explanations are investigated and the supplier is contacted if necessary to provide further clarity.

Data including country of residence prior to application, location of study and previous educational institution are used in processing to determine whether or not the student is a migrant.

HESA data are matched to the previous year's dataset using a combination of name, date of birth and HESA student identifier. This combination of variables is also used alongside expected length of study, date of commencement of study and end date of course to a) resolve the dataset so that there is only a single record for a student (students can be listed in the dataset multiple times as they can enrol on more than one course at the same time), b) to determine whether that student is an international student and if so, c) whether they should be deemed a long- or short-term migrant.

HESA data are matched to other datasets using a combination of date of birth, sex and postcode (either term-time or home address could be used depending on the data being matched).

The principal variables of interest are date of birth, sex, postcode of term-time address, postcode of home address and whether or not the person is deemed to be a long-term or short-term migrant. This information is then fed into a matching process to determine a distribution of long- and short-term migrants, with long-term migrants feeding into the mid-year population estimates and short-term migrants feeding into short-term international migration.

More sense checks along the lines of those outlined above are then carried out.

Short-term international migration

The short-term international migration team receive data from the team carrying out the work on the LA distribution of migrants, who have extracted and processed the data from the raw HESA dataset. As such the short-term international migration team do not work with the raw data themselves. The short-term international migration team receive a table detailing local authority numbers and their distribution, they conduct checks to ensure that the:

- data are complete

- data seem plausible

- data are consistent with time-series trends

- distribution by local authority provided sums correctly

6.6. Comparison with other data sources

Security requirements means that data with identifiers (name, address, date of birth) cannot be compared with other sources, or with older (or newer) versions of the same data, until the data have been hashed. A further security measure of separation of duties means the team undertaking the pre-processing are not able to undertake analysis on the data once they have been hashed. Analysis is undertaken by separate teams. This substantially limits the degree of comparison possible.

Metrics from one year to another can be compared, such as geo-referencing rates as above. Other comparisons need to take place on data where only hashed (that is encrypted) data are available.

A comparison with data from year to is made as part of internal acceptance testing by the internal analysis team (see Acceptance testing).

Other than the decennial Census, there are no natural comparators to HESA: it is the only notable data source which provides details of those attending higher education within the UK. Given the nature of the age structure and the geographic concentration of the dataset, it is only possible to compare HESA data linked to other data sources with those data sources not linked to HESA.

6.7. Distortive effects of performance measurements and targets

Timing of data

Data are provided to HESA by universities in a bid to secure teaching funding from higher education funding councils, this can give rise to issues because universities have an incentive to provide information on student enrolment as close to the start of the academic year as possible, before students begin dropping out, to maximise funding received.

Over-coverage

This means that those dropping out of university are likely to be recorded in the HESA dataset when, for statistical purposes, they should not. As a result they will be recorded as being resident at their term-time address, rather than their home address where they are more likely to be resident for the majority of the year: inflating the student population in the university local authority and reducing home local authorities, depending on when the move takes place. This assumes that the university drop out either a) failed to register for GP services at university or b) has registered for GP services at university and then failed to re-register following their move home; either of these means that there is no indicator on other sources used to correctly identify the area of actual residence.

Timeliness of address

HESA data does not always accurately reflect the student's term-time address for the majority of the year (the basis for internal migration statistics), due to the student not updating details with their university either at the start of the academic year or following a move within the academic year. As a result, population changes due to the student moving their term-time residence across geographical boundaries may be missed. This is an issue for several university areas where campuses or halls of residence are close to geographical boundaries. An example of this is the University of Leicester where there is a hall of residence in a neighbouring local authority (Oadby and Wigston).

6.8. Impacts for producing statistics

Strengths

The HESA data:

- should capture every student registered at a covered higher education institution

- provides student level data covering key demographic data

- provides a good coverage source for students: an age group known not to update their GP registration in good time, making up shortfalls in other sources

- provides a home address so we know where students have come from

- provides a good coverage source for international students starting on higher education courses and where they are based

- includes length of study so the difference between those believed to be long- and short-term international students can be determined

- allows longitudinal linkage so that the characteristics by age and sex of those entering and leaving higher education are known

Limitations

The HESA data:

- are limited to those engaged in higher education study at specific institutions

- does not always reflect the student's current term-time address, due to the student not updating details either at the start of the academic year or following a move within the academic year; as a result population changes due to residence changes across geographical boundaries may be missed

- only details those coming into and currently engaged with the higher education system, it does not provide details on what happens to drop outs or graduates

- only allows one home address and as such students who have more than one parental home may be moved from the wrong location or not matched correctly with other data sources

6.9. Risk

The HESA dataset, in common with many administrative sources, is subject to change beyond the control of statistical producers that can occur as a result of changes to underpinning rules, classifications, administrative definitions, systems, policy, interpretation and other external effects.

The team that undertakes the pre-processing of the data are in the process of expanding the range of quality assurance that is undertaken, and embedding quality into the coding and procedures used to pre-process the data. Nevertheless the need for improved quality assurance is recognised, and there is a risk that errors can occur. This element of risk has been reduced and will be further reduced. It is unlikely that a major issue will escape detection, however there remains a possibility that an unknown and undetected issue flows through into the data that may have a smaller impact on published statistics.

6.10. Quality issues

The table below summarises the key issues from the administrative source that have an impact on the quality of our statistics, based on HESA student data.

Quality Issues

Issue	Description	Impact	Mitigation
Over-coverage	Students who join a course and then drop-out are included provided they complete two weeks of study. This is correct for the administrative purpose but does not accord with the target usual residence definition used by ONS.	Estimates of internal migration flows to student areas will be overstated to a degree.	Currently no mitigation for this issue is used. However, there is ongoing research into the use of activity data and the timeliness of information across sources, but this should correct itself over time.
Timeliness of address information	There is anecdotally a tendency for students to move a short period into an academic year, but after they have registered their term-time address with the institution.	Internal migration moves may be missed and international migration estimates (long- and short-term) may be allocated to the wrong local authority area.	In many areas moves can be a short distance and may not cross an LA boundary so have no impact. However this remains a larger risk for some areas such as institutions close to or on a border (for example University of Warwick) and for some large cities, especially London. Research is under way into activity and the timeliness of information that may help somewhat in the future. However this remains an issue that will cause some inaccuracy in the statistics.
Data Item – and Matching	Errors in the name or date of birth fields can mean failure to match students. These may be due to issues like the use of diminutives (Bob instead of Robert), transposing day and month, typos and other transcription errors.	Missing a match can result in an internal move being missed, inflating the population estimate in the 'home' local authority and reducing the term-time location's estimate.	Research is under way into activity and the timeliness of information that may help somewhat in the future. However this remains an issue that will cause some inaccuracy in the statistics, but this should correct itself over time.
Date of birth	An incorrect date of birth, such as a transposed month and day or typo, may result in an incorrect age being calculated.	This may mean that the age information assumed for internal migrants is incorrect.	This is thought to be a minor issue, with a limited impact on the age distribution of population and migration estimates.

Notes for Producer's quality assurance investigations and documentation:

1. This is a standard process for SRE datasets and all of the error types listed have been identified at different times across the range of datasets that are acquired for the SRE.

2. Matchkeys are alphanumeric strings generated within processing from other data held on each record. For work on HESA data they are based on combinations of key demographic data (name, date of birth, sex, etc.). They are generated to enable automatic matching of individuals across different datasources.

3. Hashing, used here as shorthand for applying a cryptographic hash function. This is a function that is a mathematical algorithm that maps data of arbitrary size to a bit string of a fixed size (a hash function) which is designed to also be one-way function, that is, a function which is infeasible to invert. Hashing is essentially the one-way encryption of data.

7 . Conclusion

Higher Education Statistics Agency (HESA) data provide an accurate data source of information on students in higher education. These data can be used effectively to estimate migration and population, providing a timeliness of information that is not always present in other sources.

There are limitations of the data, such as counting students in the data after two weeks attendance or failure to capture latest address following a student's move. These are features of the data and the way it is collected, which can have an impact on population statistics.

Whilst these issues do cause some inaccuracy in the statistics, the magnitude of this is small and the data meet information gaps in other data, thus providing a valuable contribution to the accuracy of the statistics for which they contribute.

7.1. Annex 1

List of Higher Education Establishments

The following is a list of higher education establishments reporting to HESA as of academic year 2014 to 2015.

England

Anglia Ruskin University
Aston University
Bath Spa University
The University of Bath
University of Bedfordshire
Birkbeck College
Birmingham City University
The University of Birmingham
University College Birmingham
Bishop Grosseteste University
The University of Bolton
The Arts University Bournemouth
Bournemouth University
The University of Bradford
The University of Brighton
The University of Bristol
Brunel University London
Buckinghamshire New University
The University of Cambridge
The Institute of Cancer Research
Canterbury Christ Church University
The University of Central Lancashire
University of Chester
The University of Chichester
The City University
Conservatoire for Dance and Drama
Courtauld Institute of Art
Coventry University
Cranfield University
University for the Creative Arts
University of Cumbria
De Montfort University
University of Derby
University of Durham

The University of East Anglia
The University of East London
Edge Hill University
The University of Essex
The University of Exeter
Falmouth University
The National Film and Television School[1]
University of Gloucestershire
Goldsmiths College
The University of Greenwich
Guildhall School of Music and Drama
Harper Adams University
University of Hertfordshire
Heythrop College
The University of Huddersfield
The University of Hull
Imperial College of Science, Technology and Medicine
The University of Keele
The University of Kent
King's College London
Kingston University
The University of Lancaster
Leeds College of Art
Leeds Beckett University
The University of Leeds
Leeds Trinity University
The University of Leicester
The University of Lincoln
Liverpool Hope University
Liverpool John Moores University
The Liverpool Institute for Performing Arts
The University of Liverpool
Liverpool School of Tropical Medicine[1]
University of the Arts, London
London Business School
University of London (Institutes and activities)
London Metropolitan University
London South Bank University
London School of Economics and Political Science
London School of Hygiene and Tropical Medicine
Loughborough University
The Manchester Metropolitan University
The University of Manchester
Middlesex University
University of Newcastle-upon-Tyne
Newman University
The University of Northampton
University of Northumbria at Newcastle
Norwich University of the Arts
University of Nottingham
The Nottingham Trent University
The Open University
Oxford Brookes University
The University of Oxford
Plymouth College of Art[1]
University of Plymouth
The University of Portsmouth
Queen Mary University of London
Ravensbourne
The University of Reading
Roehampton University
Rose Bruford College
Royal Academy of Music
Royal Agricultural University
Royal College of Art
Royal College of Music

The Royal Central School of Speech and Drama[1]
Royal Holloway and Bedford New College
Royal Northern College of Music
The Royal Veterinary College
St George's Hospital Medical School
St Mary's University, Twickenham
The University of Salford
The School of Oriental and African Studies
Sheffield Hallam University
The University of Sheffield
Southampton Solent University
The University of Southampton
Staffordshire University
University of St Mark and St John
University Campus Suffolk
The University of Sunderland
The University of Surrey
The University of Sussex
Teesside University
Trinity Laban Conservatoire of Music and Dance
University College London[1]
The University of Warwick
University of the West of England, Bristol
The University of West London
The University of Westminster
The University of Winchester
The University of Wolverhampton
University of Worcester
Writtle College
York St John University
The University of York

Wales

Aberystwyth University
Bangor University
Cardiff University
Cardiff Metropolitan University
Glyndr University
Swansea University
University of Wales Trinity Saint David
University of South Wales

Scotland

The University of Aberdeen
University of Abertay Dundee
The University of Dundee
Edinburgh Napier University
The University of Edinburgh
Glasgow Caledonian University
Glasgow School of Art
The University of Glasgow
Heriot-Watt University
Queen Margaret University, Edinburgh
The Robert Gordon University
The University of St Andrews
SRUC
The University of Stirling
The University of Strathclyde
University of the Highlands and Islands
The University of the West of Scotland

Northern Ireland

The Queen's University of Belfast
St Mary's University College
Stranmillis University College
University of Ulster

Notes for Conclusion:

1. See [HESA note](#).

8 . Annex 1

List of Higher Education Establishments

The following is a list of [higher education establishments reporting to HESA](#) as of academic year 2014 to 2015.

England

Anglia Ruskin University
Aston University
Bath Spa University
The University of Bath
University of Bedfordshire
Birkbeck College
Birmingham City University
The University of Birmingham
University College Birmingham
Bishop Grosseteste University
The University of Bolton
The Arts University Bournemouth
Bournemouth University
The University of Bradford

The University of Brighton
The University of Bristol
Brunel University London
Buckinghamshire New University
The University of Cambridge
The Institute of Cancer Research
Canterbury Christ Church University
The University of Central Lancashire
University of Chester
The University of Chichester
The City University
Conservatoire for Dance and Drama
Courtauld Institute of Art
Coventry University
Cranfield University
University for the Creative Arts
University of Cumbria
De Montfort University
University of Derby
University of Durham
The University of East Anglia
The University of East London
Edge Hill University
The University of Essex
The University of Exeter
Falmouth University
The National Film and Television School[1]
University of Gloucestershire
Goldsmiths College
The University of Greenwich
Guildhall School of Music and Drama
Harper Adams University
University of Hertfordshire
Heythrop College
The University of Huddersfield
The University of Hull
Imperial College of Science, Technology and Medicine
The University of Keele
The University of Kent
King's College London
Kingston University
The University of Lancaster
Leeds College of Art
Leeds Beckett University
The University of Leeds
Leeds Trinity University
The University of Leicester
The University of Lincoln
Liverpool Hope University
Liverpool John Moores University
The Liverpool Institute for Performing Arts
The University of Liverpool
Liverpool School of Tropical Medicine[1]
University of the Arts, London
London Business School
University of London (Institutes and activities)
London Metropolitan University
London South Bank University
London School of Economics and Political Science
London School of Hygiene and Tropical Medicine
Loughborough University
The Manchester Metropolitan University
The University of Manchester
Middlesex University
University of Newcastle-upon-Tyne
Newman University

The University of Northampton
University of Northumbria at Newcastle
Norwich University of the Arts
University of Nottingham
The Nottingham Trent University
The Open University
Oxford Brookes University
The University of Oxford
Plymouth College of Art[1]
University of Plymouth
The University of Portsmouth
Queen Mary University of London
Ravensbourne
The University of Reading
Roehampton University
Rose Bruford College
Royal Academy of Music
Royal Agricultural University
Royal College of Art
Royal College of Music
The Royal Central School of Speech and Drama[1]
Royal Holloway and Bedford New College
Royal Northern College of Music
The Royal Veterinary College
St George's Hospital Medical School
St Mary's University, Twickenham
The University of Salford
The School of Oriental and African Studies
Sheffield Hallam University
The University of Sheffield
Southampton Solent University
The University of Southampton
Staffordshire University
University of St Mark and St John
University Campus Suffolk
The University of Sunderland
The University of Surrey
The University of Sussex
Teesside University
Trinity Laban Conservatoire of Music and Dance
University College London[1]
The University of Warwick
University of the West of England, Bristol
The University of West London
The University of Westminster
The University of Winchester
The University of Wolverhampton
University of Worcester
Writtle College
York St John University
The University of York

Wales

Aberystwyth University
Bangor University
Cardiff University
Cardiff Metropolitan University
Glyndr University
Swansea University
University of Wales Trinity Saint David
University of South Wales

Scotland

The University of Aberdeen
University of Abertay Dundee
The University of Dundee
Edinburgh Napier University
The University of Edinburgh
Glasgow Caledonian University
Glasgow School of Art
The University of Glasgow
Heriot-Watt University
Queen Margaret University, Edinburgh
The Robert Gordon University
The University of St Andrews
SRUC
The University of Stirling
The University of Strathclyde
University of the Highlands and Islands
The University of the West of Scotland

Northern Ireland

The Queen's University of Belfast
St Mary's University College
Stranmillis University College
University of Ulster

Notes for Conclusion:

1. See [HESA note](#).