Department for Education

Methodology Section: Graduate outcomes using the Longitudinal Education Outcomes (LEO) data

19 March 2020

Contents

1. Introduction 2
   - Background to the Longitudinal Educational Outcomes (LEO) dataset 2
2. Data quality 2
   - Employment and earnings data 2
   - Coverage 3
3. Methodology 4
   - Time period 4
   - Employment outcomes 4
   - Annualised earnings 7
4. Definitions 8
   - Domicile categories 8
   - Subject areas 10
   - Residence 11
   - Prior Attainment 13
5. Data matching and match rates 14
   - Overall match rates 15
   - Match rate by graduate characteristic 16
6. Get in touch 17
   - Media enquiries 17
   - Other enquiries/feedback 17

About this section

This document provides further background on the data used in this release and outlines key defintions used in the publication as well as data quality and methodological changes since the last publication in the series.

Contact: Email: HE.LEO@education.gov.uk Press office: 020 7783 8300 Public enquiries: 0370 000 2288

1. Introduction

Background to the Longitudinal Educational Outcomes (LEO) dataset

The Small Business, Employment and Enterprise Act 2015 enabled government, for the first time, to link higher education and tax data together to chart the transition of graduates from higher education into the workplace[1]. One of the advantages of linking data from existing administrative sources is that it provides a unique insight into the destinations of graduates without imposing any additional data collection burdens on universities, employers or members of the public. Compared to existing sources of graduate outcomes data, it is also based on a considerably larger sample, does not rely on survey methodology, and can track outcomes across time to a greater extent than was previously possible.

The LEO dataset links information about students, including

- personal characteristics such as sex, ethnic group and age
- education, including schools, colleges and higher education institution attended, courses taken and qualifications achieved
- employment and income
- benefits claimed

It is created by combining data from the following sources:

- the National Pupil Database (NPD), held by the Department for Education (DfE)
- Higher Education Statistics Agency (HESA) data on students at UK publicly funded higher education institutions and some alternative providers, held by DfE
- Individualised Learner Record data (ILR) on students at further education institutions, held by DfE
- employment data from the Real Time Information System (RTI). RTI contains information formerly collected on the P45 and P14 forms, held by Her Majesty's Revenue and Customs (HMRC)
- data from the Self-Assessment tax return, held by HMRC
- the National Benefit Database, Labour Market System and Juvos data, held by the Department for Work and Pensions (DWP)

By combining these sources, we can look at the progress of higher education leavers into the labour market.

The privacy notice explaining how personal data in this project is shared and used can be found here.

2. Data quality

Employment and earnings data

The **employment data** covers those with P45 and P14 records submitted through the Pay As You Earn (PAYE) system. These figures have been derived from administrative IT systems that, as with any large-scale recording system, are subject to possible errors with data entry and processing. While some data cleaning was necessary, the resulting data looks to provide a good reflection of an individual's employment and earnings for the year.

For the purposes of collecting taxes only the tax year of employment is needed, accurate start and end dates within the tax year are not required. For this reason, issues encountered with the employment data included records with duplicate dates and records with dates which were invalid for our intended use (for example, where an employment start date occurred after the end date).

Additionally, a number of returns are found to have missing start dates due to the employer not forwarding a timely P45. The default dates recorded in the dataset are either 6 April (the first day of the tax year) or, where only an end date is known, the day before that end date. Similarly, for records where the

[1] For more information on the legal powers governing the dataset please see section 78 of the Small Business, Enterprise and Employment Act 2015 and sections 87-91 of the Education and Skills Act 2008.

employment is known to have come to an end within a tax year but the end date is not known, the record is given a default 5 April end date, the last day of the tax year.

Individuals can also have overlapping spells of employment. Before carrying out analysis, the P45 and P14 records for each individual were cleaned and then merged into a single record to give a longitudinal picture of their employment and a total sum of their earnings in each tax year.

Before cleaning, the dataset contained just under 73 million P45 records. Of these, just over 6.5 million invalid records were removed (the majority were duplicate records). Of the remaining records, around 20% had an uncertain start date and around 20% an uncertain end date. For each uncertain date, we used dates from other employment or benefits records for that individual to create a merged employment spell with a known start and end date.

Example 1: Two employment spells

Spell A |---------|

Spell B |----------|----------|

Merged result |----------|

In example 1, the start date of spell B is uncertain with its possible range shown in blue. In this instance we can merge the two records resulting in an employment spell with the start date of spell A and an end date from spell B.

Any remaining uncertain dates were imputed through random sampling of gap lengths from a frequency distribution that was constructed from gaps with a known length.

Coverage

Beginning in April 2013, the P45 reporting system was phased out in favour of the Real Time Information (RTI) system, which requires employers to submit information to HMRC each time an employee is paid. This system has now reached full deployment. RTI offers substantial improvements to the P45 system in terms of data coverage, since employers must now provide information on all their employees if even one employee of the company is paid above the Lower Earnings Limit. The move to RTI will mean that data coverage is high for the 2014/15 to 2017/18 tax years used in this publication.

As well as employment data for those who pay tax through PAYE, the employment data now additionally includes those who pay tax through self-assessment.

Self-assessment forms are completed by a range of people who for example are self-employed, have received income from investments, savings or shares and by people who have complicated tax affairs. A list of people who are required to complete a self-assessment return can be found at www.gov.uk/self-assessment-tax-returns/who-must-send-a-tax-return. We have recently obtained a new self-assessment earnings dataset from HMRC, which contains variables on:

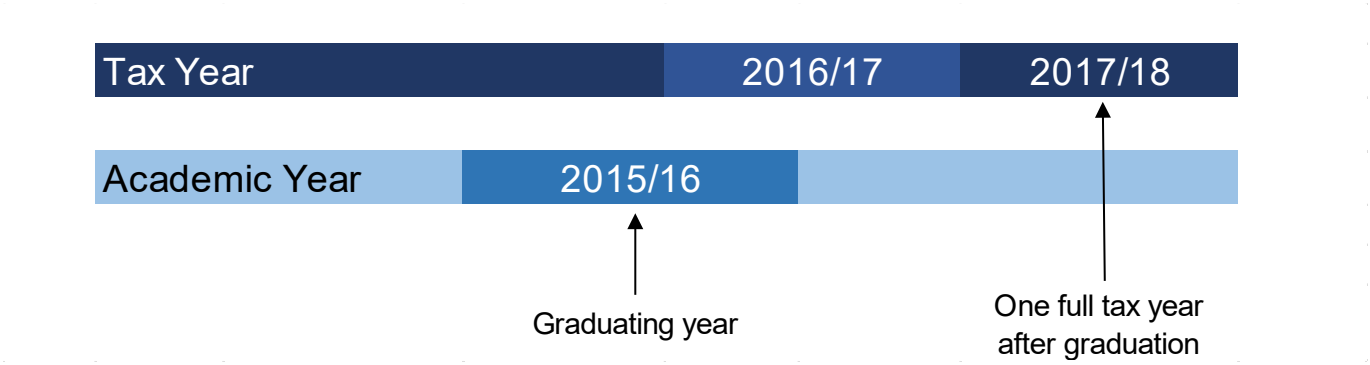
- Earnings received through employment (PAYE)
- Income from partnership enterprises
- Income from sole-trader enterprises
- Total earnings for the tax year from the self-assessment form.

We have used the income from partnership enterprises and income from sole-trader enterprises to ascertain graduates who are self-employed and their earnings from self-employment enterprises. We have taken a sum of these two variables, and where the sum of these is greater than £0, graduates are classified as self-employed. Where self-employment earnings are used, the earnings amount is the sum of these two variables.

3. Methodology

Time period

The earliest time period for which employment and earnings data is reported is one year after graduation. This refers to the first full tax year after graduation. Hence, for the 2015/16 graduation cohort, the figures one year after graduation refer to employment and earnings outcomes in the 2017/18 tax year. This time period was picked as using the tax year that overlaps with the graduation date would mean that graduates are unlikely to have been engaged in economic activity for the whole tax year.



In this publication, we look at one, three, five and ten years after graduation, focussing on the 2017/18 tax years with some comparative analysis with 2014/15 to 2016/17 tax years. Thus we look at employment and earnings outcomes in the 2017/18 tax year for graduates from the 2006/07, 2011/12, 2013/14 and 2015/16 academic years. For 2015/16 tax year graduates from the 2004/05, 2009/10, 2011/2012 and 2013/2014 academic years and the other tax years are calculated using this method.

Employment outcomes

We refer to a graduate as **matched** if they have been successfully matched to the Department for Work and Pensions' Customer Information System (CIS) or if they have been matched to a further study instance on the HESA Student Record. Graduates who have not been matched to CIS or a further study record are referred to as **unmatched**. These graduates were not found on DWP's Customer Information System (CIS), either because they had never been issued with a National Insurance number or because the personal details provided from the HESA data did not fulfil the matching criteria. **These graduates are excluded from calculations performed for UK domiciled populations**. In particular, they are not included in outcomes categories in Tables 1 to 14 and 20 to 32.

UK domiciled graduates who have been **matched** are then placed in one of five outcomes categories. These are:

1. Activity not captured
2. No sustained destination
3. Sustained employment only
4. Sustained employment with or without further study
5. Sustained employment, further study or both.

Unmatched graduates are **included** in the denominator when calculating employment outcomes for non-UK domiciled graduates (Tables16, 17 and 33) and are placed in a separate 'unmatched' outcome category. For these populations the match rates are much lower and non-UK graduates are much more likely to leave the UK after graduation. Including these graduates in the calculations means we get a better indication of the proportion of graduates who have stayed in the UK to work or study after graduation, making it easier to compare countries with very different match rates.

For non-UK domiciled graduates the employment outcome categories should not be used as an indication of success in finding employment after graduation, it is likely that the majority of these graduates who are 'unmatched' or in 'activity not captured' are employed outside of the UK.

More information on match rates is given in section 6: Data matching and match rates. If a graduate is unmatched on the CIS but has a further study record for the tax year in question, then they are counted as being in further study, and hence are not in the unmatched category.

Activity not captured
Graduates in this category have been successfully matched to CIS but do not have any employment, out-of-work benefits or further study records in the tax year of interest. Reasons for appearing in this category include: moving out of the UK after graduation for either work or study, voluntarily leaving the labour force or death.

No sustained destination
Graduates who have an employment or out-of-work benefits record in the tax year in question but were not classified as being in 'sustained employment' and do not have a further study record.

Sustained employment defined by P45 data
The 'sustained employment' measure aims to count the proportion of graduates in sustained employment in the UK following the completion of their course. The definition of sustained employment is consistent with the definition used for 16-19 accountability and the outcome based success measures published for adult further education[2]. This definition looks mainly at employment activity in the six month October to March period of each tax year. A graduate needs to be in paid employment for at least one day in five out of six months between October and March of a given tax year to be classified as being in 'sustained employment' in the given tax year. If they are not employed in March, they must additionally have at least one day in employment in the April of the same calendar year to be counted as being in sustained employment.

For instance, a graduate employed from 1st October 2017 to 5th January 2018 and then again from 30th March 2018 onwards would be classed as being in sustained employment in 2017/18 as although they are not employed in February 2018 they are employed in the other five months in the period from October 2017 to March 2018.

However a graduate employed from 1st October 2017 to 28th February 2018 but not employed in March 2018, would not be considered as being in sustained employment unless they had a day in employment April 2018.

Sustained employment defined by self-assessment data
This publication incorporates self-assessment data into measures of sustained employment. Self-assessment data captures the activity of individuals with income that is not taxed through PAYE, such as income from self-employment, savings and investments, property rental, and shares[3]. Currently, only data for the from 2013/14 onwards tax year is available for inclusion in LEO. For this reason we have only published employment and earnings outcomes for these tax years in this publication.

For the purposes of this publication, individuals are classed as being in sustained employment in the tax year if they meet our definition of sustained employment based on PAYE **or** have returned a self-assessment form stating that they have received income from self-employment and their earnings from a Partnership or Sole-Trader enterprise are more than £0 (profit from self-employment). These individuals may or may not have an additional PAYE record. Individuals who have received income through self-assessed means other than self-employment, such as through rental of property, and do not have a PAYE record, are not classed as being in employment (either sustained or unsustained). Those who have made a loss from self-employment are currently excluded from sustained employment as we are unable to

[2] https://www.gov.uk/government/statistics/adult-further-education-outcome-based-success-measures
[3] A full list of income sources that must be declared through a self-assessment return can be found here, https://www.gov.uk/self-assessment-tax-returns/who-must-send-a-tax-return

distinguish between those who made a loss and those who submitted self-assessment returns for other reasons at this moment in time.

Further study
A graduate is defined as being in further study if they have a valid higher education study record at any UK HEI on the HESA Student Record or designated English Alternative Provider (AP) on the AP HESA Student Record that overlaps the relevant tax year. Further study undertaken at further education colleges is not currently reflected in these figures but we will review this in future publications. The further study does not have to be at postgraduate level to be counted. The purpose of this category is to identify how students spent their time in the relevant tax year and as such cannot be used to calculate the proportion of graduates who go on to postgraduate study. We have not counted instances lasting 14 days or less, a change from previous publications. Additionally, students enrolled on further education courses, on some initial teacher training enhancement, booster and extension courses, whose study status is dormant or who were on sabbatical are excluded from this indicator in line with our previous methodology.

As a tax year overlaps with two academic years, some students would be coming to the end of their further study in the tax year in question and some would be starting their further study. For example, those who graduated in the 2015/16 academic year and went straight on to a one-year masters course would be counted as being in further study in the 2017/18 tax year (one year after graduation) as their course would finish in July 2017. If a graduate from 2015/16 waited a year before starting their one-year masters course then they would typically be counted as being in further study in the 2017/18 tax year (one year after graduation) if their course started in September 2017 for instance.

Sustained employment only
Graduates are considered to be in sustianed employment **only** if they have a record of sustained employment (as defined either via the P45 or self assessment data) but no record of further study (as defined above).

Sustained employment with or without further study
Sustained employment with or without further study includes **all** graduates with a record of sustained employment (defined either via the P45 or self assessment data), regardless of whether they also have a record of further study (as defined above).

Sustained employment, further study or both
Sustained employment, further study or both includes all graduates with a record of sustained employment **or** further study. This category includes all graduates in the ‘sustained employment with or without further study’ category as well as those with a further study record **only**.

It is important to note that our definition of sustained employment does not distinguish between the different types of work that graduates are engaged in and so cannot provide an indication of the proportion of graduates who are employed in graduate occupations. Furthermore, we cannot distinguish between full-time and part-time employment.

The below table summarises the type of activity people may have to be unmatched or to fall into one of the five outcomes categories.

Table A: Classification of graduate outcomes

LEO category	Further study	Sustained employment	Any employment	Out-of-work Benefits
Unmatched	×	Unmatched to CIS	Unmatched to CIS	Unmatched to CIS
Activity not captured	×	×	×	×
No sustained destination	×	×	✓	×
	×	×	×	✓
	×	×	✓	✓
Sustained employment only	×	✓	✓	×
	×	✓	✓	✓
Sustained employment (with or without further study)	✓	✓	✓	×
	✓	✓	✓	✓
	×	✓	✓	×
	×	✓	✓	✓
Sustained employment, further study or both	✓	Unmatched to CIS	Unmatched to CIS	Unmatched to CIS
	✓	×	×	×
	✓	×	✓	×
	✓	×	×	✓
	✓	×	✓	✓
	✓	✓	✓	×
	✓	✓	✓	✓
	×	✓	✓	×
	×	✓	✓	✓

Annualised earnings

Earnings figures are only reported for those classified as being in sustained employment via PAYE and where we have a valid earnings record from the P14 **or** where they are self-employed and have reported income of over £0 for that tax year. Those in further study are excluded, as their earnings would be more likely to relate to part-time jobs. Note that our publications prior to December 2017 did not include earnings from self-assessment. Under the new methodology, some graduates will have increased earnings if they have PAYE earnings as well as self-employment earnings. However, there are also more graduates included in the earnings calculations – those who have self-employment earnings but do not have qualifying PAYE earnings. This group typically has lower earnings than graduates with PAYE earnings. Thus, the reported median earnings under the new methodology is not necessarily higher under the new

methodology compared to the old methodology. See our December 2017 publication[4] for more details on the effect of this methodology change.

Under our new methodology, PAYE and earnings from self-employment are treated differently.

For each graduate who has been paid through the PAYE system, the earnings reported for them for a given tax year are divided by the number of days recorded in the employment spell in that same tax year. This provides an average daily wage, which is then multiplied by the number of days in the tax year to create their annualised earnings[5].

This calculation has been used to maintain consistency with figures reported for further education learners after study. It provides students with an indication of the earnings they might receive once in stable and sustained employment.

For earnings from self-employment, raw earnings are used. Due to the nature of the Self-Assessment tax return, dates of self-employment are not required and therefore are not available to annualise the self-employment earnings in the same way that PAYE earnings are annualised. We are therefore assuming that the Self-Assessment tax return relates to activity that took place over the full tax year.

Where a graduate has income from both sustained employment paid through PAYE and though self-employment, the earnings used for this graduate is the sum of their annualised PAYE earnings and their raw earnings from self-employment. It should be noted that a graduate with a PAYE records (that does not reach the 'sustained' criteria) **and** a self-employment earnings record will be counted as being in 'sustained employment' but we do not include their earnings in the earnings calculation. This is to avoid the risk of annualising PAYE data that could be based on a very short earnings spell.

The annualised earnings calculated are slightly higher than the raw earnings reported in the tax year. This is because the earnings of those who did not work for the entire tax year will be higher when annualised. The difference between the annualised and raw figures decreases as time elapses after graduation. Overall median annualised earnings one year after graduation are around £650 higher than the overall median raw earnings reported in the data. Five years after graduation, the overall median annualised earnings are less than £300 higher than the overall median raw earnings. The trend follows for both graduates who are in PAYE employment only and graduates who earnt income from both PAYE employment and self-employment.

Information provided on the Self-Assessment tax return includes a field on earnings through PAYE employment, which we have used only where P14 earnings is not present.

All earnings presented are nominal. They represent the cash amount an individual was paid and are not adjusted for inflation (the general increase in the price of goods and services).

4. Definitions

Domicile categories

Domicile categories have been based upon graduates' domicile prior to the start of their course, as recorded in the HESA student record for graduates from HEIs and as recorded in the ILR for graduates from FECs. Graduates have been categorised into three top-level categories – UK, EU and Other Overseas. Due to data quality issues with the domicile variable on the ILR in the 2003/04 and 2004/05

[4] Graduate oubcomes (LEO): including self-employment earnings data.
[5] Note we do not know the actual number of days worked just the length of the employment spells, so this method does not adjust for part-time workers. For example, if a graduate is employed for the full tax year we will use 365 days (or 366 in leap year) in the calculation.

academic years, we have not included non-UK domiciled graduates from FECs in the tables for these years.

UK domiciled refers to graduates domiciled in England, Scotland, Wales or Northern Ireland prior to the start of their course. Tables 1 to 14 and 20 to 32refer only to UK domiciled graduates.

EU domiciled refers to graduates domiciled in the EU other than in England, Scotland, Wales or Northern Ireland. As such, graduates domiciled in Gibraltar have been classed as EU domiciled. Over the period covered by this publication, the membership of the EU has expanded and hence different graduating cohorts consist of different sets of countries. Graduates have been classed as EU domiciled if their recorded country of domicile was a member of the EU at the start of their year of graduation. Table B below details for which cohort(s) each country has been designated as part of the EU domiciled category. Countries listed include all of their European Union territories; for instance, Finland includes the territory of the Åland islands.

Table B: Countries and territories included in the European Union category by graduating cohort

Country/Territory	**Graduating cohorts in which domicile is counted as EU domiciled**
Austria Belgium Denmark Finland France Germany Gibraltar Greece Ireland Italy Luxembourg Netherlands Portugal Spain Sweden	All graduating cohorts (2003/04, 2004/05, 2008/09, 2009/10, 2010/11, 2011/12, 2012/13, 2013/14, 2014/15, 2015/16)
Cyprus Czech Republic Estonia Hungary Latvia Lithuania Malta Poland Slovakia Slovenia	2004/05, 2008/09, 2009/10, 2010/11, 2011/12, 2012/13, 2013/14, 2014/15, 2015/16
Bulgaria Romania	2008/09, 2009/10, 2010/11, 2011/12, 2012/13, 2013/14, 2014/15, 2015/16
Croatia	2013/14, 2014/15, 2015/16

Overseas domiciled refers to graduates domiciled in countries/territories not belonging to the European Union. The Crown Dependencies of Jersey, Guernsey and the Isle of Man are not part of the UK or of the European Union and thus they have been included this category.

Table 19 in the accompanying tables gives employment and earnings outcomes for the 20 largest countries of domicile[6] within our data. We have followed methodology used by HESA in defining country of domicile, for instance aggregating together the various territories of France in the France total but keeping China and Hong Kong separate.

Note that country of domicile is not the same as nationality (as recorded on the HESA student record). For instance, in 2012/13, 91% of UK domiciled graduates were UK nationals, while 7% of EU domiciled graduates and about 4% of overseas domiciled graduates were UK nationals.

Subject areas

The Higher Education Statistics Agency (HESA) are changing the way they report subjects from the 2019/20 academic year; the current Joint Academic Coding System (JACS) is being replaced by the Higher Education Classification of Subjects (HECoS). HESA have produced the Common Aggregation Hierarchy (CAH) which bridges between the two systems, and to maintain consistency across years were are using level 2 of the CAH to report breakdowns by subject area.

The number of subject categories increases to 35, compared with 23 using the previous JACS groupings. In many cases the CAH categories map exactly to a JACS category (e.g. Medicine and dentistry, Mathematical sciences, Creative arts and design) ; in the remainder of cases, the CAH categories just provide a more detailed split compared with JACS groups (e.g. the JACS group 'Engineering & Technology' is now split into 'Engineering' and 'Materials and technology' separately; similarly for 'Historical and Philosophical Studies' split into 'History and archaeology' and 'Philosophy and religious studies'). More information on HECoS and CAH can be found here: https://www.hesa.ac.uk/innovation/hecos

CAH Code	Subject
CAH01-01	Medicine and dentistry
CAH02-02	Pharmacology, toxicology and pharmacy
CAH02-04	Nursing and midwifery
CAH02-05	Medical sciences
CAH02-06	Allied health
CAH03-01	Biosciences
CAH03-02	Sport and exercise sciences
CAH04-01	Psychology
CAH05-01	Veterinary sciences
CAH06-01	Agriculture, food and related studies
CAH07-01	Physics and astronomy
CAH07-02	Chemistry
CAH07-04	General, applied and forensic sciences
CAH09-01	Mathematical sciences

[6] The table includes the top 20 non-UK countries of domicile as defined by the numbers of first degree graduates from English institutions in 2013/14.

CAH10-01	Engineering
CAH10-03	Materials and technology
CAH11-01	Computing
CAH13-01	Architecture, building and planning
CAH15-01	Sociology, social policy and anthropology
CAH15-02	Economics
CAH15-03	Politics
CAH15-04	Health and social care
CAH16-01	Law
CAH17-01	Business and management
CAH19-01	English studies
CAH19-02	Celtic studies
CAH19-04	Languages and area studies
CAH20-01	History and archaeology
CAH20-02	Philosophy and religious studies
CAH22-01	Education and teaching
CAH23-01	Combined and general studies
CAH24-01	Media, journalism and communications
CAH25-01	Creative arts and design
CAH25-02	Performing arts
CAH26-01	Geography, earth and environmental studies

It is important to note that, even with these additional splits, each CAH subject area can still include a diverse range of subjects, some of which will lead to significantly different employment and earnings outcomes. For example, 'subjects allied to medicine not otherwise specified' contains courses ranging from nutrition and dietetics to biomedical sciences. We have not attempted to split the CAH codes down further as this would lead to an increase in the amount of data being suppressed.

Residence

Residence information is based on term time accommodation recorded in the HESA student record/ the ILR. Note that a student's residence status may potentially change during the course of their studies; we use their status during their graduating year, this may be different to their residence status in their earlier years. Collection of this variable is mandatory for full-time students and those on sandwich courses; coverage is lower for part-time and other modes of study – see Table C below.

Table C: Coverage of residence data by Mode of Study

Coverage: Young (under 21 at start of course) UK domiciled male and female first degree graduates from English HEIs and FECs
Cohorts: 2003/04, 2004/05, 2008/09, 2009/10, 2010/11, 2011/12, 2012/13, 2013/14

Mode of Study	Coverage of Residence data (%)
Full-time	94

Mode of Study	Coverage of Residence data (%)
Sandwich Degree	95
Part-time	69
Other	31

We have presented residence information in three categories: living at parental/guardian home, living elsewhere and not known. The 'living elsewhere' category consists of a variety of different living arrangements – see Table D for the breakdown of residence into term time accommodation in 2009/10, and the proportion in each of these finer categories. In this table, the 'not in residence at institution' category includes students on an industrial placement or language year abroad, 'own residence' includes a student's permanent residence, which may be either owned or rented by them, and 'other rented accommodation' refers to a more temporary arrangement, including renting in a flat share on a yearly basis. Note that the proportion in each category vary slightly each year, as does the categorisation used by HESA.

Table D: Breakdown of term time accommodation in 2009/10
Coverage: Young (under 21 at start of course) UK domiciled male and female first degree graduates from English HEIs and FECs
Cohorts: 2009/10

Residence	Term time accommodation	Proportion (%)
Living at parental/guardian home	Living at parental/guardian home	23
Living elsewhere	Institution maintained property	9
	Private-sector halls	3
	Not in attendance at the institution	1
	Own residence	11
	Other rented accommodation	40
	Other	4
Not known	Not known	8

Current Region

The current region geographical location data is based on the latest address that DWP has recorded for each individual on their Customer Information System (CIS). The LEO dataset does not contain the actual address or postcode for each individual, we currently have data on the Government Office Region (GOR) and Local Authority District where the individual lives at the end of each tax year.

The CIS is primarily updated when an individual notifies DWP or HMRC of a change of address or through the individual interacting with a tax or benefit system. Individuals who have not been matched to the CIS will not have geographical information. This does not have an adverse effect on the data analysis as 'unmatched' graduates are excluded from employment and earnings outcomes.

For those matched to CIS, address data is available in nearly all cases (over 99.8%), however for those who are not in receipt of benefits or contributing to the tax system then this information could be out of date. Even when contributing to the tax system, employee address is not a mandatory field in the data submitted to HMRC via employers HR systems. It is also possible that in the years soon after leaving university

graduates may still use their parents address if they are moving frequently between rented accommodation. More work is needed to try and understand how big an impact this has on the address data held on CIS.

Prior Attainment

Prior attainment is the attainment of students prior to commencing their higher education course. We have calculated prior attainment based on key stage 5 qualifications recorded in the National Pupil Database (NPD), which contains data about pupils in schools and colleges in England. Due to the coverage of the NPD, we are unable to provide prior attainment breakdowns for our earliest cohorts (graduates in 2003/04 and 2004/05) or for mature students. Note also that coverage for graduates domiciled in Scotland, Wales and Northern Ireland is significantly lower than for those domiciled in England, since only those who took their KS5 qualifications in England are included.

We have used a more detailed categorisation of prior attainment than previous publications. The majority of categories are based on point scores in A levels. For these categories we have included Applied A levels and Vocational A levels alongside traditional A levels. Prior to the 2009/10 academic year, the available grades from an A level were A, B, C, D, E, N and U, with A being the highest, E being the lowest passing grade and N and U being considered fails. From 2009/10 onwards, an A* grade, higher than an A, was also available. Among our graduate cohorts, only graduates from 2012/13 and 2013/14 would typically have had A* available to them. In order to keep our categories comparable across years, our categorisation does not distinguish between A and A* grades.

We use the following categories, listed in order of preference (i.e. if an individual satisfies the criteria of two or more categories, they are included only in the first of those categories):

4 As or more
360 points
300-359 points
240-299 points
180-239 points
Below 180 points
1 or 2 A level passes (and no other qualifications other than AS levels)
BTEC (regardless of grade)
Other (this includes mixtures of A levels and other qualifications)

For category ’4 As or more’, it is important to note that 3 A levels is usually enough for entry to most universities. Hence many students who might be capable of attaining four A grades would only take 3 A levels. Indeed, some schools only offer 3 A levels to their students.

For categories ’360 points’ to ‘Below 180 points’, we use the conversion between A level grades and points listed in Table E and only consider graduates with at least 3 A level passes. Hence ’360 points’ requires three grades of A or A*, while the threshold for ’300-359 points’ is equivalent to three Bs and the threshold for ’240-299 points’ is equivalent to three Cs.

Note also that graduates with one or two A level grades as well as a BTEC National Diploma would be included in the ‘BTEC’ category. This was chosen as a BTEC is considered a “full” set of qualifications while one or two A levels is usually not enough to be considered for entry to most first degree courses.

Table E: Conversion between A level grades and point scores

A level grade	Point Score
A or A*	120
B	100
C	80
D	60
E	40
N	Not counted as one of top 3 A levels
U	

5. Data matching and match rates

The HESA student records are matched to DWP's Customer Information System (CIS)[7] using an established matching algorithm based on the following personal characteristics: National Insurance Number (NINO), forename, surname, date of birth, postcode and sex. Some of these characteristics are simplified to make the matching process less time-intensive and allow more matches, for instance if a surname is misspelt in one of the datasets. Only the first initial of the forename is used, the surname is encoded using an English sound-based algorithm called SOUNDEX[8], and for most matches only the sector of the postcode is used. The National Insurance Number is not present on the HESA student record itself and has been matched on where possible by fuzzy matching with personal data from the Student Loans Company. This process increases the likelihood of finding a match with CIS. Accordingly groups less likely to take a student loan, for instance international students who are not eligible for one, are likely to have lower match rates.

All records accessed for analysis are anonymous so that individuals cannot be identified. The personal identifying records used in the actual matching process are accessed under strict security controls.

There are five match processes carried out, ranging from the highest quality and most likely to be accurate (Green) to the lowest quality and most likely to be a false match (Red-Amber). Table G shows the criteria for each match type.

Once the HESA records have been matched to the CIS the corresponding tax and benefits records for that individual can then be linked to their HESA record.

All match rate analysis in this chapter is restricted to the HESA population covered in this publication, that is, UK domiciled, first degree graduates from UK Higher Education Institutions.

[7] The CIS is a computer system used by the Department for Work and Pensions to store basic identifying information about customers and provides information on all individuals who have ever had a national insurance number.
[8] SAS function that turns a surname into a code representing what it sounds like, which allows some flexibility for different spellings. For example Wilson=Willson

Table F: Criteria for each type of match

Match quality	NINO (National Insurance number)	Forename (initial)	Surname (soundex)	Date of birth	Sex	Postcode (sector)
1. Green	✓	4 or 5 ✓✓✓✓✓				
2. Amber	✓	3 ✓✓✓				
3. Green-Amber	×	✓	✓	✓	✓	✓
4. Amber-Red	×	✓	✓	✓	1 ✓	
5. Red-Amber	×	×	×	✓	✓	✓ (full postcode)

Overall match rates

In this section we consider match rates to the CIS spine. This differs slightly from the match rates displayed in the main tables of this publication, which include also those without a CIS match but with a record of further study in the given year.

Table H shows the overall CIS match rates for graduates who studied full-time as well as the proportion with a tax or benefit record. Potential reasons for not being able to find a P45 record, despite having a match to the CIS spine, include: earning below the Lower Earnings Limit (LEL), self-employment, moving abroad and death.

Table G: Match rates for UK domiciled first degree graduates at English HEIs, by year of graduation

Coverage: UK domiciled male and female first degree graduates from English HEIs.

Cohorts: 2003/04, 2004/05, 2005/06, 2006/07, 2007/08, 2008/09, 2009/10, 2010/11, 2011/12, 2012/13, 2013/14, 2014/15, 2015/16

Academic year	Matched to tax/ benefit record (%)	Matched to CIS spine (%)
2003/04	0.95	0.95
2004/05	0.95	0.96
2005/06	0.95	0.96
2006/07	0.96	0.97
2007/08	0.97	0.97
2008/09	0.97	0.97
2009/10	0.97	0.98
2010/11	0.97	0.98
2011/12	0.98	0.98
2012/13	0.98	0.99
2013/14	0.98	0.99
2014/15	0.98	0.99
2015/16	0.98	0.99

Table H shows that the match rate was very high for the most recent cohorts: 99% of full-time graduates in 2015/16 were matched using the CIS, and almost all of these had at least one tax record or out-of-work benefit record. This compares to a match rate of 95% of graduates in 2003/04. The higher match rates for more recent cohorts is at least partly explained because the CIS holds the most recent names and addresses for individuals, and so if the details change after someone graduates there is less chance that they will be matched.

These rates are marginally higher than previous publications, particularly in 2010/11 and 2011/12 where the increase was by around 1 percentage point.

Match rate by graduate characteristic

Table I shows match rates by sex. The match rate for females is slightly lower in the earlier years than for males, but this difference is negligible or non-existent in recent cohorts. As the CIS holds the latest information about an individual, anyone that has changed their name since graduation will have a different name on the CIS compared to their HESA record. This particularly affects females, due to a higher likelihood than males of changing their name upon marriage.

Table H: CIS match rate by sex

Coverage: UK domiciled male and female first degree graduates from English HEIs.

Academic year	Female (%)	Male (%)
2003/04	0.93	0.98
2004/05	0.94	0.98
2005/06	0.94	0.98
2006/07	0.95	0.98
2007/08	0.96	0.98
2008/09	0.97	0.98
2009/10	0.97	0.98
2010/11	0.97	0.98
2011/12	0.98	0.98
2012/13	0.99	0.99
2013/14	0.99	0.99
2014/15	0.99	0.99
2015/16	0.99	0.99

The match rates were also compared for different ethnic groups out of the UK-domiciled students. There was little consistent difference between the groups, the only exception being graduates whose self-declared ethnicity was Chinese, where the match rate was 95% in 2015/16. Further investigation showed that this was most likely due to the ethnically Chinese forenames and surnames being switched on one of the databases. This is more common for Chinese names, because the family name traditionally comes before the individual name. This hypothesis is further corroborated by the fact that ethnically Chinese students with common English names have match rates that are very similar to graduates from other ethnic groups.

The number of forenames or surnames an individual has can affect the match rate, because with multiple names it is more likely that they will not all be recorded, or there may be forenames recorded as surnames or vice versa. Analysis of the match rates showed that those with at least two surnames had a slightly lower match rate than those with only one.

Match rates are noticeably lower for non-UK domiciled graduates – see the international publication for more details. The main reason for this is that graduates will only appear on the CIS spine if they have been issued with a National Insurance Number. Students who have no intention of working or claiming benefits in this country, for instance if they do not intend to work during study and to leave the country immediately after study, are unlikely to apply for a National Insurance Number.

Other reasons for lower match rates among non-UK domiciled graduates include higher likelihoods of misspelling of names and lower take up of/eligibility for student loans, meaning we would not be able to attach NINO to the HESA data to aid the matching process.

6. Get in touch

Media enquiries

Press Office News Desk, Department for Education, Sanctuary Buildings, Great Smith Street, London SW1P 3BT.

Tel: 020 7783 8300

Other enquiries/feedback

Simon Childs, Higher Education Analysis, Department for Education, Sanctuary Buildings, Great Smith Street, London SW1P 3BT.

Tel: 07920 594501Email: HE.LEO@education.gov.uk

Department for Education

About this publication:

enquiries Simon Childs, Higher Education Analysis, Department for Education, Sanctuary Buildings, Great Smith Street, London SW1P 3BT.

Tel: 07920 594501 Email: HE.LEO@education.gov.uk

https://www.gov.uk/government/statistics/graduate-outcomes-leo-inclusion-of-self-employment-earnings-data