



Experimental statistics: Graduate outcomes using the Longitudinal Education Outcomes (LEO) data – inclusion of self-employment earnings

SFR 76/2017, 20 December 2017

This is the fourth in a series of higher education data releases from the Department for Education’s new Longitudinal Educational Outcomes (LEO) dataset. Previous releases have looked at graduate earnings outcomes based on PAYE earnings only. While providing useful insight into graduate earnings patterns, a lack of self-employment earnings data was seen as a limitation.

The DfE has now incorporated self-employment earnings data into the LEO dataset. This release looks at what impact this has on our previous findings, focusing on graduate earnings in the 2014/15 tax year for the 2008/09, 2010/11 and 2012/13 cohorts (one, three and five years after graduation).

Our analysis focuses on a comparison of four sets of earnings statistics:

- total earnings (all earnings);
- annualised earnings for graduates with PAYE employment only;
- earnings for graduates with self-employment only;
- earnings for graduates with income from both PAYE and self-employment.

Chart 1 shows how these differ in the 2014/15 tax year for the 2012/13, 2010/11 and 2008/09 graduate cohorts who were, respectively, one, three and five years post-graduation. On average, those with self-employment earnings earn less than those with PAYE earnings only. This is particularly so for those with self-employed only earnings, who earn less than half of what those with PAYE earnings only earn.

Chart 1: Median earnings one, three and five years after graduation, by income type (2014/15 tax year).



The low level of graduate self-employment in the period considered means that the inclusion of these earnings has little impact on total median earnings compared to that suggested by PAYE only data. The LEO data set is not able to identify whether a graduate is working full-time or part-time and therefore the statistics presented in this release do not control for this or any other potential difference in the PAYE only and self-employed populations.

Contents

1.	Introduction.....	4
2.	Background to the Longitudinal Education Outcomes (LEO) dataset.....	5
	Coverage	5
	Years after graduation	6
	Subject areas.....	6
	Employment outcomes	6
	Contextual Information.....	6
3.	Results	7
	Subject level findings	9
4.	Proposal for future publication of earnings figures	12
5.	Accompanying tables.....	12
6.	Glossary	12
7.	Data quality.....	14
	Employment and earnings data	14
	Coverage	15
8.	Methodology	15
	Time period.....	15
	Subject areas.....	16
	Employment outcomes	16
	Annualised earnings	19
9.	Data matching and match rates	20
	Overall match rates.....	21
	Match rate by graduate characteristic.....	22
10.	Experimental Statistics	22
11.	Feedback.....	23
12.	Next steps.....	23
13.	Get in touch	23
	Media enquiries	23
	Other enquiries/feedback.....	23

About this release

This experimental statistical first release provides a comparison of graduate earnings from the Longitudinal Educational Outcomes (LEO) dataset for graduates who are paid via PAYE, graduates who are self-employed and graduates who received income from both PAYE and self-employment. This follows on from our previous releases, for which we were able to provide graduate earnings from PAYE only. This release focuses on graduate outcomes and earnings in the 2014/15 tax year for the 2008/09, 2010/11 and 2013/13 cohorts one, three and five years after graduation respectively.

This is the fourth statistical release looking at the employment and earnings outcomes of university graduates from the LEO dataset.

Feedback

As these statistics are currently experimental we welcome feedback on this release and the data presented within it. Contact details are provided in [section 11: Feedback](#).

In this publication

The following tables are included in this release:

Main tables (SFR76_2017_SelfEmployment_earnings_subject_table.xls)

Table 1a: Activity of graduates by subject and sex one year after graduation (2012/13 graduating cohort)

Table 1b: Activity of graduates by subject and sex three years after graduation (2010/11 graduating cohort)

Table 1c: Activity of graduates by subject and sex five years after graduation (2008/09 graduating cohort)

Underlying data

CSV format (SFR76_2017_SelfEmployment_earnings_subject_data.csv)

Metadata (SFR76_2017_SelfEmployment_earnings_subject_metadata.txt)

1. Introduction

This release explores the effect of including self-employment earnings data in the Higher Education strand of the Longitudinal Educational Outcomes (LEO) dataset.

Our release in December 2016 showed that the proportion of graduates submitting a Self Assessment tax return with a record of self-employment varied across institution and subject of study. This suggested differences in earnings outcomes might in part be due to differences in the likelihood of a subject or institution's graduates receiving self-employment earnings, either alongside or instead of PAYE earnings. This release updates this and presents earnings from self-employment to look at graduate earnings patterns in the aggregate and by subject. Our intention is to include self-employment earnings in all future LEO releases, including by provider.

A Self Assessment tax return is completed by a range of people, including those who 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 return can be found on the [self-assessment pages of the gov.uk website](#).

The Department for Education has recently obtained from Her Majesty's Revenue and Customs (HMRC) a dataset on income from the Self Assessment tax return which contains the following four variables:

- earnings received through employment (PAYE);
- profit from partnership enterprises;
- profit from sole-trader enterprises;
- total earnings for the tax year from the Self Assessment tax return.

We have used the profit from partnership enterprises and profit from sole-trader enterprises to identify graduates who are self-employed and have earnings from self-employment. 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.

The figures we have taken as self-employment earnings are not directly comparable to PAYE income. PAYE income generally refers to the amount paid to the employee (salary, wages), whereas income from self-employment relates to income declared for tax purposes and so allowable expenses will have been deducted from these earnings.

We have used this information to present four sets of earnings in the tables accompanying this release:

- total earnings (all earnings);
- annualised earnings for graduates with PAYE employment only;
- earnings for graduates with self-employment only;
- earnings for people with income from both PAYE and self-employment.

Where median earnings include income from PAYE, the PAYE earnings are annualised. Where median earnings include income from self-employment, the self-employment earnings have not been annualised and therefore are raw earnings (meaning that we are assuming that the self-employment activity relates to the full tax year). Where median earnings include both, the PAYE element has been annualised and been added to the raw self-employment earnings. More information on our methodology can be found in [section 8: Methodology](#).

These four earnings definitions can be used to compare how earnings from PAYE and self-employment differ from one another for higher education graduates.

The addition of self-assessment data has also allowed us to refine the way in which we identify whether a graduate has a sustained employment outcome. In our last release, we classified graduates as being in sustained employment if they met our definition of PAYE sustained employment or if they had any record of self-employment on the self-assessment record. In this release, we have decided to classify people as being in sustained employment if they meet our definition of PAYE sustained employment or if they have self-assessed earnings greater than £0 from partnership and/or sole-trader enterprises.

2. Background to the Longitudinal Education 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 workplace¹. 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 focus of this release is on the effect recently obtained self-employment earnings data has had on employment outcomes and earnings. Please see [section 7: Data Quality](#) for consideration of the strengths and weaknesses of the LEO dataset, [section 9: Data matching and match rates](#) for further information on matching processes, and the [glossary](#) for the definitions used throughout this publication.

The privacy notice explaining how personal data in this project is shared and used can be found [here](#).

Coverage

This publication looks at those who graduated with a first degree qualification from higher education institutions (HEIs) in England. First degrees are also known as bachelor's degrees. We have only looked at those classified as UK domiciled prior to entry to higher education. Figures are presented for all graduates and have not been split by full-time or part-time mode of study. On average, we have been able to link over 95% of each graduate cohort to tax and/or benefit data (see [section 9: Data matching and match rates](#)).

Designated alternative providers are not included in this publication as they were not required to return student level data to HESA prior to the 2015/16 academic year. The University of Buckingham has historically returned HESA data so is included in the publication.

All figures are based on UK tax, benefit and student records only, activity of those who move abroad to work or study after graduating is not reflected in the employment or further study figures. Instead, these individuals are categorised as 'activity not captured'.

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.

The methodology for defining employment and earnings outcomes is set out in [section 8: Methodology](#).

Years after graduation

The time periods used in this publication are one, three and five years after graduation, which refers to the first, third and fifth full tax year after graduation, respectively. For instance, for the 2012/13 graduation cohort, the figures one year after graduation refer to employment and earnings outcomes in the 2014/15 tax year. This time period was picked as graduates are unlikely to have been engaged in economic activity for the whole tax year that overlaps with the graduation date. This is displayed graphically in figure 1 below.

Figure 1: Relationship between academic year, tax year and definitions of ‘years after graduation’ used in this publication

		Tax year				
		2010/11	2011/12	2012/13	2013/14	2014/15
Academic year of graduation	2008/09	1 year	2 years	3 years	4 years	5 years
	2009/10		1 year	2 years	3 years	4 years
	2010/11			1 year	2 years	3 years
	2011/12				1 year	2 years
	2012/13					1 year

Cohorts in **bold** are covered in this publication

The cohorts covered in this release mirror those used in the third higher education data release from the LEO dataset, published in June 2017.

Subject areas

The subjects covered by this publication are based on version 3.0 of the Joint Academic Coding System (JACS) and covered the same 23 subjects covered in the third higher education data release from the LEO dataset, published in June 2017. Further information on these subjects can be found in the [section: 8 methodology](#) as well as in the [glossary](#).

Employment outcomes

Outcomes are presented for graduates who 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 been matched are then placed into one of five outcomes categories. These outcomes categories remain the same as in the third higher education data release from the LEO dataset, published in June 2017. Further information can be found in [section 8: methodology](#).

With the addition of the new self-employment earnings data, there has been a change to the sustained employment definition in this publication. For the purposes of this publication, individuals are classed as being in sustained employment in the 2014/15 tax year if they meet our definition of sustained employment based on PAYE **or** have returned a Self Assessment tax return 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, from here on known as ‘earnings 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 distinguish between those who made a loss and those who submitted a Self Assessment tax return for other reasons at this moment in time.

Contextual Information

There are a number of factors that can influence the employment and earnings outcomes of graduates beyond the subject and institution attended. The outcomes presented in this release are ‘raw’ outcomes,

they **do not control for differences in the characteristics of students** that might influence graduate employment outcomes. This should be borne in mind when making comparisons across subjects.

3. Results

Having included earnings from self-employment for the first time, we found a number of graduate cases which had been previously flagged as self-employed but either did not have an earnings record in the new self-employment earnings data or which did have a record but did not have earnings over £0. We no longer classify these graduates as self-employed. As the data used in this publication was extracted from HMRC's database over a year after the data used in our previous publication, there are also a number of records which we had not previously classified as self-employed but now fit our definition of self-employment: these graduates have now been reclassified as self-employed. Table 1 summarises the extent of this change.

Figures in the following tables have been rounded, but differences between figures are based on unrounded data. Therefore, changes discussed in the text or presented in the tables may be slightly different to results obtained by calculating differences using the rounded figures in the tables.

Table 1: Changes to self-employment categorisations following inclusion of earnings from self-employment.

	Previous methodology	New methodology	Change in self-employment status	Count	Percentage
1 Year after graduation (2012/13 cohort)	Not self-employed	Not self-employed	No	264,660	94.5%
	Self-employed	Self-employed	No	12,175	4.3%
	Not self-employed	Self-employed	Yes	1,265	0.5%
	Self-employed	Not self-employed	Yes	2,000	0.7%
3 Years after graduation (2010/11 cohort)	Not self-employed	Not self-employed	No	232,955	93.2%
	Self-employed	Self-employed	No	13,400	5.4%
	Not self-employed	Self-employed	Yes	1,160	0.5%
	Self-employed	Not self-employed	Yes	2,310	0.9%
5 Years after graduation (2008/09 cohort)	Not self-employed	Not self-employed	No	211,935	92.3%
	Self-employed	Self-employed	No	14,225	6.2%
	Not self-employed	Self-employed	Yes	1,115	0.5%
	Self-employed	Not self-employed	Yes	2,380	1.0%

There were some changes to classifications of self-employment. However, the majority of cases did not change. As the above table shows, the amount of records where there was a change accounted for 1.2% for those one year after graduation (2012/13 cohort), 1.4% for those three years after graduation (2010/11 cohort) and 1.5% for those five years after graduation (2008/09 cohort). In all groups, the number of records where people have gone from being self-employed to not self-employed is greater than the number of records which are now classified as self-employed which weren't previously.

Table 1 also highlights how the percentage in self-employment increases with time after graduation: 4.8% 1 year after graduation (2012/13 cohort); 5.8% 3 years after graduation (2010/11 cohort); and 6.7% 5 years after graduation (2008/09 cohort).

Table 2 summarises the impact this has had on earnings split by cohort and gender.

Table 2: The effect of including self-employment earnings data on total median earnings by sex and years after graduation.

	Median Earnings (all types of sustained employment)		
	Excluding earnings from self-employment	Including earnings from self-employment	Difference
Female + Male			
One year after graduation (2012/13 cohort)	18,600	18,500	-100
Three years after graduation (2010/11 cohort)	22,500	22,400	-100
Five years after graduation (2008/09 cohort)	25,600	25,400	-200
Female			
One year after graduation (2012/13 cohort)	18,000	18,000	-100
Three years after graduation (2010/11 cohort)	21,500	21,400	-100
Five years after graduation (2008/09 cohort)	24,500	24,300	-200
Male			
One year after graduation (2012/13 cohort)	19,500	19,400	-100
Three years after graduation (2010/11 cohort)	24,000	23,800	-200
Five years after graduation (2008/09 cohort)	27,500	27,200	-300

Table 2 shows that inclusion of earnings from self-employment have caused a slight decrease in median earnings. For both female and male graduate groups there was a greater negative difference in median earnings five years after graduation than at three years after graduation and the same for male graduates when comparing three years after graduation to one year after graduation. The negative difference three and five years after graduation is greater for male graduates than it is for female graduates.

The year-on-year increase in median earnings does not increase as dramatically in self-employment-only earnings as it does in PAYE-only earnings or even median earnings for people who earn income from both PAYE and self-employment. Chart 2 shows how the self-employed median earnings do not increase as steeply across years as the other types of earnings do. This could be explained by the nature of self-employment and how working as self-employed differs from working as an employee.

Chart 2: Median earnings one, three and five years after graduation, by income type



Chart 2 shows four sets of median earnings one, three and five years after graduation. The medians are:

- total median earnings (all earnings);
- median annualised earnings for graduates with PAYE employment only²;
- median earnings for graduates with self-employment only;
- median earnings for people with income from both PAYE and self-employment.

The total median earnings when earnings from self-employment are included differs only slightly from earnings for those with PAYE only. However, the inclusion of this data means that we are able to provide a more complete and accurate picture of graduate earnings, and are able to reflect the different types of employment graduates are engaged in post-graduation.

Subject level findings

Due to different employment outcomes from different courses, the self-employment earnings data impacts some subjects more than others. Table 3 shows the percentage of students with earnings from self-employment by subject.

Table 3: Percentage of graduates with earnings from self-employment by subject studied one, three and five years after graduation.

Subject	Earnings from self-employment (%)		
	One year after graduation (2012/13 cohort)	Three years after graduation (2010/11 cohort)	Five years after graduation (2008/09 cohort)
Agriculture & Related Subjects	6%	7%	8%
Architecture, Building & Planning	5%	6%	7%
Biological Sciences (excluding Psychology)	5%	5%	6%
Business & Administrative Studies	3%	3%	4%
Combined	7%	7%	6%
Computer Science	4%	5%	5%
Creative Arts & Design	13%	16%	18%
Economics	2%	2%	3%
Education	3%	3%	3%
Engineering & Technology	4%	4%	5%
English Studies	4%	5%	6%
Historical & Philosophical Studies	4%	4%	5%
Languages (excluding English Studies)	5%	5%	6%
Law	3%	4%	5%
Mass Communications & Documentation	6%	7%	7%
Mathematical Sciences	2%	3%	3%
Medicine & Dentistry	7%	10%	17%
Nursing	1%	2%	2%
Physical Sciences	3%	3%	3%
Psychology	3%	3%	3%
Social Studies (excluding Economics)	2%	3%	3%
Subjects Allied to Medicine (excluding Nursing)	9%	10%	10%
Veterinary Science	3%	5%	8%

² Graduates with a PAYE record and a self-employment income above £0 are counted as being in sustained employment. However, as we annualise PAYE data when reporting earnings we have only included graduates with a sustained employment record of earnings (under PAYE) in this category. To do otherwise would risk annualising PAYE data that could potentially only relate to a few weeks' earnings.

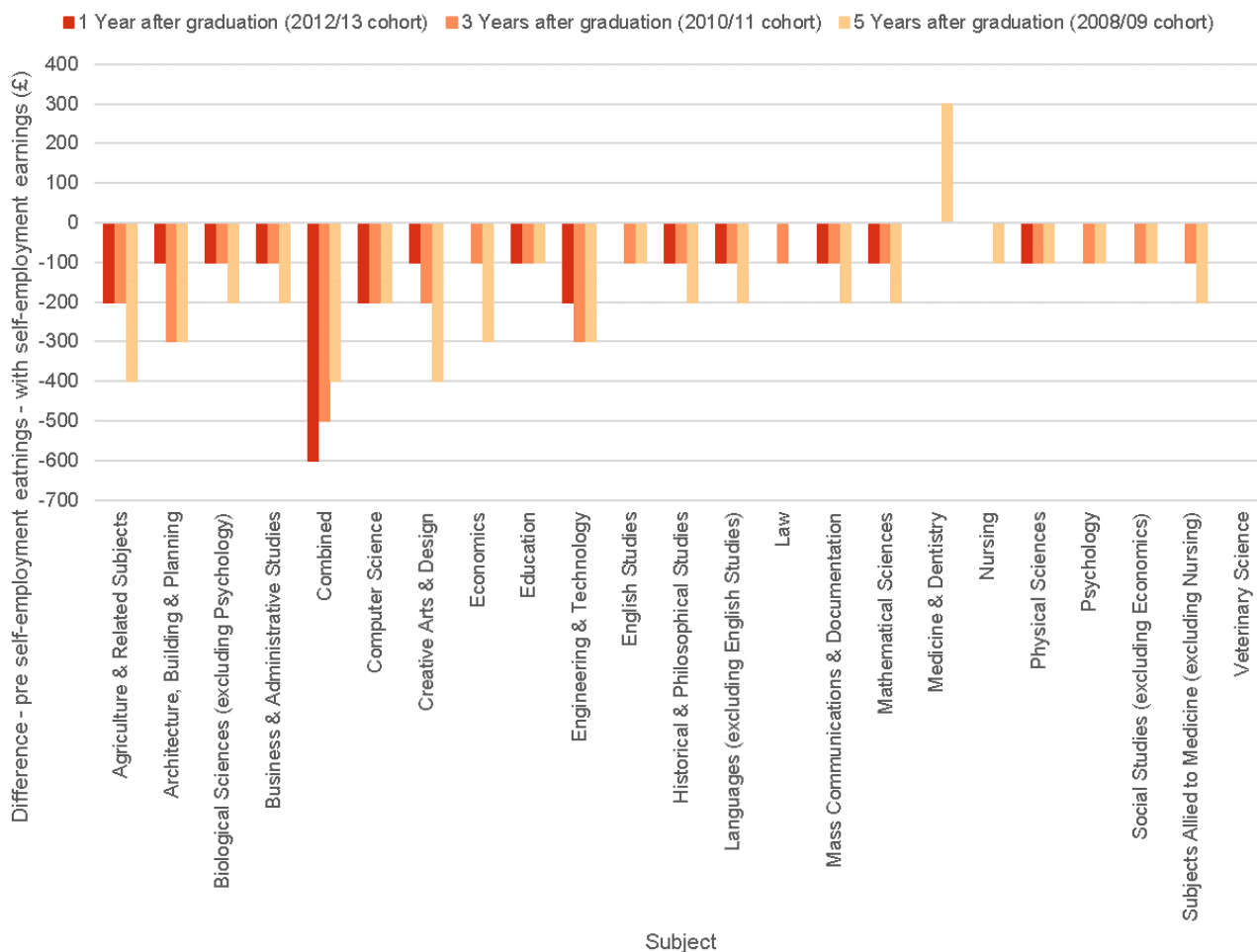
At one, three and five years after graduation creative arts & design, subjects allied to medicine and medicine & dentistry are consistently the top three subjects with the highest unrounded percentage of graduates with earnings from self-employment. Mathematical sciences, economics and nursing consistently have the lowest unrounded percentage of self-employed graduates across one, three and five years after graduation. Creative arts & design has the highest percentage of self-employed graduates at all three points after graduation, and nursing also has the lowest self-employed percentage at all three points.

Table 4: The effect of self-employment earnings data on total median earnings by subject studied and years after graduation.

Subject	1 Year after graduation (2012/13 cohort)			3 Years after graduation (2010/11 cohort)			5 Years after graduation (2008/09 cohort)		
	Pre self-employment earnings data	With self-employment earnings data	Difference	Pre self-employment earnings data	With self-employment earnings data	Difference	Pre self-employment earnings data	With self-employment earnings data	Difference
Agriculture & Related Subjects	16,800	16,600	-200	19,500	19,300	-200	22,000	21,500	-400
Architecture, Building & Planning	23,000	23,000	-100	27,500	27,200	-300	29,600	29,300	-300
Biological Sciences (excluding Psychology)	15,800	15,700	-100	20,700	20,600	-100	24,300	24,100	-200
Business & Administrative Studies	19,100	19,000	-100	23,000	22,900	-100	26,400	26,200	-200
Combined	19,500	18,900	-600	22,400	21,900	-500	24,300	23,900	-400
Computer Science	20,900	20,700	-200	24,600	24,400	-200	27,400	27,100	-200
Creative Arts & Design	14,100	14,000	-100	17,600	17,400	-200	20,200	19,700	-400
Economics	24,000	24,000	-	31,000	30,900	-100	37,500	37,200	-300
Education	18,800	18,700	-100	21,700	21,600	-100	24,600	24,500	-100
Engineering & Technology	24,800	24,700	-200	28,900	28,600	-300	31,500	31,200	-300
English Studies	15,900	15,900	-	21,100	20,900	-100	24,000	23,900	-100
Historical & Philosophical Studies	16,600	16,500	-100	21,600	21,500	-100	24,900	24,700	-200
Languages (excluding English Studies)	18,500	18,500	-100	23,500	23,400	-100	26,900	26,700	-200
Law	16,500	16,500	-	21,000	20,900	-100	25,100	25,100	-
Mass Communications & Documentation	15,500	15,400	-100	19,400	19,300	-100	22,400	22,200	-200
Mathematical Sciences	22,300	22,200	-100	27,800	27,700	-100	33,000	32,800	-200
Medicine & Dentistry	35,900	35,900	-	42,800	42,800	-	46,700	47,000	300
Nursing	25,600	25,600	-	27,300	27,300	-	28,700	28,600	-100
Physical Sciences	19,100	19,000	-100	23,700	23,500	-100	26,800	26,700	-100
Psychology	15,700	15,700	-	19,900	19,900	-100	22,300	22,200	-100
Social Studies (excluding Economics)	17,700	17,600	-	21,500	21,400	-100	24,700	24,600	-100
Subjects Allied to Medicine (excluding Nursing)	20,900	20,800	-	24,200	24,100	-100	26,500	26,300	-200
Veterinary Science	27,500	27,500	-	32,100	32,100	-	36,300	36,200	-

Chart 3 shows the difference between median earnings before and after the incorporation of self-employment earnings by subject.

Chart 3: The difference between median earnings before and after the incorporation of self-employment earnings by subject one, three and five years after graduation.



Total earnings

One year after graduation the subject most affected is 'combined' with total earnings £600 lower than when only PAYE earnings are considered. This is followed by agriculture & related subjects, computer science and engineering & technology, all with a difference of -£200. Three years after graduation, the biggest difference is still for combined courses, with a difference of -£500, followed by architecture, building & planning and engineering & technology, both with a difference of -£300. The largest difference five years after graduation was for agriculture & related subjects, combined courses and creative arts & design, all with total earnings £400 lower than when only PAYE earnings are considered. The only positive increase to median earnings was for medicine & dentistry five years after graduation with an increase of £300. Five years after graduation, creative arts & design and medicine & dentistry were the two subjects with the highest percentage of graduates with earnings from self-employment.

Earnings from self-employment only

As the accompanying tables show, medicine & dentistry is the subject with the highest median earnings from self-employment for both sexes. This is also the case for PAYE median earnings, median earnings from both PAYE & self-employment and for the overall median earnings.

Economics has the lowest median earnings from self-employment one and three years after graduation. Combined is the lowest at five years after graduation. For female graduates, the lowest median earnings from self-employment are for physical sciences one year after graduation, engineering & technology three years after graduation and architecture & related subjects five years after graduation. For male graduates, the lowest median earnings from self-employment are for economics one and three years after graduation, and combined five years after graduation.

Earnings from self-employment and PAYE

One year after graduation, the lowest median earnings for female graduates who are both a PAYE employee and self-employed was for creative arts & design, three years after graduation was for agriculture & related subjects and five years after graduation was creative arts & design. For female graduates the subjects with the lowest median earnings from PAYE & self-employment were creative arts & design one year after graduation, agriculture & related subjects three years after graduation and combined five years after graduation. For male graduates this was creative arts & design one year after graduation, psychology three years after graduation and creative arts and design five years after graduation.

4. Proposal for future publication of earnings figures

When calculating national-level earnings figures we propose to display the total earnings figures only. Providing several different earnings figures can be confusing for users and the difference between PAYE only figures and the total is small in most cases.

When producing figures at university level, we have received feedback that it would be useful to distinguish between the earnings of the different employment categories, as some universities have a high proportion of graduates in self-employment,. However, splitting out earnings for those in self-employment would result in a large number of figures being suppressed due to the small numbers involved. Our proposal is that we release the following:

- a. Total median earnings for those included in earnings calculations.
- b. Median PAYE earnings for those in PAYE sustained employment only.

We would welcome feedback on this proposal (see [section 11: Feedback](#)).

5. Accompanying tables

The following tables are available in Excel format on the department's statistics website (<https://www.gov.uk/government/statistics/graduate-outcomes-leo-including-self-employment-earnings-data>):

Main tables (SFR76_2017_SelfEmployment_earnings_subject_table.xls)

Table 1a: Activity of graduates by subject and sex one year after graduation (2012/13 graduating cohort)

Table 1b: Activity of graduates by subject and sex three years after graduation (2010/11 graduating cohort)

Table 1c: Activity of graduates by subject and sex five years after graduation (2008/09 graduating cohort)

Underlying data

CSV format (SFR76_2017_SelfEmployment_earnings_subject_data.csv)

Metadata (SFR76_2017_SelfEmployment_earnings_subject_metadata.txt)

6. Glossary

Academic year: Runs from 1 August to 31 July. For example, the 2012/13 academic year ran from 1 August 2012 to 31 July 2013.

Customer Information System (CIS): A computer system used by the Department for Work and Pensions that contains a record for all individuals that have been issued with a National Insurance (NINO) number. It contains basic identifying information such as name, address, date of birth and NINO.

First degree qualification: This covers qualifications commonly known as bachelor's degrees and includes postgraduate bachelor's degrees at H level. Not all undergraduate courses are included: for example, the Professional Graduate Certificate in Education (PGCE), foundation degrees and Higher National Diplomas (HND) are excluded. For further information on this classification, please refer to the

Higher Education Statistics Agency at the link referenced below.³ Note that it does not necessarily mean that the degree is the first higher education qualification undertaken by the student.

Further study: The **HESA** Student Record is used to identify instances of further study in higher education. 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. Each tax year spans two academic years; therefore, graduates will be flagged as being in further study if they have a HESA record in one of these two academic years.

HEI: Higher Education Institution.

Higher Education Statistics Agency (HESA): collects data from universities, higher education colleges and other specialist providers of higher education. In this publication, we have used the HESA Student Record to identify our graduate base population and higher education further study instances.

Individualised Learner Record (ILR): used by the further education (FE) and skills sector in England to collect data about learners in the system and the learning undertaken by each of them.

Joint Academic Coding System (JACS): a standardised way of classifying academic subjects and modules, maintained by HESA and the Universities and Colleges Admissions Service (UCAS). In this publication, we group subjects using JACS high-level subject groupings. Previous research has shown that the earnings profile of economics graduates differs substantially from graduates of other social sciences; therefore, in this publication we have presented Nursing separately from other Subjects Allied to Medicine, Psychology separately from other Biological Sciences, economics separately from other social sciences, and English Studies separately from other Language.

JACS code	Subject
1	Medicine & Dentistry
2	Subjects Allied to Medicine (excluding Nursing)
B7	Nursing
3	Biological Sciences (excluding Psychology)
C8	Psychology
4	Veterinary Science
5	Agriculture & Related Subjects
6	Physical Sciences
7	Mathematical Sciences
8	Computer Science
9	Engineering & Technology
A	Architecture, Building & Planning
B	Social Studies (excluding Economics)
L1	Economics
C	Law
D	Business & Administrative Studies
E	Mass Communications & Documentation
F	Languages (excluding English Studies)
Q3	English Studies
G	Historical & Philosophical Studies
H	Creative Arts & Design
I	Education
J	Combined

³ Under 'Field Explanations': <https://www.hesa.ac.uk/support/definitions/students#level-study-qualification-obtained>

National pupil database (NPD): contains information about pupils in schools and colleges in England. It includes information on test and exam results, prior attainment and pupil progress, as well as pupil characteristics such as sex, ethnicity and eligibility for free school meals.

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

PAYE: HMRC's system to collect Income Tax and National Insurance from employment. The LEO project uses information from the **P45** to ascertain employment spell length and from the **P14 (P60)** to determine annual earnings. See [section 7: data quality](#) for a discussion of how PAYE data has been used in LEO.

Self-assessment: 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.

Sustained employment: A learner is counted in sustained employment if they were recorded as being employed in 5 out of the 6 months between October and March in the tax year, for example, 5 out of 6 months between October 2010 and March 2011 for the 2010/11 tax year. Additionally, graduates are counted in sustained employment if they have returned a Self Assessment tax return 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).

Sustained annualised earnings: The calculated average daily wage across the tax year grossed up to the equivalent annual figure. This is only calculated where the learner was in sustained employment. Self-assessed earnings have not been annualised. Earnings figures in this publication are **nominal**.

Tax year: Runs from 6 April to 5 April the following year. For example, the reference period covered by the 2014/15 tax year runs from 6 April 2014 to 5 April 2015.

UK Domiciled: indicates that the student was domiciled in England, Scotland, Wales or Northern Ireland prior to entry of the course. Students who were domiciled in the Channel Islands or the Isle of Man are not considered to be UK domiciled in this publication.

7. 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 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



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 tax year 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:

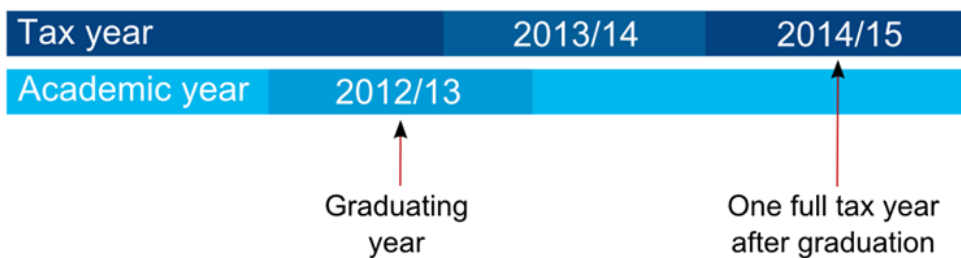
- Earnings received through employment (PAYE)
- Income from partnership enterprises
- Income from sole-trader enterprises
- Total earnings for the tax year from the self-assessment form.

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.

8. 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. So, for the 2012/13 graduation cohort the figures one year after graduation refer to employment and earnings outcomes in the 2014/15 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.



Subject areas

The subjects covered by this publication are based on version 3.0 of the Joint Academic Coding System (JACS). In previous publications, we presented results for economics separately from the other subjects in social studies. Following consultation on the subject splits, we are additionally extracting nursing from the other subjects allied to medicine, psychology from the biological sciences and English studies from the other languages. A list of the 23 subjects covered can be found in [section 6: Glossary](#).

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

Employment outcomes

Outcomes are presented for graduates who 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. In this publication, these individuals are referred to as **matched**. 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 any calculations.** More information on match rates is given in [section 9: 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 will be moved out of the 'unmatched' category and into the 'further study' category.

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.

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, earning below the Lower Earnings Limit or voluntarily leaving the labour force.

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

The 'sustained employment' measure aims to count the proportion of graduates in sustained employment 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 education⁴. This definition looks at employment activity in the six month October to March period starting from the first tax year after graduation. A graduate needs to be in paid employment for five out of six months between October and March to be classified as being in 'sustained employment'.

For example, those who graduated in the 2012/13 academic year would be counted as being in sustained employment one year after graduation if they were in paid employment for at least one day a month in the five out of six months between October 2014 and March 2015. If they are employed in all five months from October to February, but do not have an employment record for March, then they must have an additional employment record in April to be considered as being in sustained employment.

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 shares⁵. Currently, only data for the 2014/15 tax year is available for inclusion in LEO. For this reason, we have included the graduating cohorts of academic years 2008/09, 2010/11 and 2012/13 in this publication, since for these cohorts, the tax year 2014/15 corresponds to five years, three years and one year after graduation, respectively.

For the purposes of this publication, individuals are classed as being in sustained employment in the 2014/15 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 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 in the relevant tax year. Further study undertaken at further education colleges is not reflected in these figures as this information is not collected by HESA. 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. 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.

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 2012/13 academic year and went straight on to a one-year masters course would be counted as being in further study in the 2014/15 tax year (one year after graduation) as their course would finish in September 2014. If a graduate from 2012/13 waited a year before starting their one-year masters course then they would be counted as being in further study in the 2014/15 tax year (one year after graduation) as their course would start in September 2014.

We have not used a sustained definition when defining further study. The majority of higher education courses last longer than 6 months and dropout rates tend to be low, especially at postgraduate level.

Sustained employment only

Graduates are considered to be in sustained employment if they were employed for at least one day for five out of the six months between October and March of the tax year in question or if they had a self-

⁴ <https://www.gov.uk/government/statistics/adult-further-education-outcome-based-success-measures>

⁵ 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>

employment record with earnings over £0 (profit) from either a Partnership or Sole-Trader enterprise in that tax year. To be in the sustained employment **only** category, graduates must not have a record of further study in the tax year in question.

Sustained employment with or without further study

Sustained employment with or without further study includes **all** graduates with a record of sustained employment, regardless of whether they also have a record of 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 database in the relevant tax year. The further study does not have to be at postgraduate level to be counted. Further study undertaken at further education colleges is not reflected in these figures as this information is not collected by HESA.

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	x	Unmatched to CIS	Unmatched to CIS	Unmatched to CIS
Activity not captured	x	x	x	x
No sustained destination	x	x	✓	x
	x	x	x	✓
	x	x	✓	✓
Sustained employment only	x	✓	✓	x
	x	✓	✓	✓
Sustained employment (with or without further study)	✓	✓	✓	x
	✓	✓	✓	✓
	x	✓	✓	x
	x	✓	✓	✓
Sustained employment, further study or both	✓	Unmatched to CIS	Unmatched to CIS	Unmatched to CIS
	✓	x	x	x
	✓	x	✓	x
	✓	x	x	✓
	✓	x	✓	✓
	✓	✓	✓	x
	✓	✓	✓	✓
	x	✓	✓	x
	x	✓	✓	✓

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. This is the first time earnings from self-assessment data have been included.

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 wages which is then multiplied by the number of days in the tax year to create their annualised earnings⁶.

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.

Information provided on the Self Assessment tax return includes a field on earnings through PAYE employment. Where this differs from the earnings reported on the P14, we have taken the earnings from PAYE reported through self-assessment over the earnings from the P14. This methodology has been used for this publication because the self-assessment data extract is more up-to-date than the PAYE extract that we have been using for the past year. In the future we intend to keep the PAYE earnings reported on the P14 form.

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 earned income from both PAYE employment and self-employment. However, the difference between raw earnings and annualised earnings is smaller for graduates who are employed from PAYE employment only than it is for those who receive income from both PAYE employment and self-employment.

For earnings from self-employment, raw earnings are presented. 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 through 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.

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).

9. 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. Only the first initial of the forename is used, the surname is encoded using an English sound-based algorithm called SOUNDEX⁸, and for most matches only the sector of the postcode is used.

⁶ 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.

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

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 B 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.

Table B: 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	x	✓	✓	✓	✓	✓
4. Amber-Red	x	✓	✓	✓	1 ✓	
5. Red-Amber	x	x	x	✓	✓	✓ (full postcode)

Overall match rates

Table C 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 C: Match rates for UK domiciled first degree graduates at English HEI's, by year of graduation

Academic year	Matched to tax/ benefit record (%)	Matched to CIS spine (%)
2003/04	94	94
2004/05	94	95
2005/06	95	95
2006/07	95	96
2007/08	96	97
2008/09	96	97
2009/10	97	97
2010/11	96	96
2011/12	96	97
2012/13	98	99

Table C shows that the match rate was very high for the most recent cohorts: 99% of full-time graduates in 2012/13 were matched using the CIS, and almost all of these had at least one tax or out-of-work benefit record. This compares to a match rate of 94% 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.

Due to improvements in the matching process since our first publication, our match rates increased slightly from our first to our second publication. This is particularly apparent for the 2010/11 and 2011/12 cohorts, which had a dip in the match rate in our first publication compared to other cohorts. Match rates for these cohorts are now more comparable with other cohorts. Other than the addition of earnings data from Self Assessment, the dataset used for this publication is identical to that in our second and third publications.

Match rate by graduate characteristic

Table D 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 D: CIS match rate by sex

Academic year	Female (%)	Male (%)
2003/04	92	97
2004/05	93	97
2005/06	93	98
2006/07	95	98
2007/08	96	98
2008/09	96	98
2009/10	97	98
2010/11	96	97
2011/12	96	97
2012/13	99	99
2013/14	99	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 91% in 2012/13. 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.

10. Experimental Statistics

Experimental statistics are new official statistics that are undergoing evaluation. These statistics are being published as experimental statistics in order to involve users and stakeholders in their development and as means to further improve the use of the data in the future.

The Department has a set of [statistical policies](#) in line with the Code of Practice for Official Statistics.

11. Feedback

We welcome feedback on this publication. Contact details can be found in [section 13: Get in touch](#).

12. Next steps

We welcome user feedback on the data contained in this release, and those wishing to provide comments should send them to HE.LEO@education.gov.uk. As we develop the LEO data further, we will look to publish further experimental statistical releases as well as establish a regular cycle of publications covering graduate outcomes to inform public understanding of the higher education system and improve the information available to students when deciding on higher education institutions and subjects.

13. 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

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

Tel: 07384 211 094 Email: HE.LEO@education.gov.uk



Department for Education

© Crown copyright 2015

This publication (not including logos) is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated. Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned.

To view this licence:

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

email psi@nationalarchives.gsi.gov.uk

write to Information Policy Team, The National Archives, Kew, London, TW9 4DU

About this publication:

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

Tel: 07384 211 094 Email: HE.LEO@education.gov.uk

<https://www.gov.uk/government/statistics/graduate-outcomes-leo-including-self-employment-earnings-data>

Reference: [SFR 76/2017]



Follow us on Twitter:
[@educationgovuk](https://twitter.com/educationgovuk)



Like us on Facebook:
facebook.com/educationgovuk