Department for Education

Technical Report for Education and Labour Market Pathways of Individuals (LEO)

Technical report

May 2021

Oliver Anderson and Moira Nelson, Department for Education

Government Social Research

Contents

Background to the Longitudinal Educational Outcomes (LEO) dataset

1. The Small Business, Employment and Enterprise Act 2015 enabled government, for the first time, to link education and tax data together to chart the transition of individuals from schools, colleges and higher education institutions 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 individuals without imposing any additional data collection burdens on schools, colleges, universities, employers or members of the public.

2. The LEO dataset links information about individuals, including

- personal characteristics such as gender, ethnic group, special educational needs, free school meals eligibility
- education, including schools, colleges and higher education institutions 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 Individual Record (ILR) data on students at further education institutions, held by DfE
- PAYE employment data 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)

3. By combining these sources, we can look at the progress of individuals through post-compulsory education into the labour market.

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

[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.

Data sources

Employment and earnings data

5. The employment data comes from 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.

6. The employment and earnings administrative dataset covers those who pay tax through PAYE or through completing a self-assessment tax form. The core purpose of this process is to collect tax from those who are eligible to pay it through this mechanism and so there is not complete coverage. Employers are not required to supply information to HMRC for individuals who earn below the tax threshold, although for large employers these individuals are thought to be included due to methods of data transfer. Further, HMRC started to implement Real Time Information (RTI) in April 2013 which includes nearly all such individuals. 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 means that data coverage is higher for the most recent financial years.

7. The employment data is subsequently cleaned and manipulated within the Department for Education.

8. Any earnings for individuals with no employment spells in a tax year are excluded from this analysis. Similarly, any employment spells for individuals with no earnings in a tax year are removed for the purposes of this analysis.

Self-assessment data

9. As well as employment data for those who pay tax through PAYE, the employment data includes those who pay tax through self-assessment. This is currently only available to the DfE from the 2013-14 tax year. This is **not included in the analysis** in this report, as there is not a time series of data to be used. Main activities, pathways and earnings analysis all use data from 2003-04 to 2017/18 tax years and because self-assessment data is not available for these years then it is not included in the analysis. Figure 1 shows that the self-assessment data only exists for all cohorts of individuals for tax years 11 to 15.

Figure 1: Mapping of tax years and years after GCSEs for which self-assessment data is available (and for different academic year KS4 cohorts)

KS4 cohorts 2001/02 to 2006/07

		Tax year														
		Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Year 9	Year 10	Year 11	Year 12	Year 13	Year 14	Year 15
Academic year cohort	2001/02	2003/04	2004/5	2005/6	2006/7	2007/8	2008/9	2009/10	2010/11	2011/12	2012/13	2013/14	2014/15	2015/16	2016/17	2017/18
	2002/03	2004/5	2005/6	2006/7	2007/8	2008/9	2009/10	2010/11	2011/12	2012/13	2013/14	2014/15	2015/16	2016/17	2017/18	
	2003/04	2005/6	2006/7	2007/8	2008/9	2009/10	2010/11	2011/12	2012/13	2013/14	2014/15	2015/16	2016/17	2017/18		
	2004/05	2006/7	2007/8	2008/9	2009/10	2010/11	2011/12	2012/13	2013/14	2014/15	2015/16	2016/17	2017/18			
	2005/06	2007/8	2008/9	2009/10	2010/11	2011/12	2012/13	2013/14	2014/15	2015/16	2016/17	2017/18				
	2006/07	2008/9	2009/10	2010/11	2011/12	2012/13	2013/14	2014/15	2015/16	2016/17	2017/18					

10. Including self-assessment in the analysis would make a difference to the proportion in employment, but not to average earnings. Table 1 shows that if individuals that completed a self-assessment were included, the employed numbers would be around six to seven percentage points higher than using employment only. This is in terms of the main activities and using the hierarchy (see Figure 5). This would have a knock on effect to those assigned a 'no sustained activity' and 'no activity' category. The notable difference is a significant drop in those with 'no activity' due to self-assessment activity (five percentage points in years 14 and 15). Average earnings for years 11 to 15 (Table 2) are very similar when calculated from employment only and employment and self-assessment earnings.

Table 1: Effect of including self-assessment data in employment main activity categorisation

KS4 cohorts 2001/02 to 2005/06

Year after completion after GCSEs	Year 11	Year 12	Year 13	Year 14	Year 15
employed including self-employment	64%	65%	66%	66%	67%
no sustained activity	9%	8%	8%	8%	8%
no activity	10%	11%	12%	12%	13%
employed (only)	58%	59%	59%	60%	60%
no sustained activity	11%	10%	9%	9%	9%
no activity	14%	15%	16%	17%	18%
percentage point difference employment	6	7	7	7	7
percentage point difference sustained activity	-2	-2	-1	-1	-1
percentage point difference no activity	-4	-4	-4	-5	-5

Table 2: Effect of including self-assessment data in average earnings calculations

KS4 cohorts 2001/02 to 2005/06

Year after completion after GCSEs	Year 11	Year 12	Year 13	Year 14	Year 15
Average earnings employment only	£21,853	£22,556	£23,168	£23,609	£23,983
Average earnings including self-assessment	£21,861	£22,586	£23,208	£23,655	£24,030

Out of work Benefits data

11. Benefits data are taken from the underlying DWP payments systems and are supplemented by the information entered by Jobcentre advisers.

12. The benefits classed as out-of-work for this analysis were:

- Jobseekers Allowance (JSA)
- Jobseekers Training Allowance (JTA)
- Employment and Support Allowance (ESA)
- Incapacity Benefit (IB)
- Income Support (IS)
- Passported IB (PIB)
- Severe Disablement Allowance (SDA)
- Pension Credit (PC)
- State (Retirement) Pension (RP)
- Carers Allowance (Invalid Carers Allowance – ICA)
- Attendance Allowance (AA)
- Universal Credit – Searching for Work (UAA)
- Universal Credit – No Work requirements (UBC)
- Universal Credit – Preparing for work (UCE)
- Universal Credit – Planning for work (UDF)

13. In the most recent tax years, there has been a move to Universal Credit (UC) for example replacing JSA, ESA and IS. For more information on benefits please see Benefits on gov.uk.

National Pupil Database (NPD)

14. The NPD is a collection of linked education datasets held by DfE. Linking between education datasets is believed to be very high quality, particularly for pupils formerly in state-funded schools, however, it is accepted that a small proportion of incorrect matches may have been made. These are administrative datasets and as such there may be errors and differences between datasets. This includes school census data and attainment datasets (key stage 4 and key stage 5).

Higher Education Statistics Agency (HESA) data

15. HESA collects data from universities, higher education colleges and other specialist providers of higher education. Data on higher education in institutions in the UK is received from HESA and matched to the NPD data. There are some limitations to this matching in that the specification of those individuals who are matched changed over time. For example, 16 and 17 year olds in the HESA data were not matched until the 2014/15 academic year. This may underestimate the number of individuals in the first few years who had participated in HE at these ages.

Individualised Individual Record (ILR) data

16. This dataset covers all age Traineeships, and adult (19+) Further Education (FE) and Skills individuals that completed an eligible FE learning aim funded by the Education and Skills Funding Agency (ESFA). It covers all colleges, FE providers and specialist post-16 institutions in England.

Data matching

17. Education records from key stage 4 and key stage 5 attainment data, school census, pupil referral unit (PRU) census and alternative provision (AP) census, HESA and ILR are **matched** to DWP's Customer Information System (CIS)[2] 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. This process increases the likelihood of finding a match with CIS. NINO is collected on the ILR, but is not present on school or attainment records.

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

19. 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). Figure 2 below shows the criteria for each match type.

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

Figure 2: Matching process and accuracy

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)

21. The matching algorithm relies on a number of fields being accurately populated across education and CIS datasets. Not all individuals will necessarily have any record

[2] 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

of employment or benefits for legitimate reasons, so the match rate will never reach 100%.

22. For those individuals which have found a match to CIS, we then apply a person resolution process to reconcile the various education identifiers from across the phases of education against the LEO identifier to uniquely identify an individual linked to their labour market outcomes. Only those where there is a consistent link are retained in the LEO data used for analysis.

23. Table 3 shows match rate by GCSE cohort. The individuals without a match are not included in analysis used in report or data tables, i.e. the 'matched individuals' form the sample and percentages are derived from this these figures. As can be seen, this match rate is improving with later cohorts.

Table 3: Individuals matched to LEO data by KS4 cohort

Key stage 4 cohorts: 2001/02 to 2006/7

KS4 cohort	Individuals in cohort	Matched Individuals	Individuals not matched to DWP/HMRC data	Unresolved Individuals	% matched	% without a match	% unresolved
2001/02	589,516	547,747	25,928	15,841	93%	4%	3%
2002/03	621,676	580,383	23,715	17,578	93%	4%	3%
2003/04	637,169	602,492	19,497	15,180	95%	3%	2%
2004/05	641,516	610,239	17,468	13,809	95%	3%	2%
2005/06	654,730	627,467	15,890	11,373	96%	2%	2%
2006/07	660,746	637,152	11,158	12,436	96%	2%	2%

Methodology

Definition of individual cohorts

24. Individuals are assigned to the latest academic year they appear in the key stage 4 data. This academic year is referred to as the 'cohort'. Most individuals will be academic age 15, but a minority are older or younger than this depending on their circumstances. Individuals are included in the cohorts regardless of the type of school they attended, so that we have pupils from:

- state-funded mainstream schools (including academies and free schools, city technology colleges, local authority maintained schools and FE colleges with 14-16 provision)
- state-funded special schools (including local authority maintained, academies and free schools)
- alternative provision (AP) (including pupil referral units, AP schools, hospital schools and other AP settings)
- non-maintained special schools and independent schools (including special schools).

25. The earliest cohort in the dataset finished key stage 4 in the 2001/02 academic year. This is the earliest pupil-level data available in the NPD. The first time period for which employment and earnings data is reported is one year after finishing key stage 4. This refers to the first full tax year after leaving school. Hence, for the 2001/02 key stage 4 cohort, the figures one year after key stage 4 refer to employment and earnings outcomes in the 2003/04 tax year. This time period was selected as using the tax year that overlaps with finishing school would mean that individuals were unlikely to have been engaged in economic activity for the whole tax year.

26. Figure 3 shows how an individual doing a 'classic' pathway of KS5, HE and employment would be coded in this analysis. It helps to show age and the mapping of tax years and academic years.

Figure 3: Exemplar mapping of age, academic and tax years and education and labour market activities

academic year	example individual	tax year	Year	activity	age
2001/02	last year key stage 4				15
2002/03	first year key stage 5	2002-03	n/a		16
2003/04	second year key stage 5	2003-04	year 1	KS5	17
2004/05	first year HE	2004-05	year 2	HE	18
2005/06	second year HE	2005-06	year 3	HE	19
2006/07	third year HE	2006-07	year 4	HE	20
2007/08	employed	2007-08	year 5	No sustained activity	21
2008/09	employed	2008-09	year 6	employed	22
2009/10	employed	2009-10	year 7	employed	23

27. The latest employment and earnings data are for the 2017/18 tax year, which means that the 2001/02 cohort has 15 full tax years of outcomes data and the 2006/07 cohort has 10 full tax years of data.

Combining cohorts of individuals

28. Cohorts of individuals who completed key stage 4 in England between 2001/02 and 2006/07 have been combined to produce a more representative and robust picture of individuals' education and labour market activities, pathways and outcomes. When looking at all six cohorts of individuals, this consists of 3.6 million individuals. This is particularly important when looking at smaller sub-groups of interest. Combining cohorts of individuals completing their GCSEs at the same time means any changes or patterns are more likely to be real differences (and not say, something randomly different about a certain group from a certain year).

29. For example when comparing labour market outcomes of individuals of certain minor ethnic groups we can have more confidence that any differences are actual and there is not just something randomly different about the individuals from these minor ethnic groups from a given school year (cohort).

30. Figure 4 shows what this means in terms of tax years and different cohorts of data. For years 1 to 10 all cohorts of individuals are included in the analysis, however for year 15 it is only the 2001/02 cohort. For years 11 to 14 the number of cohorts reduces

from five to two gradually. This means that year 1, and all subsequent years, for each of the different cohorts is a different tax year. For this reason, earnings are adjusted to make comparisons like for like (see adjusted earnings section below).

Figure 4: Mapping of tax years and years after GCSEs for different academic year KS4 cohorts

KS4 cohorts 2001/02 to 2006/07

AY KS4 cohort		Labour market/ education variable year 1	Labour market/ education variable year 2	Labour market/ education variable year 3	Labour market/ education variable year 4	Labour market/ education variable year 5	Labour market/ education variable year 6	Labour market/ education variable year 7	Labour market/ education variable year 8	Labour market/ education variable year 9	Labour market/ education variable Year 10	Labour market/ education variable Year 11	Labour market/ education variable Year 12	Labour market/ education variable Year 13	Labour market/ education variable Year 14	Labour market/ education variable Year 15
2001/02		2003/04	2004/5	2005/6	2006/7	2007/8	2008/9	2009/10	2010/11	2011/12	2012/13	2013/14	2014/15	2015/16	2016/17	2017/18
2002/03		2004/5	2005/6	2006/7	2007/8	2008/9	2009/10	2010/11	2011/12	2012/13	2013/14	2014/15	2015/16	2016/17	2017/18	
2003/04		2005/6	2006/7	2007/8	2008/9	2009/10	2010/11	2011/12	2012/13	2013/14	2014/15	2015/16	2016/17	2017/18		
2004/05		2006/7	2007/8	2008/9	2009/10	2010/11	2011/12	2012/13	2013/14	2014/15	2015/16	2016/17	2017/18			
2005/06		2007/8	2008/9	2009/10	2010/11	2011/12	2012/13	2013/14	2014/15	2015/16	2016/17	2017/18				
2006/7		2008/9	2009/10	2010/11	2011/12	2012/13	2013/14	2014/15	2015/16	2016/17	2017/18					

Education and labour market activity definitions

31. Education and labour market activities are calculated for each individual in key stage 4 cohorts 2001/02 to 2006/07.

Table 4: definitions

Tax years: 2003-04 to 2017-18

Employment	the individual has been in paid employment for at least one day in each of the 12 months of the tax year. If the individual has a spell of employment but no income in the tax year (e.g. career break) then the individual is not counted as being employed.
Key stage 5	the individual was academic age 16-18 and entered for one or more approved level 3 qualifications in a state-funded mainstream or special school, independent school, FE college or other FE provider in the tax year which overlaps the start of the academic year. This is taken as a proxy for sustained participation as the individual must have been in KS5 long enough for the school or college to enter them for exams. The latest year that the individual appears in the KS5 data is taken as the final year of KS5. Any years between the end of key stage 4 and the end of key stage are counted as sustained KS5.
Other education	Years 1 to 2 (i.e. 16 to 18 years old) only. The individual appeared in the Individualised Learning Record (ILR) aims data (in England) for at

	least one day in each of 6 consecutive months of the tax year. This includes classroom learning level 2 or below and apprenticeships any level.
Adult FE (19+)	Years 3 to 15, i.e. 19 years old plus, only. The individual appeared in the Individualised Learning Record (ILR) aims data (in England) for at least one day in each of 6 consecutive months of the tax year. This includes classroom learning and apprenticeships (both) any level.
HE	the individual appears in the HESA Student Record data (UK HE institutions) for at least one day in each of 6 consecutive months of the tax year, studying for a qualification of at least level 4.
Out of work benefits	the individual was claiming out-of-work benefits (see out of work benefits data for more information) for at least 1 day in each of (at least) 6 consecutive months of the tax year
No sustained activity (**mixed** in pathways analysis)	the individual had some paid employment, participated in some learning or claimed some benefits in the tax year, but did not fulfil the requirements for any of the sustained definitions. The transition from education to employment often results in no sustained activity due to the changeover from academic to tax year. For example, if someone finishes learning in the summer and starts employment in September they would not meet the sustained employment or sustained learning for that tax year (April to April)
Activity not captured (**unknown** in pathways analysis)	the individual was matched to CIS data but could not be found in any of the applicable labour market or education datasets for that tax year. Some reasons for appearing in this category include not working but not claiming out-of-work benefits, earning below the Lower Earnings Limit or working outside of the UK. Self-employed individuals (who do not have other activities) will be in this category.

Definitions for sub-groups

32. Sustained activities are provided for all individuals and also broken down by a number of sub-groups. The sources and definitions of these are shown below.

a) Gender – This is taken from key stage 4 NPD data and is available for all individuals regardless of school type. For individuals in school types without a pupil level census (e.g. independent schools) this is taken from data provided by awarding organisations.

b) Free school meals (FSM) eligibility in year 11 – This is those who are eligible for (and are claiming) FSM from the NPD for the academic year the individual finished key stage 4. When data is not available for individuals (e.g. school types without pupil level data collections such as independent schools) they are not included in the analysis. Please note that eligibility criteria for FSM has changed over this time period.

FSM are available to children from low-income families whose parents are claiming certain benefits types[3] and hence it is used as a proxy for lower socioeconomic status or social disadvantage.

c) Special Educational Needs (SEN) in year 11 – This is taken from NPD data for the academic year the individual finished key stage 4. As with FSM those where SEN status is unknown/not captured are excluded from analysis.

   SEN analysis is broken down into three sub-groups: 1) individuals with statement of SEN, 2) individuals with SEN without a statement and 3) individuals not identified as SEN. This uses the SEN Code of Practice[4] which came into effect on 1 January 2002. A child or young person has a statement (now a Education, Health and Care (EHC) plan) when a formal assessment has been made. A document is in place that sets out the child's need and the extra help they should receive. This is issued by the Local Education Authority after an assessment which involves parental, education, medical, psychological and social services input. The categories School Action and School Action Plus are combined into SEN without a statement.

d) Ethnicity – Major and minor ethnic group are taken from NPD data for the academic year the individual finished key stage 4. This is not available for those individuals in school types without pupil level data collections (e.g. independent schools) and therefore not included in the analysis.

   In the 2001/02 academic year, schools were using two different classification systems. The majority of schools were using an older classification which did not have any mixed ethnicities. Instead these would be classed as Any Other Ethnic Group. Furthermore there was no breakdown for Traveller of Irish Heritage or Gypsy/Roma in 2001/02 (other White).

e) First language – First language is taken from NPD data for the academic year the individual finished key stage 4. Again, as with other variables stated above, this is not available for those individuals in school types without pupil level data collections (e.g. independent schools) and those 'not known/not collected' are excluded from the analysis.

f) School type – the school type for each individual is taken from the key stage 4 attainment data. This is therefore available for all individuals. Many school types are newer than the oldest cohorts in this data, for example sponsored academies and free schools. We combine the earliest sponsored academies (from 2002/03 onwards) with City Technology Colleges and Local Authority maintained schools to form one mainstream (non selective) state school group. Special schools are excluded from this. Selective (i.e. grammar) schools are removed from this group.

   Selective (or grammar) schools are shown as a separate school type to enable comparisons between them and mainstream state schools (non-selective) maintained and independent (i.e. private) schools.

g) Key stage 4 attainment – this is taken from the key stage 4 attainment file and is therefore available for all individuals. The measure used is achievement of (at

[3] https://www.gov.uk/apply-free-school-meals

[4] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/273877/special_educational_needs_code_of_practice.pdf

least) 5 GCSEs (or equivalent) at A*-C. The equivalents which counted in each academic year may differ depending on performance tables rules at the time. Although this is not the current headline measure for key stage 4 (School performance measures: about the data) it is a consistent measure in place at the time for all cohorts.

h) Region of school – this is derived from the administrative local authority for the school the individual attended when they were in key stage 4 and is available for all individuals. This is the region of the school and may not be where the pupil lived (e.g. pupil residing in Scotland or Wales who cross borders).

i) IDACI – the Income Deprivation Affecting Children Index (IDACI) is one of the English Indices of Deprivation[5] and is taken from NPD data for the academic year the individual finished key stage 4. This is not available for those individuals in school types without pupil level data collections (e.g. independent schools). Individuals without a valid IDACI score are excluded from the IDACI tables. Individuals are assigned to quintiles (top 20 per cent, bottom 20 per cent and middle 60 per cent) based on their IDACI score. The top quintile contains individuals from the most deprived areas.

j) Graduates – in this publication, an individual is classed as a graduate if they have achieved a qualification of at least level 6 (https://www.gov.uk/what-different-qualification-levels-mean/list-of-qualification-levels) in a UK higher education institution. This can be at any point after GCSEs, i.e. up to year 15 (year 2017-18). This is taken from HESA data: latest year 2017/18.

   Table 5 shows the proportion of graduate individuals completing their degrees in each tax year after GCSEs. It can be observed that just over a third (37%) graduate 5 years after GCSEs, i.e. when they are around 21. This means that almost two thirds (63%) of graduates do not do the 'classic' two years of A-levels, 3 year degree back to back. Those graduating from year 6 onwards could have taken longer to complete KS5, taken a gap year or studied over a longer period than 3 years (e.g. sandwich years, 4 years courses etc). There is a peak around year 6 and numbers and proportions drop sharply around year 7/9, but we still see individuals graduating after year 10.

   This table is extremely important to note when comparing analysis from this report with graduate statistics and analysis. In fact, it shows that direct comparisons would not be wise. For example, comparing graduate outcomes one year after graduation with graduate individuals 6 years after GCSEs is not valid. Almost two thirds of graduate individuals are still studying in year 6 and if any of these were working part time this would be included in the analysis in this report.

k) Non-graduates level 3 or above: this is individuals that achieve a highest level qualification of level 3 to 5. (Full) Level 3 is at least two A-level passes or a qualification of the same size or more. It includes a range of academic and technical qualifications.

[5] https://www.gov.uk/government/collections/english-indices-of-deprivation

l) Non-graduates level 2 or below: this sub-group is formed of individuals with highest level of educational achievement level 2 or below. A full level 2 is achieving at least five GCSEs A* to C or a qualification of the same size.

Table 5: Proportion of graduate individuals completing their degrees each (tax) year after GCSEs

Tax years: 2003-04 to 2017-18

(Tax) Years after KS4/GCSEs	Proportion of graduate individuals completing degree in that year
1	0%
2	0%
3	0%
4	0%
5	37%
6	34%
7	14%
8	7%
9	3%
10	2%
11	1%
12	1%
13	0%
14	0%
15	0%

Source: Longitudinal Education Outcomes dataset

Earnings trajectories

Annualised earnings

33. For each individual who has been paid through the PAYE system, the earnings reported for a given tax year are divided by the number of days recorded in the employment spells 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[6].

[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 an individual is employed for the full tax year we will use 365 days (or 366 in leap year) in the calculation.

Adjusted earnings

34. So that earnings are comparable across tax years, the **annualised earnings** have been adjusted for inflation using 2017-18 as a base year. For earlier years, earnings were adjusted using CPIH (Consumer Prices Index Including Owner Occupiers' Housing Costs) inflation rates in each tax year relative to the previous tax year. Please see https://www.ons.gov.uk/economy/inflationandpriceindices for further information.

35. Earnings figures are only reported for those classified as being in **sustained employment** and where we have a valid PAYE earnings record for that tax year. Annualised, adjusted earnings for all cohorts (2001/02 to 2006/07) are combined and medians calculated for each year after finishing key stage 4. For year 1, this is data from 15 cohorts of individuals, for year 15 this is data from a single cohort. These tables are presented for all individuals for each sub-group and additionally for graduates and non-graduates within each sub-group.

Main activity charts

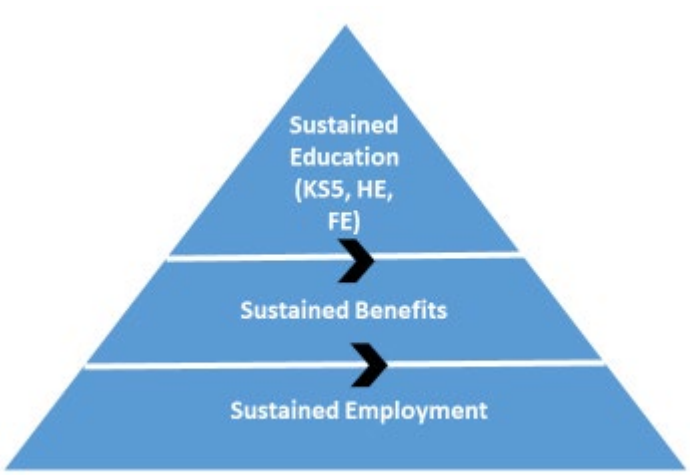
36. The analysis behind the main activities tables, shown throughout the main report for all individuals and different key groups of interest, has been undertaken on the first six academic year key stage cohorts: from 2001/02 to 2006/07. This allows tracking of activities of between 10 and 15 full tax years. There around 3.6m individuals in this combined cohort. See Table 3 in data matching section for more details.

37. The aim of the main activities analysis is to show the education and labour market activities over time. It enables tracking of individuals and comparisons between groups of interest, giving further detail on the evolution of individuals' activities over time.

38. For each tax year an individual is assigned a 'main activity' based on their education and labour market activities. If they meet the criteria for one of the definitions above they are assigned that activity as their main activity. Thus they are categorised as sustained employment, sustained benefits, sustained FE, sustained HE or sustained KS5, 'no sustained activity' (mixed), or 'no activity captured' (unknown) for a given tax year.

39. Though the definitions above tend to be mutually exclusive, is it possible, and perfectly reasonable, for an individual to meet the criteria for more than one definition in a tax year. Some examples could be: those that work part time whilst doing key stage 5 or HE; someone employed but on low income and claiming out of work benefits; or an apprenticeship (which is FE but they are also employed). In a situation where an individual meets the criteria for more than one sustained activity in a tax year a hierarchy is applied to assign a main activity. The hierarchy is shown in Figure 5 below.

Figure 5: Hierarchy – Assumptions for assigning main activity when individual meets more than one education and labour market definition
Tax years: 2003-04 to 2017-18



40. The rationale and assumptions behind the hierarchy are:

- Education and employment: assign main activity to be education. Most important to capture educational information as likely to be doing education as investment in human capital (i.e. skills). Individuals in employment and education in early years tend to have lower average earnings, perhaps suggesting they were working part time. This is particularly the case for KS5 and HE, whilst FE is likely to include employment for apprenticeships and vocational education. Though average earnings increase over time and individuals in education and employment in year 10 may be working full time, this analysis opts to capture education as the main activity as they are likely to be doing this to improve their labour market outcome.

- Education and claiming out of work benefits: assign main activity to be education. Important to capture education information as could be vehicle for changing situation. There are strict criteria for being able to claim out of work benefits whilst in education is quite strict and our analysis shows that average earnings is low for individuals in this category. It is envisaged that those in education and claiming out of working benefits are looking to improve their labour market outcomes, e.g. find employment.

- Employment and claiming out of work benefits: assign main activity to be claiming out of work benefits. The criteria for claiming out of work benefits whilst being in employment is quite strict and our analysis shows that average earnings is low for individuals in this category. As those in employment are claiming out of working benefits, they are likely to be looking to change their labour market outcomes (for example longer hours and/or better pay).

41. Similarly, apply the following hierarchy if in more than one sustained education category: KS5 > HE > Adult (FE) 19+. As above it is possible to meet the criteria for more than one education activity in the same tax year and hence we choose a 'main activity' The rationale for the hierarchy is:

- KS5 and HE: Assign KS5 over HE. It is usually necessary to have level 3 qualifications to start HE and there is likely to be little gain in doing level 3 qualification whilst doing HE (which is defined as level 4 qualifications and above). There are very few of these cases as level 3 needs to be achieved to start higher education.
- HE and Adult FE (19+): Assign main activity as HE. If in both the ILR and HESA data, either for the same learning or something different then choose higher level of learning, which is HE (level 4 and above).

42. There is no crossover between KS5 and other education as this is a deliberate methodological decision. By using the DfE definition of KS5 and putting it at the top of the (education) hierarchy it means that individuals doing KS5 at FE colleges or sixth form colleges, an important part of FE, are classed as KS5. Rather than under or misrepresent FE (in the first two years) we use another categorisation. This is why FE classroom learning level 2 or below and apprenticeships is classified as 'other education'.

43. Readers are reminded key stage 5 can also include vocational qualifications, as long as they are the equivalent of at least one A-level (pass). Table 6 below shows that around two fifths of KS5 takes place in FE or sixth colleges in year 1 and almost seventy percent in year 2. Thus a large amount of what could be deemed FE learning at level 3 is assigned a main activity of KS5.

Table 6: Number of individuals doing KS5 in FE and sixth colleges
Tax years: 2003-04 and 2004-05

Full tax years after completion after GCSEs	Main activity KS5	Of which FE level 3
1	1,712,934	742,892
2	321,706	222,820

Pathways

44. The main activities analysis assigns between 9 and 15 tax years of activities. An example of a classic pathway, where an individual does KS5 (A-levels) immediately after KS4 over a 2 year period, followed by HE for 3 years and then becomes employed and stays employed up until the latest data is shown below.

PMR	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Year 9	Year 10	Year 11	Year 12	Year 13	Year 14	Year 15
Individual X	KS5	HE	HE	HE	mixed	employed	employed	employed	employed	employed	employed	employed	employed	employed	employed

45. The next step of the methodology aims to simplify the main activities data and show pathways by turning tax years into activities. This is done by collapsing multiple tax years of the same activity into one activity. First activity is the main activity for year 1. Only when the activity is different in the following year is the next activity created, but if the activity in the next tax year is the same this is ignored. This process is undertaken over the (10 to) 15 years of data. The example above would therefore become:

PMR	First activity	Second activity	Third activity	Fourth activity
Individual X	KS5	HE	Mixed'	employed

46. This approach allows us to examine more easily and in greater depth the pathways of individuals. The number of individuals (by sub-group) taking each unique pathway is determined. Due to the number of individuals in the cohort, this can result in a very large number of pathways of differing complexity and popularity.

Sankey charts

47. To aid visualisation of these pathways in the commentary, Sankey charts are used. These are a type of flow diagram in which the length of coloured bands or 'nodes' (representing the activities) indicate the total numbers passing through them, and the depth of the 'flows' between the nodes reflects the numbers moving from one activity to the next. Therefore, the width/size of the flows and nodes are important. These charts are a useful tool in distilling complex information and particularly the direction and size of flows within a system.

48. When an education or labour market activity name in the Sankey chart includes a '2', this means that it is the second period of that activity for a group of individuals. For example if the chart includes the nodes 'HE' and 'HE 2' this means that some of the individuals from the sub-group in question had two separate spells in higher education i.e. they did at least one other activity between HE and HE 2. Similarly '_final' signifies a latest activity, for example 'employed_final' means that the individuals represented by this node had sustained employment as their latest main activity.

49. In the examples shown throughout the report, the Sankey charts show the most common education and labour market pathways of individuals after they leave compulsory education and where they end up (in the latest data). As illustrated in the analysis, these pathways are extremely complex and this makes it extremely difficult to depict visually. For this reason, the top 50 pathways were selected.. The representativeness of the Sankey charts vary for different sub-groups due to the number of individuals taking the most popular pathways and the number and complexity of pathways for that group. Table 7 shows that the representativeness ranges from 23 to 56 per cent but generally the Sankey charts show the most popular pathways representing around a third of the group in question.

Table 7: Percentage of individuals represented by top 50 pathways for different sub-groups

Tax years: 2003/04 to 2017/18

Sub-group	Percentage of individuals represented by top 50 pathways
All individuals	33%
Male	32%
Female	34%
FSM eligible	23%
FSM not eligible	34%
Not identified as SEN	35%
SEN with statement	38%
SEN without statement	23%
Asian	39%
Black	27%
Chinese	56%
Mixed	30%
White	32%
Bangladeshi	31%
Indian	50%
Pakistani	30%
Black African	33%
Black Caribbean	24%
White British	32%
First language English	32%
First language Other	36%
Achieved five passes A* to C at KS4	45%
Did not achieve five passes A* to C at KS4	24%
Mainstream school	32%
Independent school	54%
Selective school	57%
North east	30%
East Midlands	33%
London	33%

Data quality

50. The LEO dataset is derived from administrative data and as such there are limitations to the activities that we can assign individuals to.

- Key stage 4 cohorts – these are based on those who appear in the performance tables data. It captures all those undertaking KS4, i.e. GCSEs, in English schools. However, there will be children who do not appear in this dataset (e.g. home-schooled, not on roll or in independent schools but not entering exams). This is not envisaged to be many, but estimating is difficult
- Key stage 5 as an activity – those individuals aged 16-18 and in school sixth forms and special schools who do not enter for any level 3 qualifications will not be classed as in sustained learning. Their other activities (if any) will determine the sustained activities they are assigned to. Those of this age group in sixth form colleges and further education colleges who do not meet this criterion will be included in sustained FE (if they meet the sustained criteria)
- Further education outside England
- Higher education outside UK
- School based Initial Teacher Training (ITT): Depending on setting those in teacher training may not be picked up. For example some school based teacher training may not be identified, though the majority is partnered with a university or FE college and will appear in HESA /ILR data
- Deceased individuals are not currently identified in this analysis and will be in the activity not captured
- Individuals who have moved abroad will be in activity not captured unless they are paying tax in the UK

Publications using LEO data

51. A number of DfE Official Statistics publications use LEO data to produce statistics on education and labour market outcomes. The definitions and methodology used in these publications do not all align with each other or with this publication. In most cases, this is due to either the purpose of the publication (for accountability purposes) or for timeliness.

Destinations of KS4 and KS5 pupils: https://www.gov.uk/government/collections/statistics-destinations

52. This data is produced as part of the School and College Performance tables and hold schools and colleges to account for their success in helping young people continue in education, employment or apprenticeships.

Further Education: Outcome based success measures: https://www.gov.uk/government/collections/statistics-outcome-based-success-measures

53. This publication presents employment and learning outcomes of Further Education (FE) individuals. It includes all age apprenticeships, all age traineeships, and adult (19+) FE and Skills individuals that completed an Education and Skills Funding Agency (ESFA) funded aim.

Graduates outcomes (LEO): https://www.gov.uk/government/collections/statistics-higher-education-graduate-employment-and-earnings

54. This is a series of publications on the employment and earnings outcomes of higher education graduates. The information is provided at institution level as well as national.

55. The similarities and differences between these publications and this research report (Post 16 Education and Labour Market Activities, Outcomes and Pathways (LEO)) are outlined in the box below.

Differences in definitions between education and labour market pathways (LEO) research report and other LEO publications

It should be clear to the reader that the objectives of this research report are quite different to the other publications mentioned above. The aim is to show the employment and labour market activities over time and compare this for different sub-groups. Education level (graduates and non-graduates, non-graduates level 3 or above, level 2 or below) is introduced to show how this changes for these sub-groups when factoring in education level. Therefore the methodology and definitions are tailored to this report.

This report tracks cohorts of individuals, i.e. people who left school at the same time, whilst graduate statistics tracks cohorts people who left university at the same time, outcome based success measures publication tracks those completing FE at the same time. Whilst destination measures follows individuals it has a short term focus and has a very different purpose. Most of the other publications are undertaken for accountability purposes.

One key element is differences in the definition of employment and median earnings. This report uses a definition of being in employment for 12 months of the tax year, whilst other publications use 5 out of 6 months between October and March. In median earnings calculations, this report includes all those who meet the employment definition, whereas other publications exclude those in education. Additionally self-employment income is included in other publications but not in this report. These are all driven by the objectives of this report.

When showing education and labour market activities and earnings trajectories over time, we wanted to only show those that were in employment for the whole tax year. This allows us to see if someone has been consecutively employed over several (tax) years. Whilst those who are in education have their main activity captured as education, they are not removed from average earnings calculations as long as they meet the 12 month employment definition. We do not have hours worked or sector information and hence think it is important to include all employment that meets the definition. As covered earlier, self assessment data is only available from 2013/14 and therefore it would not be possible to include this in the analysis, which covers years 2003/4 to 2017/18.

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

Oliver Anderson, Central Research Division, Department for Education, Sanctuary Buildings, Great Smith Street, London, SW1P 3BT

Email: oliver.anderson@education.gov.uk

Department for Education

Reference: DFE-RR1125

ISBN: 978-1-83870-261-8

For any enquiries regarding this publication, contact us at:
oliver.anderson@education.gov.uk or www.education.gov.uk/contactus

This document is available for download at www.gov.uk/government/publications