

Apprenticeship Attainment and Economic Outcomes in London Data Publication – Methodology Information

November 2026

About this document

This document provides a range of information on the methodology used to produce Apprenticeship Attainment in London and Apprenticeship Economic Outcomes in London reports. It provides an overview of the data used, cohorts of students analysed and methodology for identifying an individual's highest level of education by age 24.

Contents

Aim of the analysis	1
The Longitudinal Education Outcome dataset	2
Key Stage 4 Cohort	3
Demographic Data	6
Economic Data	6
Get in Touch	8

Aim of the analysis

The Apprenticeship Attainment and Economic Outcomes in London reports are based on rich administrative data.

The analysis presented in Apprenticeship Attainment in London allows for comparison of backgrounds between those who have different levels of qualifications at age 24.¹ Such analysis can help inform questions around progression in education, and the educational pathways different types of learners take. For example, analysing “what is the difference in the share of learners who were eligible for free school meals (FSM) between those who have a Level 2 and Level 3 apprenticeships” can provide insights into whether learners who were eligible for free school meals are less likely to progress onto a Level 3 apprenticeship. This release is the first of several outputs using LEO. The data will also be used to select a comparison group for estimating early earning returns.²

¹ By 24 most individuals have completed their formal education.

² Note that any estimation of earning returns using LEO data are limited to only a few years after completing a qualification.

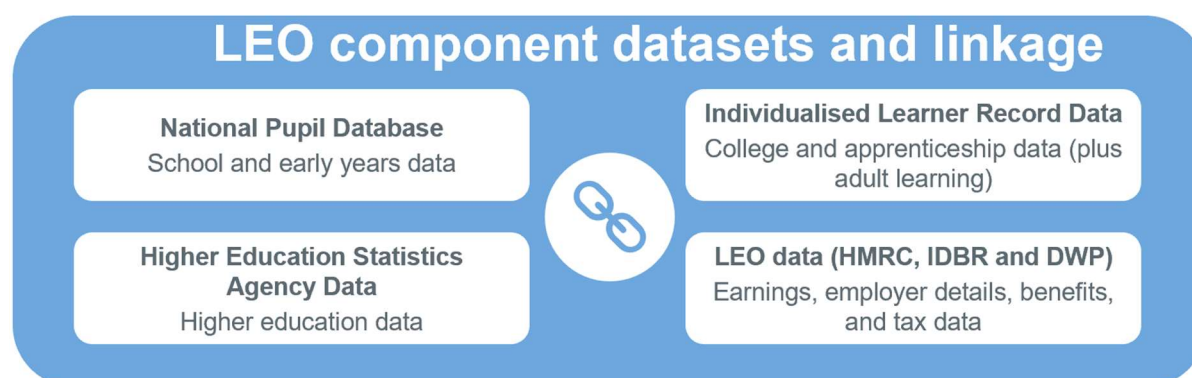
The statistics presented show the background characteristics (demographics, prior attainment) of those whose highest qualification at age 24 (9 years after finishing Key Stage 4 (KS4)) is a Level 2 or Level 3 apprenticeship. The data is split by those whose home location during KS4 (by age 15 at the start of the academic year) was in London compared to the rest of England.

The Apprenticeship Economic Outcomes in London report shows descriptive data on earnings, employment, claiming unemployment related benefits and sector by highest qualification at age 24. Similarly to the first report, the data is split by the highest qualification at age 24, but grouped by academic, vocational and apprenticeships at Level 2 and 3, and for all qualifications at Level 1.³ This allows understanding of the implications of different attainment levels in London. Many factors beyond someone's highest qualification level will impact their earnings, so further analysis is needed to account for variations in earnings caused by individual's background characteristics and attainment.

The Longitudinal Education Outcome dataset

The analysis is based on a dataset called the Longitudinal Education Outcomes (LEO) dataset. This dataset combines administrative data from Department for Education (DfE), Higher Education Statistics Agency (HESA), Department for Work and Pensions (DWP) and HM Revenue and Customs (HMRC).⁴ This analysis uses the standard release of LEO from the ONS Secure Data Service (SRS).

Figure 1: Datasets included in LEO utilised for the analysis.



- The DfE data includes schooling data from the National Pupil Database (NPD) which provides background characteristics of learners and grades achieved in school. This analysis focuses on grades achieved at the end of Key Stage 2, (by academic age 10) demographics and grades at the end of Key Stage 4 (by academic age 15) and grades at the end of Key Stage 5 (by academic age 17⁵). Complete data is only available for those who attended an English state-funded school.⁶

³ As Level 1 qualifications are generally introductory qualifications, they are less naturally split into academic, vocational and apprenticeships.

⁴ Further information on the [LEO dataset is available from DfE](#).

⁵ Some individuals will be academic age 18 if they have retaken the year.

⁶ Alternative provision was added into the school census from 2010. So, the 2007/08 cohort does not include individuals from alternative provision. Alternative provision is schooling for those who have been excluded or are at high risk from exclusion from mainstream provision, or because they have significant health needs.

- The Individualised Learner Record (ILR) data from the DfE/ESFA which includes provision of post 16 further education (FE) in England by colleges, FE providers and Specialist post-16 institutions. The ILR covers 16-18 and 19+ vocational/technical education, including apprenticeships and other adult skills programmes. Therefore, the ILR is used to identify individuals whose highest qualification is an apprenticeship.
- HESA data covers higher education provision across the UK, and includes enrolments, student characteristics, subjects studied, and qualifications achieved.
- The DWP data covers timing and duration of benefit claims for out of work benefits. It includes out of work benefits such as Jobseeker's Allowance (JSA), Incapacity Benefit (IB), Income Support (IS), Employment Support Allowance (ESA) and out of work conditional Universal Credit (UC).
- HMRC data includes data on employee earnings and employment spells based on tax records. It also includes self-assessment data for self-employed individuals, with an indicator of if someone was self-employed in the financial year and their total self-employed earnings in the year.
- IDBR data includes employer data for those in employment, specifically for someone's main employment the sector of work and the size of the business.

HMRC, DWP and IDBR data are not used in the Apprenticeship Attainment statistical release but are used in the Apprenticeship Economic Outcomes statistical release.

Key Stage 4 Cohort

The descriptive analysis on apprenticeships is based on cohort analysis. This means looking at a group of individuals who finish KS4 schooling (qualifications at age 16, predominantly GCSE) in the same year.⁷ The analysis then tracks these individuals over time, matching them to other education records. The two cohorts used in this analysis are for individuals whose academic age was 15 (their age at the beginning of the academic year, 31 August) in 2007/08 and 2010/11, finishing KS4 in that academic year. As shown by Table 1, the cohort is then tracked through the KS5, ILR and HESA data to get an individual's highest qualification at academic age 24.

For the Apprenticeship Economic Outcomes statistical release, only the 2007/08 cohort is shown to allow for a longer tracking of economic outcomes.

Academic Year: Cohort 1	Academic Year: Cohort 2	Common Stage	Academic Age	KS4 Data	KS5 Data	ILR (FE and Apprenticeships data)	HESA (University and HE data)
2007/08	2010/11	Last year of GCSE	15	Results from Year 11			
2008/09	2011/12	AS Level	16				If achieved a qualification in
2009/10	2012/13	A level	17				

⁷ The majority of the cohort will be academic age 15 at the end of KS4, but some who have retaken the year will be academic age 16.

2010/11	2013/14	1 st Year of University	18		KS5 data from Year 13 plus next year to include retakes	Highest qualification in the ILR between 2008/09 and 2016/17 (up until academic age 24) for Cohort 1 and between 2011/12-2019/20 for Cohort 2.	the HESA data between 2008/09 and 2016/17 (up until academic age 24) for Cohort 1 and between 2011/12-2019/20 for Cohort 2.
2011/12	2014/15	2 nd Year of University	19				
2012/13	2015/16	3 rd Year of University	20				
2013/14	2016/17		21				
2014/15	2017/18		22				
2015/16	2018/19		23				
2016/17	2019/20		24				

Table 1: The data used to analyse cohort 2007/08 (born in 1991/92) and cohort 2010/11 (born in 1994/95)

Note: Across all years the academic age of the individual is defined by an individual's age at the 31st August as the academic year begins. Data from YPMAD at age 24 is also included for retake data and academic level 2 and level 3 data not captured in KS4 and KS5 data.⁸

⁸ Due to differences in how full level and 3 are defined in the ILR and YPMAD, YPMAD data is only included for academic qualifications.

Identifying an individual's highest level of qualification

An individual's highest level of qualification is identified by combining the qualifications achieved in different datasets:

- **Key Stage 4** (from NPD) to identify if an individual has achieved either a Level 1 or Level 2 (passed 5 GCSEs at A*-C or equivalent) qualification. Level 2 qualifications at KS4 are all categorised as academic.
- **Key Stage 5** (from NPD) to identify if someone has achieved a full Level 3 qualification (equivalent to 2 A levels). The qualification is categorised as either academic or vocational based on the largest volume of entries. Applied A levels, A levels and AS Levels are categorised as academic, and BTEC/OCR and NVQ/VRQ are categorised as vocational.
- **Young Person's Matched Administrative Dataset (YPMAD)** (from NPD) which includes some data that is not included in KS4, KS5 and ILR datasets. In particular, it captures a small amount of Level 2 learning (such as GCSE retakes) that is done by 16-18 years old that is not captured by the ILR or KS5 (which only includes Level 3 learning).
- **Individualised Learner Record** which shows all further education qualifications. A learner is recorded as having a Level 2 or 3 qualification if they have achieved the equivalent to a full Level 2 or Level 3 (either by one qualification or by a combination of partial aims). Qualifications in the ILR are categorised as vocational unless the following applies:

Table 2: Defining qualification group for Level 2 and 3 qualifications in ILR.

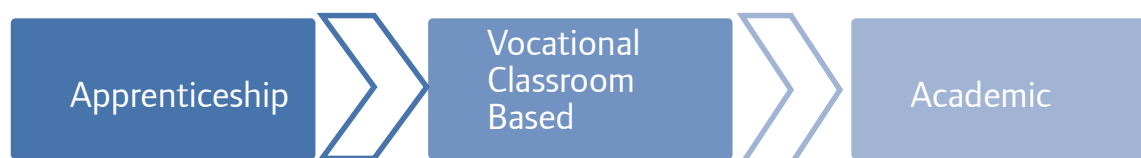
Definition from ILR	Qualification Group
Qualification Type = Advanced Subsidiary Level	Academic
Qualification Type = Advanced level	Academic
Qualification Type = International Baccalaureate	Academic
Qualification Type = Applied Advanced Level	Academic
Qualification Type = Applied Advanced Level (Double Award)	Academic
Qualification Type = Applied Advanced Subsidiary Level	Academic
Qualification Type = Access to Higher Education Diploma	Academic
Level of Learning = Level 2 Apprenticeship	Apprenticeship ⁹
Level of Learning = Level 3 Apprenticeship	Apprenticeship

- **Higher Education Data** (from HESA) this is used to identify those who have obtained a Level 4 or above academic qualification, as well as a small number of those who have done qualifications below Level 4 at a higher education institution. For categorising

⁹ Apprenticeships are defined using the apprenticeship indicator in the ILR. This identifies the main programme aim for apprenticeships.

Level 4 qualifications, foundation degrees are assumed to be academic, and higher national certificates and higher national diplomas are assumed to be vocational.

Where a learner has achieved multiple aims (courses) at the same level, qualifications are ranked by qualification group (apprenticeship, classroom vocational or academic) where:



Then qualifications are ranked by which qualification was achieved most recently and finally by aim rank in year.¹⁰

For those who have achieved multiple partial aims to make up their full Level 2 and Level 3, the width (the percentage the qualification makes up a full Level 2 or 3) of a combined qualification group is calculated to estimate the main qualification type, followed by qualification group, achievement data and aim rank in year. For example, if an individual has taken 2 BTEC qualifications with a width 50% each and an academic qualification with width 50%, the individual is categorised into vocational classroom-based qualifications. Then the two BTEC qualifications are ranked (predominantly to get a unique subject sector area) based on which was achieved most recently and aim rank in year.

Demographic Data

A learner's demographic data is predominantly sourced from the Spring School Census data (from NPD) during KS4. This data includes:

- Ethnicity
- Free School Meals eligibility
- Special Educational Needs provision and type
- Gender
- Home location used for:
 - Identifying students from London
 - Used as a proxy for deprivation (as measured by IDACI)¹¹

Economic Data

To track economic outcomes overtime and to see what differences in outcomes appear to persist we measure economic outcomes for the 2007/08 cohort in 5 different financial years: 2013/14 (when individuals will be aged 21 or 22), 2015/16 (age 23/24), 2017/18 (age 25/26), 2018/19 (age 26/27) and 2020/21 (age 28/29).

¹⁰ Aim rank is a derived variable in the ILR which ranks qualifications based on level, achievement date and aim sequence number (a unique number for recording aims in the ILR). Hence, in this case an individual who start two qualifications in the same qualification group, at the same level on the same date, the qualification with the lowest aim sequence number will be chosen.

¹¹ Income Deprivation Affecting Children Index. This indicator is created by the Ministry of Housing, Communities & Local Government and measures the proportion of all children aged 0 to 15 living in income deprived families, it is a subset of the Index of Multiple Deprivation (IMD).

Employment proportions are recorded as the share of individuals who either are recorded as self-employed¹² during the financial year or are an employee for at least half the days in the year.

The earnings data combines employee earnings and earnings from self-employment during the financial year. The earnings are then adjusted by inflation (CPIH) to show real earnings in 2020/21 prices to allow the analysis to focus on non-inflation related changes to median earnings over time. Individuals are removed from the median earnings calculation if they are either: studying a higher education qualification in that year (either as recorded in HESA or a Level 4 or above in the ILR)¹³, or if their earnings are recorded as zero. Data on part-time work is not available, so an individual's earnings being lower could either be due to a lower salary or lower number of hours worked.

Benefit shows the percentage of individuals who are claiming at least one sort of out of work benefit for at least half of the days in the year. Similarly to earnings, it excludes those who are studying a higher education qualification in that year.

Individuals may have multiple jobs or be claiming multiple unemployment related benefits over a financial year, all this information is summarised into one overall employment spell, total earnings over the year and one overall benefit spell.

To show the data by gender, the individual's gender recorded in the Spring Census in the NPD is used. This allows it to be consistent with the demographic data in the Apprenticeship Attainment in London release.

Data on sector shows for those who are employees in the HMRC earning data, the sector of their employer for their main job. The analysis excludes those whose sector is Activities of extraterritorial organisations and bodies, dormant and activities of households as employers – undifferentiated goods and services – producing activities of household for own use. The data also combines smaller sectors together due to sample size limitations where Production and Agriculture are defined as:

- Mining and Quarrying
- Agriculture
- Forest and Fishing
- Electricity gas steam and air conditioning supply, water supply – sewerage, waste management and remediation activities)

And all other services are defined as:

- Arts, Entertainment Recreation
- Finance & Insurance
- Digital
- Real Estate
- Transport & Service
- Professional, Scientific & Technical
- Public Admin & Defence.

¹² Due to limitations in the data, we don't know for what duration an individual was working as self-employed during the year.

¹³ Due to limitations in data availability, individuals cannot be excluded if they are in higher education in 2020/21.

Get in Touch

Media enquiries

All media enquiries should be emailed, and we will respond as quickly as possible.

Email: mayorspressoffice@london.gov.uk

For urgent media enquiries outside office hours and at weekends please call: 020 7983 4000 and ask for the duty press officer.

For technical enquiries please email Ruth.Moxon@london.gov.uk