

Data Explained

Longitudinal Education Outcomes (LEO)

Identifying and studying young people who are Not in Education, Employment or Training (NEET)

Authors: Peter Urwin, Dave Thomson, Min Zhang, Dave Bibby, Emma Gorman and Luyang Chen

Date: December 2025

This Data Explained summarises experiences and learning from working with Longitudinal Education Outcomes data in the course of producing research into the post-16 pathways of lower attaining young people. This publication is intended to help guide future researchers using this data and to provide feedback into future dataset development and documentation.

The administrative data discussed in this Data Explained was made securely available to support research funded by the Nuffield Foundation. The data used in this research project comes from the Department for Education (DfE), HM Revenue and Customs (HMRC) and the Department for Work and Pensions (DWP) and was accessed through the Office for National Statistics (ONS) Secure Research Service. The data was not originally collected for research and it is expected that there are gaps and inconsistencies in its recording, a number of which are detailed in the following.

Project details

This project analysed different groups of learners between 2011 and 2019, to shed new light on the education and labour market experiences of lower attaining young people, focusing on their post-16 pathways. The investigation considered 260,000 young people who in 2011 achieved Grade 3 ('D') or below in Maths and/or English GCSE in the year they turned 16 (Key Stage 4); with a particular focus on young people within this group whose KS4 achievement in 2011 suggests a more substantial educational challenge, concentrating on 47,000 pupils who achieved grades of E, F or G in Maths and English GCSE. Consideration of the 2011 KS4 cohort allowed us to estimate the labour market value of post-16 education pathways taken by lower attainers up to the age of 24 (2019). Many lower attainers have 'fractured' post-16 learning pathways and high levels of drop-out from education. A key challenge is therefore the identification and analysis of NEET spells experienced by these young people, especially when there is no information in the administrative data.

Initial research questions

- (1) For young people who find themselves with a grade 3¹ or below in Maths and/or English at KS4, which post-16 education pathways secure favourable education and labour market outcomes?
- (2) How have recent policy initiatives impacted the post-16 education and training pathways taken by these young people, and the outcomes they achieve? How do such impacts vary by social background and has this changed over time?
- (3) What is the causal impact of different post-16 education pathways taken by young people who do not achieve grades 9 to 4 in English and/or Maths at 16 and how do these estimated impacts [on later life outcomes] vary according to social background?

Research methodology

1. Using descriptive statistics and visualisation tools (e.g. Sankey/Alluvial Diagrams), we describe the flows of young people who leave school having not achieved grades 9-4 (A*-C) in English and/or Maths along different post-16 pathways, and how this translates into further study and/or entry to the labour market. We focus particularly on describing how patterns of post-16 engagement vary according to Free School Meal (FSM) status.

2. Natural variation in the data allows us to use quasi-experimental methods to estimate the impacts of different post-16 learning choices on the probability of obtaining Level 3 learning by age 21; and employment and earnings outcomes by age 24.

¹ Grade D or below prior to 2017

Key variables

1. The population of interest was all state school pupils in England who were 15 years old at the beginning of the 2010-11 academic year - i.e. young people who turned 16 during the 2010-2011 academic year – we refer to this as the 2011 Key Stage 4 (KS4) Cohort.
2. Determining whether a young person is in education is the first step in the process of categorising their post-16 activity status. The 2011 KS4 cohort are considered to be in a post-16 educational institution during the year they turn 17, if they are recorded as being on roll at a state-funded school in the Autumn School Census in National Pupil Database (NPD) or have an active learning aim in the Individualised Learner Record (ILR) on (or around) enumeration dates in October 2011, January 2012 and May 2012². An accompanying Data Explained describes the detailed process of data linking and manipulation we undertake to analyse *Post-16 Learning Spells and Qualification Aims*, for young people who are in a post-16 educational institution during the year they turn 17.
3. This process is then repeated for subsequent years, bringing in another dataset, the HESA student record, to include enrolments at higher education institutions.
4. We also consider, (ii) HMRC data on employment and earnings (including self-employment); (iii) DWP data on ‘active benefit claimants’ such as Job Seekers Allowance (JSA), Job Training Allowance (JTA) and some Universal Credit (UC) claimants³; the Employment and Support Allowance (ESA) Work-Related Activity Group and those on ‘inactive benefits’.
5. Young people who are not identified as being either (i) in education or (ii) employment; but are recorded as being on (iii) some form of active or inactive benefits in DWP data, are considered as NEET active or NEET inactive (respectively)⁴. Young people who have no administrative record across all these datasets for the relevant year, are in a group that can be considered ‘Not known/ missing (none of the above)’. They could be NEET, but not claiming benefits; they may have emigrated; there may have been errors in matching⁵; we may have individuals with earnings records but no corresponding employment records; and young people may no longer be in England or be deceased.

² These dates in October, January and May are used as they are dates on which learning providers are mandated to provide ‘census’ returns.

³ As is often the case when analysing the earnings and employment returns to learning, we focus on historical cohorts of learners (here the 2011 KS4 cohort). This allows us to observe a significant number of years after learning has been completed, but it also means we are focusing on cohorts prior to Raising of the Participation Age (RPA) and prior to the full roll out of UC, so historical benefits and employment were an option for this cohort of school leavers.

⁴ It is worth noting that ‘active’ and ‘inactive’ here have different meanings to the terms used when considering all age groups in the wider labour market – for instance, a NEET inactive young person is not in education – the category of ‘inactive’ in the wider labour market includes those who are in education.

⁵ Matching errors may arise due to under-matching between NPD and HMRC/DWP; matching between NPD, ILR and HESA may suffer from over or under matching and the same is true for matching between DWP and HMRC. For certain groups of young people (for instance, those with English as an Additional Language, EAL in education datasets) match rates to DWP/HMRC records can drop below 90%, whereas match rates for non-EAL pupils exceed 95%.

6. Introduction of National Client Caseload Information System (NCCIS) data to LEO has improved our ability to identify the status of many of these young people who we would previously have considered in the 'not known/missing' group. NCCIS data collection is the responsibility of local authorities and is used by DfE to estimate [participation in education, training and NEET age 16 to 17 by local authority](#). It only covers young people up to the age of 18 and some of the variability across regions may be due to varying practices in collection. However, it allows us to identify activity status for many young people who were previously missing from the main administrative data records, including those in custody. Combining this information with an approach that differentiates between those who have no administrative data record in any of the years beyond KS4 (including NCCIS); and those who we see in at least one year after KS4, helps in the estimation of activity status between the ages of 18 and 21.

The table below includes variables used in the process of determining a young person's post-16 status and identifying NEET spells. This includes education records and for further detail of the (NPD and ILR) variables used when young people are identified as being in post-16 education, see the accompanying *Post-16 Learning Spells and Qualification Aims Data Explained*.

Dataset name	Variable Type	Variable Name	Description
Education_Records_Lookup	Identifier	Ae_id	Bridging identifier to link datasets
Learner_AE_ID_to_EDUkey_Lookup	Identifier	Edukey	Links AE_ID to DWP/HMRC person identifier
Learner_AE_ID_to_PMR_Lookup	Identifier	PMR	Links AE_ID to NPD person identifier
Individualised Learner Record (ILR)	Identifier	External_id	Main identifier used in ILR
	Spell Variables	Learning_aim_id	Identifies a qualification aim on which a student is enrolled
		Startdate	Date aim started
		Ukprn_anon	Provider at which student is enrolled
		Enddate	Date aim ended (will be null if ongoing at end of year)

Data Explained

NPD School Census	Identifier	Pupilmatchingrefanonymous	Main identifier used in NPD
	Spell Variables	Laestab_anon	School at which pupil is enrolled
		Entrydate	Date pupil joined the school
		Leavingdate	Date pupil left the school
		Onroll	Flag to indicate the pupil was on roll at the time of Census
HESA	Identifier	He_pupilmatchingrefanonymous	Identifier used in linking across data sources
	Spell Variables	He_xpsr01	Identifies the relevant population of enrolled students
		He_comdate	Date aim started
		He_enddate	Date aim ended (will be null if ongoing at end of year)
		He_xlev601	Identifies level of study
NCCIS	Identifier	Nccis_pupilmatchingrefanonymous	Identifier used in linking across data sources
		Nccis_current_activity_code	Activity code
	Spell Variables	Nccis_current_activity_start_date	Date activity started
		Nccis_acadyr	Academic year
Nccis_monthno	Month		

Data Explained

Employment	Identifier	Edukey	Identifier used in linking across data sources
	Spell variables	Startdate	Start date of employment
		Enddate	End date of employment
Benefits	Identifier	Edukey	Identifier used in linking across data sources
	Spell variables	Startdate	Start date of employment
		Enddate	End date of employment
		Out_of_work_benefit	Flag indicating out of work benefits

We calculate the following activity states based on observed dates from the sources in the table above on each enumeration date:

1. In Custody (NCCIS)
2. Enrolled at a state school (School Census)
3. Enrolled at a higher education institution (HESA)
4. Enrolled at a further education provider (ILR)
5. In employment (Employment)
6. In receipt of out-of-work benefits (Benefits)
9. Not otherwise classified

Individuals are assigned to the first category of the hierarchy they fall into.

NCCIS is then scanned for those in group 9 above. Any individual in one of the following activities is allocated to group 6 (NEET active)

- 540 - Working not for reward
- 610 - Not yet ready for work or learning
- 615 - Start date agreed (other)
- 616 - Start date agreed (RPA compliant)
- 619 - Seeking employment, education or training

Finally, we create group 7 (NEET inactive) consisting of individuals observed in the following activity states in NCCIS:

- 620 - Not available to labour market/learning - carer

- 630 - Not available to labour market/learning - teenage parent
- 640 - Not available to labour market/learning - illness
- 650 - Not available to labour market/learning - pregnancy
- 660 - Not available to labour market/learning - religious grounds
- 670 - Not available to labour market/learning - unlikely ever to be economically active
- 680 - Not available to labour market/learning - other reason

Summary of comments on specific variables

Variable name(s) (include dataset if relevant)	Comments (separate into multiple columns if appropriate)
NCCIS_Activity_Code	Provides status for many previously 'missing' young people. NCCIS collection is mandated at local authority level and 'raw' NCCIS is used at a local level to (e.g.) help create measures such as risk of NEET indicator (RONI) and inform subsequent interventions. Some NCCIS regional variability is very likely due to varying administrative data practices and local characteristics also impacts collection of NCCIS data, such as population density.

How you dealt with data limitations

As with many administrative datasets, death and emigration are not observed. For individuals in an unknown activity state, we do not necessarily know whether they are economically inactive, have emigrated or have died. In some cases, there are later observations for individuals so we might assume they were temporarily inactive. But this nevertheless leaves a number of cases where we cease observing records. Our approach has been to test the sensitivity of results to different assumptions about attrition cases.

Suggested improvements recommended to data owners

Addition of National Client Caseload Information System (NCCIS) data to NPD-LEO has significantly improved the ability of researchers to study young people at risk of NEET spells. However, there are concerns over data quality and we would recommend a dedicated investigation that can consider improvements in collection practices, manipulation and analysis. Much effort is expended by local authorities in collecting NCCIS information and this is the only information available for a population that contains some of the most vulnerable and disadvantaged young people. An investigation of what more can be done to leverage this information for policy analysis and insight would be extremely valuable.

References

Bibby, D., Gorman, E., Thomson, D., Urwin, P. and Zhang, M. (2024), *Post-16 pathways to employment for lower attaining pupils: are they working?* November, Nuffield.

<https://www.nuffieldfoundation.org/wp-content/uploads/2021/09/Post-16-pathways-to-employment-for-lower-attaining-pupils-are-they-working.pdf>

Disclaimer

This work was produced using administrative data accessed through the Office for National Statistics (ONS) Secure Research Service. The use of the data in this work does not imply the endorsement of ONS Secure Research Service or data owners in relation to the interpretation or analysis.

This work uses research datasets which may not exactly reproduce National Statistics aggregates. National Statistics follow consistent statistical conventions over time and cannot be compared to Data First linked datasets.

Acknowledgements

This work was supported by the Nuffield Foundation and ADR UK (Administrative Data Research UK). ADR UK is a partnership transforming the way researchers access the UK's wealth of public sector data, to enable better informed policy decisions that improve people's lives. ADR UK is an Economic and Social Research Council (ESRC) investment (part of UK Research and Innovation).

Contact

Name: Prof Peter Urwin

Email: urwinp@westminster.ac.uk

