



# Methodology for the Youth Jobs Gap

April 2019

CENTRE FOR  
VOCATIONAL  
EDUCATION  
RESEARCH



National  
Institute of  
Economic and  
Social Research

**I**  
Impetus

**Impetus** transforms the lives of young people from disadvantaged backgrounds by ensuring they get the right support to succeed in school, in work and in life. We find, fund and build the most promising charities working with these young people, providing core funding and working shoulder-to-shoulder with their leaders to help them become stronger organisations. In partnership with other funders we help our charities expand and we work to influence policy and decision makers so that young people get the support they need.

**The National Institute of Economic and Social Research (NIESR)** is Britain's longest established independent research institute, founded in 1938. Our mission is to carry out research into the economic and social forces that affect people's lives and to improve the understanding of those forces and the ways in which policy can bring about change. The Institute is independent of all party political interests and is not affiliated to any single university, although our staff regularly undertake projects in collaboration with leading academic institutions.

Our work with Impetus is part of NIESR's ongoing research in the Centre for Vocational Education Research (CVER). CVER was launched in March 2015, funded by the Department for Education, to create a research institution that will advance our understanding of the requirements for vocational education in the UK today, identify the challenges in provision of vocational education, and develop and strengthen the knowledge-base to enable a more agile, relevant and needs-based vocational education sector to become a driving force for economic growth and social mobility, as it is in other countries.

# Contents

**Validity of the methodology.....4**  
**Discussion of methodological issues.....5**  
**References.....10**

The technical details in this document are important to fully understand our Youth Jobs Gap briefing data. This document will be kept updated as we publish more briefings and develop our methodology. This version is published alongside the first briefing: *Establishing the Employment Gap*.

While the main briefings focus on our findings, there are some caveats and limitations. We would welcome challenge and feedback from any reader who thinks any of these issues are underplayed, or indeed suggestions for things we might have missed. We further hope that explaining them will assist other organisations to plan their own research projects involving LEO going forward. Please direct any enquiries to [info@impetus.org.uk](mailto:info@impetus.org.uk).

There is a summary of the methodology on the inside front cover of the briefing, which has sufficient detail to understand and interpret the briefing.

Given we are using a new dataset (LEO) and methodology, readers will want reassurance that the methods used are valid and sensible, and as such that the findings in the main briefing are valid. We cover this in the section on validity of the methodology.

There are a number of technical caveats that must be stated, methodological decisions we’ve made that are worth explaining, and indeed things we’ve learned about LEO which may be of interest to more technically-minded readers. These are included in the *discussion of methodological issues* on page 5.

---

# Validity of the methodology

LEO is a new dataset, and as such this methodology is a new way to measure rates of people not in education, employment and training (“NEET rates”). While the methodology has strengths and weaknesses, we are able to compare the findings to existing NEET figures calculated from the Labour Force Survey (LFS). The figures are comparable, which provides confidence that the findings are meaningful.

For example, in March 2017, we can measure the NEET rate across five cohorts aged 20–24, and we find 13% of young people are NEET. This compares to 12.7% of 19–24 year olds being NEET in the official statistics for that period.<sup>1</sup>

This difference is not unexpected, given methodological differences. The official NEET estimate is provided with a confidence interval of  $\pm 1\%$  to account for sampling variability, but this doesn’t account for potential reasons for systemic undercounting of NEET young people. For example, the LFS underpinning the official NEET figures has around a 50% non-response rate.<sup>2</sup> The assumption is that those who do respond to LFS are sufficiently representative of the overall population, and for all practical purposes this suffices, but this difference could easily account for the slightly higher figures calculated from the LEO data, which covers all young people.

That is not to say that the LEO data is perfect or definitive. This is a new approach and certainly the NEET figures are an overestimate. For example, LEO has no ability to distinguish between people who are NEET because they are actually NEET in the way commonly meant, and those who are on gap years or who have sadly died. The following section discusses methodological issues in detail.

As well as being close to the published NEET rates as a percentage of the population in related age groups, the NEET figures calculated from LEO show a similar cycle over time and within a year, i.e. a downward trend over the time period in question, and a clear seasonal peak for the September measure. This provides added reassurance that the methodology used calculating NEET figures from LEO is measuring the same underlying phenomenon.

Overall, it seems reasonable to conclude that while the NEET figures calculated using LEO are still only estimates, the limitations are no more significant than those of the LFS methodology, and certainly minor compared to the new insights this larger dataset brings. However, as the figures are covering the whole population rather than a survey population, they allow to derive much more granular NEET rates than previous studies, especially for small local areas. This allows in particular to obtain data for local stakeholders, which can be used to understand key groups in need of support and their particular labour market conditions.

---

# Discussion of methodological issues

## Technical details about the Longitudinal Education Outcomes dataset (“LEO”)

0.1 LEO is often referred to as a dataset, but in fact it is simply about linking existing administrative datasets. The DfE has published a short summary<sup>3</sup> of which the relevant points are:

0.2 LEO combines data held by:

The Department for Education (DfE)

The Department for Work and Pensions (DWP)

Her Majesty’s Revenue and Customs (HMRC)

0.3 This data includes:

- personal characteristics including gender and ethnic group
- if the young person had special education needs during their education
- household income
- if the young person qualified for free school meals
- schools, colleges and universities attended, courses taken, and qualifications achieved
- any benefits claimed
- any careers advice and training offered by the government
- employment and income

0.4 This data is then used to compare students’ levels of education to their levels of employment and earnings in later life. To do this, LEO links personal information relating to their education, employment and benefit claims in order to:

- enable comparisons of the performances of schools, colleges and universities
- provide statistical information to support education and career decisions
- evaluate and monitor the impact of education or training on outcomes
- support government decision making which improve services

## Young people in scope

1.1 We only include those young people reaching school leaving in academic years 2006–07 to 2011–12. These are six cohorts of young people leaving secondary schools in the particular academic year (e.g. 2007). These cohorts are aged 18–24 in recent years for which LEO data is available.

1.2 When choosing which cohorts to study, there is a trade-off. Earlier cohorts are now older, enabling you to look at outcomes further after school. However, by definition they also left school earlier, meaning the lessons for the current school system may be less clear. Choosing the 2007–12 cohorts enables us to look at the 18–24 age group with the most recently available data.

1.3 We only include data for those sitting GCSEs in mainstream schools in England e.g. excluding those in private or special schools, and those educated in Wales or Scotland.

1.4 The total number of pupils (rounded to the nearest 100) in scope for this research, and a comparison to the published GCSE results for state-funded mainstream schools<sup>4</sup>, is as follows:

| Cohort | In scope (LEO) | GCSE results |
|--------|----------------|--------------|
| 2007   | 601,300        | 600,700      |
| 2008   | 600,200        | 598,100      |
| 2009   | 577,800        | 578,800      |
| 2010   | 578,400        | 578,100      |
| 2011   | 566,800        | 566,900      |
| 2012   | 561,300        | 561,300      |

Disadvantage

2.1 All young people in scope are defined as either being disadvantaged, if eligible for free school meals (FSM) at some point in year 11; or not disadvantaged, if they were known not to be eligible. A small number of students where this status is “unknown” are removed.

2.2 As pupil premium wasn’t a policy when the earlier cohorts were at school, FSM is the only marker of disadvantage consistent across all the cohorts.

Geography

3.1 All young people in scope are defined as belonging to the local authority and region in which they went to school in year 11. This may be different to where they actually lived, and will be different in many cases to where they subsequently worked.

3.2 During the 2007–12 period, there were boundary changes affecting Bedford and Central Bedfordshire local authorities. To maintain consistency, we have chosen to combine these local authorities in our analysis into a single unit. We treat Cheshire East, Cheshire West and Chester as a single unit for the same reason.

3.2 We have excluded the City of London and Isles of Scilly local authorities, as they are too small for any meaningful analysis.

Qualification

4.1 Young people are grouped based on their highest qualification two years after completing school, i.e. academic age 18. There are three groups:

- High qualified – highest qualification A-levels or equivalent
- Middle qualified – highest qualification five GCSEs at grade A\*–C or equivalent
- Low qualified – fewer than 5 GCSE passes

The proportion of young people in each group changes over time (the absolute number also changes, but not in the same way, reflecting the decreasing size of cohorts in paragraph 1.4). For example, we see growth in proportion young people with high level of qualification, in line with published official statistics on 16 to 19 attainment.

4.2 These qualification groups are necessarily broad, and this means they cover a range of different qualifications. This may help to explain some of the differences between different subgroups, including between disadvantaged and non-disadvantaged. For example, we know that within the group of young people with high level of qualification, disadvantaged young people are less likely to subsequently go on to university than their better off peers. This means that a comparison of outcomes between highly qualified disadvantaged young people and their better off peers is inevitably skewed slightly by the fact the two groups are not perfectly alike. A similar comment applies to different ratios of academic and vocational qualifications, as e.g. vocational qualifications are more often taken by people from disadvantaged backgrounds, and people with middle and low level of qualification. This will be refined in future analysis.

**NEET status**

5.1 We define a young person as either NEET or EET at a point in time, based on whether LEO records any education or employment for the young person in question in the three months up to and including that point in time (EET) or not (NEET).

5.2 Our starting assumption, and indeed our main motivation, is to better understand differences in NEET rates with a view to helping young people move from NEET to EET. For some technically NEET young people, this may not actually be a relevant consideration, for example those caring for their new-born children or on gap years.

5.3 We measure NEET rates four times a year (December, September, March, June) to mirror the published NEET figures. This starts from two years after leaving school (December 2009 for the 2007 cohort), through to the spring of 2017. We then analyse NEET rates by characteristics such as whether young people come from a disadvantaged background, by levels of qualification, and regions. Different cohorts are included at different times, based on whether they are 18-24.

**Ages**

6.1 When considering a particular cohort a specific number of years after they leave school, we refer to the age of the people in that cohort for convenience. For example, the 2007 cohort generally are 15 at the start of that year (2006-07), and therefore are 19 at the start of the 2010-11 academic year. Using this definition, all young people of the same age finished compulsory education at the same time. However, throughout the year an increasing proportion have birthdays and so will be a year older. This is illustrated in the following table, for the 2007 cohort:

| Date     | Actual age of 2007 cohort                 | Age (our methodology) |
|----------|---|-----------------------|
| Sep 2009 | Almost all 18, a few turn 19 in the month | 18                    |
| Dec 2009 | Mostly 18, a sizable minority are 19      | 18                    |
| Mar 2010 | Around half are 18, half are 19           | 18                    |
| Jun 2010 | Mostly 19, though a few are still 18      | 18                    |
| Sep 2010 | Almost all 19, a few turn 20 in the month | 19                    |

---

6.2 This concept is known as academic age. Because age is pegged to academic years and the age reflects the age of the cohort at the start of the year. As the year goes on, more and more of the cohort will be one year older than their academic age. We have not undertaken an exercise to combine people of the same real age from two different cohorts, and all references to particular ages refer to the academic age and by extension a cohort. It is most sensible to consider young people's post-18 outcomes by comparing people who left school at the same time, even if their ages differ slightly; as opposed to combining people based on their actual age, which would lead to different underlying groups every three months.

## **NEET vs EET**

7.1 Our definition of NEET, no observation of earnings or learning activity in three consecutive months, is a relatively low threshold as minimal earnings would be sufficient to count as EET. This cut-off is set to mirror the existing NEET definition, which cannot be replicated exactly because it is based on answering a survey question. We believe the numbers of young people regularly working very low numbers of hours is small. Anyone who is not NEET is EET – they are mutually exclusive and exhaustive.

7.2 For the purposes of this analysis we broadly speaking present our findings with the assumption that EET is a good thing and NEET is a bad thing. This is of course simplistic. On the EET side, there are young people who are unhappy with their hours, pay, or prospects. On the NEET side will be people opting out of the labour market because of a gap year or because they care for relatives or children. Both these things are impossible to measure within the existing LEO data and are beyond the scope of this project.

7.3 All traineeships, paid or unpaid, count as EET as they are recorded separately as learning activity.

## **Point-in-time vs longitudinal analysis**

8.1 Each NEET rate (or EET rate) calculated is a measure simply at a point in time. Comparing one to the next, the analysis does not imply how much movement there is between the two groups.

## **Raising the participation age**

9.1 Participation in EET activity became compulsory for the 16–18 age group during the period under investigation. As a consequence, we exclude the 16–18 age group from the analysis presented in the briefing and look only at outcomes from 18–24.

## **Suppression due to the interaction of multiple variables**

10.1 It is not always possible to provide descriptions for very specific groups, e.g. to describe NEET rates by disadvantage and qualification within a particular local authority. This results from the size of the group described being below a minimum number of people required not to be able to identify individuals – in accordance with Government regulation, figures below certain thresholds are “supressed” as in official publications. The suppression threshold applied in preparation of this briefing is either 10 and below or 5 and below, depending on which data is being used.

---

10.2 Suppression does not affect data uniformly. Suppression is a particular significant in groups that were smaller to start with, especially small local authorities and disadvantaged young people. For example, in the small local authority of Rutland in 2012, there were only 28 GCSE pupils eligible for FSM, according to published data. Unsurprisingly, when described separately within groups of high, medium and low qualifications, and how many of that group are NEET at any given time, the number is below 10, and thus cannot be published or used in calculation. As a consequence, it is not possible to provide a complete analysis of all subgroups at the local authority level.

10.3 If only one data point is suppressed, we suppress at least one other component cell (the next smallest) to avoid calculation of suppressed values from totals. This is known as “secondary suppression”.

### Other methodological points

11.1 Underlying match rate: Over 95% of young people are matched in LEO, and so this analysis starts only with them. We effectively assume they are representative of the entire population. In reality this is not quite true, but as can be seen from paragraph 4.1 we have a sufficiently large number of young people that it is unlikely to make a difference.

11.2 Self-employment data: At the time of beginning this analysis, self-employment data was not included in the dataset, so self-employment will not be taken into account in deciding who meets the definition of EET in this analysis. Self-employment data is now available through LEO and will be included in future analysis.

11.3 “DIVE group” NEETs. Our definition of NEET includes any young person who cannot be found earning or learning at that point in time. This will include a limited number of NEETs for whom the starting assumption about moving from NEET to EET is either not applicable or particularly challenging. This includes, for example, those who have sadly died; those who have been imprisoned; those who are volunteering (including unpaid internships and unpaid gap years); and those who have emigrated or are working abroad.

11.4 A number of technical changes as LEO has developed will improve the accuracy of data overtime. For example, the introduction of real time information (RTI) means data related to employment should be more accurate from 2014 onwards; the removal of invalid edukeys improves accuracy further from 2016 onwards.

---

# References

1. Department for Education, [\*SFR 41/2017: NEET Statistics Quarterly Brief April to June 2017, England\*](#), August 2017
2. Office for National Statistics, [\*Labour force survey: Performance and quality monitoring report, April to June 2017\*](#)
3. Department for Education, [\*Longitudinal Education Outcomes study: how we use and share personal data\*](#), December 2017
4. Department for Education, [\*GCSE and equivalent results: 2011 to 2012 \(revised\)\*](#), January 2013





All young people can  
succeed at school and work  
with the right support

**Impetus**  
**Floor 4**  
**Evergreen House North**  
**Grafton Place**  
**London NW1 2DX**

**Impetus.org.uk**  
**@ImpetusPEF**  
**info@impetus.org.uk**  
**0203 474 1000**