Understanding the free school meals earnings gap: methodology

Methodological details for the derivation of unadjusted and adjusted free school meals earnings gaps and Blinder-Oaxaca decomposition.

Contact:
James Tierney
integrated.data.analysis@ons.gov.uk
+44 1633 456314

Release date:
4 August 2022

Next release:
To be announced

Table of contents

1 . Data sources

The [free school meals (FSM) earnings gap analysis](#) uses the Longitudinal Educational Outcomes (LEO) database constructed by the Department for Education (DfE). It links administrative education data from early years through to higher education from the DfE and the Higher Education Statistics Agency (HESA) with employment, benefits and earnings data from the Department for Work and Pensions (DWP) and Her Majesty's Revenue and Customs (HMRC). As of 2021, the LEO database comprised 38 million individuals who had attended education in England.

However, there are some data limitations. To be included in the LEO database with school-level education data and outcomes, individuals must have been in the English school education system during at least one academic year. Additionally, the years for which qualification and labour market data exist do not overlap exactly. Qualification data are available from each stage of education for learners who were registered for key stage 4 assessments from the 2001 to 2002 to the 2018 to 2019 academic years, and labour market outcome data are available from the 2003 to 2004 to the 2018 to 2019 tax years. There are also specific exclusions to some types of data. HMRC pay as you earn (PAYE) records do not contain earnings information for "cash in hand" payments, and the school census does not include students who were home schooled.

As outcomes in this publication focus on early adulthood, we studied individuals who turned 30 years old in the tax years ending 2017 to 2019.

This analysis uses the LEO data asset available to researchers in the Secure Research Service (SRS), also referred to as the "standard release". Further information about the LEO database is available from the [DfE](#) or the [SRS](#).

2 . Statistical methods

Raw earnings gaps

Raw earnings gaps are estimated by calculating the difference in the geometric mean of annual earnings between those eligible for free school meals (FSM) and those either in non-FSM eligible (state educated) or independent school groups, divided by the geometric mean of FSM group annual earnings as follows:

$$\left(\frac{\left(\prod_{i=1}^{n} y_i \right)^{\frac{1}{n}} - \left(\prod_{j=1}^{m} y_j \right)^{\frac{1}{m}}}{\left(\prod_{j=1}^{m} y_j \right)^{\frac{1}{m}}} \right) \times 100$$

Where:

y_i is the annual earnings of individual i at age 30 (non-FSM or independent)

y_j is the annual earnings of individual j at age 30 (FSM)

The geometric mean is used to ensure raw and adjusted earnings gaps are directly comparable.

This methodology differs from previous Office for National Statistics (ONS) analysis of pay gaps and relates to annual earnings rather than rates of pay. Readers are discouraged from directly comparing outputs.

Adjusted earnings gaps

The adjusted earnings gap is estimated using ordinary least squares (OLS) regression. This is a form of linear regression, which is a technique that models the relationship between a dependent variable and explanatory variables. The ONS has used this method to estimate [gender](#), [ethnicity](#), and [disability](#) pay gaps. It is used to calculate the earnings gap between free school meals recipients and those who did not receive free school meals (or those that attended independent schools).

The explanatory variables are:

- free school meal status or independent school attendance during key stage 4 (KS4)

- attainment - this includes highest level of qualification by age 30, KS4 total point score, English GCSE (A* to C) attainment dummy and Maths GCSE (A* to C) attainment dummy

- years of labour market experience

- region

- ethnicity

Separate models are run for males and females.

The dependent variable is the natural logarithm of pay as you earn (PAYE) earnings at age 30. Applying a log transformation helps meet the assumptions required for using OLS techniques.

The model can be expressed as:

$$ln(earnings) = \alpha + \beta_1 nonFSM + \beta_2 Independent + \delta X + u$$

Where X is the vector of explanatory variables and u is the error term. Squared terms for labour market experience variables are also included.

As the dependent variable (earnings) is log transformed, adjusted earnings gaps are measured as follows:

$$\left(e^\beta - 1\right) \times 100$$

This estimates the percentage difference in the geometric mean of earnings relative to the earnings of the FSM group after controlling for differences in education, labour market experience, region, and ethnicity.

Some predictor variables are not included in the model as they are not available in the data, for example, occupation.

Explaining the earnings gaps through decomposition

A Blinder-Oaxaca decomposition is used to divide the difference in average pay between two groups into two separate elements: one part that can be explained by the differences in observed characteristics of the two groups and a second part that quantifies the unexplained or unobserved element.

Since this technique is only able to compare differences between two groups, non-FSM (state educated) and independent school group members are grouped together into a single group of non-recipients. This differs from earlier sections of the analysis, where independent school group members were considered separately.

The following equations illustrate the two-fold Blinder–Oaxaca decomposition.

Firstly, estimate separate FSM (fsm) and non-FSM (non) OLS wage regressions for individual i. Where Y is the natural log of annual PAYE earnings, X is a vector of explanatory variables and u is the associated error term.

$$(1)\ Y_{fsm} = \beta_{fsm} X_{fsm} + u_{fsm}$$

$$(2)\ Y_{non} = \beta_{non} X_{non} + u_{non}$$

Let bfsm and bnon be, respectively, the OLS estimates of fsm and non and denote mean values using a bar over each variable. We follow the Blinder-Oaxaca methodology, as outlined in A Stata implementation of the Blinder-Oaxaca decomposition (PDF, 335 KB) by Jann in 2004, using a pooled regression. Since OLS estimates produce error terms with an expected value of zero, we have:

$$(3)\ \overline{Y}_{non} - \overline{Y}_{fsm} = b_{pooled}\left(\overline{X}_{non} - \overline{X}_{fsm}\right) + \overline{X}_{non}\left(b_{non} - b_{pooled}\right) + \overline{X}_{fsm}\left(b_{pooled} - b_{fsm}\right)$$

The first term on the right-hand side of equation (3) denotes the impact of differences in endowments (for example, education and experience) between FSM and non-FSM groups on earnings. The other terms refer to the unexplained component, often attributed to discrimination, but also reflects the potential impact of other unobserved variables on earnings.

3 . Concepts used in the analysis

Measuring disadvantage during childhood

As details of family income and social background are not yet available in the Longitudinal Education Outcomes (LEO) data, eligibility for free school meals (FSM) is used as a proxy for individual-level socio-economic disadvantage during childhood. This is a common approach for identifying whether an individual grew up in a household with a low income. Pupils are classified as FSM eligible according to their status during the academic year they started at aged 15 years.

Independent school attendees are not eligible for FSM. As there is no household income proxy for this group, they are considered separately for the raw and adjusted earnings gap analyses. The decomposition combines those who did not receive FSM and independent school attendees into a single group because of the methods used.

Measuring earnings

This analysis looks at annual earnings from employment in tax years where an individual was not also in education.Earnings from employment are calculated using pay as you earn (PAYE) records. When someone has had multiple jobs in a tax year, the sum of the earnings from those jobs is used. Only earnings greater than £0 are considered across this analysis.

Earnings from employment are distinct from full-time equivalent salary so will be affected by reduced participation in the labour market for reasons such as caring responsibilities or part-time study. Full-time equivalent salary currently cannot be derived using the LEO database because of a lack of data on hours worked.

Earnings have been adjusted for inflation using the consumer price index with housing (CPIH), with April 2019 as the base. Average age-earnings are then calculated using the mean. The mean is used as it is required for the decomposition methods used; median earnings follow a similar trend.

Determining highest level of qualification

An individual's highest level of qualification is derived using the regulated qualifications framework (RQF) and its predecessor, the qualifications and credits framework (QCF). The RQF has nine levels from entry level through to level 8, with level 8 the highest (doctorate or equivalent). This framework shows how different qualifications relate to each other with qualifications at the same level being similarly demanding but often different in content, duration, and assessment methods. For example, RQF level 3 includes A levels and BTEC level 3 national qualifications.

This analysis considers registered qualifications attained by age 30 years and recorded in the data at school level, as part of a course of further education in England, or at a UK higher education institution.

Measuring labour market experience

A range of proxies are used in the literature to estimate the relationship between experience and labour market outcomes. For example, analysis of the Annual Survey of Hours and Earnings (ASHE) uses age while other sources calculate it based on years of employment.

As we do not have data on hours worked, we sum the number of years between the ages of 18 and 29 years in which an individual has positive PAYE earnings.

We differentiate between years in which the individual earned above the annual full-time equivalent (37.5 hours per week) national minimum wage (NMW) and years in which they earned below the NMW. This acts as a proxy for whether an individual was likely to have been employed full time. Part time workers renumerated at a rate of pay higher than the NMW may also meet this definition.

Coverage and context

This publication looks at young people who meet all the following criteria:

- were in the English education system at key stage 4 (KS4), including both state and independent schools

- turned 30 years old between the start of the 2016 to 2017 tax year and the end of the 2018 to 2019 tax year

- had positive earnings from employment in the tax year when they turned 30 years old and were not in education

- had a unique pupil matching reference (PMR) number

After meeting these criteria, observations required complete data for all variables used.

Users should be aware that this sample is unlikely to have data missing completely at random (MCAR) or missing not at random (MNAR), and this may affect the accuracy of our estimates. The Department for Education (DfE) reports further details of match rates for the LEO database overall.

In total, 1,195,029 individuals are included in the analysis.

Readers should also note that estimated earnings gaps are likely subject to selection bias as we only observe earnings for people who worked.

4 . Glossary

Earnings

Earnings from employment in the Longitudinal Education Outcomes (LEO) dataset come from HM Revenue and Customs' (HMRC) pay as you earn (PAYE) system. This value is equal to total annual earnings. It is reflective of both the rate of pay and the number of hours and/or days worked. Earnings from self-employment are not included.

Ethnicity

Ethnicity is the most recently recorded group drawn from education tables in the LEO database and uses the categories from the 2011 Census. You can find more information about the categories used on GOV.UK.

Free school meals

Free school meals (FSM) are a statutory benefit available to school-aged children whose families receive other qualifying means-tested benefits from the Department for Work and Pensions (DWP). Families need to register for them and not all families entitled to FSM go on to claim them. School census data record only where learners are both eligible and register a claim. In England, universal FSM are provided for all infants in primary school.

FSM is a commonly used proxy measure for socio-economic disadvantage (including household income deprivation) during childhood. We use FSM eligibility in the academic year an individual started at aged 15 years.

Gender

"Male" or "female" as reported in the learner's key stage 4 (KS4) record.

Independent school

These are registered schools that do not receive government funding. They often charge fees for pupils to attend. Independent school attendees are not eligible for free school meals, although bursaries and other financial support may be available.

Matched record

Matched records are those in the LEO database where an individual's education records have successfully been matched to their earnings or benefits outcomes records. Unmatched records are those that appear in the education data but could not be matched with earnings or benefits. Approximately 95% of records in the LEO database have been successfully matched.

The match rate varies by ethnicity, first language and geographical region. In some cases, a lack of match will be because of administrative reasons, for example, names being changed or spelt inconsistently across data sources. In other cases, it will reflect a learner's activities; they may have moved out of the UK before ever working, have died, or not engaged with formal employment or the benefits system at any point in their life.

5 . Future developments

The authors welcome feedback and the opportunity to learn from best practices.

This methodology is part of a wider program of work exploring education and outcomes for free school meal (FSM) recipients. The [first publication](#) looking at earnings outcomes by FSM status and demographic factors was released in January 2022 and will be followed by an analysis of education qualifications later in 2022.

We will also examine how earnings outcomes are affected by interactions between factors; for example, whether the same routes through education tend to have similar outcomes for young people who received free school meals (FSM) but are from different ethnic groups.

6 . Related links

[Education, social mobility and outcomes for students receiving free school meals in England: initial findings on earnings outcomes by demographic and regional factors](#)
Article | Released 25 January 2022
Analysis of earnings outcomes relative to the living wage for those aged 25 years who received free school meals. This article explores demographic and geographic factors.

[Gender pay gap in the UK: 2021](#)
Article | Released 26 October 2021
Differences in pay between women and men by age, region, full time and part time, and occupation.

[Apply to access the Longitudinal Education Outcomes (LEO) dataset](#)
Webpage | Updated 1 August 2022
Guidance on how to apply to access the LEO standard extract through the Office for National Statistics Secure Research Service.