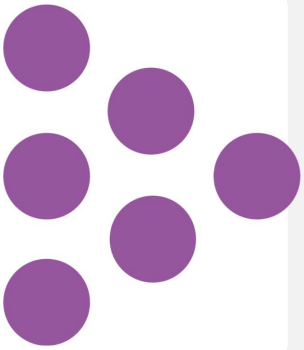
NFER
National Foundation for Educational Research

Technical Report

The Early Years Workforce in England Annual Report 2026

Methodology Appendix

National Foundation for Educational Research (NFER)



The Early Years Workforce in England Annual Report 2026

Methodology Appendix

Michael Scott, Juan Manuel Del Pozo Segura and Jack Worth

Published in February 2026

By the National Foundation for Educational Research,

The Mere, Upton Park, Slough, Berkshire SL1 2DQ

https://www.nfer.ac.uk/

© 2026 National Foundation for Educational Research

Registered Charity No. 313392

ISBN: 978-1-916567-51-1

How to cite this publication: Scott, M., Del Pozo Seguara, J.M. and Worth, J. (2026). The Early Years Staff Labour Market in England: Annual Report 2026 – Methodology Appendix. Slough: NFER.

NFER
National Foundation for Educational Research

Contents

1. Introduction 1
2. Data sources and the definition of early years workers 2
   2.1. Data sources 2
   2.2. Definition of early years workers 2
3. SCEYP: analysis of early years staff numbers in England, including by region and qualifications 4
4. ASHE: analysis of support staff pay over time and their position in the wider earnings distribution 5
5. LFS: analysis of early years workers' perceptions of their working hours, weekly pay and attitudes to work compared to those of similar workers 7
   5.1. Identifying early years workers and a suitable comparison group 7
   5.2. Variables used in the analysis 8
   5.3. The analytical approach to the LFS 8
   5.4. Sample sizes 9
6. APS – Analysis of early years workers' wellbeing and workplace perceptions 11
   6.1. Overview of the approach 11
   6.2. Variables used in the analysis 11
   6.3. Sample sizes 12

NFER
National Foundation for Educational Research

1. Introduction

This methodology appendix explains the data we used and the analysis we undertook to produce our findings in "The Early Years Workforce in England: Annual Report 2026".

Section 2 lists the data sources we used in the analysis, as well as how we defined early years workers in national surveys. The remaining sections then each cover methodological detail relevant to the analysis in the main report.

Section 3 explains how we used data and statistics from the Survey of Childcare and Early Years Providers to estimate the number of early childcare workers in England, the distribution of those workers, their turnover rates, and how numbers break down by qualification level.

Section 4 discusses how we used Annual Survey of Hours and Earnings data to estimate early years workers' pay, including how we defined early years workers in this data.

Section 5 shows how we used the Labour Force Survey to measure early years workers' working hours, weekly pay, and perceptions of their working hours. This section explains key variable and group definitions, provides sample sizes and outlines our methodology for defining a group of similar workers to compare early years workers too.

Section 6 outlines how we used the Annual Population Survey to measure early years workers' wellbeing and workplace perceptions.

This work was undertaken in the Office for National Statistics Secure Research Service using data from ONS and other owners and does not imply the endorsement of the ONS or other data owners.

2. Data sources and the definition of early years workers

2.1. Data sources

The following data sources were used to inform the analysis in this research report:

- The Survey of Childcare and Early Year Providers (SCEYP). More information here.
- Annual Survey of Hours and Earnings (ASHE). Available from Office for National Statistics (ONS). More information here.
- The Labour Force Survey (LFS). Available from UK Data Service. More information here.
- The Annual Population Survey (APS). More information here.

2.2. Definition of early years workers

The definition of early years workers in SCEYP data is set by DfE. More information can be found in the methodology of DfE's statistics release.

Commented [MS1]: Check this is true and/or amend.

ASHE, LFS and APS are all national surveys that include workers in a wide range occupations and industries. We therefore use the information in these surveys to identify early years workers for the purposes of our analysis.

We define early years workers using codes that classify a worker's occupation (SOC2020) and codes that classify a worker's industry (SIC2007). We only use 2020 data onwards because SOC categorisation changes make it difficult to identify a like-for-like group in years before 2020.

Firstly, we include everyone who works in any of the following occupations:

Table 1 Occupational groups that are classed as early years workers regardless of industry

SOC2020 Code	SOC2020 Group
2315	Nursery education teaching professionals
2324	Early education and childcare services managers
3232	Early education and childcare practitioners
6111	Early education and childcare assistants
6114	Childminders

Secondly, and in addition, we also include those who work in a wider list of occupational groups listed below **and** who work in the early years industry, as defined by SIC2007 codes. These occupations have been selected as they are likely to be child-facing roles within an early years setting. (We define these as 8891 Child day-care activities and 8510 Pre-primary education).

Table 2 Occupational groups that are classed as early years workers if the worker also works in the early years industry.

SOC2020 Code	SOC2020 Group
1172	Social services managers and directors
1232	Residential, day and domiciliary care managers and proprietors
2211	Generalist medical practitioners
2212	Specialist medical practitioners
2234	Registered nurse practitioners
2237	Other registered nursing professionals

SOC2020 Code	SOC2020 Group
2259	Other health professionals n.e.c.
2314	Primary education teaching professionals
2316	Special and additional needs education teaching professionals
2319	Teaching professionals n.e.c.
2321	Head teachers and principals
2322	Education managers
2469	Welfare professionals n.e.c.
3221	Youth and community workers
3231	Higher level teaching assistants
6112	Teaching assistants
6113	Educational support assistants
6117	Playworkers
6131	Nursing auxiliaries and assistants

Note: n.e.c. stands for 'not elsewhere classified'.

In some of the LFS and APS analysis, we also look at the differences between early years workers by level of highest qualification. SCEYP classifies people according to their highest early years or teaching qualification, whereas LFS and APS only provides the highest qualification of all types. However, the highest qualification of someone working in early years is likely to be an early years or teaching qualification, although it is a limitation of this approach that this may not necessarily be the case.

3. SCEYP: analysis of early years staff numbers in England, including by region and qualifications

We use statistics published in the SCEYP statistical release, available here.

Our analysis aims to describe the time series of early years staff numbers in the last five years, especially since 2023. However, there was a methodological change in 2025 that applied to the weighting of survey responses from group-based providers that affected the estimated number of staff. The new weighting was used for 2025, while data from the previous and new weighting was presented for 2024. No updated figures were provided for years prior to 2024.

To establish a continuous time series between the two periods, we combined the data to create an approximation. We used the data based on new weighting for 2024 and 2025. Using the previous and new weighting for 2024, we created an adjustment factor (i.e. how much the 2024 numbers had changed due to the weighting). We applied the adjustment factor to the pre-2024 data. This adjustment implicitly assumes that while the estimated overall staff numbers have changed due to the methodology change, the pre-2024 trends would have remained the same if they had been reweighted. This may not be the case if underlying trends in the variables used to create the weightings have themselves changed. However, we cannot test this as we do not have access to the underlying data for this analysis.

4. ASHE: analysis of support staff pay over time and their position in the wider earnings distribution

To analyse where the pay of early years workers sits within the earnings distribution in England and how it has changed over time, we used data from ASHE for 2021 to 2025. We only use 2021 data onwards because SOC categorisation changes make it difficult to identify a like-for-like group in years before 2021. We use 2025 because it is the latest data available. Since the ASHE is collected in April of each year, we re-aligned the data so that the 2021 ASHE represented the 2020/21 academic year while 2025 represented the 2024/25 academic year.

Our sample consisted of individuals in the ASHE working in one 'main job'. We included full-time and part-time workers and focused on **hourly** earnings because many early years workers work part-time. For those working in a 'main job' and an 'additional job', we discarded the 'additional job'. For those working in multiple part-time jobs or multiple full-time jobs, we discarded their records altogether. We also discarded anyone not on a permanent employment contract, anyone with missing earnings records or occupation / industry codes, anyone working a junior pay rate or who were on an apprenticeship and anyone whose earnings were affected by leave.[1]

The ASHE has some known limitations, such as non-coverage of those in self-employment, relatively high non-response rates and non-sampling bias (since the ASHE only samples jobs registered on a pay as you earn (PAYE) scheme). To minimise the impact of these limitations, we applied the ASHE calibration weight to our analysis. This helped to ensure that our estimates were weighted to be representative of the entire labour force in England, as per the Labour Force Survey (LFS). Finally, we identified early years workers in the sample using the definition outlined in Section 2. Since many are self-employed, our analysis is likely to systematically exclude childminders.

The annual sample sizes of individuals in our main analysis sample, as well as those who are early years workers, are provided in **Table 3**.

Using our full sample of all workers, for each year we estimated each percentile of the hourly earnings distribution (i.e. we estimated 100 percentiles so that each represented one per cent of the earnings distribution). We also estimated the median hourly earnings of the early years workers in the sample for each year (using the definitions outlined in Section 2). We then combined the two data series to find where the median earnings for early years workers is placed within the entire earnings distribution in each year.

[1] We included in the analysis those who were put on furlough during the Covid-19 pandemic but who were still paid at their full rate of pay. We excluded those who were put on furlough where it impacted their earnings. Due to small sample sizes in 2020 (when the collection of the ASHE was impacted by the pandemic), excluding all workers on furlough would have dramatically reduced sample sizes and likely also had implications for the occupational composition of the sample.

Table 3: Sample sizes for ASHE analysis

Year	Total number of workers in England in sample	Total number of early years workers in England in sample
2020	90,469	840
2021	96,195	853
2022	109,436	1,000
2023	121,437	1,035
2024	128,798	956
2025	121,394	840

Source: NFER analysis of ASHE data for 2020 to 2025.

5. LFS: analysis of early years workers' perceptions of their working hours, weekly pay and attitudes to work compared to those of similar workers

Our analysis of early years workers' working hours and perceptions of their working hours primarily uses data from the LFS. The LFS data enables us to measure how working hours, weekly pay and working hours perceptions have changed over time. It also enables us to compare how working hours and working hours perceptions compare to those in other occupations.

The analysis involved several key steps, including identifying early years workers and a suitable comparison group in the data, ensuring comparability in the two groups, and defining the key indicators for reporting.

5.1. Identifying early years workers and a suitable comparison group

Early years workers

We define these using SOC2020 and SIC2007 codes, as defined in Section 2.

Similar workers

For our comparison group, we included all other workers in the LFS in England who are not early years workers. We reweight this group so that it better resembles early years workers in terms of key characteristics: highest qualification level, age, gender, region and – in some instances – working pattern (i.e. whether the person is full-time or part-time).

We reweight the other workers group to improve comparability in these specific characteristics with the early years workers group. This ensures that the distribution of these characteristics is the same among the early years workers and the comparison group. We use a technique called entropy balancing to reweight the other workers group within each survey wave and derive a 'similar workers' group.[2] The original ONS weights are used a starting point for these adjustments.

It is important to note that this re-weighting approach does not remove all the underlying differences in characteristics and motivations between early years workers and similar workers. However, it minimises the risk that any observed differences in working conditions are driven by differences in the distribution of these key characteristics. Nonetheless, no comparison of different occupations should be interpreted as the effect of entering that occupation, although working conditions, and employees' perceptions of them, can be influenced by entering that occupation rather than another.

Looking at early years workers by sector and qualification level

Where we have broken the early years workforce group into subgroups based on the sector they work in (school-based or group-based) and their qualification level, we define a group of similar workers based on each subgroup, we have done that in much the same way as above. For the qualification sub-group analysis we do not need to include qualification level in the balancing.

[2] Hainmueller, J., & Xu, Y. (2013). ebalance: A Stata Package for Entropy Balancing. Journal of Statistical Software, 54(7), 1–18. https://doi.org/10.18637/jss.v054.i07

5.2. Variables used in the analysis

The variables in the LFS survey which we reported on are as follows:

Gross weekly pay

Average (mean) of a derived variable which calculates the gross weekly pay of the respondent in their main job, based on answers given to various questions about earnings. See LFS user guide for details.

We use LFS income weights for this variable. For all other variables listed here we used person weights.

Working hours in the reference week

Average (mean) response to a derived variable which calculated total hours worked across all jobs in the latest week.

We looked at the average hours for workers across all working patterns.

In addition, we also restricted analysis to those who worked a full-time schedule, identifying individuals who had worked at least five days during the reference week using responses to the question 'In the week ending Sunday the …, on which days were you scheduled to work?'. We also restricted this analysis to those who were not off work on those days despite being scheduled to work, using responses to the question 'In that week, did you have any days off work because you were sick or injured?'.

Proportion wanting to work fewer hours even with less pay

Average (mean) of a measure that is derived from a combination of responses and routed questions regarding whether someone is seeking to work fewer hours. This variable is created based on whether respondent is looking for a different or additional paid job, whether they would like to work fewer hours than in their current job, and/or whether they would accept less pay for shorter hours in a new job if they are seeking one. See LFS user guide for further details.

5.3. The analytical approach to the LFS

Obtaining an average

The LFS is a quarterly data collection with a longitudinal design. This means that respondents are sampled for one quarter and can then feature in up to four further quarters. For this project, we are primarily concerned with annualised estimates of the variables of interest. To obtain these estimates from quarterly data, we followed the suggested approach by ONS. For an approximation of an academic year, we look at four successive quarterly waves, starting with those in October to December, and finishing with those in July to September. This is the closest alignment with the academic year that can be achieved with the data. We then:

1. Calculate the mean of the variable of interest (see above) for early years workers and similar workers in each quarter, using the weightings obtained from entropy balancing for the latter group. We also calculate the standard error of that estimate.
2. Obtain a value for the academic year by calculating an 'average of averages' using the averages obtained from the four quarterly waves for that academic year. (Specifically, we sum the quarterly averages and divide by four).

Because respondents will feature in successive waves, estimates from each wave will be correlated. We specifically adjust our estimate of the standard errors for this using a variance formula. The variance formula for the annual estimate described above equals:

$$\mathrm{Var}\left(\frac{\hat{\theta}_4+\hat{\theta}_1+\hat{\theta}_2+\hat{\theta}_3}{4}\right)=\frac{1}{16}\left(\left(\sum_{j=1}^{4}\mathrm{Var}(\hat{\theta}_j)\right)+2\sum_{j<k,j,k=4}\rho_{jk}\sqrt{\mathrm{Var}(\hat{\theta}_j)\mathrm{Var}(\hat{\theta}_k)}\right)$$

where $\hat{\theta}_j; j=1,..,4$ are the quarter-specific estimates (averages) for a specific academic-year, $\mathrm{Var}(\hat{\theta}_j); j=1,..,4$ is the variance of each estimate and ρ_{jk} is the correlation coefficient between the estimate in the quarter j and k for a specific academic year.

Commented [MS2]: @Jack Worth - up to you if you want this level of detail.

Commented [JW3R2]: It's fine to include

This latter term means that, for example, the correlation between quarters 1 and 2, ρ_{12}, differs between the 2015/16 and 2016/17 academic years. We therefore simplify this formula by following Holmes and Skinner's[3] observations: for a given variable, we use only one representative correlation coefficient for any pair of quarters a given number of periods apart. Hence, regardless of the academic year, we use one correlation coefficient for estimates separated one period (i.e. ρ_{43}, ρ_{12}, ρ_{23} are the same across academic years), another for estimates separated two periods (i.e. ρ_{42}, ρ_{13} are the same across academic years) and another for those three periods apart (i.e. ρ_{41} are the same across academic years).

Estimation of these three coefficients is based on the autocorrelations of the quarterly estimates. In other words, the correlation coefficient for estimates separated by one period is given by the first autocorrelation coefficient across the estimates, and so on.

5.4. Sample sizes

The sample sizes in the LFS analysis are shown in **Table 4**. Sample sizes for each individual measure differ depending on the extent of missing data for each measure. The table is illustrative as it summarises the sample sizes from only one measure: **average working hours**. The sample sizes of both early years workers and similar workers have generally been falling slightly over time, which is due to falling response rates to the LFS across the whole population, although there has been a partial recovery in the later years. The sample sizes for questions about earnings is smaller because not all respondents are asked these questions.

Commented [JD4]: There is an uptick from 2023

Where we compare our key measures over time and between early years workers and similar workers, we test whether any differences were statistically significant by conducting a t-test that the difference was statistically significantly different from zero (at the five per cent level).

[3] Holmes, DJ and Skinner, CJ (2000) Variance estimation for Labour Force estimates of Level and change. *GSS Methodology Series*, No 21. https://webarchive.nationalarchives.gov.uk/ukgwa/20160105160709/https://www.ons.gov.uk/ons/guide-method/method-quality/specific/gss-methodology-series/gss-methodology-series--21--variation-estimation-for-labour-force-survey-estimates-of-level-and-change.pdf. Last retrieved: January 2026

Table 4: Sample sizes for LFS analysis

Year	Number of early years workers	Number of similar workers
2021/22	985	98,751
2022/23	688	69,910
2023/24	696	69,730
2024/25	1094	101,916

Source: NFER analysis of Labour Force Survey data for 2020/21 to 2024/25.

Commented [MS5]: Needs completing

Commented [JW6R5]: @Juan Manuel Del Pozo Segura please could you add, based on the whole sample for working hours, including FT and PT and not the full schedule (consistent with the text above)

Commented [JD7R5]: Done

Commented [MS8]: @Michael Scott - check these.

Commented [MS9R8]: @Juan Manuel Del Pozo Segura and @Jack Worth - I think we need to discuss.

Reason 1 - I don't get the same number JM's got here. I'm looking at the main project outputs. The Ns there, particularly for 'weighted and balanced', are a bit lower than here. I'm looking at the '5 var ALL' variant.

Reason 2 - I'm concerned that by using 'Total Hours' here we're mispresenting the sample size for other variables. Most notably, 'average weekly pay' has a sample size that is around half of this.

Commented [JW10R8]: OK, good point on pay, because it only uses some waves (1 and 5, is it?). But using an indicative example is fine rather than going back to the data and trying to get an underlying N, which then doesn't apply to any measures because of missingness

6. APS – Analysis of early years workers' wellbeing and workplace perceptions

6.1. Overview of the approach

The APS is comprised of aggregated data from four quarters of LFS data plus a boost sample. As the APS includes additional measures that are not available in the LFS, such as wellbeing and workplace perceptions, we also analyse data from the APS to look at how early years workers compare to similar workers.

We take the same approach to this analysis as described above, with the following exceptions:

- Instead of pooling quarterly data into academic years, we simply use the APS in calendar years. This means each wave of data only contains one observation from each individual and our variance estimates do not need adjusting as in the LFS.
- We only had APS data to the end of 2024, rather than the end of the 2024/25 academic year.

6.2. Variables used in the analysis

The variables in the APS survey that we reported on are as follows:

Proportion who agree they have opportunities for career progression

Average (mean) based on a 5-category Likert scale variable (Strongly disagree, Disagree, Neither disagree nor agree, Agree and Strongly agree) that reports workers' agreement with the following statement: 'My job offers good opportunities for career progression'. We treat 'Agree' and 'Strongly Agree' as 1 and any other response as 0.

Proportion rating managers as 'good' or 'very good' at involving employees and their representatives in workplace decision making

Average (mean) of a 5-category Likert scale variable (Very poor, Poor, Neither good nor poor, Good, Very good) that reports workers' opinion on how poor or good their managers are at involving employees and their representatives in decision making. We treat 'good' and 'very good' as 1 and any other response as 0.

Subjective wellbeing (0-10 scale)

Average (mean) of a 10-point scale variable that reports workers' response to the following questions:

- Life satisfaction: 'Overall, how satisfied are you with your life nowadays where 0 is 'not at all satisfied' and 10 is 'completely satisfied'?'
- Feelings things done in life are worthwhile: 'Overall, to what extent do you feel that the things you do in your life are worthwhile where 0 is 'not at all worthwhile' and 10 is 'completely worthwhile'?'
- Happiness: 'Overall, how happy did you feel yesterday where 0 is 'not at all happy' and 10 is 'completely happy'?'
- Anxiety: 'On a scale where 0 is 'not at all anxious' and 10 is 'completely anxious', overall, how anxious did you feel yesterday?'

You can find out more about these questions here.

Commented [MS11]: @Juan Manuel Del Pozo Segura - Don't we turn this into a binary variable? I thought we did.

Agree/Strongly Agree = 1
Else = 0

Same with the next one?

Commented [JS12R11]: You are right: this applies for both variables

Commented [JD13]: @Michael Scott
Technically. There is no variable called "Average wellbeing". I understand you refer to "How satisfied you are with your life"
This is what I am describing here

Commented [MS14R13]: Thanks JM - have fixed.

6.3. Sample sizes

The sample sizes in the APS analysis are shown in **Table 5**. Sample sizes for each individual measure differ depending on the extent of missing data for each measure. The table is illustrative as it summarises the sample sizes from only one measure: **the proportion who agree they have opportunities for career progression**. Sample sizes in the APS appear smaller than the LFS data because each APS wave strictly contains one response per person, whereas in the LFS data the same person could be featured up to four times – see LFS section for more details.

Table 4: Sample sizes for APS analysis

Year	Number of early years workers	Number of similar workers
2021	286	32,395
2022	237	27,037
2023	172	20,155
2024	210	23,440

Source: NFER analysis of Annual Population Survey data for 2021 to 2024.

Commented [JW15]: @Juan Manuel Del Pozo Segura and this one please

Commented [JD16R15]: Check the MS Teams message I sent on Wednesday 18/02/2026

Commented [JW17R15]: Sorry, can't find this message, but as above, just use indicative Ns from one of the measures that is typical.

NFER
National Foundation for Educational Research

Evidence for excellence in education

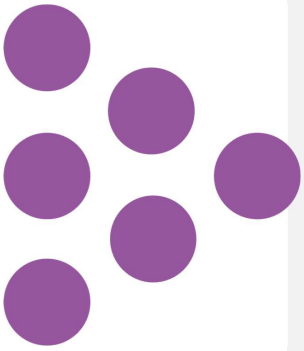
Public

© National Foundation for Educational Research 2026

All rights reserved. No part of this document may be reproduced or transmitted in any form or by any means, electronic, mechanical, photocopying, or otherwise, without prior written permission of NFER.

The Mere, Upton Park, Slough, Berks SL1 2DQ
T: +44 (0)1753 574123 • F: +44 (0)1753 691632 • enquiries@nfer.ac.uk

www.nfer.ac.uk

NFER ref. NUWC

ISBN. 978-1-916567-51-1