



Education
Endowment
Foundation

EDUCATION
POLICY
INSTITUTE



Growing apart

The evolution of the
disadvantage gap

Emily Hunt, Maxime Perrott, Dr Eva
Jiménez and David Robinson

June 2026

About the Education Policy Institute

The Education Policy Institute (EPI) is an independent, impartial, and evidence-based research institute that promotes high quality education outcomes, regardless of social background. We achieve this through data-led analysis, innovative research and high-profile events.

About the Education Endowment Foundation

The Education Endowment Foundation (EEF) is an independent charity dedicated to breaking the link between family income and education achievement. We support schools, colleges, and early years settings to improve teaching and learning for 2 to 19-year-olds through better use of evidence. We do this by:

- **Summarising evidence.** Reviewing the best available evidence on teaching and learning and presenting in an accessible way.
- **Finding new evidence.** Funding independent evaluations of programmes and approaches that aim to raise the attainment of children and young people from socio-economically disadvantaged backgrounds. Putting evidence to use.
- **Putting evidence to use.** Supporting education practitioners, as well as policymakers and other organisations, to use evidence in ways that improve teaching and learning.

We were set-up in 2011 by the Sutton Trust partnership with Impetus with a founding £125m grant from the Department for Education. In 2022, we were reendowed with an additional £137m from government, allowing us to continue our work until at least 2032.

About the data used for this research

This publication includes analysis of the National Pupil Database (NPD):
<https://www.gov.uk/government/collections/national-pupil-database>

The Department for Education (DfE) is responsible for the collation and management of the NPD and is the Data Controller of NPD data. Any inferences or conclusions derived from the NPD in this publication are the responsibility of EPI and not the DfE.

This work was undertaken in the Office for National Statistics (ONS) Secure Research Service using data from ONS and other owners and does not imply the endorsement of the ONS or other data owners. This work uses research datasets which may not exactly reproduce National Statistics aggregates.

About the authors

Emily Hunt, Associate Director for Social Mobility and Vulnerable Learners at the Education Policy Institute (EPI). Emily has published widely since joining EPI in 2018 and recent publications include EPI's 2025 Annual Report, *Breaking Down the Disadvantage Gap*, and *Local Disadvantage Gaps in England*. Prior to EPI, Emily worked at the Department for Education (DfE) as an economist for 14 years.

Maxime Perrott is a researcher at EPI and is nearing completion of a PhD in Advanced Quantitative Methods at the University of Bristol, with research focusing on how month of birth affects early childhood outcomes and socio-emotional development. Prior to joining EPI, Maxime has worked as a graduate teacher at the University of Bristol and as a research assistant at both the University of Bristol and the Education Endowment Foundation.

Dr Eva Jiménez is a former researcher at EPI, holding a PhD in psychology from the University of Warwick. Eva's previous publications include EPI's 2025 Annual Report and *Breaking Down the Disadvantage Gap*. Prior to joining EPI, Eva worked as a postdoctoral researcher in Warwick's Institute of Advanced Study as part of her Early Career Fellowship.

David Robinson, Director of Post-16 and Skills at EPI. David joined EPI in June 2017. David's background includes six years at the DfE as lead analyst first on school and college accountability and then on capital funding. Since joining EPI, David has authored a broad range of reports on post-16 education in England.



Contents

Executive summary	5
Introduction	10
Main findings	21
Early Years	21
Primary school	26
Secondary school.....	33
16-19 education	42
Further findings by prior attainment	47
Conclusions	57
References	60
Annex A: Additional analyses.....	63
Annex B: Model variables	76

Executive summary

This report provides a thorough examination of the socio-economic disadvantage attainment gap in England in 2023/24, exploring how and why educational inequalities emerge and widen across different phases of education. Exploiting the rich potential of the National Pupil Database (NPD), the research updates and extends previous work by the Education Policy Institute (EPI).¹

It takes a significant step forward by extending our earlier work's research design to incorporate a much richer set of drivers of the disadvantage gap. Through detailed linkage of individuals' attainment, demographic backgrounds, Special Educational Needs and Disabilities (SEND) histories and experiences of disrupted schooling, we construct dynamic measures to explore how inequalities accumulate from the early years through to post-16 education. In contrast to earlier work, we also introduce new cohort and institution-level indicators – capturing peer composition, collective experiences of absence and exclusions, and whole-school contexts – to understand how the environments in which children and young people learn to shape their outcomes over and above individual characteristics.

By combining these new measures with statistical decomposition models, we aim to provide a more in-depth assessment of the mechanisms through which disadvantage operates. This enables us to quantify the extent to which the gap can be explained by differences in disadvantaged students' characteristics and the extent to which it arises from differential impacts of these characteristics, including factors not captured in administrative data. Together, these innovations offer the most comprehensive national breakdown to-date of the 2023/24 disadvantage gap.

Research

Our research provides new insights into the key influences on educational inequalities by addressing two key questions. First, what factors are associated with lower attainment among disadvantaged children and young people at ages five (Early Years Foundation Stage [EYFS]), 11 (key stage 2 [KS2]), 16 (key stage 4 [KS4]) and in 16-19 education? Second, how much of the disadvantage gap at each key stage in 2023/24 can be attributed to different factors at a student, cohort and institutional level?² Although the effects in our analysis are associative rather than causal, our research helps identify the most important factors in shaping the national gap and inform priorities for where policy development should be focused.

Early years

- On average, disadvantaged pupils are already 4.6 months behind their peers by the end of EYFS.³ Differences in pupils' (observable) characteristics – related to sex, SEND status, ethnicity, first language and months of birth – explain only 30 per cent of this early year's gap. Likely key factors missing from the model include the home and family environment, early education and care, and interactions with children's services.
- Being identified as having SEND is the single most important contributor to the disadvantage gap at age five, adding 2.4 months, noting that this association is not necessarily causal as EYFS assessments can themselves inform SEND identification. Not only are disadvantaged pupils more likely to be identified with SEND (adding 1.2 months to the gap), but the attainment penalty associated with SEND is larger for disadvantaged pupils (adding another 1.1 months to the gap).
- Cohort and school characteristics each make comparatively small contributions to the early years gap, of under 0.3 months (or under five per cent). For example, disadvantaged pupils are more likely to be in a reception year with a higher proportion of peers with identified SEND and this narrows the gap by 0.2 months; they are also more likely to attend a school with a higher overall absence rate which adds 0.1 month to the gap.

Primary school

- By the end of primary school, disadvantaged pupils are 10.1 months behind. The majority (73 per cent) of this gap can be explained by differences in pupils' (observable) characteristics.
- The lower prior attainment of disadvantaged pupils is the largest contributory factor, adding 4.4 months to the gap, noting that prior attainment encapsulates wide-ranging influences, including early development needs, health and unidentified SEND.
- The size of the KS2 gap is strongly associated with disadvantaged pupils having already fallen behind in reception year in literacy, mathematics, and communication and language.
- The higher prevalence of identified SEND for disadvantaged pupils contributes another 1.3 months to the gap. Together this highlights how the accumulation of inequalities by early primary school continues to affect children's attainment outcomes at the end of primary school, though only some of these may be amenable to education policy.
- Disrupted schooling also plays an important role, notably pupil absence. On average, pupils with even moderate levels of absence are 1.9 months behind their peers with below average absence.⁴ This overall 'absence penalty' is even greater for disadvantaged pupils, who are also much more likely to be absent than their peers. Taken together, absence contributes 2.2 months (around one-fifth) to the KS2 gap.
- Cohort and school characteristics are less important influences on the KS2 gap than student characteristics, all making modest contributions of under 0.4 months (or under four per cent).

Secondary school

- By the time they take their GCSEs, disadvantaged students are 17.9 months behind their peers. Much of this gap (87 per cent) can be explained by differences in the characteristics of disadvantaged students.
- By far the largest contributor to the KS4 gap is the lower prior attainment (at age 11) among disadvantaged pupils, adding 6.8 months to the gap. Prior attainment at age seven (key stage 1 [KS1]) also contributes (1.6 months) over and above KS2 attainment, reinforcing that early intervention has a 'protective' factor against gap-widening later on.
- Differences in disadvantaged pupils' attainment in core subjects (mathematics and English) at KS2 are strongly associated with the size of the KS4 gap.
- Disrupted schooling, principally student absence, is the second most important influence. Disadvantaged students both miss more school than other students and experience a greater attainment penalty when they do. Together, this accounts for 6.1 months (or one-third) of the gap. Other forms of disruption – suspensions, permanent exclusions and school moves – are also each associated with lower attainment (with each incidence lowering pupil attainment by at least one month) but because these affect far fewer students, they make a smaller contribution to the national gap.
- Ethnicity emerges as another important contributor, accounting for around 1.5 months of the gap. The association between ethnicity and GCSE attainment differs between disadvantaged and non-disadvantaged students, with disadvantaged White British students making some of the slowest progress during secondary school.
- We also consider the attainment gap for those students who are persistently disadvantaged. These are the subset of disadvantaged students who have been eligible for free school meals (FSM) for at least 80 per cent of their time in school. This gap is larger still at 21.3 months but the contributors to the gap are broadly

16-19 education

- By the end of their 16-19 education disadvantaged students were 3.5 grades behind their peers across their best three qualifications, with 79 per cent of this gap explained by differences in the characteristics of disadvantaged students.⁵
- The majority of the 16-19 disadvantage gap reflects inequalities that have accumulated earlier in the education system. Differences in prior attainment up to age 16 explain around 70 per cent of the total gap, underlining the lasting impact of attainment gaps that emerge in the early years, primary and secondary school.
- Within KS4 subjects, the ‘open-slot Attainment 8’ measure of prior attainment contributes most to the 16-19 gap, even more than KS4 English or mathematics.
- Cohort characteristics also contribute to the gap. Disadvantaged students are more likely to study in settings with lower average prior attainment and higher concentrations of disadvantage and SEND, which is associated with lower overall outcomes. This likely reflects the greater level of selection in the 16-19 phase. These differences in context account for around 10 per cent of the total gap.
- As in earlier phases, ethnicity contributes to the gap, with evidence that the relationship between ethnicity and attainment differs for disadvantaged students. In the 16-19 phase, this pattern is driven largely by lower progress among disadvantaged White British students.

Cross-phase themes

Several themes emerge when considering the gap across phases. Overall, the story is one of disadvantage accumulating as children and young people progress through education and, while data inconsistencies mean it is difficult to make direct comparisons across phases, the nature of what is driving the gap appears to evolve:

- Across all phases after reception year, prior attainment is the single biggest contributor to the disadvantage gap. While prior attainment at age five contributes 40 per cent of the KS2 gap, by the time pupils take their GCSEs, earlier attainment (across ages 5, 7 and 11) collectively accounts for 44 per cent of the KS4 gap. In the 16-19 phase – when we can additionally account for GCSE attainment, prior attainment explains the majority of the gap. This underscores the extent to which post-16 inequalities are carried forward from earlier schooling.
- Although it is difficult to isolate the exact contribution of any single subject to the gap – as students who perform well in one subject tend to do well in others – attainment in both mathematics and English-related subjects is strongly associated with the gap across phases.
- Disrupted schooling – notably pupil absence – plays a progressively larger role as children progress from primary to secondary school. Absence contributes around one-fifth (21 per cent) of the gap at KS2, rising to one-third (34 per cent) by KS4.
- Identified SEND emerges as the single largest measurable contributor to the EYFS gap, but this partly reflects the limited administrative data available on young children’s early environments. By KS2, our richer SEND measure based on pupils’ histories remains a key contributor (accounting for 15 per cent of the gap). While its direct role diminishes by KS4 and post-16, this is partly because earlier SEND-related needs have already shaped prior attainment, which becomes the dominant pathway through which SEND contributes.

- Across phases, there is important variation in attainment outcomes *within* the group of SEND learners. When children with SEND have timely identification and sustained support, their attainment outcomes can be at least as good as their peers with similar prior attainment and other characteristics.
- In contrast to later ages, disadvantaged pupils attain better at ages 5 and 11 when they are in disadvantaged settings and this helps to slightly narrow the gap. This suggests that, other things being equal, it is disadvantaged pupils in less disadvantaged primary schools who could be at more risk of low attainment than other disadvantaged pupils.
- Other cohort-level characteristics make only small (or non-existent) contributions to the gap across phases including size of year group, the percentage of the cohort with English as an Additional Language and the percentage of the cohort in lower-attaining ethnic groups.
- Similarly, institution-level characteristics make only small (or non-existent) contributions to the gap, noting that higher school absence rates make small contributions to the gap at ages 5, 11 and 16, over and above a pupil's own absence rate.
- Across all phases, cohort and institution-level measures make smaller contributions than individual-level factors, though cohort effects grow in importance with age, consistent with developmental evidence that peer influences strengthen in adolescence. Notably, differences in the composition of 16-19 cohorts – in relation to disadvantage, SEND and prior attainment – collectively contribute around 10 per cent to the 16-19 gap.
- Ethnicity contributes to the gap at every phase. This pattern is driven in part by the comparatively lower progress of disadvantaged White British pupils, an effect that persists throughout the system.

¹ <https://epi.org.uk/publications-and-research/breaking-down-the-gap/>

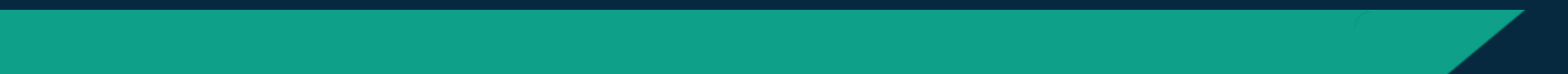
² We present attainment gaps using a 'months of learning' measure up to KS4 and a grade gap measure in the 16-19 phase, using established methodologies as defined in EPI's 2025 Annual Report: <https://epi.org.uk/annual-report-2025/>

³ The gaps presented in this report may differ slightly from those in EPI's 2025 Annual Report (<https://epi.org.uk/annual-report-2025/>), as they are based on the subset of individuals who have complete data for all variables in our models. Further details can be found in Annex B.

⁴ Moderate absence is defined as an absence rate between the national average but below 10 per cent – the level at which a pupil is considered persistently absent in official statistics.

⁵ EPI uses a separate methodology for the 16-19 gap which is measured in A levels equivalent grades rather than months. We calculate a student's total point score across all level 1-3 qualifications (summing their grades), but cap this at the best combination equivalent to three A levels in size. For further details, see Tuckett et al., 2021.

Introduction and methodology



Introduction

Through this research, we undertake an in-depth exploration of the disadvantage gap in 2023/24 across four key stages of education: the Early Years Foundation Stage (EYFS); key stage 2 (KS2); key stage 4 (KS4); and in 16-19 education. Our aim is to provide the clearest picture yet of how the disadvantage gap emerges and widens as children and young people move through the English education system.

Our starting point is to update the previous research undertaken by the Education Policy Institute (EPI) which decomposed the 2022/23 disadvantage gap, by using the most recent cohort available in the National Pupil Database (NPD) at the time of our analysis i.e. those completing education in 2023/24.⁶ We extend our previous exploration of the gap in several ways by building new measures to take a much richer account of the factors shaping the educational outcomes of disadvantaged children and young people at an individual, cohort and institution level.⁷

Our new measures at a cohort and institution level help to capture the wider learning environment that shapes a student's experiences but is not visible through individual characteristics alone. These contextual indicators allow us to identify whether aspects like peer composition or whole-school absence compound or mitigate disadvantage and potentially offer insights into mechanisms that operate beyond the influence of individual students.

In terms of the structure of the report, in the rest of this section we provide an overview of our new measures, the wider context for our research and our research questions. In subsequent chapters, we set out our full methodology – including data sources and limitations – in Section 2, our main findings (by key stage) in Section 3, further analysis by prior attainment in Section 4, and our conclusions in Section 5. We provide further detail on our findings and variables in the Annexes.

New measures

Individual

The NPD contains rich individual-level data in relation to attainment at each key stage, disadvantage (based on free school meals [FSM] eligibility), as well as demographic characteristics relating to sex, month of birth, ethnicity and first language.

A key contribution of this research has been the use of the NPD to develop detailed, historic measures of pupils' prior attainment, Special Educational Needs and Disabilities (SEND) histories and experiences of disrupted schooling. These dynamic measures aim to provide insights on when and how gaps emerge which can be masked in snapshot measures.

For prior attainment, we have been able to track the cohort completing KS4 in 2023/24 all the way back to reception year. This means for the first time, we have been able to include measures of prior attainment at several earlier stages not just the previous key stage. While our headline findings use composite

⁶ <https://epi.org.uk/publications-and-research/breaking-down-the-gap/>

⁷ Throughout this report we use 'cohort' to refer to individuals in the same year group in the same institution.

measures of prior attainment, we supplement these with additional models which use more granular measures of prior attainment by subject. These have some limitations (see Methodology and Section 4 for details) but provide some sense of the most important mechanisms through which prior attainment is operating.

For SEND, our previous research⁸ differentiated between pupils with no identified needs compared to children on SEND Support or having an Education, Health and Care Plan (EHCP) but did not consider the duration of support or type of need – both of which may be associated with disadvantage. We have therefore used pupils' histories to develop two measures of SEND in our new analysis – one combining duration and complexity to assign pupils into different SEND categories; and a second based on (contemporaneous) primary need type.

Through this research, we also recognise the variety of ways in which children experience disruptions to schooling. While our previous research considered pupil absence as a driver of the disadvantage gap, this research goes a step further by considering other instances of missing out on school through suspensions, permanent exclusions and school moves. Each of these measures is cumulative (i.e. aggregated over the final three years of primary school or secondary school) to capture potential persistence in their effects on attainment, while avoiding the earlier period of pupils' schooling directly affected by the pandemic.

Cohort

In our previous research,⁹ we took into account limited cohort characteristics related to the average prior attainment of the year group, as well as the size of the year group. Through this research, we have explored whether peer groups have a wider influence on the gap. Our new measures vary by phase but broadly relate to the demographic characteristics of the year group, as well as their collective experiences of disrupted schooling – for example, our GCSE gap model (where we have the richest data) includes:

- percentage of the cohort who are disadvantaged;
- percentage of the cohort who have identified SEND;
- percentage of the cohort who arrive late to the English state school system (i.e. in Years 10 or 11) with English as an Additional Language (EAL);
- percentage of the cohort who are in a lower-attaining ethnic group;¹⁰
- average absence rate of the cohort;
- average suspensions rate of the cohort; and
- average permanent exclusions rate of the cohort.

⁸ <https://epi.org.uk/publications-and-research/breaking-down-the-gap/>

⁹ <https://epi.org.uk/publications-and-research/breaking-down-the-gap/>

¹⁰ For further details on all these measures, including the definition of lower-attaining ethnic groups, see Annex B.

Institution

This research also includes new variables at an institution level. Alongside institution type, we have also included (where data allows) whole-school measures related to the proportion of pupils who are absent, suspended or permanently excluded. This allows us to investigate if there are factors contributing to the gap that operate over and above individual- and cohort-level influences.

We initially included a wider set of variables at an institutional level, though found in several cases that their effects were already captured statistically through cohort measures in our pre-16 models. This led us to exclude school-level indicators of the same measure in our final models – see Methodology and Annex B for further details.

Context

Our focus on disrupted schooling reflects its prominence as a key, and growing, challenge for the education system in recent years. Rates of overall and persistent absence remain well above pre-pandemic levels, and while suspensions have fallen from their 2023/24 peak, they are still around twice the level seen before the pandemic.^{11, 12} Permanent exclusions also reached their highest level in 2023/24, followed by a modest decline the following year.¹³ The effects of disrupted schooling are not felt evenly: disadvantaged pupils and pupils with SEND face much higher risks of persistent absence, being suspended or being excluded, placing them at heightened risk of disengagement from learning.^{14, 15} Consequently, disrupted schooling is not just a challenge for families and schools, but integral to addressing educational inequalities and narrowing the disadvantage gap.

Research questions

Specifically, we aim to address the following research questions:

1. What factors are associated with lower attainment among disadvantaged children and young people at the EYFS, KS2, KS4 and in the 16-19 phase?
2. How much of the disadvantage gap at each key stage in 2023/24 can be attributed to factors at an individual, cohort, institution and regional level?

¹¹ The overall absence rate – defined as the total number of sessions missed due to absence for all pupils as a percentage of the total number of possible sessions for all pupils – fell to 6.6 per cent in combined Autumn and Spring Terms 2024/25, down from 6.9 per cent in the previous Autumn and Spring Terms but were consistently below five per cent prior to the pandemic. The persistent absence rate – defined as the percentage of pupil enrolments that miss 10 per cent or more of their possible sessions – fell to 17.6 per cent in combined Autumn and Spring Terms 2024/25 (from 19.2 per cent in autumn and spring 2023/24) but remained well above the pre-pandemic rate of around 10 per cent.

¹² There were 4.02 suspensions per 100 pupils in autumn 2024/25 compared with 2.17 in autumn 2019/20.

¹³ The permanent exclusions rate peaked in autumn 2023/24 at 0.05 per 100 pupils (equivalent to five permanent exclusions for every 10,000 pupils) and then fell to 0.04 per 100 pupils in autumn 2024/25.

¹⁴ <https://www.gov.uk/government/statistics/pupil-absence-in-schools-in-england-autumn-2024-and-spring-2025>

¹⁵ <https://explore-education-statistics.service.gov.uk/find-statistics/suspensions-and-permanent-exclusions-in-england/2024-25-autumn-term>

3. Are these factors different for the persistent disadvantage gap at KS4 in 2023/24?

To address our first research question, we use regression models to estimate both descriptive disadvantage gaps and ‘modelled’ gaps that control for a rich set of observable factors in the NPD. These models allow us to identify the dominant mechanisms through which disadvantage operates.

For our second and third research questions, we use the ‘Oaxaca-Blinder’ decomposition technique to break down the overall disadvantage gap into differences in the characteristics of disadvantaged students and their peers, and differences in the effects of these characteristics on disadvantaged students’ attainment, relative to their peers. This distinction matters because while policymakers have to take certain characteristics as a given, they may be able to directly influence how these subsequently shape the life outcomes of young people and break their link with disadvantage. By quantifying how much individual characteristics (or their associated impacts) contribute to the national gap, we provide a clearer sense of where policy and interventions could be targeted to help narrow the gap, while recognising that the relationships identified are associative rather than causal.



Methodology

Data sources

All findings presented in this report are derived from analysis of data held in the NPD – specifically, the EYFS profile, KS2, KS4, and 16-19 education.

We combine student attainment data with student characteristics from the pupil-level school census in relation to sex, ethnicity, first language, SEND and month of birth. Further details are outlined below.

For KS2 and KS4 only, we also analyse pupil-level data in relation to absence and suspensions derived from the half-termly school census; for KS4 only, we also include data on permanent exclusions. Issues with data quality and availability mean this is not possible for pupils in reception year or in the 16-19 phase.

Coverage

We focus on children and young people attending state-funded schools and colleges in England in 2023/24. Our analysis includes local authority-maintained schools, sponsored academies, converter academies, free schools and special schools. It excludes students whose sole (or main) registration is in alternative provision, pupil referral units, or hospital schools, as well as those attending independent schools.

To be included in our sample, individuals need to have complete data across all variables. In cases where individuals lack full histories of their prior attainment – which might include pupils arriving late to the English state school system such as asylum seekers and other migrants, pupils who frequently move schools, those missing from education or are not on-roll for other reasons – these individuals will not be captured in our results.

For the 16-19 phase, the analysis is necessarily restricted to young people who continue in education after age 16. Given that disadvantaged students are more likely to exit education or training at this point, this selective attrition means our estimates are likely to understate the true scale of educational inequalities affecting disadvantaged young people.

Defining disadvantage

We define disadvantage based on eligibility for FSM. Specifically:

- In reception, disadvantage is based on pupils eligible for FSM in the current academic year.
- At the end of KS2 and KS4, disadvantage is based on pupils who are eligible for FSM at any point in the previous six years.
- At the end of KS4, we also identify the subgroup of disadvantaged pupils who are ‘persistently disadvantaged’ based on being FSM-eligible for at least 80 per cent of a pupil’s time in school.
- In the 16-19 phase, where FSM data tends to be less reliable, disadvantage is based on whether a student was eligible for FSM at any point during the previous six years, as recorded at the end of secondary school.

Methodological approach

Through this research, we address the following research questions:

1. What factors are associated with lower attainment among disadvantaged children and young people at the EYFS, KS2, KS4 and 16-19 phase?
2. How much of the disadvantage gap at each key stage in 2023/24 can be attributed to factors at an individual, cohort, institution and regional level?
3. Are these factors different for the persistent disadvantage gap at KS4 in 2023/24?

For research question 1, we estimate both descriptive gaps (i.e. gaps that do not control for any characteristics) alongside ‘modelled’ gaps based on series of ordinary least squares regression models.¹⁶

Our regression models use student attainment as the dependent variable. Our attainment measures are consistent with those used in EPI annual reports, although this report focuses on a single academic year (2023/24). These are based on statutory assessments taken at key stages using point scores or grades (rather than threshold measures) to better capture the attainment of individuals across the distribution, with further details found in Robinson et al. (2025).¹⁷

We then include a rich set of factors – at an individual, cohort, institution and regional level – as independent variables. We build up our models sequentially to allow us to examine how the gap changes, as we introduce additional controls.

We have accounted for background characteristics as consistently as possible between phases, but there are some notable differences.

In the case of prior attainment, we control for this at all key stages except early years. We standardise our prior attainment variables (by centring the mean at zero and setting the standard deviation to one) to compare variables on the same scale, within and across phases. Unlike our dependent variable (which is a single score averaged across more than one subject), our measure of prior attainment comprises individual scores across multiple subjects. This is because in predicting current attainment we are aiming for as granular measure of a pupil’s prior attainment as possible.

We initially tested a ‘maximalist’ modelling approach, incorporating a wide range of variables at the student, cohort and institution levels. We then conducted statistical robustness tests to identify and address multicollinearity among predictors.¹⁸

As many school-level variables were already captured through year group measures in our pre-16 models, in some cases these tests led us to exclude school-level indicators of the same measure which was sufficient to deal with correlated predictors. For this reason, our final models did not include: the

¹⁶ Our results are not sensitive to whether we cluster standard errors at a school level.

¹⁷ <https://epi.org.uk/annual-report-2025/>

¹⁸ Multicollinearity occurs when two or more predictor variables in a regression model are highly correlated, making it difficult to isolate their individual effects. This can lead to imprecise coefficient estimates and affect interpretability of the model’s results.

percentage of school pupils who are disadvantaged; the percentage of school pupils who have identified as SEND; the percentage of school pupils who have EAL; and the percentage of school pupils who are in a lower-attaining ethnic group.¹⁹

In our post-16 models, we found that cohort prior attainment, cohort disadvantage, and the proportion of pupils with SEND were highly correlated with each other, reflecting the more selective nature of post-16 education. To address this, we used principal component analysis to combine these variables into a single composite indicator of ‘cohort disadvantage’. Further details on our final models can be found in Annex B.

Where students did not have a complete set of data for a given phase, they were dropped from our models allowing us to have the same individuals across all specifications which consequently removed any compositional differences that could affect our results, while approximately losing up to one-fifth of the original sample at each key stage.²⁰

It is also important to emphasise that all the estimates in our analysis are associative rather than causal.

For research question 2, we decomposed the disadvantage gap at each phase to identify the individual contribution of different characteristics to the size of the gap.²¹ This investigates whether differences in the *prevalence* of certain characteristics, or differences in the ‘*effect*’ of those characteristics with attainment make a substantive contribution to the disadvantage gap.

For example, how much does the higher level of absence of disadvantaged students, and the effect of those absences on attainment for disadvantaged pupils, contribute to the gap? Though note in this instance when we say effect, we mean statistical effect, rather than causal effect.

¹⁹ Note that we retained school-level variables in cases where these appeared to make an important contribution to the gap (based on their statistical significance) and did not breach an adjusted generalised variance inflation factor (GVIF) threshold of two. We used an adjusted GVIF as several categorical variables have multiple levels, allowing comparisons across predictors with differing degrees of freedom. See Annex B for further details.

²⁰ The imposition of complete data for each individual causes us to lose more of the sample as pupils move through the education system. At EYFS and KS2, we lose roughly one-eighth of the sample, for KS4 and in 16-19 education, around one-fifth.

²¹ Note that a predictor can have a large association (positive or negative) with attainment in the regression model but nonetheless account for a small share of the overall national gap, if few pupils differ between groups on that factor.

Decomposition analysis

To understand how different factors contribute to the disadvantage gap, we have used the Oaxaca-Blinder decomposition. Factors can contribute to the gap in two ways:

Explained contribution - this is where differences in the characteristics of disadvantaged and non-disadvantaged students contribute towards the gap. For example, as disadvantaged students tend to have relatively lower attainment in earlier education phases, differences in prior attainment often contribute to the explained part of the gap.

Unexplained contribution - this is where the statistical effects of a characteristic (sometimes referred to as a 'return' or 'penalty') differ between disadvantaged and non-disadvantaged students. For example, if an increase of 10 scaled score points in pupils' KS2 results was associated with an increase of one GCSE grade for non-disadvantaged pupils, and only 0.8 grades for disadvantaged pupils, then prior attainment would be making a contribution to the unexplained part of the gap.

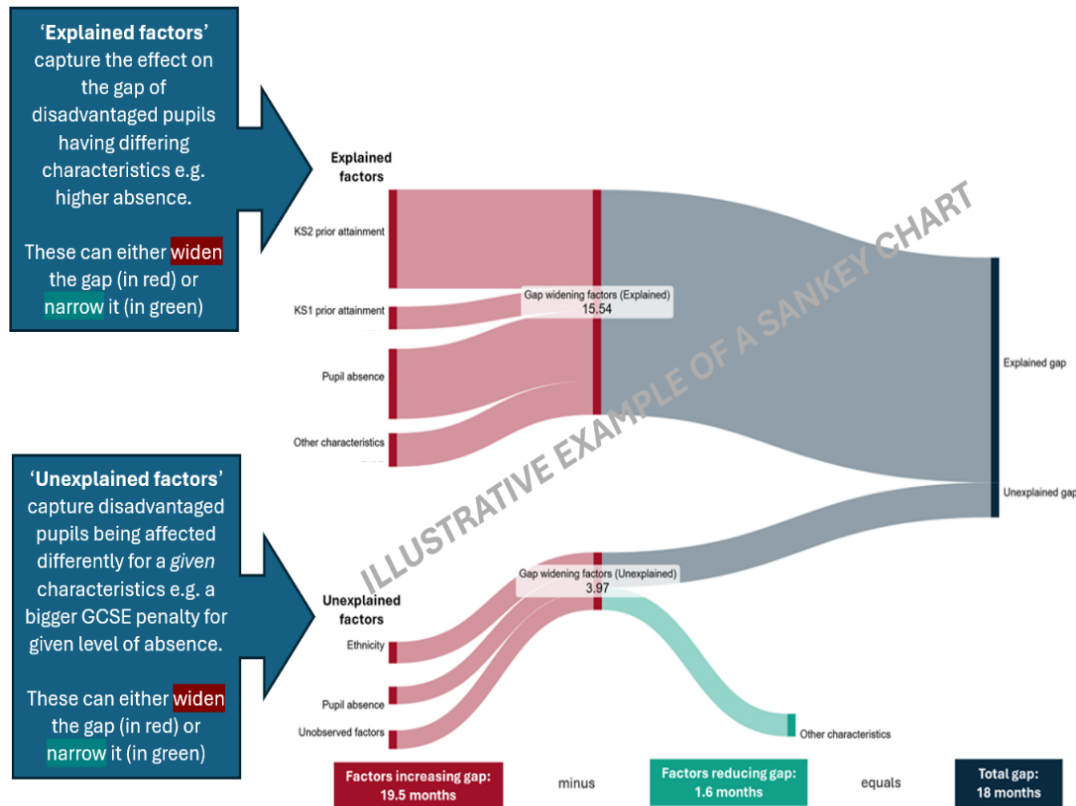
The unexplained component also includes the constant, that is, the portion of the attainment gap not accounted for by the observed characteristics included in the model. We label this as 'unobserved factors'. This component may reflect differences associated with disadvantage after conditioning on observed characteristics, as well as the influence of other factors not captured in the data, such as parental education.

Both explained and unexplained factors can either increase the gap or reduce the gap. The sum of these positive and negative factors will equal the size of the overall gap.

We used the two-fold Oaxaca-Blinder decomposition for research question 2 – further details can be found in the 'Decomposition analysis' box. We used the same sample of students and set of variables for research questions 2 and 1, though we standardised all our continuous variables (not just prior attainment), to ensure the decomposition results were meaningful.¹ Given that prior attainment measures across different subjects at the same key stage tend to be correlated, we grouped the granular measures of prior attainment into a single composite prior attainment construct at a given key stage to support more meaningful interpretation of the Oaxaca-Blinder results in our main findings. We supplemented our main models in Section 3 with further analysis using disaggregated models of prior attainment in Section 4.

We represent the decompositions in two ways:

- In the findings section of the report, we use Sankey charts to highlight key factors contributing to the gap, with an illustrated example shown below. The Sankey charts show factors increasing the gap (in red) and decreasing the gap (in green) and break these down further into factors contributing to the explained gap (top left), and factors contributing to the unexplained gap (bottom left). These then flow into the total explained and unexplained components of the gap on the right. Only statistically significant factors contributing five per cent or more to the gap (increasing or decreasing) are shown. Factors contributing less than this are aggregated into an 'other characteristics' category separately for the explained and unexplained components.
- In Annex A, we show the full breakdown of all factors in our decompositions, showing separately the factors that were aggregated together in the Sankey charts.
- For both research questions 1 and 2, we present attainment gaps using a 'months of learning' measure up to KS4 and a grade gap measure in the 16-19 phase, using established methodologies as defined in EPI's 2025 Annual Report.¹



Limitations

While our models draw on unusually rich administrative data across all phases of education, several important limitations should be borne in mind when interpreting the findings.

First, SEND is imperfectly observed in administrative data and is identified in waves over time. Many children's needs are not recognised until later in primary or even secondary school, and the likelihood of identification itself has a socio-economic gradient. This means that at earlier stages – especially the EYFS – unidentified SEND operates as an unobserved factor, likely inflating the 'unexplained' component of the gap. It also means that, at later stages, some of the variation attributed statistically to prior attainment or disrupted schooling is, in practice, capturing unmet or later identified SEND rather than independent influences.

Second, prior attainment is measured imperfectly and captures a wide footprint of developmental, health and contextual factors. Assessments contain noise and – in the case of the EYFS – rely on teacher judgements and therefore vary in how fully they capture pupils' underlying capabilities. As a result, some of the apparent 'catch-up' among disadvantaged pupils with the same EYFS or KS2 scores in the later education phases could reflect regression-to-the-mean rather than genuine differences in progress. This, in turn, would mean that part of the 'unexplained' gap-narrowing associated with prior attainment reflects a statistical artefact rather than real educational progress.

While prior attainment is included to improve comparability between disadvantaged and non-disadvantaged pupils, it may also act as a mediator of disadvantage (i.e. part of the pathway through which disadvantage influences later attainment) rather than a pure confounder. As a result, estimates from specifications that condition on prior attainment should be interpreted as gaps *within each phase*

i.e. net of earlier achievement (or the portion of the gap not explained by prior attainment), rather than the total association between disadvantage and later outcomes. In this sense, some of the contribution attributed to prior attainment reflects inequalities that have already emerged earlier in the education system, and ‘controlling for’ it may understate the overall extent to which disadvantage shapes attainment across phases. This important caveat also applies to our findings on how SEND characteristics influence later attainment outcomes where these estimates control for prior attainment alongside other factors (figures 6, 7, and A13).

Third, ethnicity related findings must be interpreted in light of missing data for late arrivals to the English state education system. Because our models require complete linked records back to age five, pupils who enter the system late – disproportionately recent migrants and asylum seekers from vulnerable minority ethnic groups – are necessarily excluded. As the report notes elsewhere, our ethnicity coefficients therefore reflect the outcomes of the less vulnerable subset of minority ethnic pupils for whom full data exist, and our estimates likely understate the attainment penalties associated with minority groups experiencing more late arrival and earlier trauma.

Finally, disrupted schooling measures are limited by data availability, and by the way different forms of disruption intersect. Absence, suspensions, permanent exclusions and school moves are all highly intertwined with SEND, health, poverty and wider family circumstances. Because some of these underlying drivers are unobserved, disruption variables absorb more variation in this model than they would in a dataset with richer information on health, well-being, family environment and children’s services involvement. This means their estimated contributions partly reflect broader unmet need rather than disruption per se.

Taken together, these limitations do not undermine the core findings of the report, but they do place boundaries around the precision and interpretation of our decomposition results. They also point clearly to areas where future research – particularly using linked health, social care and pre-school data – could provide deeper insights still into the origins and evolution of the disadvantage gap.

Findings



Main findings

In this section, we consider the disadvantage gap for children and young people across all compulsory phases of education. We first present our descriptive disadvantage gap findings, looking at the size of the gap in the 2023/24 academic year, the latest available at the time of our analysis.²²

We then consider the ‘modelled’ gap which uses statistical modelling to examine the size of the disadvantage gap, after controlling for a rich range of factors at a student, cohort, institutional and regional level. By sequentially controlling for these other characteristics, we can draw insights on the nature of underlying gaps for disadvantaged students.

Together, these findings on descriptive and modelled gaps address our first key research question on the factors associated with lower attainment among disadvantaged children and young people.

We then decompose the disadvantage gap at each phase to identify the individual contributions of different factors to the size of the gap. This investigates how much of the overall gap is ‘explained’ by the different characteristics of disadvantaged students, how much is ‘unexplained’ due to the differing impacts of these characteristics on disadvantaged students’ attainment, and how much remains unobserved.

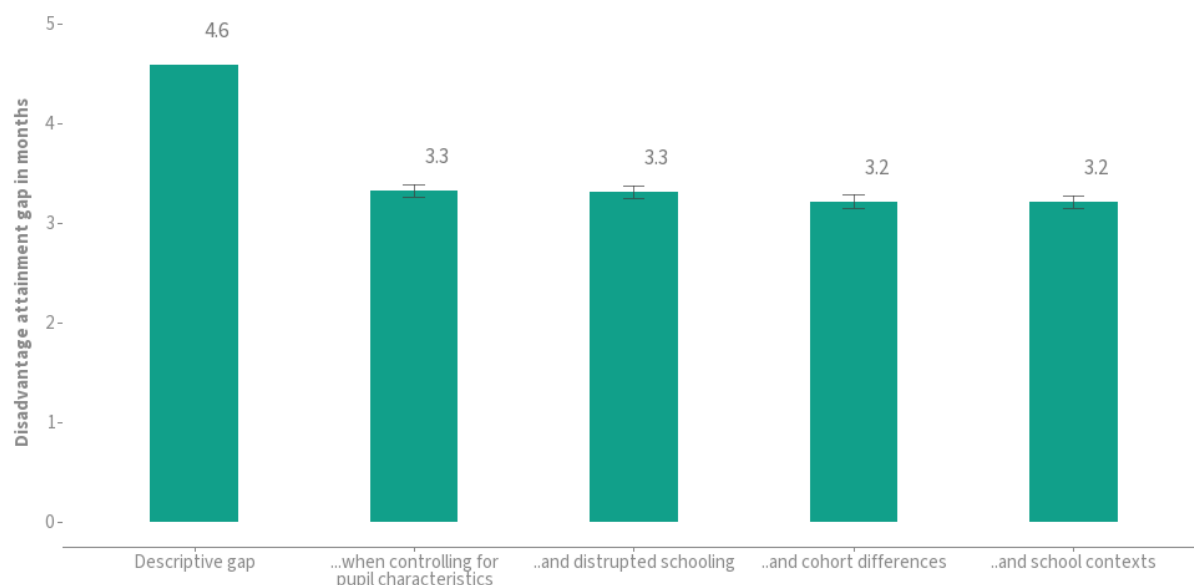
These findings address our second research question on how much of the disadvantage gap can be attributed to factors at an individual, cohort and institution level.

Early Years

Nearly one-fifth (19 per cent) of reception pupils were disadvantaged in 2023/24, based on being eligible for FSM. On average, these pupils were 4.6 months behind their peers on the EYFS profile (figure 1).

²² The descriptive gaps presented here may differ from those presented in EPI’s 2025 Annual Report (<https://epi.org.uk/annual-report-2025/>), as they are based on the subset of students who have complete data for all the variables in our models.

Figure 1: The disadvantage gap at age five in 2023/24, including for models with pupil, cohort, school and regional controls



Notes: The descriptive gap includes no controls; subsequent modelled gaps sequentially include controls for a range of characteristics. Bars show 95 per cent confidence intervals. N = 517,020.

This (descriptive) gap reduces by over one month (to 3.3 months) once we control for pupil characteristics.²³ In particular, differences in disadvantaged pupils' SEND status and ethnicity partly explain their lower EYFS scores compared to their peers. Among disadvantaged pupils in reception year, 22 per cent are identified with SEND compared with 12 per cent of their peers. Disadvantaged pupils are also over-represented in ethnic groups with some of the lowest (conditional) scores in the EYFS, such as Gypsy, Roma and Traveller of Irish Heritage, and Black Caribbean pupils.

Unlike at later ages – where we can control for pupils' histories of absence, permanent exclusions and school moves – there is more limited data for reception-aged children who have just started school. While we can control for the number of times that reception-aged children are suspended, this makes no difference to the national gap, as this affects relatively few children.²⁴

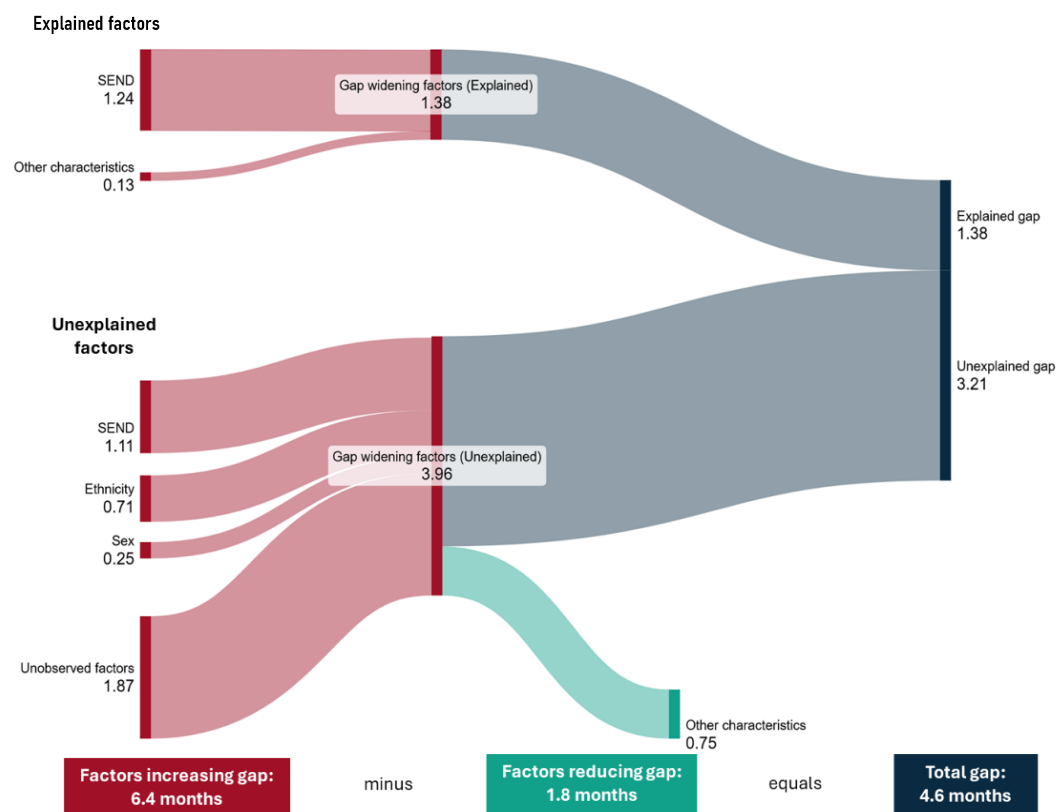
²³ Pupil characteristics control for: sex; SEND status; ethnicity; first language; and month of birth. Unlike at KS2 and KS4, we cannot control for SEND duration or primary need type in reception year.

²⁴ In our EYFS model, 'disrupted schooling' simply controls for the number of suspensions in reception year. We find that each suspension is associated with lower EYFS attainment by 1.5 months – roughly half the size of the disadvantage penalty (3.2 months) in our fully specified model – noting that very few pupils are suspended and these estimates are not causal.

Including further controls for the characteristics of the cohort in reception year also makes little difference, with the gap reducing by only 0.1 month (to 3.2 months).²⁵ There is no further reduction if we additionally control for school characteristics and region.²⁶ This indicates that differences in the characteristics of disadvantaged pupils compared to their peers can explain some of the gap at age five, while differences in their cohorts and schools appear to make only small differences overall. In our fully specified model, disadvantaged pupils remain 3.2 months behind their peers at age five even when they have similar (observable) characteristics, similar peers and attend similar schools.

In our Oaxaca-Blinder decomposition model (figure 2), we breakdown the total (descriptive) disadvantage gap at age five to understand how much can be ‘explained’ by the characteristics and settings of disadvantaged pupils compared to their peers. And how much of the gap is ‘unexplained’, reflecting the differing impacts of these characteristics on disadvantaged students’ attainment relative to their peers.

Figure 2: Decomposition of the disadvantage gap (in months) at age five in 2023/24



Notes: Other characteristics includes all factors contributing < 5 per cent to the gap, netting off both positive and negative factors. The full breakdown can be seen in the annex.

N = 517,020.

²⁵ Cohort differences control for: the number of pupils in reception year (plus squared-term); percentage of reception pupils who are FSM-eligible; percentage of reception pupils with identified SEND; percentage of reception pupils in lower-attaining ethnic groups; and percentage of reception pupils with EAL.

²⁶ School contexts control for: school type; school absence rate; school suspension rate; and region.

Explained gap

Of the 4.6-month gap at age five, 1.4 months (or 30 per cent) can be explained by the differing characteristics of disadvantaged pupils, including the characteristics of their cohort, their school and the region where they live. This means that most (70 per cent) of the gap is unexplained by measurable differences that are captured in our model – this ‘unexplained’ gap is equivalent to the fully modelled gap in figure 1.

The most important (measurable) factor – albeit in a model with data limitations – is the higher prevalence of SEND among disadvantaged pupils relative to their peers, which contributes 1.2 months to the gap. This association is not necessarily causal, as EYFS assessments can inform SEND identification and these are often undertaken by the same teachers. Additionally, having a child with SEND can affect parental employment and income (plus, parents of disabled children are more likely to be disabled themselves), increasing the likelihood of being eligible for FSM.

Other differences in characteristics and contexts contribute little to the explained gap and are grouped together under ‘other characteristics’ (and shown in more detail in Annex A, figure A1).

Unexplained gap

Figure 2 shows that SEND also makes an important contribution to the ‘unexplained’ gap: even when disadvantaged pupils have the same SEND status as their non-disadvantaged peers, they are 1.1 months behind by the end of reception year. Previous EPI research has found that although disadvantage (at a pupil, school, neighbourhood and local authority level) is associated with much higher odds of being identified with SEND, disadvantaged children in those deprived schools and areas are *less* likely to be identified than their peers.²⁷ This apparent rationing of provision within disadvantaged contexts may mean that disadvantaged pupils’ SEND needs must be more pronounced to be identified and supported in the first place, relative to their SEND peers from non-disadvantaged backgrounds.

Identified SEND is therefore a key contributor to the age-five disadvantage gap, reflecting both its higher prevalence among disadvantaged pupils and its larger EYFS penalty for disadvantaged pupils. This warrants further research into the additional challenges facing low-income parents of children with SEND, including the identification of need, accessing support and finding a suitable school place.

The next most important contributor to the unexplained gap is ethnicity, adding 0.7 months. This suggests that ethnic background is associated with different EYFS outcomes for disadvantaged and non-disadvantaged pupils. We return to this later in this report where we see similar findings for the disadvantage gaps in later phases.²⁸

Looking in more detail at the ‘unexplained’ gap (figure A2 in Annex A) reveals a modest role for some cohort and school factors. Cohort disadvantage and school type both help narrow the unexplained gap (each by 0.2 months). Analysis of our regression outputs suggests that while being in a more disadvantaged cohort correlates with lower EYFS scores overall, this is not the case for disadvantaged pupils who achieve slightly higher EYFS scores in these settings.²⁹ This helps narrow the age-five gap (by

²⁷ <https://epi.org.uk/wp-content/uploads/2025/02/SEND-Final-Report-version-FINAL-04.02.2024-2.pdf>

²⁸ Among non-disadvantaged pupils at age five, most ethnic groups have lower EYFS scores on average than White British pupils (with the exceptions of Chinese, White and Asian, and White Irish pupils); but among disadvantaged pupils, most ethnic groups have higher EYFS scores than White British pupils (except, Gypsy, Roma and Traveller of Irish Heritage, White Irish, and Any Other ethnic group).

²⁹ Further details of our regression models are included in the Methodology section; results are available on request.



0.2 months) and aligns with evidence that disadvantaged pupils often perform better, or have smaller gaps, in more disadvantaged settings.^{30, 31, 32}

While this pattern may seem counterintuitive, several mechanisms are discussed in the wider literature including: the additional funding available to more disadvantaged schools via the Pupil Premium; being well placed to adopt universal approaches towards narrowing gaps; having strong ecologies of expertise around disadvantaged children and families; and having strong support systems in place to identify and meet additional needs.

It is striking that the single largest component of the age-five gap is unobserved factors (part of the unexplained gap), contributing 1.9 months. This likely reflects the limitations of the NPD, which does not capture key influences on early child development such as parenting, the home learning environment and access to children's services.³³

³⁰ https://assets.publishing.service.gov.uk/media/5a80bcd2ed915d74e6230292/DFE-RR411_Supporting_the_attainment_of_disadvantaged_pupils.pdf

³¹ <https://www.instituteforgovernment.org.uk/sites/default/files/2025-04/school-readiness-policy-making-left-behind-groups.pdf>

³² https://static1.squarespace.com/static/68c02a5904df7c39a9c27a73/t/68c9c4cbf8692051ec0df946/1758053579502/Funding_Futures_Summary_Report.pdf

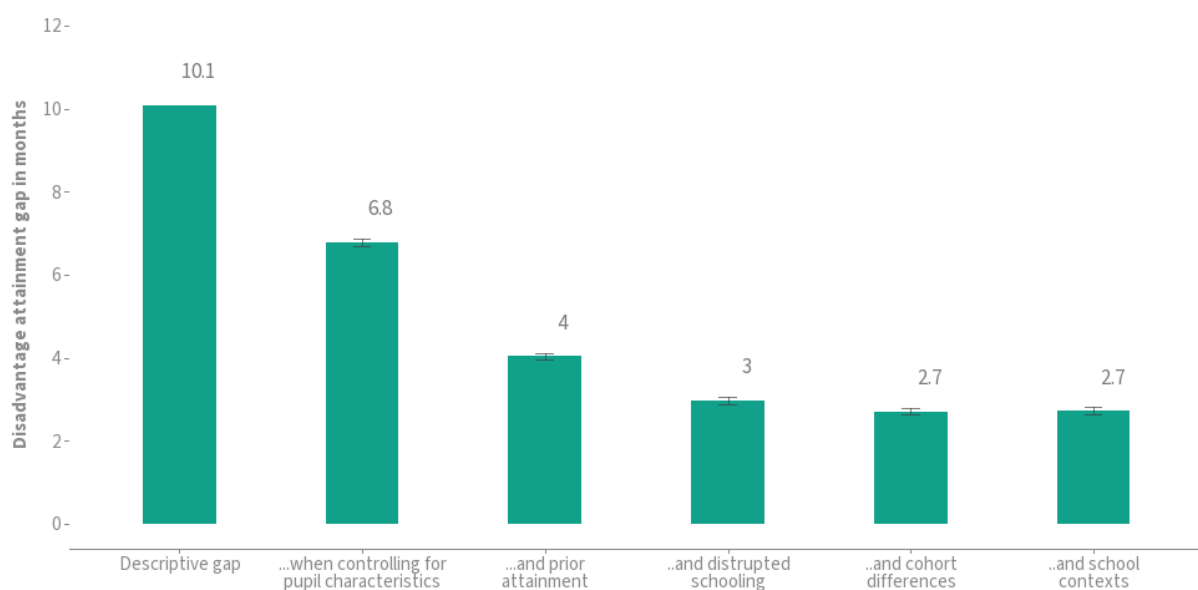
³³ <https://epi.org.uk/wp-content/uploads/2018/07/EPI-Annual-Report-2018-Lit-review.pdf>

Primary school

By the end of primary school, 30 per cent of pupils have been eligible for FSM at some point in the previous six years. We use this as our measure of disadvantage at KS2.

Figure 3 shows that the (descriptive) KS2 gap was 10.1 months in 2023/24. This compares to 4.6 months at reception (figure 1) and, while these are different cohorts, it indicates that disadvantaged pupils tend to fall further behind their peers as they progress through school.

Figure 3: The disadvantage gap at KS2 in 2023/24, including for models with pupil, cohort, school and regional controls



Notes: The descriptive gap includes no controls; subsequent modelled gaps sequentially include controls for a range of characteristics. Bars show 95 per cent confidence intervals. N = 578,807.

The KS2 gap narrows substantially by 3.3 months (to 6.8 months) once we control for pupil characteristics, indicating that disadvantaged pupils' characteristics differ in ways that matter for their KS2 attainment.³⁴ In particular, we find that by age 11, disadvantaged pupils:

- are over twice as likely as their non-disadvantaged peers to have been issued an EHCP at some point (8.4 per cent compared to 4.0 per cent);
- have spent longer on SEND Support (with 4.2 per cent having been on SEND Support for the duration of their time in primary school compared with 2.0 per cent of non-disadvantaged pupils);³⁵ and

³⁴ Pupil characteristics control for: sex; SEND duration and primary need type; ethnicity; first language; and month of birth.

³⁵ This is based on being on SEND Support in each annual Spring Census from reception to Year 6.

- are roughly twice as likely to have been identified with certain primary need types including social, emotional and mental health; moderate learning difficulties; and speech, language and communication needs.

The gap further reduces by 2.8 months (to 4.0 months), after accounting for prior attainment at age five.³⁶ This indicates that lower attainment at age 11, at least partly reflects disadvantaged pupils already lagging behind at age five – and once we account for this, the age 11 gap substantially reduces. To the extent that early year’s policy and interventions can narrow the age-five gap – not just in early years settings but covering all the wider influences on children’s first five years of life including children’s services and welfare – is critical to mitigating disparities at later ages too.

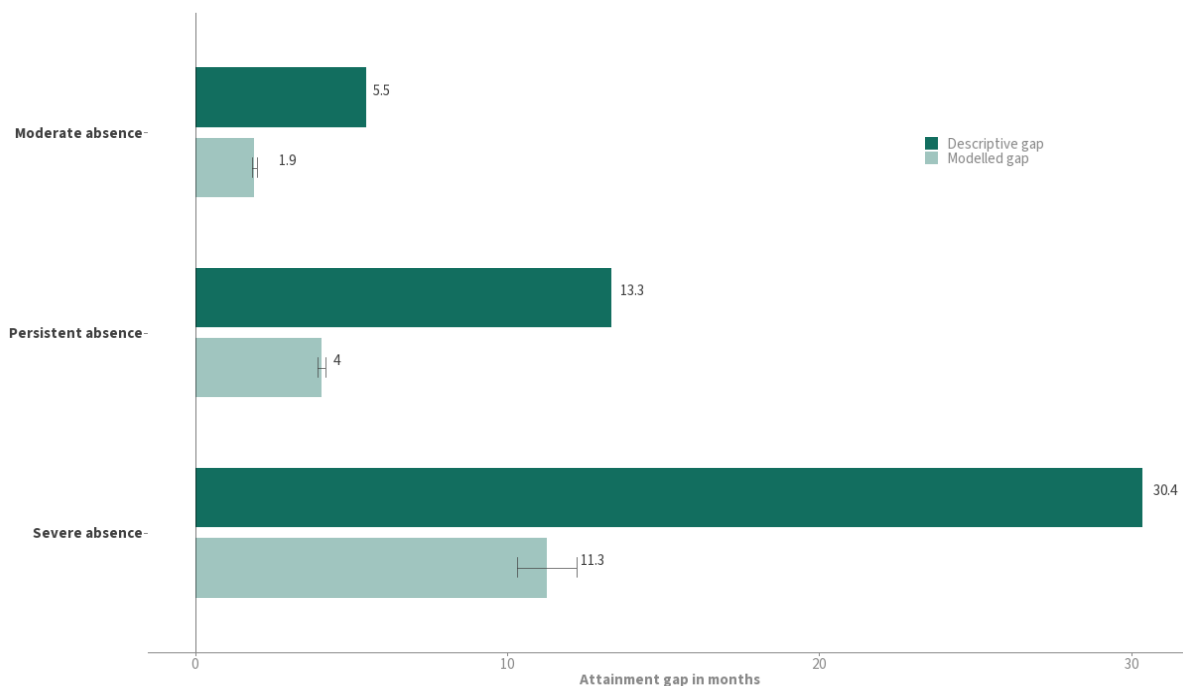
Unlike at age five, by age 11 we can exploit the NPD to account for factors beyond pupil characteristics and prior attainment, notably pupils’ histories of disrupted schooling. We find that once we control for pupils’ patterns of absence, number of suspensions and number of schools moves over Years 4 to 6, the gap further reduces by one month (to 3.0 months). This highlights the significant role that disrupted schooling plays in widening the gap, with disruptions in all their forms associated with lower KS2 attainment.³⁷

Pupil absence over Years 4 to 6 is particularly associated with lower KS2 attainment. Figure 4 shows that, conditional on all other observable characteristics, even moderate levels of pupil absence are associated with lower attainment at KS2 (by 1.9 months) compared to pupils whose absence levels are below the national average. This attainment penalty grows to 4.0 months among persistently absent pupils and to 11.3 months among severely absent pupils, noting that these estimates are associative not causal, with pupil absence often symptomatic of other barriers to attaining well, such as unidentified SEND, poor mental health, school disengagement and poverty.

³⁶ Prior attainment controls for a pupil’s average (standardised) attainment at age five – specifically, a pupil’s EYFS score in each of the seven learning goals.

³⁷ Conditional on all other observables, each school move is associated with 0.6 months lower KS2 attainment and each suspension with 0.5 months lower attainment. Note these estimates are associative not causal and relate to those pupils who are retained in our sample. The vast majority (97 per cent) of Year 6 pupils have not been suspended in the previous three years and most (86 per cent) have remained at the same school.

Figure 4: KS2 attainment gap in 2023/24 by levels of pupil absence over Years 4 to 6 (relative to pupils with below-average absence), including with full pupil, cohort, school and regional controls



Notes: The descriptive gap includes no controls; the modelled gap includes the full set of pupil, cohort, school and regional controls. Persistently absent pupils miss 10 per cent of school sessions (but below 50 per cent); severely absent pupils miss 50 per cent or more. Bars show 95 per cent confidence intervals. N = 578,807.

Controlling for cohort characteristics in Year 6 reduces the gap slightly (by 0.3 months, to 2.7 months), while school-level controls have no effect overall.^{38,39} As at age five, pupil-level factors remain the strongest influences on the disadvantage gap, consistent with wider evidence on educational inequalities.⁴⁰

Taken together, disadvantaged pupils remain 2.7 months behind their peers at the end of primary school even when they have similar (observable) characteristics, similar school histories, similar peers and attend similar schools. These unexplained differences between disadvantaged pupils and their peers are explored further in our Oaxaca-Blinder decomposition model below.

Figure 5 illustrates the decomposition of the age 11 disadvantage gap. Of the total 10.1-month gap, 7.4 months (almost three-quarters) is 'explained' by differences in the characteristics and settings of disadvantaged pupils compared to their peers, leaving 2.7 months (27 per cent) 'unexplained'. This

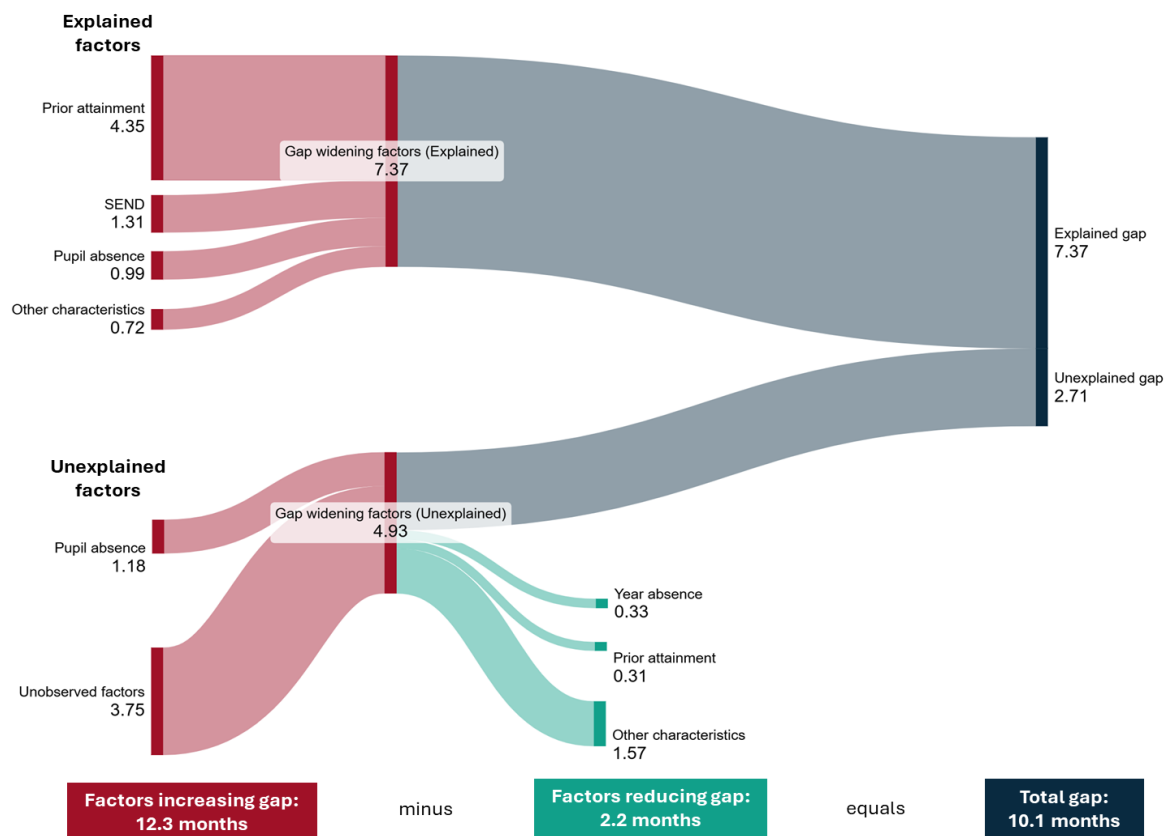
³⁸ Cohort differences control for: average (standardised) prior attainment of the cohort at age five; number of Year 6 pupils (plus squared-term); percentage of Year 6 pupils who are FSM-eligible; percentage of Year 6 pupils with identified SEND; percentage of Year 6 pupils in lower-attaining ethnic groups; percentage of Year 6 pupils arriving late with EAL; Year 6 absence rate; and Year 6 suspension rate.

³⁹ School contexts control for: school type; school absence rate; school suspension rate; and region.

⁴⁰ https://assets.publishing.service.gov.uk/media/66e2cd5a61763848f429d58f/Factors_influencing_primary_school_pupils_educational_outcomes.pdf

reflects how these characteristics and settings have different impacts on disadvantaged pupils' attainment relative to their peers and is equivalent to the fully modelled gap in figure 3.

Figure 5: Decomposition of the disadvantage gap (in months) at KS2 in 2023/24, statistically significant factors contributing $\geq 5\%$ of the gap



$N = 578,807$.

Explained gap

Lower prior attainment among disadvantaged pupils at age five is the largest contributor, adding 4.4 months to the explained gap. This underscores how much of the age 11 gap has already become entrenched by age five and, as noted above, the critical role of early years policies and interventions during children's first year at school.

It is worth remembering that prior attainment at age five is itself shaped by developmental needs, especially SEND. Because disadvantaged children are more likely to have SEND, EYFS scores are partly channelling both socio-economic and developmental inequalities, making it a 'catch-all' proxy for multiple, overlapping needs. Additionally, some of the unobserved gap reflects unidentified SEND and – while this is the case at every key stage – its magnitude, and the types of SEND that dominate, will shift across phases.

Other contributors to the explained gap include SEND characteristics capturing duration and primary need type (adding 1.3 months) and pupils' absence histories over Years 4 to 6 (adding 1.0 month). In analysing absence histories as a contributor to the disadvantage gap, we include both the level of pupil absence and the percentage of overall of absences that are unauthorised.⁴¹

It is worth noting that school disruptions are intertwined with SEND: previous EPI research has found that persistent absence, suspensions and frequent school moves are all risk factors for children not having their needs identified.⁴²

Conversely, our analysis shows that when children with SEND have timely identification and sustained support, their KS2 outcomes can be at least as good as their peers with similar prior attainment and other characteristics. Figure 6 shows that, when Year 6 pupils have been on SEND Support since Year 1 or reception, their KS2 outcomes are slightly better than their peers with no identified needs, once we control for all other factors affecting attainment. This is necessarily subject to selection bias as it only includes those pupils on state school rolls with complete data, including on prior attainment, but it shows how early and effective support for SEND can prevent needs escalating and even gaps emerging.⁴³

Our detailed analysis also finds that Year 6 pupils with certain primary need types have much smaller KS2 attainment gaps than other need types, once we account for other factors. Figure 7 has the same important caveat as figure 6 – that it is subject to selection bias – but among this subset of pupils, those with visual or hearing impairments have a relatively small (two to three months) KS2 attainment gap compared to their peers with no identified needs, once we control for prior attainment and other factors. Conversely, pupils with moderate learning difficulties are almost 11 months behind on average, and those with speech, language and communication needs, around seven months behind.

So, while it is clear that SEND characteristics collectively make a meaningful contribution to the overall KS2 disadvantage gap, there is important variation in outcomes *within* the group of SEND learners.

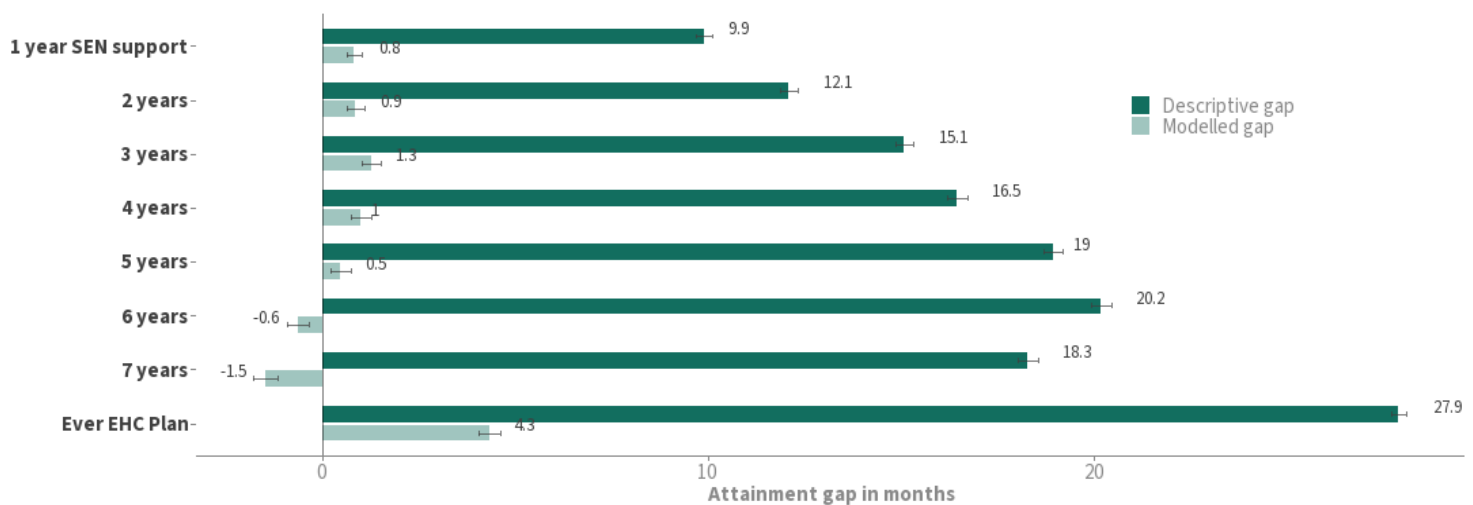
We also see this pattern of heterogeneity among SEND learners at KS4.

⁴¹ We find in our regression models that a higher proportion of absence that is unauthorised (for a given level of overall absence) is associated with a small, negative effect on attainment at both KS2 and KS4.

⁴² <https://epi.org.uk/wp-content/uploads/2025/02/SEND-Final-Report-version-FINAL-04.02.2024-2.pdf>

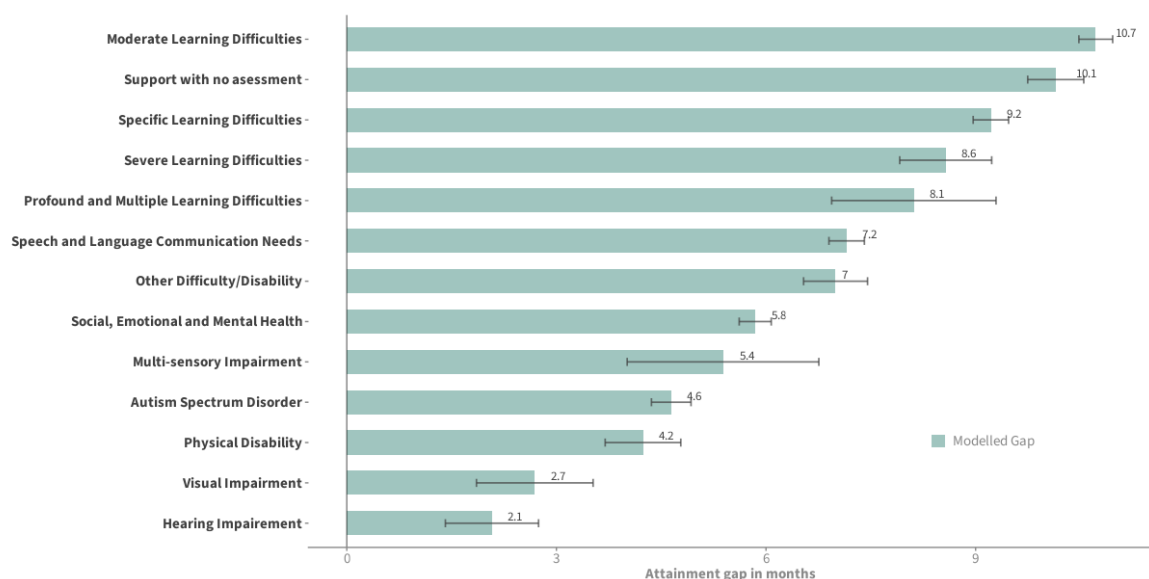
⁴³ Our requirement to have full data to be included in our models means that we lose a disproportionate number of pupils who have had an EHCP at any point during primary school. Our results, therefore, need careful interpretation given that those retained within our analysis may not be representative of all pupils with SEND, particularly those with more complex needs.

Figure 6: KS2 attainment gap in 2023/24 by SEND status over the primary school phase compared to pupils with no identified SEND



Notes: The descriptive gap includes no controls; the modelled gap includes the full set of pupil, cohort, school and regional controls. Within the pupil controls, this includes primary type of SEND need in Year 6. Bars show 95 per cent confidence intervals. N = 578,807

Figure 7: KS2 attainment gap in 2023/24 by SEND primary need type in Year 6 compared to pupils with no identified SEND



Notes: The modelled gap includes the full set of pupil, cohort, school and regional controls. Within the pupil controls, this includes the duration and complexity of SEND during the primary school phase. Bars show 95 per cent confidence intervals. N = 578,807.

Cohort- and school-level factors are less important contributors to the explained gap than pupil-level factors; these are grouped under 'other characteristics. When we explore these in more detail (figure A3 in Annex A), we find that being in schools with higher absence rates further increases the gap (by 0.3

months, over and above pupil-level absence), while being in a year group with higher prior attainment at age five helps narrow the gap (by 0.4 months).

Unexplained gap

Turning to the unexplained gap, the biggest single contributor is pupil absence: even when disadvantaged pupils have similar absence histories to their peers, it appears to be more detrimental to attainment (increasing the gap by 1.2 months). However, it is also possible that there are unobservable differences between absent pupils who are disadvantaged or not (e.g. being more physically or mentally unwell) that could account for this.

Further analysis (figure A4 in Annex A) reveals other factors that make a more modest contribution to the unexplained gap. This includes ethnicity which contributes 0.3 months. As we saw at age five, this suggests that the same ethnic background appears to be associated with different KS2 outcomes depending on whether pupils are disadvantaged or not.⁴⁴ We return to this in the next section on the GCSE disadvantage gap.

Figure A4 in Annex A also shows that the effect of pupils moving school contributes 0.2 months to the (unexplained) gap. Analysis of our regression outputs suggests that school mobility is more detrimental to the KS2 attainment of disadvantaged pupils, though as with all the factors in our model, this association is not necessarily causal. This may be because the circumstances surrounding school mobility may be different for disadvantaged families compared to their non-disadvantaged mobile peers – for example, less planned moves driven by economic pressures or negative school experiences.

The unexplained gap would be larger still if not for some factors appearing to be more beneficial for disadvantaged pupils' attainment. This includes prior attainment at age five. For a given starting point at age five, disadvantaged pupils appear to make *more* progress by the end of primary school than would be expected (given other observable characteristics), and this narrows the unexplained KS2 gap (by 0.3 months). So, while disadvantaged students are more likely to have lower prior attainment (contributing to the explained gap), they are also more likely to make greater progress from a given starting point (contributing to the unexplained gap).

It is possible that some of the contextual factors that shape disadvantaged pupils' EYFS scores (such as pupils' familiarity with school routines, confidence or communication skills) attenuate naturally by KS2, so disadvantaged pupils with the same EYFS score as their peers may have more academic capability than that early score implies. However, other explanations are also possible – for example, if some disadvantaged pupils are 'under-assessed' at EYFS due to teacher judgements or noise in the data, then their later attainment appears to 'catch-up' statistically simply due to measurement error in the baseline.

Additionally, cohort disadvantage in Year 6 helps lower the (unexplained) gap by 0.3 months. Just as we found at age five, analysis of our regression outputs suggests that cohort disadvantage is associated with lower pupil attainment overall, but not for disadvantaged pupils for whom it is associated with higher KS2 attainment.

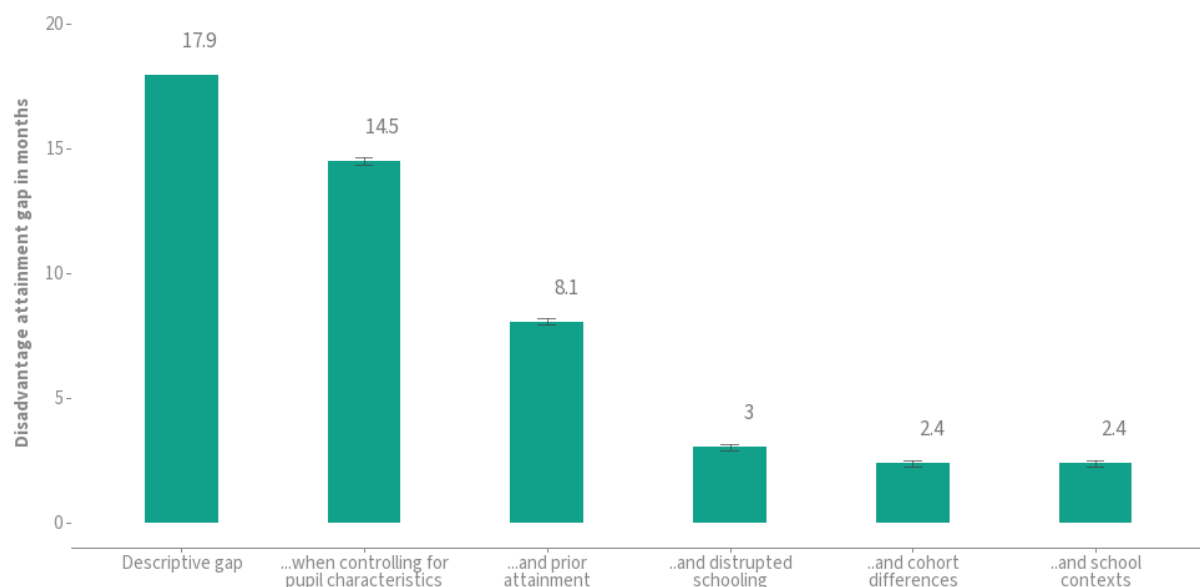
⁴⁴ Disadvantaged pupils from most ethnic backgrounds have higher average KS2 attainment than White British pupils conditional on all other characteristics; however, there are several ethnicities where this advantage is smaller than for the non-disadvantaged pupils of the same background including White Irish, White and Asian, and Indian pupils.

Secondary school

By the end of secondary school, 24 per cent of pupils have been eligible for FSM at some point in the previous six years. We use this as our measure of disadvantage at KS4.

Figure 8 shows that the (descriptive) GCSE gap was 17.9 months in 2023/24 (compared to a gap of 10.1 months at the end of primary school and 4.6 months in reception year). As noted earlier, these relate to different cohorts, but the pattern of gap-widening highlights the cumulative effects of disadvantage as children and young people progress through education.

Figure 8: The disadvantage gap at KS4 in 2023/24, including for models with pupil, cohort, school and regional controls



Notes: The descriptive gap includes no controls; subsequent modelled gaps sequentially include controls for a range of characteristics. Bars show 95 per cent confidence intervals. $N = 509,308$.

The GCSE gap substantially reduces by 3.4 months (to 14.5 months) after controlling for pupil characteristics.⁴⁵ As at KS2, the SEND characteristics of disadvantaged pupils markedly differ from their non-disadvantaged peers in ways that matter for GCSE attainment. For example, by Year 11, disadvantaged pupils:

- are over twice as likely as their non-disadvantaged peers to have had an EHCP at some point during their time in primary or secondary school (4.0 per cent compared to 1.9 per cent);
- have spent longer on SEND Support (with 3.2 per cent having been on SEND Support for 11 years compared with 1.5 per cent of non-disadvantaged pupils); and

⁴⁵ As at KS2, pupil characteristics at KS4 control for: sex; SEND duration and need type; ethnicity; first language; and month of birth.

- are roughly twice as likely as their non-disadvantaged peers to have been identified with certain primary need types including social, emotional and mental health; moderate learning difficulties; and speech, language and communication needs. As we saw at KS2, there is important variation in GCSE outcomes *within* the group of SEND learners by primary need type – see Annex A (Figure A13) for further details.

The GCSE gap narrows by a further 6.4 months (to 8.1 months) once we control for prior attainment at ages 5, 7 and 11.⁴⁶ This underscores the critical role of early years and primary-phase policy and interventions in narrowing the GCSE gap, by boosting disadvantaged pupils' attainment before and during primary school.

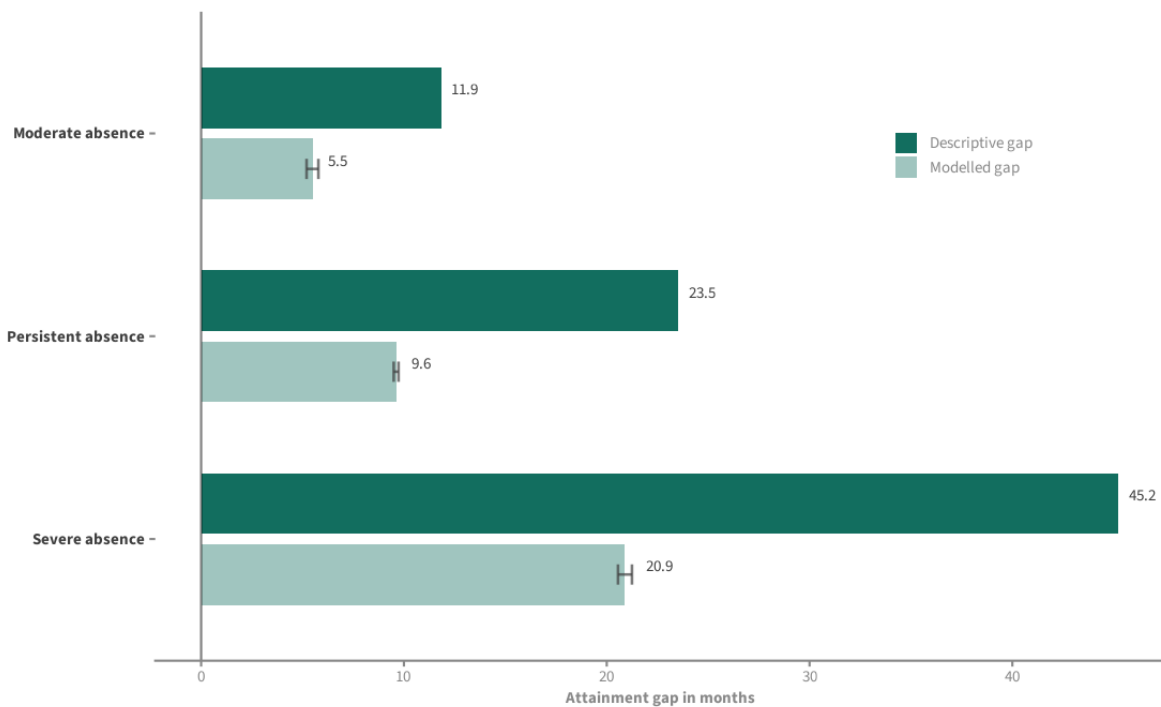
Controlling for pupils' histories of absence, suspensions, permanent exclusions and school moves during the final three years of secondary school further reduces the GCSE gap by 5.1 months (to 3.0 months). This highlights that the greater disruption to schooling experienced by disadvantaged pupils is a major influence on the GCSE gap and raises an important question as to its cause.

In particular, disadvantaged pupils are much more likely to be absent and, as we saw at KS2, increased pupil absence is associated with lower attainment. Figure 9 shows that, conditional on all other observable characteristics, even moderate levels of pupil absence over Years 9 to 11 are associated with lower GCSE attainment at KS4 (by 5.5 months) compared to pupils whose absence levels are below the national average. This attainment penalty grows to 9.6 months among persistently absent pupils and to 20.9 months among severely absent pupils.

As previously noted, these are associative not causal estimates, with pupil absences often symptomatic of other barriers to attaining well, such as unidentified SEND, poor mental health, school disengagement and poverty. This matters because without identifying and addressing the underlying causes of absence, policy responses risk not only being ineffective in improving attendance but potentially compounding the challenges facing vulnerable learners.

⁴⁶ Prior attainment controls for a pupil's average (standardised) attainment at ages 5, 7 and 11 – specifically: EYFS score in each of the seven learning goals; KS1 score in each of mathematics, reading and writing; and KS2 score in each of mathematics, reading, and grammar, punctuation and spelling.

Figure 9: KS4 attainment gap in 2023/24 by levels of pupil absence over Years 9 to 11 (relative to pupils with below-average absence), including with full pupil, cohort, school and regional controls



Notes: The descriptive gap includes no controls; the modelled gap includes the full set of pupil, cohort, school and regional controls. Persistently absent pupils miss 10 per cent of school sessions (but below 50 per cent); severely absent pupils miss 50 per cent or more. N = 509,308.

Adding controls for cohort characteristics in Year 11 has modest effects on the gap, reducing it by 0.6 months (to 2.4 months), while further controls for school characteristics collectively make no difference to the gap.^{47, 48} This reinforces wider evidence that pupil-level factors exert a stronger influence on the disadvantage gap than cohort or school-level factors – a finding that is consistent across our models – though, as we see in our 16-19 model in the following section, the influence of peers seems to become more important as pupils become older.⁴⁹

In our fully specified model, disadvantaged pupils remain 2.4 months behind their peers at the end of secondary school even when they have similar (observable) characteristics, similar recent school histories, similar peers and attend similar schools.

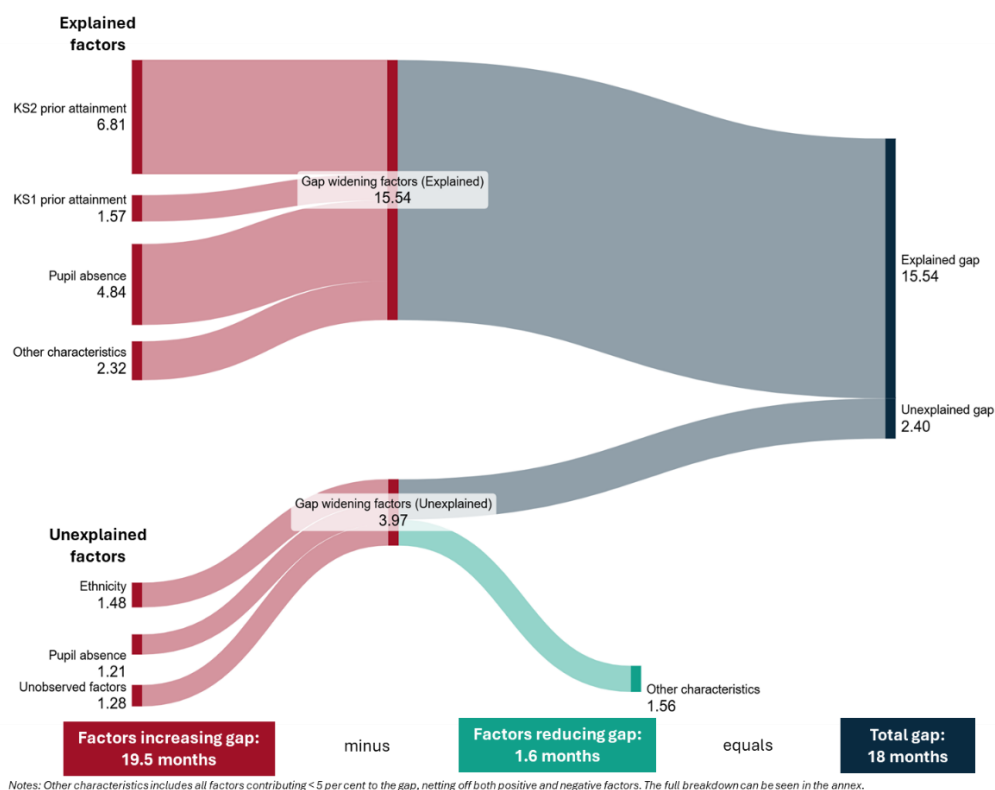
⁴⁷ Cohort differences control for: average (standardised) prior attainment of Year 11 cohort at age 11; number of Year 11 pupils (plus squared-term); percentage of Year 11 pupils who are FSM-eligible; percentage of Year 11 pupils with identified SEND; percentage of Year 11 pupils in lower-attaining ethnic groups; percentage of Year 11 pupils with EAL; Year 11 absence rate; Year 11 suspension rate; and Year 11 permanent exclusions rate.

⁴⁸ School contexts control for: school type; school absence rate; school permanent exclusions rate; and region.

⁴⁹ https://assets.publishing.service.gov.uk/media/66e4006e3f1299ce5d5c3e11/Factors_influencing_secondary_school_pupils_educational_outcomes.pdf

Figure 10 illustrates the decomposition of the GCSE disadvantage gap. Of the total 17.9 months gap, 15.5 months (87 per cent) is ‘explained’ by differences in the characteristics and contexts of disadvantaged pupils compared to their peers, leaving just 2.4 months (13 per cent) ‘unexplained’. This reflects the differential impacts of these characteristics and contexts on disadvantaged pupils’ attainment and is equivalent to the fully modelled gap in figure 8.

Figure 10: Decomposition of the disadvantage gap (in months) at KS4 in 2023/24, statistically significant factors contributing $\geq 5\%$ of the gap



$N = 509,308$.

Explained gap

It is notable that more of the GCSE gap is explained by differences in (observable) characteristics between disadvantaged and non-disadvantaged pupils than at younger ages.⁵⁰ This includes granular measures of prior attainment (at ages 5, 7 and 11), plus pupils’ experiences of disrupted schooling, which are both key mechanisms through which disadvantage leads to lower GCSE attainment.

By far the largest contributor is prior attainment at KS2. This adds 6.8 months to the (explained) gap. Notably, attainment at KS1 and in the EYFS still contributes over and above this (adding 1.6 months and 0.3 months, respectively – see figure A5 in Annex A). This indicates that earlier attainment measures

⁵⁰ The explained gap accounts for 30 per cent of the overall gap at age five and 73 per cent of the overall gap at age 11.

capture independent components of disadvantage: disparities emerging as early as ages five to seven still exert a meaningful effect on the GCSE gap over and above differences in later KS2 attainment. That the foundations for later disparities are laid early in children's lives underscores that early intervention is critical (as discussed previously), as attaining well during primary school seems to provide a protective factor against later gap-widening during secondary school.

However, 'prior attainment' should not be interpreted as a narrow academic measure which exerts a separate influence. Particularly by age 11 when it is a well-measured construct, prior attainment encapsulates the cumulative effects of inequalities established years prior to taking assessments, including early development needs, health, and the extent to which SEND was identified and supported at younger ages. Much of what appears as the 'effect' of KS2 attainment on later attainment is, in practice, a wide, overlapping set of influences, only some of which may be amenable to education policy.

The next most important contributor is pupil absence during Years 9 to 11 which adds 4.8 months to the (explained) gap. Its contribution to the national disadvantage gap is so sizeable – much more so than other aspects of disrupted schooling which also harm GCSE attainment to a lesser degree – because incidences of pupil absence are widespread, particularly among disadvantaged pupils, by Years 9 to 11.

Other factors are less important to the explained gap, each contributing under one month (figure A5 in Annex A). Within these, pupils' histories of suspensions contribute 0.8 months, while cohort- and school-level factors all make smaller contributions. Gaps tend to be slightly wider in schools with higher rates of absence and permanent exclusions, though slightly narrower in cohorts with higher rates of suspension in Year 11 (even though suspended pupils do worse individually).⁵¹

Unexplained gap

Turning to the unexplained gap, the most important factor is ethnicity, contributing 1.5 months. As we saw at EYFS and KS2, this suggests that the association between ethnicity and attainment differs for disadvantaged pupils, with White British disadvantaged pupils making some of the slowest progress during secondary school. However, an important caveat is that pupils arriving late to English secondary schools are out-of-scope because of missing data. Because our models require complete linked records back to age five, pupils who enter the system late –disproportionately recent migrants and asylum seekers from vulnerable minority ethnic groups –are necessarily excluded. Our ethnicity coefficients, therefore, reflect the outcomes of the less vulnerable subset of minority ethnic pupils for whom full data exist, and our estimates likely understate the attainment penalties associated with minority groups experiencing more late arrival and earlier trauma.

Other studies have also found that White British disadvantaged students consistently make slower academic progress than most minority ethnic groups, after controlling for prior attainment and other factors.⁵² In his 2014 analysis of the original longitudinal study of young people in England (LSYPE), Professor Steve Strand attributes the greater resilience of disadvantaged minority pupils to four key factors:

⁵¹ On average, each additional suspension over Years 9 to 11 is associated with lower KS4 attainment of 1.0 month; each permanent exclusion is associated with 2.7 months lower attainment; each school move is associated with 1.4 months lower attainment. These estimates are associative rather than causal; full regression results are available on request.

⁵² https://www.education.ox.ac.uk/wp-content/uploads/2021/05/Strand_2021_Report-to-CRED.pdf



- high educational aspirations among parents, and students themselves, to continue in full-time education post-16 and to attend university, placing education in a central role for achieving future goals;
- pupils' strong academic self-concept;
- high levels of motivation and homework completion; and
- lower levels of truancy and exclusion..⁵³

As we saw at KS2, pupil absence is another key contributor to the unexplained gap (adding 1.2 months) suggesting that it is even more detrimental for disadvantaged pupils to be absent from school.⁵⁴ Policies and interventions to reduce pupil absence and mitigate its effects are therefore important, not just in their own right but as part of a wider strategy to narrow primary and secondary school disadvantage attainment gaps.

Further analysis (figure A6 in Annex A) shows that there are also some factors that help narrow the unexplained gap, notably KS2 prior attainment. For a given starting point at age 11, disadvantaged pupils appear to make *more* progress in their GCSEs than would be expected (given other observable characteristics), and this narrows the unexplained gap (by 0.9 months).

This initially seems counterintuitive, though it mirrors what we saw at KS2 and several reasons are possible. One is that some disadvantaged pupils have underlying capabilities that were not fully captured in their KS2 scores – for example, if earlier assessments understated their skills due to measurement error, missed learning, or contextual barriers that lessen over time. It is also possible that some additional needs (such as social, emotional and mental health) influence early scores but are better identified and supported in secondary school.

As we found at earlier ages, cohort- and school-level factors tend to make only small contributions of under 0.2 months, with cohort disadvantage helping to narrow the unexplained gap slightly (by 0.1 month). Analysis of our regression outputs shows that, in contrast to the early years and KS2, cohort disadvantage is associated with lower pupil attainment for non-disadvantaged and disadvantaged pupils alike, though the GCSE 'penalty' is greater for non-disadvantaged pupils causing the (unexplained) gap to narrow slightly.

Persistent disadvantage at secondary school

Previous research by EPI has shown that persistently disadvantaged pupils have some of the widest attainment gaps, falling nearly one year behind at KS2 and nearly two years behind at KS4.⁵⁵ These pupils are eligible for FSM for at least 80 per cent of their time in school, yet the school funding system (including the Pupil Premium) does not distinguish between persistent and short-term disadvantage, meaning there is no additional support or policy focus for these pupils.

In this report, we focus on the persistent disadvantage gap at KS4 to better understand the scale and nature of the gap for these pupils in long-term poverty. Consistent with previous EPI research, figure 11

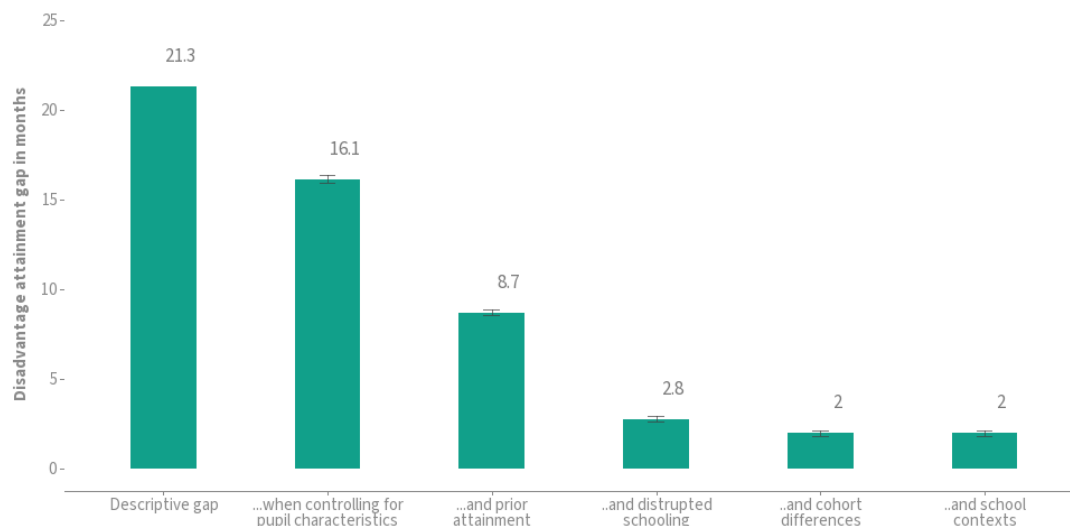
⁵³ https://www.researchgate.net/publication/264412254_Ethnicity_gender_social_class_and_achievement_gaps_at_age_16_Intersectionality_and_'Getting_it'_for_the_white_working_class

⁵⁴ However, as discussed in our KS2 findings, it is also possible that there are unobservable differences between absent pupils who are disadvantaged or not (e.g. being more physically or mentally unwell) that could account for this.

⁵⁵ <https://epi.org.uk/annual-report-2025/>

confirms a large (descriptive) persistent disadvantage gap at GCSE of 21.3 months, compared to 17.9 months for all disadvantaged pupils. These pupils likely face additional challenges including poor mental health (own and caregivers'), physical health, and worse longer-term outcomes after leaving compulsory education.^{56, 57, 58}

Figure 11: The persistent disadvantage gap at KS4 in 2023/24, including for models with pupil, cohort, school and regional controls



Notes: The descriptive gap includes no controls; subsequent modelled gaps sequentially include controls for a range of characteristics. Bars show 95 per cent confidence intervals. N = 441,855.

This gap substantially reduces by 5.2 months (to 16.1 months) after controlling for pupil characteristics. As for all disadvantaged pupils, SEND is a key factor associated with lower attainment among persistently disadvantaged pupils. We find that 47 per cent of persistently disadvantaged pupils have been SEND at some point over the previous 11 years, compared to 42 per cent of all disadvantaged pupils and 25 per cent of non-disadvantaged pupils. Persistently disadvantaged pupils are also more likely to have primary need types related to social emotional and mental health (8.7 per cent compared to 7.3 per cent of all disadvantaged pupils and 3.0 per cent of non-disadvantaged pupils) and moderate learning difficulties (4 per cent compared to 3.2 per cent of all disadvantaged pupils and 1.5 per cent of non-disadvantaged pupils).

The gap narrows by a further 7.4 months (to 8.7 months) once we control for prior attainment at ages 5, 7 and 11. This is a (one month) larger reduction than we saw in figure 8 for all disadvantaged pupils, reflecting that the subset of pupils in long-term poverty tend to have the lowest levels of prior attainment.

⁵⁶ <https://www.sciencedirect.com/science/article/pii/S1054139X23003993>

⁵⁷ <https://adc.bmj.com/content/archdischild/early/2019/05/15/archdischild-2018-316702.full.pdf?ijkey=hKrHArIAeu7gg9F&keytype=ref>

⁵⁸ https://view.officeapps.live.com/op/view.aspx?src=https%3A%2F%2Fffteducationdatalab.org.uk%2Fwp-content%2Fuploads%2F2023%2F01%2Ffinal_report_inc_exec_summary.docx&wdOrigin=BROWSELINK

Even by the end of reception year, we find that just 37 per cent of persistently disadvantaged pupils reached a ‘good level of development’ on the EYFS compared to 42 per cent of all disadvantaged pupils and 59 per cent of non-disadvantaged pupils.⁵⁹

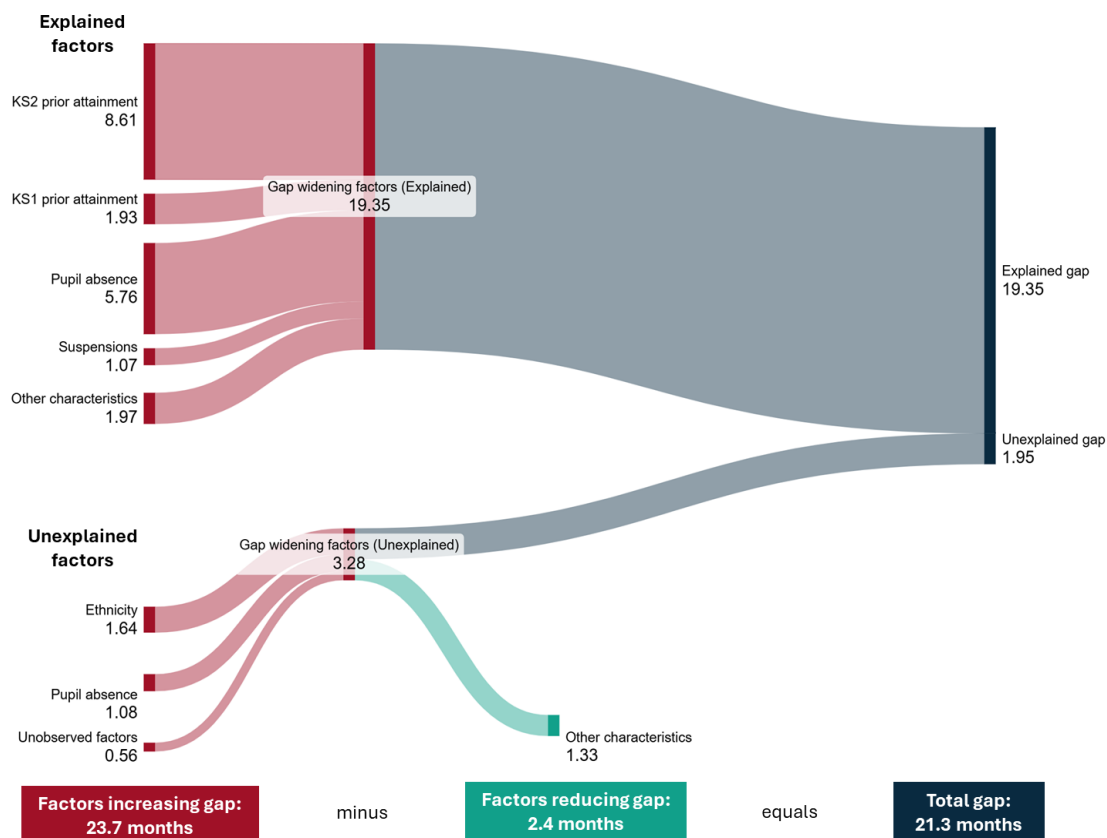
Controlling for pupils’ histories of absence, suspensions, permanent exclusions and school moves, during the final three years of secondary school, further reduces the persistent disadvantage gap by 5.9 months. The resulting gap (2.8 months) is a *smaller* gap than when we include an equivalent set of controls for all disadvantaged pupils (3.0 months). This indicates that persistent disadvantage primarily operates through observable pathways captured in our model (i.e. pupil characteristics, lower prior attainment and their experiences of disrupted schooling), rather than through unmeasured factors.

Adding additional controls for peer group characteristics in Year 11 has modest effects on the persistent disadvantage gap, reducing it by 0.8 months (to 2.0 months), while further controls for school characteristics collectively make no difference to the gap. In our fully specified model, persistently disadvantaged pupils remain 2.0 months behind their peers at the end of secondary school conditional on their (observable) characteristics, recent school histories, peer and school characteristics.

Figure 12 shows the decomposition of the persistent disadvantage gap. Of the total 21.3 months gap, 19.4 months (91 per cent) is ‘explained’ by differences in the characteristics and settings of persistently disadvantaged pupils compared to their peers, leaving just 2.0 months (nine per cent) ‘unexplained’.

⁵⁹ Persistently disadvantaged pupils score less well, on average, across all seven areas of learning in the EYFS than disadvantaged pupils as a whole but for ease of interpretation, here we use the Department for Education’s threshold measure of a ‘good level of development’. This relates to reaching the expected level in 12 of the 17 early learning goals (within five of the seven areas of learning).

Figure 12: Decomposition of the persistent disadvantage gap (in months) at KS4 in 2023/24, statistically significant factors contributing $\geq 5\%$ of the gap



$N = 441,855$.

Explained gap

The persistent disadvantage gap is shaped by many of the same mechanisms that underpin the overall KS4 disadvantage gap but in some cases, their relative influence is more pronounced for pupils in long-term poverty. As in the KS4 model, prior attainment at KS2 remains the dominant contributor to the (explained) gap but its effect is larger (both in relative and absolute terms), adding 8.6 months – or 40 per cent – to the total persistent disadvantage gap (compared with a contribution of 6.8 months or 38 per cent to the overall KS4 gap).⁶⁰

The next biggest contributor after prior attainment is pupil absence, contributing 5.8 months to the (explained) gap – this is one month larger than for the KS4 model (4.8 months), though not in relative terms (each contributing 27 per cent).

⁶⁰ This compares to 6.8 months – or 38 per cent – to the overall KS4 gap of 17.9 months.

Pupil suspensions over Years 9 to 11 contribute 1.1 months to the (explained) persistent disadvantage gap – again slightly larger than for the overall KS4 gap (0.8 months), though the same contribution (of five per cent) in relative terms.

Other factors at a cohort- and school-level are less important to the explained gap, each contributing under one month (figure A7 in Annex A).

Unexplained gap

Within the unexplained gap, ethnicity is the largest contributor, accounting for 1.6 months, similar to the KS4 model (1.5 months). It indicates that even after controlling for extensive pupil-, cohort- and school-level characteristics, some ethnic groups attain less well at GCSE, including White British pupils. As noted previously, an important caveat here is that pupils arriving late to English secondary schools are out-of-scope because of missing data. Our findings reflect the less vulnerable subset of ethnic minority pupils for whom we have full data, including their prior attainment going back to age five.

As discussed within our KS4 findings, some ethnic minority groups appear to benefit from ‘protective’ factors related to educational aspirations, academic self-concept, and specific behaviours around homework and attendance.⁶¹ By contrast, the profoundly poor educational outcomes for persistently disadvantaged White British pupils likely reflect the unmitigated effect of intergenerational poverty.

As for the overall KS4 gap, further analysis (figure A8 in Annex A) shows that there are also some factors that help narrow the (unexplained) persistent disadvantage gap, notably KS2 prior attainment. Consistent with our finding for all disadvantaged pupils, for a given starting point at age 11, persistently disadvantaged pupils appear to make *more* progress in their GCSEs than would be expected (given other observable characteristics), and this narrows the unexplained gap (by 1.1 months). Once again, cohort- and school-level factors tend to make only small contributions, with cohort disadvantage helping to narrow the unexplained gap slightly (by 0.3 months).

Taken together, pupils in long-term poverty encounter some of the most entrenched educational inequalities, notably in relation to much lower prior attainment than their non-disadvantaged peers. Policies and interventions that improve pupil attainment during primary school should not only help narrow the overall disadvantage gap but particularly benefit pupils in long-term poverty.

16-19 education

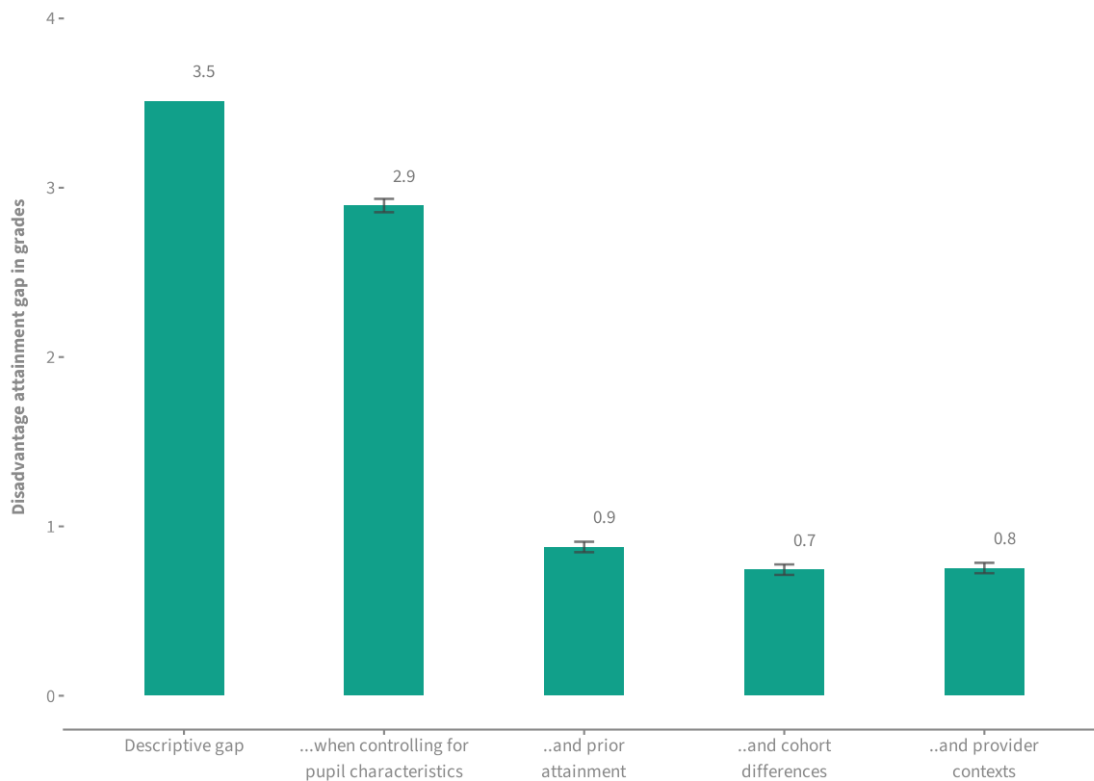
Figure 13 shows that the (descriptive) gap in 16-19 attainment was 3.5 grades in 2023/24. The gap reduces just over half a grade (to 2.9 grades) after controlling for pupil characteristics. The gap narrows substantially once we control for prior attainment at ages 5, 7, 11 and 16, to just under a grade. As at KS4, this highlights the importance of early and sustained interventions in narrowing attainment gaps.

Adding further controls for cohort (other students completing their 16-19 education at the same time) characteristics has a smaller impact on the gap, reducing by around a fifth of a grade, to 0.7 grades. However, adding controls for provider context (provider type and region) slightly increases the gap. This is likely because disadvantaged students are over-represented in London, where outcomes tend to be

⁶¹ https://www.researchgate.net/publication/264412254_Ethnicity_gender_social_class_and_achievement_gaps_at_age_16_Intersectionality_and_'Getting_it'_for_the_white_working_class

higher, even after accounting for student characteristics. This can be observed from our other regression outputs, as well as from other sources of evidence.⁶²

Figure 13: The 16-19 disadvantage gap in 2023/24, including for models with pupil, cohort and provider controls



Notes: The descriptive gap includes no controls; subsequent modelled gaps sequentially include controls for a range of characteristics. Bars show 95 per cent confidence intervals. N = 433,580.

Figure 14 illustrates the decomposition of the 16-19 disadvantage gap in 2024, showing how differences in the characteristics of disadvantaged and non-disadvantaged students, and the attainment penalty associated with those characteristics, contribute to the overall gap. Of the total 35-point gap (equivalent to 3.5 A level grades), 28 points (almost 80 per cent) is ‘explained’ by observed differences in student and provider characteristics, while the remaining 20 per cent is ‘unexplained’. This unexplained component is equivalent to the fully modelled gap shown in figure 13.

⁶² <https://epi.org.uk/annual-report-2025/>

Explained gap

Prior attainment is by far the largest contributor to the gap, with prior attainment across all earlier phases accounting for 89 per cent of the explained gap, and 70 per cent of the total gap. This reflects the substantial differences in pre-16 attainment between disadvantaged students and their peers and underscores the cumulative influence of attainment earlier in the education system.

Cohort characteristics also make a notable contribution. The composite ‘cohort disadvantage’ indicator (derived from cohort prior attainment, SEND and FSM) accounts for over a third of a grade. Specifically, students in cohorts that have lower prior attainment, with higher levels of SEND and disadvantage have lower attainment, and disadvantaged students are more likely to be in these cohorts.

Cohort characteristics appear to become more influential on the gap at older ages. At age five and age 11, all cohort variables contribute a negligible proportion of the explained gap. By age 16 this increases to seven per cent. In the 16-19 phase differences in cohort characteristics contribute 11 per cent to the gap. This is consistent with other research that shows peer influence increases, and adult influence decreases in this stage of development.⁶³

No other factor contributes more than four per cent to the explained gap (either narrowing or widening). However, when these remaining factors are aggregated, they make a net contribution to a narrowing of the gap of 0.15 grades (four per cent).⁶⁴

Unexplained gap

Almost eight points, or around a fifth, of the gap remains after controlling for differences in the observable characteristics of disadvantaged students. This may include the direct effects of economic disadvantage itself, or other unobserved factors not captured in administrative data. For example, it may partly reflect that our post-16 measure of disadvantage is (necessarily) based on pre-16 eligibility for FSM. As in earlier stages of education, this unexplained component indicates that disadvantaged students fall further behind their peers during 16-19 education, even when they begin with similar prior attainment and have otherwise similar characteristics.

As in KS4, ethnicity also contributes to the unexplained gap, accounting for 10 per cent (around a third of a grade) of the overall gap. This suggests that the association between ethnicity and attainment differs for disadvantaged students. Analysis of regression coefficients from separate disadvantaged and non-disadvantaged models suggests that this is largely driven by the lower progress of disadvantaged White British students during the 16-19 phase.

KS4 prior attainment helps to narrow the unexplained gap, reducing it by almost a quarter of a grade. Although disadvantaged students have both lower attainment and prior attainment, the association between prior attainment and 16-19 attainment is stronger than it is for non-disadvantaged students. Several factors may drive this, including the greater influence of prior attainment on disadvantaged students’ progress, differences in GCSE entry patterns (both subject choices and number of qualifications), and differences in post-16 qualification choices. For example, disadvantaged students

⁶³ <https://doi.org/10.1016/j.adolescence.2017.07.002>

⁶⁴ The remaining factors include: provider type; early years and KS1 prior attainment; month of birth; gender; ethnicity; cohort ethnicity; cohort size; region; and SEND. More detail can be seen in figure A9 in Annex A.

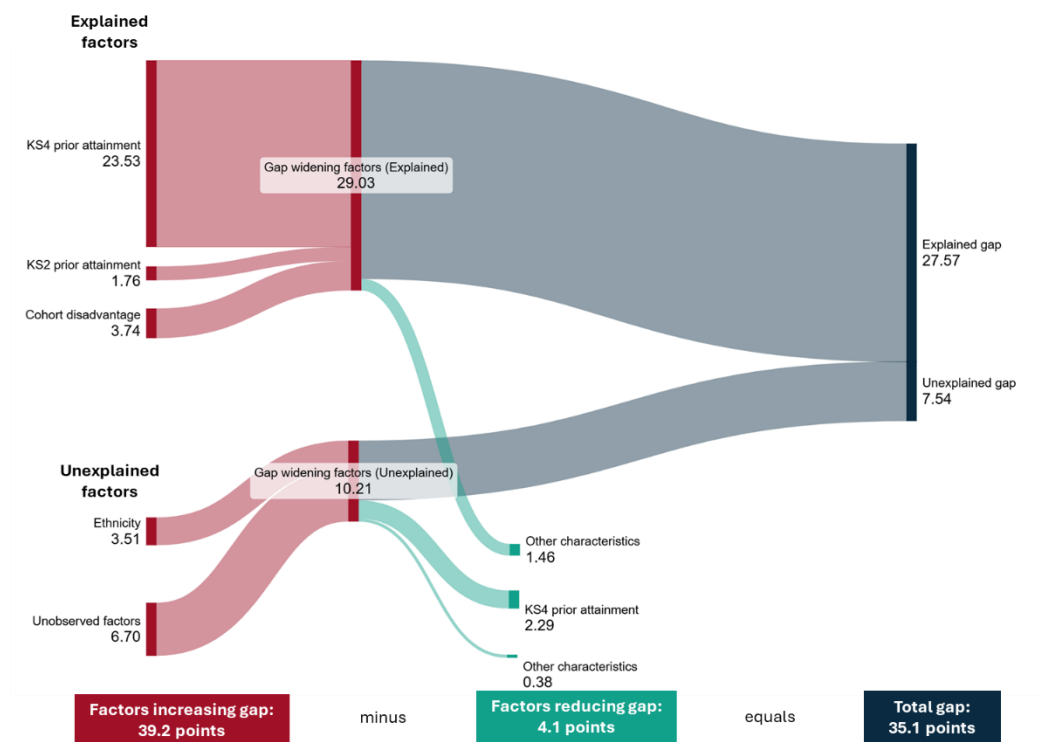
are more likely to take technical or vocational qualifications than non-disadvantaged students, and at a given level of prior attainment these qualifications are likely to secure more points in our 16-19 attainment measure (driven by our assumption of equivalence between academic and other qualifications).

Figure A11 and figure A12 in Annex A show additional decompositions of the 16-19 gap, for students in further education (FE) colleges and school sixth forms, respectively. They suggest that the drivers of the 16-19 disadvantage gap differ in important ways between FE colleges and school sixth forms, reflecting differences in student composition, qualification pathways and institutional context:

- In both FE colleges and school sixth forms, cohort characteristics make a smaller contribution to the disadvantage gap than in the pooled 16-19 analysis. This largely reflects the strong correlation between institution type and cohort disadvantage at this stage: once analysis is conducted within institution types, much of the variation associated with cohort composition is absorbed by institutional sorting rather than operating independently within settings.
- Within the unexplained component, cohort disadvantage in FE colleges contributes to a reduction in the unexplained gap. Analysis of our regression outputs suggests this reflects a stronger association between cohort disadvantage and lower attainment for non-disadvantaged students than for disadvantaged students, resulting in a larger attainment ‘penalty’ for the former. Alternatively, it may be that colleges with higher proportions of disadvantaged students are better adapted to supporting those students, mitigating the potential negative impact of being in a more disadvantaged setting.
- In school sixth forms, SEND appears to make a more significant contribution to the unexplained gap, of around a quarter of a grade. This indicates that the attainment penalty associated with SEND is larger for disadvantaged pupils in school sixth forms than in other settings, such as FE colleges. This may reflect differences in institutional provision types: school sixth forms tend to be more academically orientated, which may amplify attainment penalties for disadvantaged pupils with SEND relative to FE settings. In contrast, FE colleges typically offer a broader mix of vocational, applied and lower-level qualifications, which may mitigate the disadvantage-related SEND penalty for this group.
- Taken together, these patterns underline that post-16 disadvantage gaps are shaped not only by prior attainment but also by institutional sorting and pathway choice, with FE colleges and school sixth forms serving distinct roles within the 16-19 system.



Figure 14: Decomposition of the 16-19 disadvantage gap (in points) in 2023/24, statistically significant factors contributing $\geq 5\%$ of the gap



Notes: 10 points = 1 A level grade. Other characteristics includes all factors contributing < 5 per cent to the gap, netting off both positive and negative factors. The full breakdown can be seen in the annex.

$N = 433,580$.

Further findings by prior attainment

In this section, we break down our headline disadvantage gaps using more detailed measures of prior attainment. Our aim is to better understand the relative influence that different subjects may have on the overall gap.

However, our supplementary analysis shows that attainment across individual subjects can be highly correlated, as students who perform well in one subject tend to do well in others. Therefore, it is difficult to isolate the exact influence of any single subject with precision, as subjects share variance when the model apportions the gap to different factors.

For this reason, it is important not to overinterpret these subject-specific findings. To ensure our findings remain robust, our earlier headline results combined these granular measures into a single overall measure of prior attainment at each key stage.

Primary school

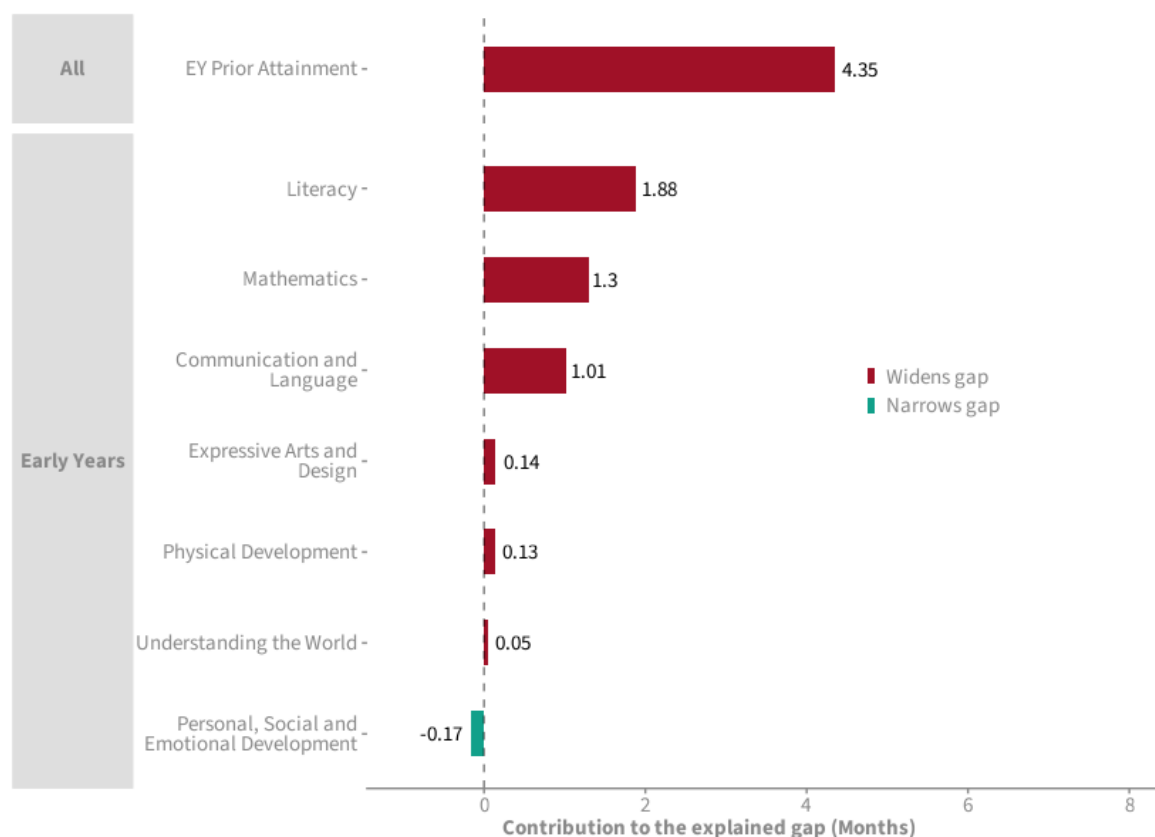
By the end of primary school (KS2) disadvantaged pupils were 10.1 months behind their non-disadvantaged peers. Our analysis found that 73 per cent of this gap (7.4 months) is ‘explained’ by differences in the characteristics and settings of disadvantaged pupils compared to their peers. Within this explained contribution, prior attainment at age five is the most important factor, contributing 4.4 months.

To understand this contribution further, we disaggregate the seven areas of learning in the EYFS measure, while noting that the EYFS is teacher assessed and not externally validated.

Explained gap

Figure 15 shows that three of the seven areas of learning appear to make a meaningful contribution to the KS2 explained gap: literacy (which adds 1.9 months), mathematics (adding 1.3 months) and communication and language (adding 1.0 month). The contribution of each of the other four areas is small at under 0.2 months (less than two per cent of the total gap). This reflects that while disadvantaged pupils tend to achieve less well than their peers across the board, the largest EYFS gaps emerge in literacy, mathematics, and communication and language.

Figure 15: Contribution of prior attainment to the ‘explained’ disadvantage gap (in months) at KS2 in 2023/24



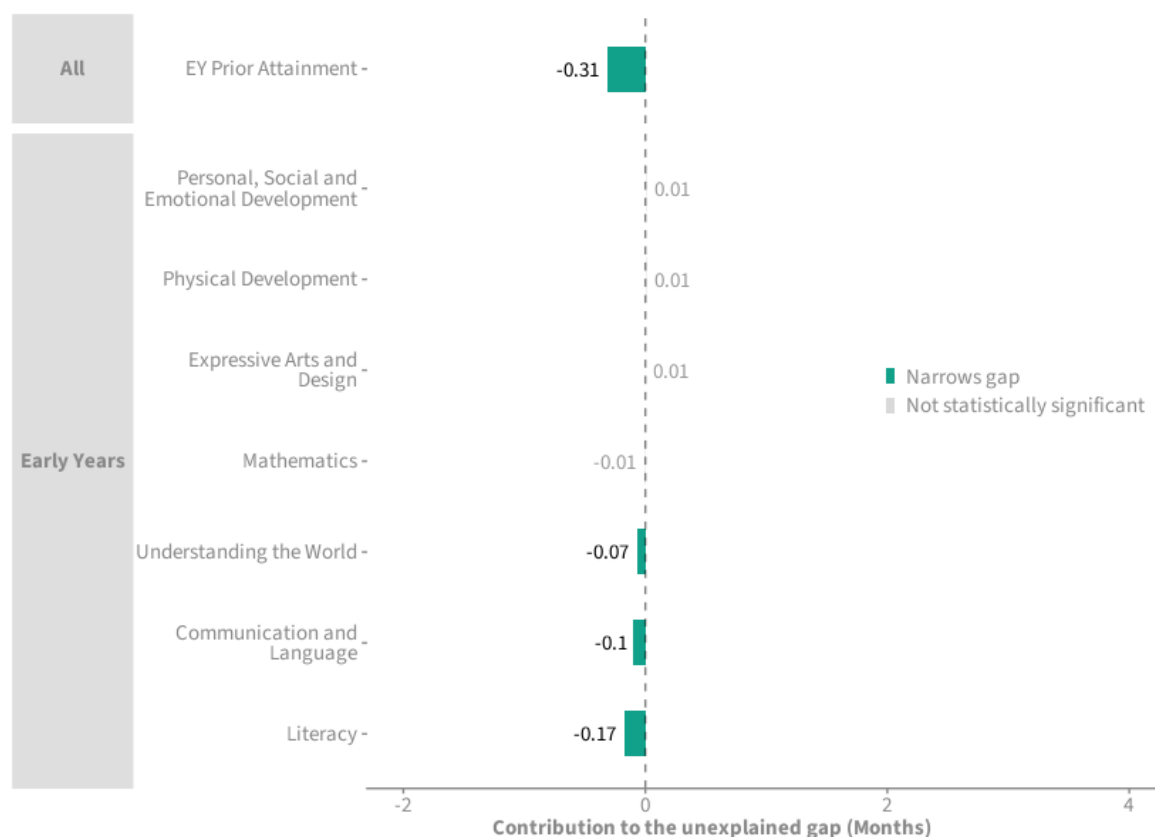
N = 578,807.

Unexplained gap

Turning to the ‘unexplained’ KS2 gap, our main analysis found that disadvantaged pupils remain 2.7 months behind their peers at the end of primary school even when they have similar (observable) characteristics, similar school histories, similar peers and attend similar schools.

Our earlier discussion highlighted that this 2.7 month ‘unexplained’ gap would be 0.3 months larger still if not for prior attainment at EYFS. Our disaggregated analysis in figure 16 shows that it is in literacy that disadvantaged pupils appear to make the most progress over primary school (relative to what would have been expected given other observable characteristics), which helps narrow the unexplained gap by 0.2 months. Other areas of learning appear to make little or no difference.

Figure 16: Contribution of prior attainment to the ‘unexplained’ disadvantage gap (in months) at KS2 in 2023/24



N = 578,807.

As discussed earlier in the report, interpreting the role of prior attainment in the unexplained gap is not straightforward. Prior attainment is measured imperfectly and captures a wide footprint of developmental, health and contextual factors. Assessments contain noise and – in the case of the EYFS – rely on teacher judgements and therefore vary in how well they capture pupils’ underlying capabilities. As a result, some of the apparent ‘catch-up’ among disadvantaged pupils with the same EYFS scores in later phases could reflect regression-to-the-mean rather than genuine differences in progress. This, in turn, would mean that part of the ‘unexplained’ gap-narrowing associated with prior attainment reflects a statistical artefact rather than real educational progress.

Overall, figure 15 and figure 16 aim to provide some sense of the most important mechanisms through which prior attainment shapes the KS2 gap. This seems to indicate that it is primarily through disadvantaged pupils having already fallen behind in reception year in literacy, mathematics, and communication and language, and if disadvantaged pupils achieved as well as their peers in these subjects, the KS2 gap would be substantially smaller.

However, it is important not to overstate the precision of the individual contributions. Our statistical tests found that some of the EYFS predictors have moderate multicollinearity meaning that some shared variance exists when the model apportions the gap to specific factors, reducing their precision.

Specifically, when using the adjusted GVIF, there is moderate multicollinearity for communication and language, and literacy, and it is borderline for mathematics and ‘understanding the world’.

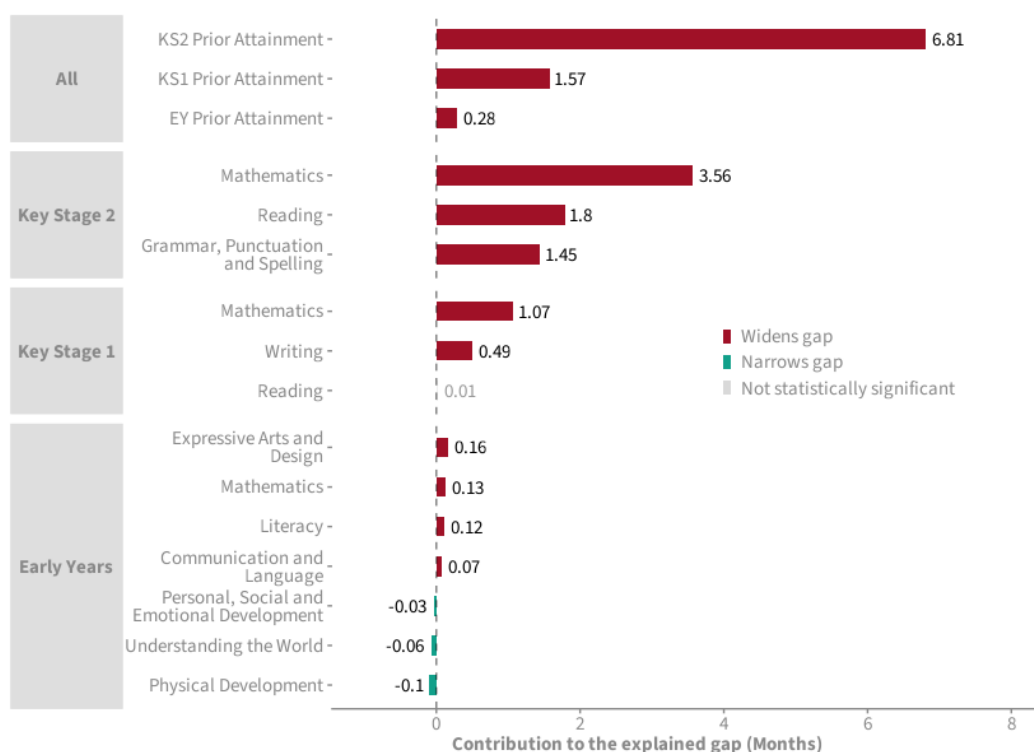
Secondary school

In our KS4 analysis we found that of the total 17.9-month gap, 15.5 months (or 87 per cent) is ‘explained’ by differences in the characteristics and settings of disadvantaged pupils compared to their peers. Within this, KS2 prior attainment is by far the most important factor, contributing 6.8 months. Yet attainment at KS1 and in the EYFS still contributes over and above this, adding 1.6 months and 0.3 months, respectively. This indicates that early intervention may act as a protective factor against a drop-off in progress during secondary school, over and above its role in narrowing the gap by boosting pupils’ attainment on entry to secondary school.

Explained gap

Here we further disaggregate the KS4 gap using more granular measures of prior attainment. Figure 17 shows that within KS2, mathematics appears to make the biggest contribution to the explained KS4 gap, adding 3.6 months, followed by reading (1.8 months), and ‘grammar, punctuation and spelling’ (1.5 months), reflecting that disadvantaged pupils tend to score less well than their peers across the KS2 curriculum.⁶⁵

Figure 17: Contribution of prior attainment to the ‘explained’ disadvantage gap (in months) at KS4 in 2023/24



N = 509,308.

⁶⁵ KS2 writing and science are not included on the basis that these categorical variables are scored differently to the other KS2 subjects, complicating interpretation of subjects’ relative contributions.

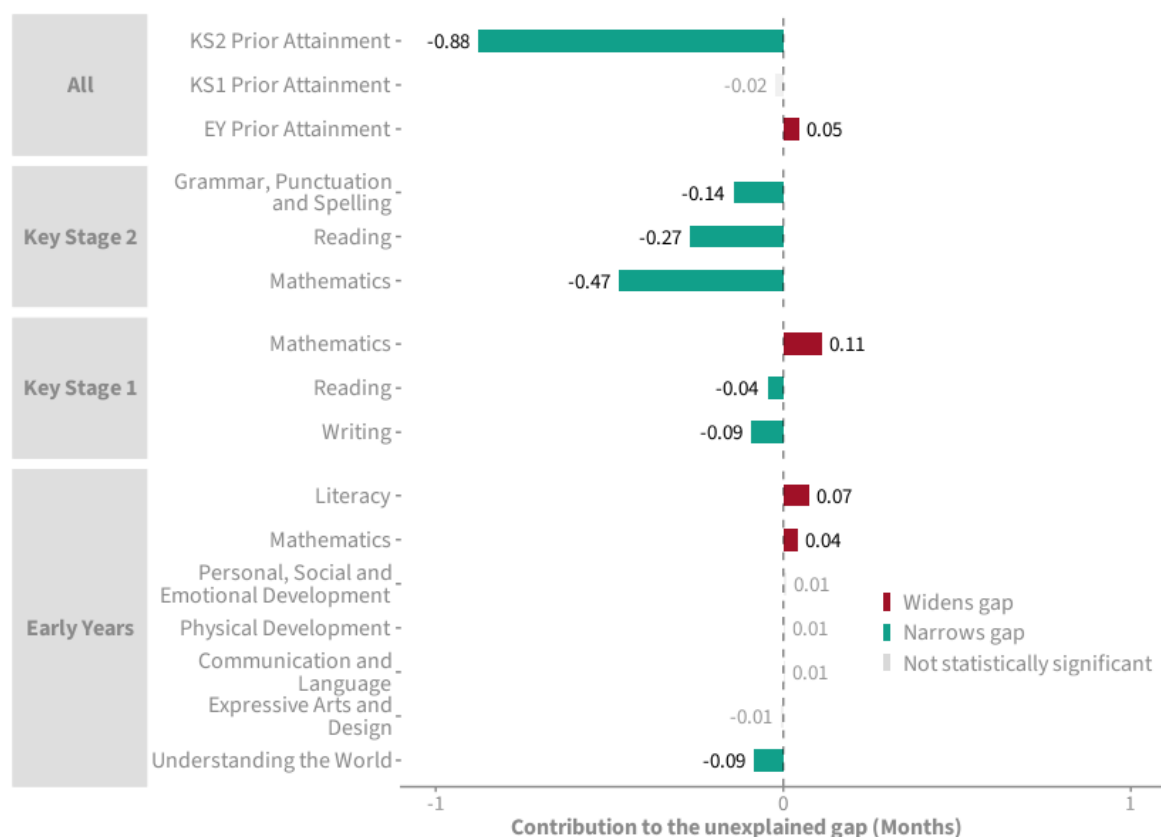
KS1 mathematics, and to a lesser extent writing, also make meaningful contributions, adding 1.1 months and 0.5 months to the KS4 explained gap, respectively. By contrast, KS1 reading appears to make no difference to the gap, while the contributions of individual EYFS areas of learning are small at under 0.2 months.

Unexplained gap

Turning to the ‘unexplained’ KS4 gap, our main analysis found that disadvantaged pupils remain 2.4 months behind their peers at the end of secondary school even when they have similar (observable) characteristics, similar school histories, similar peers and attend similar schools.

This, 2.4 months unexplained gap, would be 0.9 months larger still if not for prior attainment at KS2. Our disaggregated analysis in figure 18 shows that it is in KS2 mathematics, followed by reading, that disadvantaged pupils appear to make the most progress (relative to what would have been expected given other observable characteristics). These help narrow the unexplained gap by 0.5 months and 0.3 months, respectively. Other subjects at KS2, KS1 and at the EYFS appear to make little (i.e. up to 0.1 month) or no difference to the unexplained KS4 gap, conditional on other measures of prior attainment.

Figure 18: Contribution of prior attainment to the ‘unexplained’ disadvantage gap (in months) at KS4 in 2023/24



N = 509,308.

Overall, figure 17 and figure 18 suggest that differences in disadvantaged pupils' attainment in core subjects (mathematics and English-related domains) at KS2 – and to a lesser extent KS1 – are strongly associated with the size of the KS4 gap. Taken at face value, closing the KS2 mathematics gap appears to be associated with narrowing the KS4 gap, though several caveats apply.

First, these subject-level breakdowns reflect the structure of statutory assessments rather than conceptually distinct domains of child development. For example, KS2 reading and 'grammar, punctuation and spelling' can both reasonably be considered as related to English; taken together, their combined contribution is similar in magnitude to that of KS2 mathematics.

Second, the estimates should be interpreted as *conditional* contributions to the gap i.e. over and above the other predictors in the model. Where attainment in one subject operates partly through other outcomes (e.g. reading supporting writing, or later reading performance), some of its association with the KS4 gap may be captured by those related measures. This helps explain why more recent prior attainment measures tend to have larger coefficients: they may be absorbing effects from earlier attainment, rather than implying that earlier attainment itself is unimportant for the gap.

Third, as at KS2, it is important not to overstate the precision of the contributions of individual subjects. Our robustness tests indicate borderline multicollinearity for several predictors – specifically KS2 grammar, punctuation and spelling, KS1 reading, and EYFS communication and language. This shared variance makes it difficult for the model to apportion the gap uniquely across these measures, reducing the precision and stability of the subject-specific estimates. For this reason, in our main KS4 model we aggregate subject-level measures into a single prior attainment construct at each key stage.

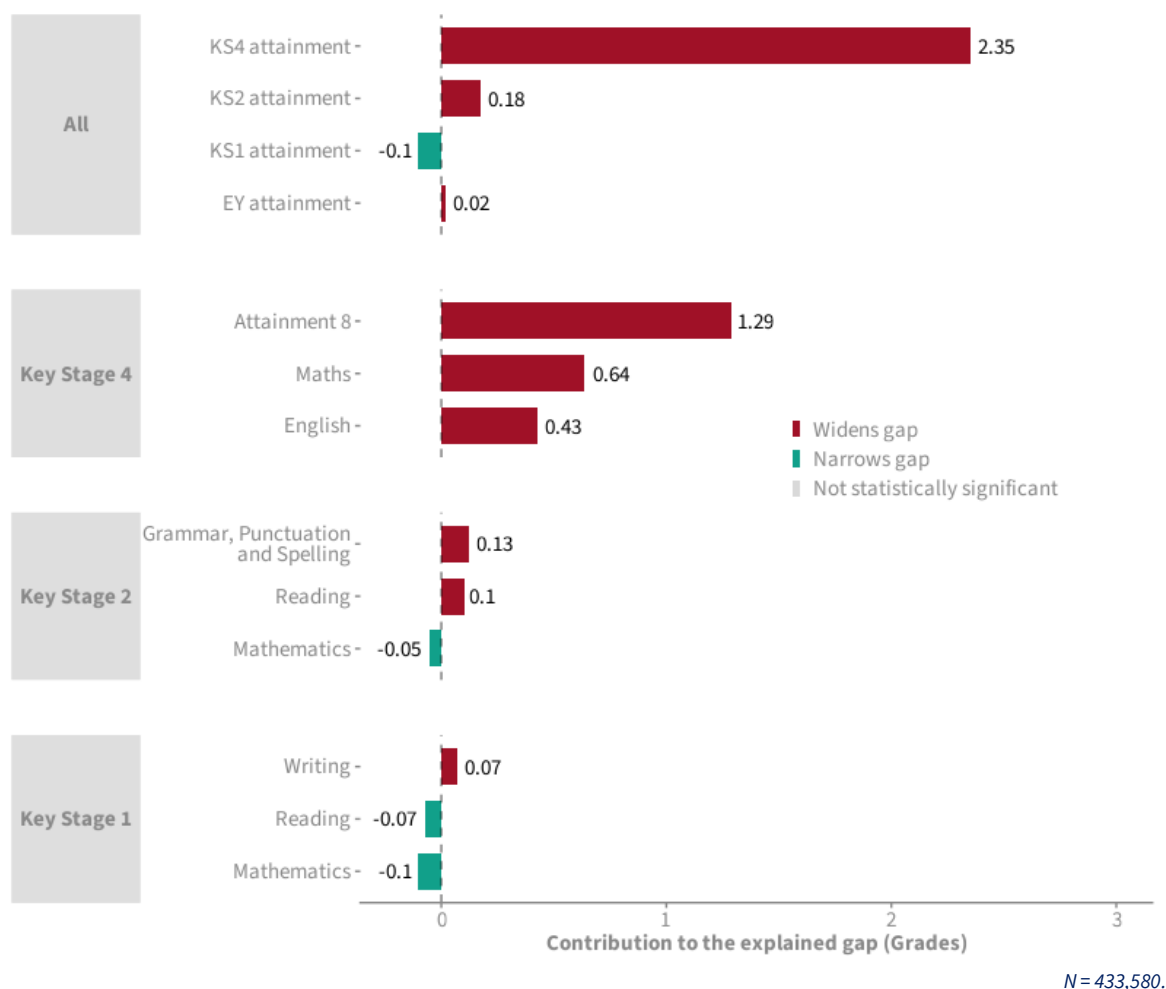
16-19 education

In our 16-19 analysis we found that of the total 3.5 grade gap, 2.8 grades (around 80 per cent) is 'explained' by differences in the observed characteristics of disadvantaged students compared to their peers, leaving 0.7 grades (20 per cent) 'unexplained'. As in earlier phases, the single biggest contributor to the explained component is prior attainment (70 per cent of the total gap), reflecting the substantial differences in attainment that have accumulated before students enter post-16 education.

Explained gap

Figure 19 breaks down the contribution of prior attainment to the explained 16-19 disadvantage gap, focusing on attainment at ages 7, 11 and 16. Overall attainment at age five is also included but is not separated into individual components due to high levels of multicollinearity. The figure shows that GCSE attainment (age 16) accounts for 96 per cent of the explained prior-attainment contribution, with earlier attainment measures making much smaller, residual contributions.

Figure 19: Contribution of prior attainment to the ‘explained’ disadvantage gap (in grades) in the 16-19 phase in 2023/24



Within KS4 prior attainment, the open-slot Attainment 8 measure contributes more to the explained gap than English or mathematics because disadvantaged and non-disadvantaged pupils differ substantially on this measure, and it has the strongest independent association with 16-19 attainment in the pooled model.⁶⁶ Because the three KS4 prior-attainment measures are correlated, the exact allocation of explained gap between English, mathematics and open-slot Attainment 8 should be interpreted cautiously. However, it does appear that the open-slot measure is a substantial contributor. There are several potential reasons for this:

⁶⁶ The open-slot Attainment 8 component measures pupils' achievement across the three additional qualifications that contribute to Attainment 8 after English, mathematics and the English Baccalaureate (EBacc) subjects. Because these slots can be filled by a wide range of academic and vocational qualifications, the measure captures both attainment and the extent to which pupils complete a full KS4 curriculum.



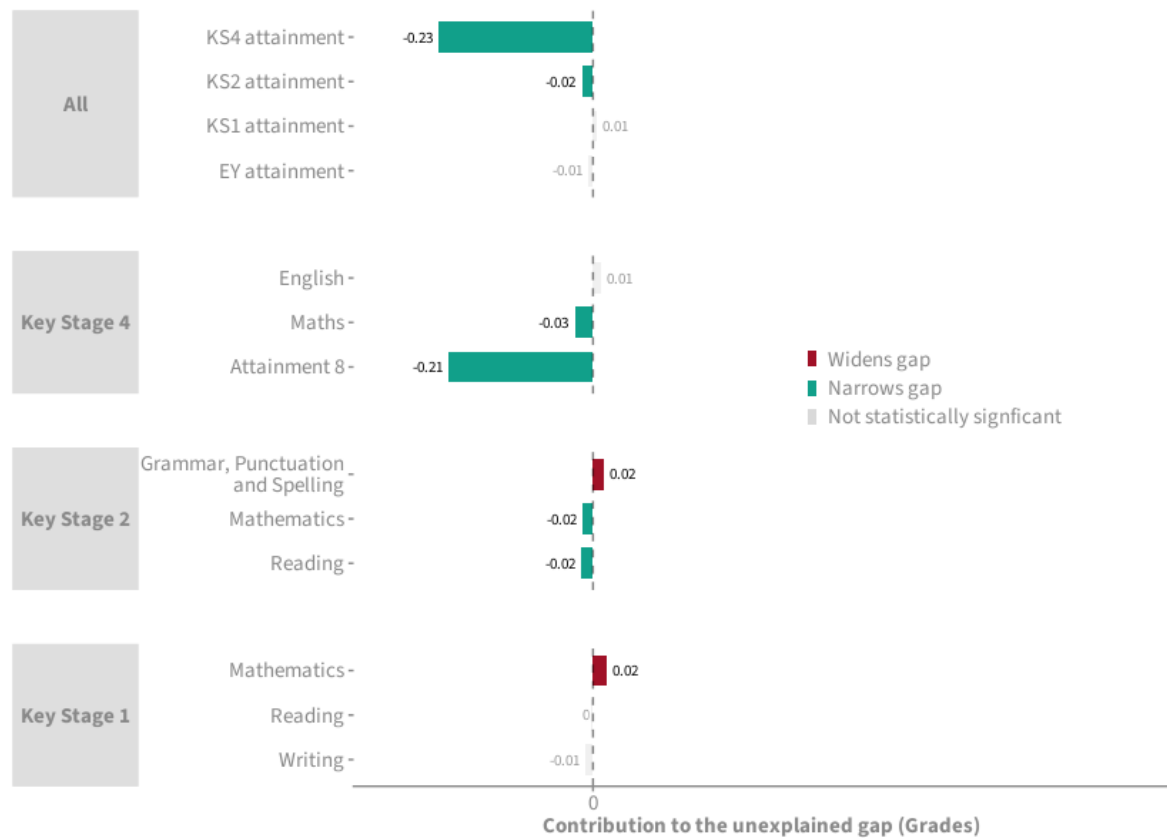
- While almost all pupils take English and mathematics, not all pupils will fill all three open slots in Attainment 8. This may be more common among disadvantaged pupils. As a result, the open-slot measure captures both the number of eligible qualifications taken and attainment within those qualifications, which may create a larger mean difference between disadvantaged and non-disadvantaged pupils than for English or mathematics alone.
- Both 16-19 study and the open slots of Attainment 8 reflect performance across a broader mix of subjects and qualification types than English and mathematics alone. This broader curriculum alignment may make open-slot Attainment 8 more closely associated with later 16-19 attainment.
- The open-slot measure may also proxy access to more advanced 16-19 pathways. Pupils who filled more open slots and achieved well in them may be more likely to progress to Level 3 study, and higher-level qualifications generally contribute more to the 16-19 attainment measure. However, English and mathematics GCSEs are also important access requirements for Level 3 study, so this is unlikely to be the only explanation.
- The open slots may also be strongly correlated with the volume of 16-19 qualifications taken, especially as both measures are a mix of average attainment and the number of qualifications taken.
- The open-slot measure may capture a broader dimension of prior attainment than English or mathematics alone. It reflects performance across a wider range of subjects and may therefore, pick up differences in general academic preparedness, curriculum breadth and achievement across the wider KS4 curriculum.

The large contribution of open-slot Attainment 8 may reflect one or more of these factors, rather than a single mechanism.

Unexplained gap

Figure 20 presents the same breakdown for the ‘unexplained’ gap, showing how the association between prior attainment and 16-19 outcomes differs between disadvantaged and non-disadvantaged students. As with prior attainment at earlier phases, KS4 prior attainment acts to narrow the unexplained gap. This should not be interpreted as prior attainment narrowing the gap overall: as shown in the explained component, lower prior attainment among disadvantaged students is a major driver of the gap. Rather, this is because open-slot Attainment 8 prior attainment tends to be less strongly linked to later outcomes for disadvantaged pupils. As a result, although disadvantaged pupils are more likely to start from a lower level on this measure, this weaker link means they face a smaller additional penalty for having lower prior attainment, which partly offsets the overall gap.

Figure 20: Contribution of prior attainment to the ‘unexplained’ disadvantage gap (in grades) in the 16-19 phase in 2023/24



N = 433,580.

One possible interpretation is that what disadvantaged students’ study in KS4 currently matters less for the pathways they take post-16, when compared with non-disadvantaged students. For example, they may be more inclined towards lower-level post-16 qualifications or dropping out, irrespective of their broader KS4 performance, when compared with non-disadvantaged students. This in turn may reflect broader barriers faced by disadvantaged students. This would be consistent with our findings on 16-19 mismatch. However, as set out in the main findings, it is not straightforward to interpret the role of prior attainment in the unexplained component, which can also reflect measurement and selection effects, as well as genuine differences in progress.

Conclusions



Conclusions

Through this research, we have undertaken an in-depth examination of the disadvantage gap in 2023/24 across four key stages of education. We set out to provide the clearest picture yet of how the disadvantage gap emerges and widens as children and young people move through the English education system. Exploiting the potential of the NPD, we linked students' attainment, backgrounds, school histories and peer and institutional contexts across four key stages to build a much richer account of the factors shaping their attainment outcomes.

Using new measures of prior attainment, SEND, and disrupted schooling, we applied decomposition techniques to quantify how student-, cohort- and institution-level characteristics each contribute to the disadvantage gap. This approach provides the first national breakdown of the 2023/24 disadvantage gap and, while recognising that the relationships identified are associative not causal, our analysis provides a sense of where policy and interventions could be targeted to help narrow the gap.

We find that identified SEND is the most important contributor to the age five disadvantage gap – and the second most important at age 11 – reflecting both its higher prevalence among disadvantaged pupils and its larger attainment penalty for disadvantaged pupils. However, we also note that this association is not necessarily causal, as teachers' assessments are often used to inform SEND identification.

At later ages, when we can account for more factors, prior attainment becomes the single largest contributor to the gap. We quantify how much of the gap has already become entrenched by the ages of 5, 7 and 11 and, in doing so, highlight the critical role of early intervention in mitigating later inequalities. We find that attaining well during primary school is not just a springboard for boosting pupils' starting points on entry to secondary school; it appears to provide a protective factor against drop-off during secondary school. Although it is difficult to isolate the exact contribution of any single subject to the gap – as students who perform well in one subject tend to do well in others – attainment in both mathematics and English-related subjects is strongly associated with the gap across phases.

However, 'prior attainment' should be understood not just as a narrow academic measure from an earlier phase. Prior attainment will also reflect the accumulated footprint of broader developmental, SEND, health and contextual experiences that shape learning long before formal assessments. To the extent that attainment is a 'catch-all' marker for a wide, overlapping set of influences – often beyond the school gates – a narrow focus on prior attainment risks masking the need for multi-faceted solutions.

We also show that early intervention is not enough: disadvantaged children and young people continue to fall behind during every phase of compulsory education. Looking in detail at pupils' experiences over the final three years of primary and secondary school, we find that disadvantaged pupils are much more likely to experience disrupted schooling in all its forms. While it is unsurprising that these disruptions are associated with worse pupil attainment, a key contribution of this research is to quantify their contribution to the disadvantage gap.

Chief among these disruptions is pupil absence. Its importance to the national disadvantage gap reflects the widespread nature of pupil absence and that even moderate levels of absence seem to harm attainment. These challenges are particularly acute for disadvantaged pupils, who are both more likely to be absent and more affected when they are. We also find that having a higher proportion of pupils who are absent within a school independently contributes to the gap, and this raises further research questions around the underlying mechanisms. Understanding the role of school climate, workforce capacity and leadership are likely critical, as these underpin the systems, expectations, relationships

and support structures that influence pupils' engagement in school and schools' ability to both prevent and mitigate absenteeism.

We know from previous EPI research that school disruptions are closely intertwined with SEND, with persistent absence, suspensions and repeated school moves all risk factors for children's needs going unidentified.⁶⁷ However, we also find cause for optimism: when children with SEND receive timely identification and sustained support, their attainment can be at least as good as their peers. We have previously made the case for using indicators such as persistent absence, repeated school moves and experiences of suspensions during primary school as automatic triggers for assessing children's needs; this could play a valuable role in improving outcomes for vulnerable pupils.

Our findings around ethnicity align with previous research suggesting significant educational underachievement among disadvantaged White British students, noting the important caveat that our analysis necessarily excludes the more vulnerable subset of ethnic minority pupils for whom we lack full data. While investigating the underlying reasons is beyond the scope of this study, previous research has suggested proximal factors – related to educational aspirations, academic self-concept, and specific behaviours related to homework and attendance – could be useful areas where school interventions might have the best chance of impacting on attainment at age 16. Future research could use longitudinal, administrative (Education and Child Health Insights from Linked Data [ECHILD]) data linking health and education records to examine the drivers of ethnic gaps with unprecedented data depth, including the role of wider children's services related to health, social care and maternity services.

We find that pupils growing up in long-term poverty face some of the most entrenched educational inequalities. Persistently disadvantaged pupils are less likely than their peers (including those in short-term poverty) to have achieved a good level of development at age five and are more likely to be identified with SEND by age 16. But while the persistent gap at GCSE is over 21 months, we find that it is largely shaped by the same pathways as the overall KS4 disadvantage gap. Policies and interventions that improve pupil attainment during primary school should not only help narrow the overall gap but particularly benefit pupils in long-term poverty.

Throughout our analysis, we find that the characteristics of peer groups and institutions are less important than pupil-level factors, though the contribution of peer groups to the gap grows at later ages, reaching nine per cent in the 16-19 phase. In contrast to later ages, we find that disadvantaged pupils attain better at age five and age 11 when they are in disadvantaged settings and this helps to slightly narrow the gap. This suggests it is disadvantaged pupils in less disadvantaged settings who could be at more risk of low attainment than other disadvantaged pupils.

Despite using granular and wide-ranging measures across our models, there remains a part of the gap that is not tied to any observed factor. This is particularly the case for the age five gap, where we lack measures relating to children's homelives, early childhood, or even pre-natal health. Additionally, many cases of SEND have yet to be identified. But even for disadvantage gaps at later ages, we are potentially missing key contributors – such as children and young people's health, workforce characteristics, school climates, and the characteristics of the local areas where children and young people grow up – which could comprise fruitful avenues for future research.

⁶⁷ <https://epi.org.uk/wp-content/uploads/2025/02/SEND-Final-Report-version-FINAL-04.02.2024-2.pdf>

References

References

Adjei, Nicholas Kofi, Daniela K. Schlüter, Gabriella Melis, Viviane S. Straatmann, Kate M. Fleming, Sophie Wickham, Luke Munford, et al. "Impact of Parental Mental Health and Poverty on the Health of the Next Generation: A Multi-Trajectory Analysis Using the UK Millennium Cohort Study." *Journal of Adolescent Health* 74, no. 1 (2024): 60–70. <https://www.sciencedirect.com/science/article/pii/S1054139X23003993>.

Beynon, Katie, and Dave Thomson. *The Long-Term Consequences of Long-Term Disadvantage*. Report to the Northern Powerhouse Partnership. FFT Education Datalab, 2022. https://view.officeapps.live.com/op/view.aspx?src=https%3A%2F%2Fftteducationdatalab.org.uk%2Fwp-content%2Fuploads%2F2023%2F01%2Ffinal_report_inc_exec_summary.docx&wdOrigin=BROWSELINK.

Crenna-Jennings, Whitney. *Key Drivers of the Disadvantage Gap Literature Review: Education in England Annual Report 2018*. Education Policy Institute, 2018. <https://epi.org.uk/wp-content/uploads/2018/07/EPI-Annual-Report-2018-Lit-review.pdf>.

Department for Education. *Pupil Absence in Schools in England, Autumn 2024 and Spring 2025*. Department for Education, 2025. <https://www.gov.uk/government/statistics/pupil-absence-in-schools-in-england-autumn-2024-and-spring-2025>.

Department for Education. *Suspensions and Permanent Exclusions in England*. Department for Education, 2025. <https://explore-education-statistics.service.gov.uk/find-statistics/suspensions-and-permanent-exclusions-in-england/2024-25-autumn-term>.

Funding Futures. *Funding Futures: Targeting Early Investment so That All Children Can Reach Their Potential*. Funding Futures, 2025. https://static1.squarespace.com/static/68c02a5904df7c39a9c27a73/t/68c9c4cbf8692051ec0df946/1758053579502/Funding_Futures_Summary_Report.pdf.

Harland, Jennie, Caroline Sharp, Lillian Flemons, et al. *Factors Influencing Secondary School Pupils' Educational Outcomes*. National Foundation for Educational Research, NatCen Social Research, and National Children's Bureau, 2024. https://assets.publishing.service.gov.uk/media/66e4006e3f1299ce5d5c3e11/Factors_influencing_secondary_school_pupils_educational_outcomes.pdf.

Harland, Jennie, Caroline Sharp, Michelle Judkins, et al. *Factors Influencing Primary School Pupils' Educational Outcomes*. National Foundation for Educational Research, NatCen Social Research, and National Children's Bureau, 2024. https://assets.publishing.service.gov.uk/media/66e2cd5a61763848f429d58f/Factors_influencing_primary_school_pupils_educational_outcomes.pdf.

Hunt, Emily, Sam Tuckett, Eva Jiménez, and David Robinson. *Breaking Down the Gap: The Role of School Absence and Pupil Characteristics*. Education Policy Institute, 2025. <https://epi.org.uk/publications-and-research/breaking-down-the-gap/>.

Hutchinson, Jo, Johnny Downs, and Tamsin Ford. *Identifying SEND: Final Report on Special Educational Needs & Disabilities and Contact with CAMHS*. Education Policy Institute, 2025. <https://epi.org.uk/wp-content/uploads/2025/02/SEND-Final-Report-version-FINAL-04.02.2024-2.pdf>.

Knoll, Lisa, Leung, Jovita, Foulkes, Lucy, Blakemore, Sarah-Jayne. "Age-Related Differences in Social Influence on Risk Perception Depend on the Direction of Influence." *Journal of Adolescence* 60 (2017): 53–63. <https://doi.org/10.1016/j.adolescence.2017.07.002>.

Lai, Eric T. C., Sophie Wickham, Catherine Law, Margaret Whitehead, Benjamin Barr, and David Taylor-Robinson. "Poverty Dynamics and Health in Late Childhood in the UK: Evidence from the Millennium Cohort Study." *Archives of Disease in Childhood* (2019): 1–7.
<https://adc.bmj.com/content/archdischild/early/2019/05/15/archdischild-2018-316702.full.pdf?ijkey=hKrHArIAeu7gg9F&keytype=ref>.

Macleod, Shona, Caroline Sharp, Daniele Bernardinelli, Amy Skipp, and Steve Higgins. *Supporting the Attainment of Disadvantaged Pupils: Articulating Success and Good Practice*. National Foundation for Educational Research, Ask Research, and Durham University, 2015.
https://assets.publishing.service.gov.uk/media/5a80bcd2ed915d74e6230292/DFE-RR411_Supporting_the_attainment_of_disadvantaged_pupils.pdf.

Metcalfe, Sophie, and Davison, Nehal.. *School Readiness - How Can Government Start Closing the Opportunity Gap in Early Years Education?* Institute for Government, 2025.
<https://www.instituteforgovernment.org.uk/sites/default/files/2025-04/school-readiness-policy-making-left-behind-groups.pdf>.

Robinson, David, Emily Hunt, Sam Tuckett, and Eva Jiménez. *EPI Annual Report 2025*. Education Policy Institute, 2025. <https://epi.org.uk/annual-report-2025/>.

Strand, Steve. "Ethnicity, Gender, Social Class and Achievement Gaps at Age 16: Intersectionality and 'Getting It' for the White Working Class." *Research Papers in Education* 29 (2014): 131–171.
https://www.researchgate.net/publication/264412254_Ethnicity_gender_social_class_and_achievement_gaps_at_age_16_Intersectionality_and_'Getting_it'_for_the_white_working_class.

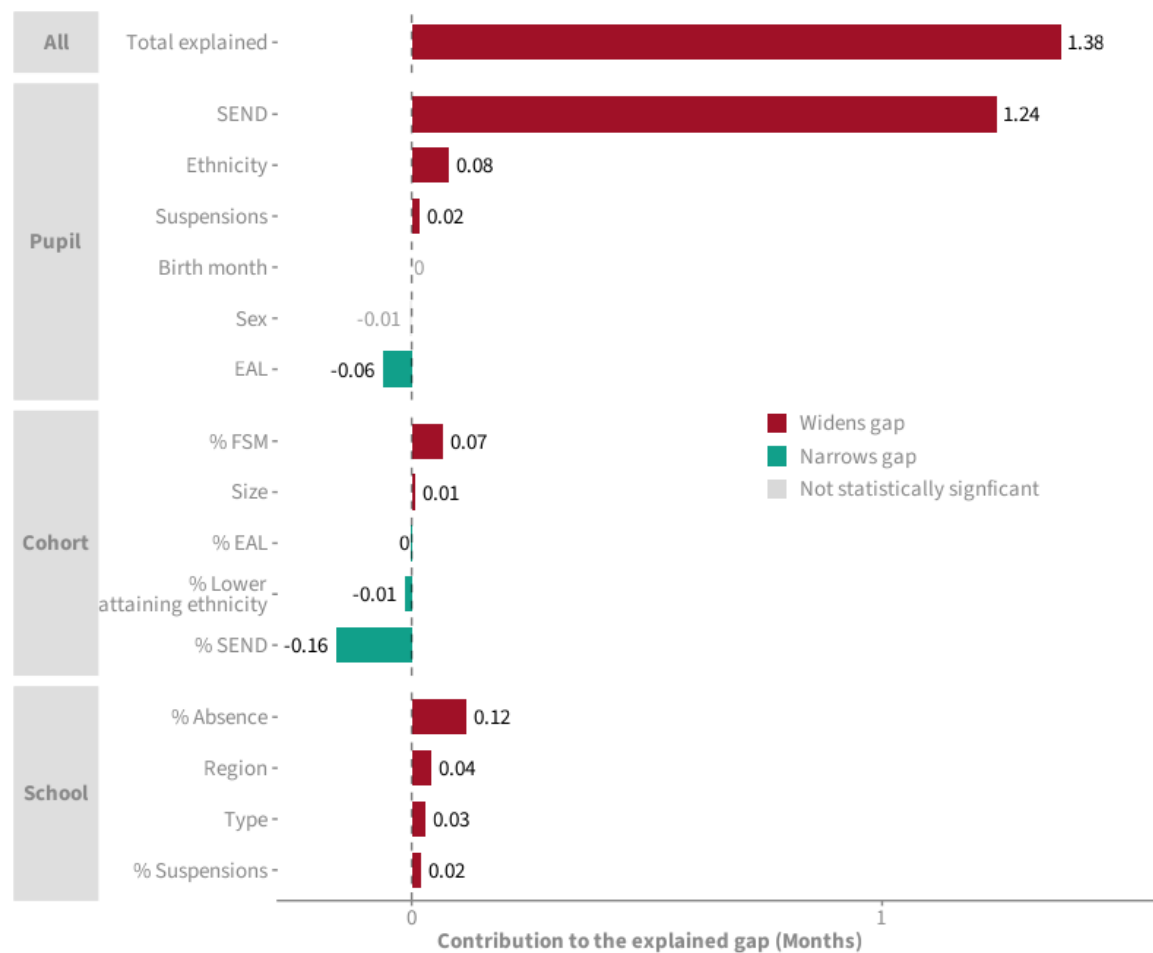
Strand, Steve. *Ethnic, Socio-Economic and Sex Inequalities in Educational Achievement at Age 16: An Analysis of the Second Longitudinal Study of Young People in England (LSYPE2)*. Report for the Commission on Race and Ethnic Disparities (CRED). University of Oxford, Department of Education, 2021. https://www.education.ox.ac.uk/wp-content/uploads/2021/05/Strand_2021_Report-to-CRED.pdf.

Tuckett, Sam, David Robinson, and Felix Bunting. *Measuring the Disadvantage Gap in 16-19 Education*. Education Policy Institute, 2021. <https://epi.org.uk/publications-and-research/measuring-the-disadvantage-gap-in-16-19-education/>.

Annexes

Annex A: Additional analyses

Figure A1: Factors contributing to the ‘explained’ disadvantage gap (in months) at age five in 2023/24



N = 517,020.

Our reference categories are:

Sex = Female.

Birth month = September.

SEND = No identified need.

Ethnicity = White British.

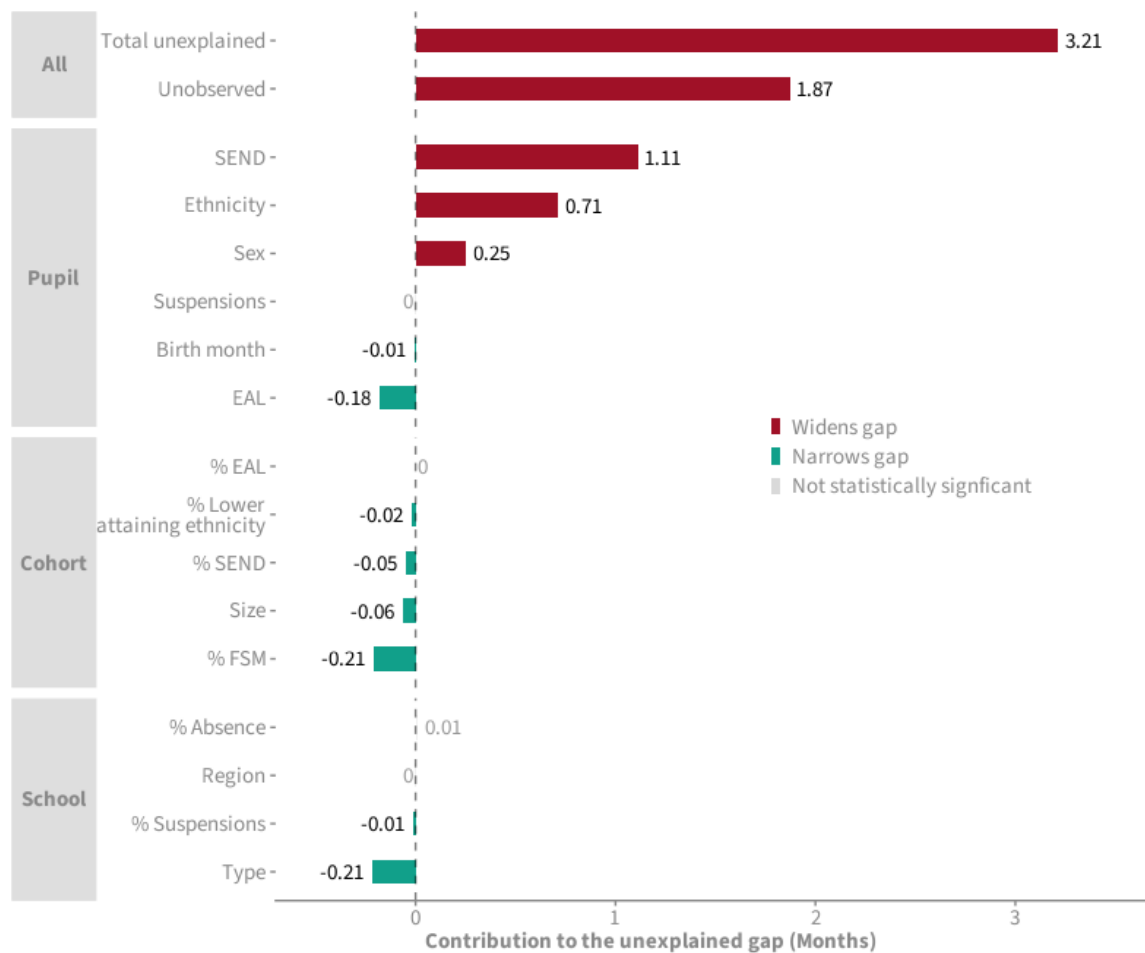
EAL = English as a first language.

School type = Local authority-maintained school.

Region = London.



Figure A2: Factors contributing to the ‘unexplained’ disadvantage gap (in months) at age five in 2023/24



N = 517,020.

Our reference categories are:

Sex = Female.

Birth month = September.

SEND = No identified need.

Ethnicity = White British.

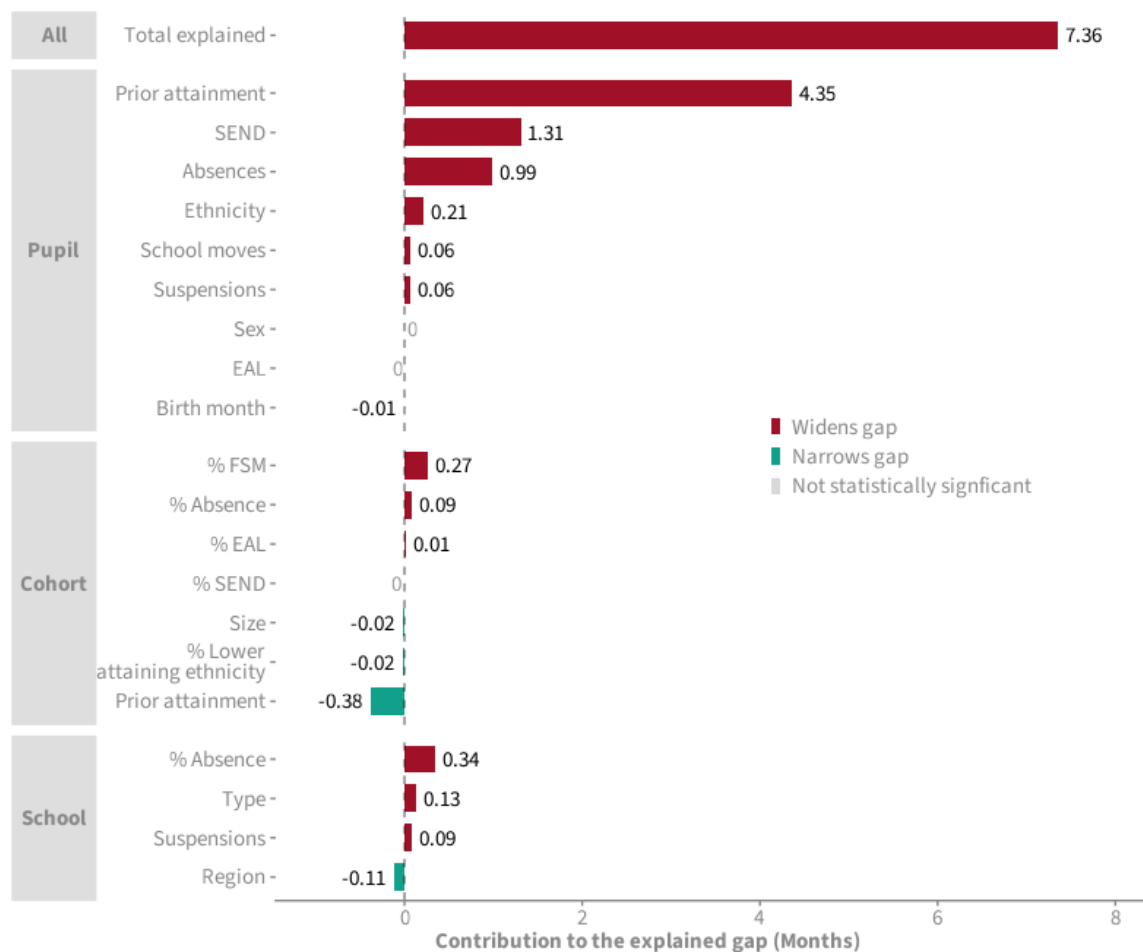
EAL = English as a first language.

School type = Local authority-maintained school.

Region = London.



Figure A3: Factors contributing to the ‘explained’ disadvantage gap (in months) at KS2 in 2023/24



N = 578,807.

Our reference categories are:

Sex = Female.

Birth month = September.

SEND = No identified need.

Ethnicity = White British.

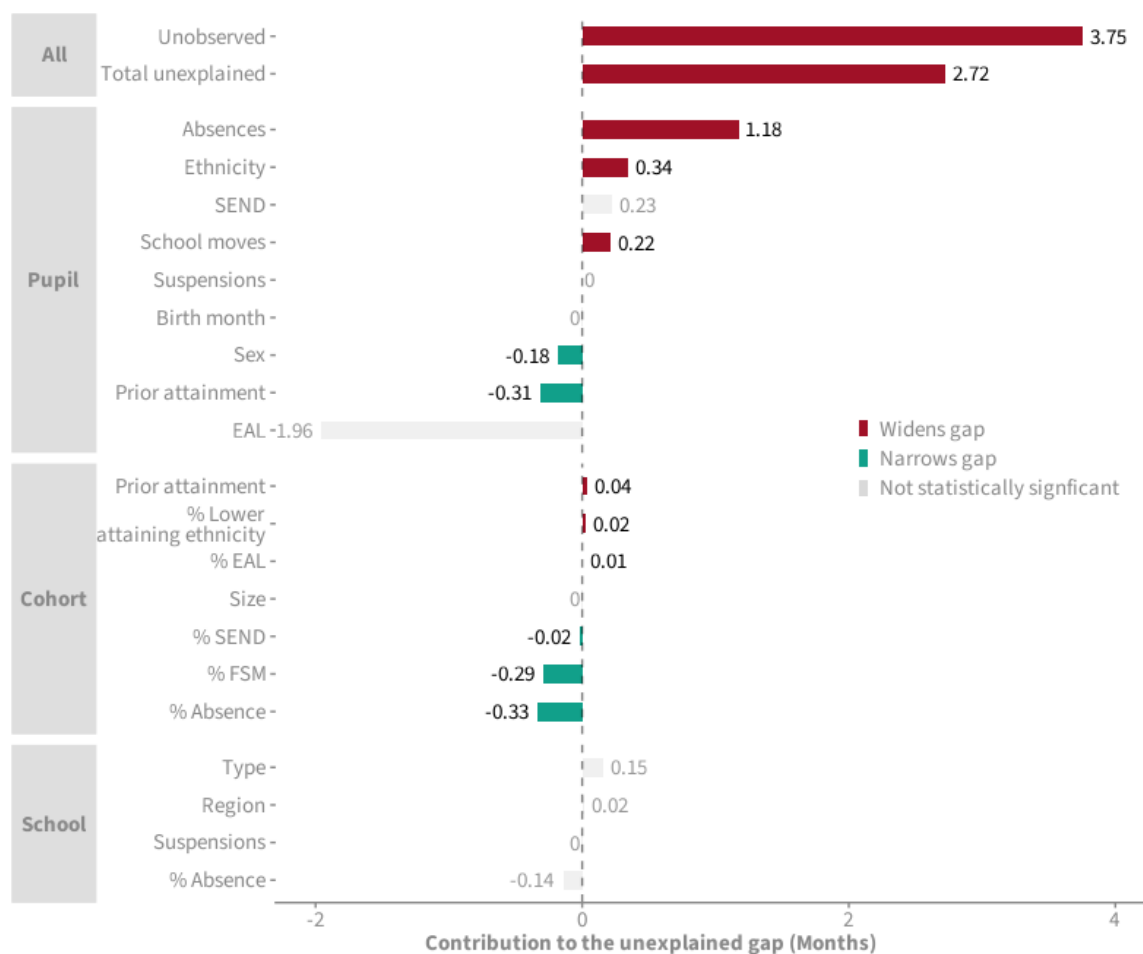
EAL = English as a first language.

School type = Local authority-maintained school.

Region = London.



Figure A4: Factors contributing to the ‘unexplained’ disadvantage gap (in months) at KS2 in 2023/24



N = 578,807.

Our reference categories are:

Sex = Female.

Birth month = September.

SEND = No identified need.

Ethnicity = White British.

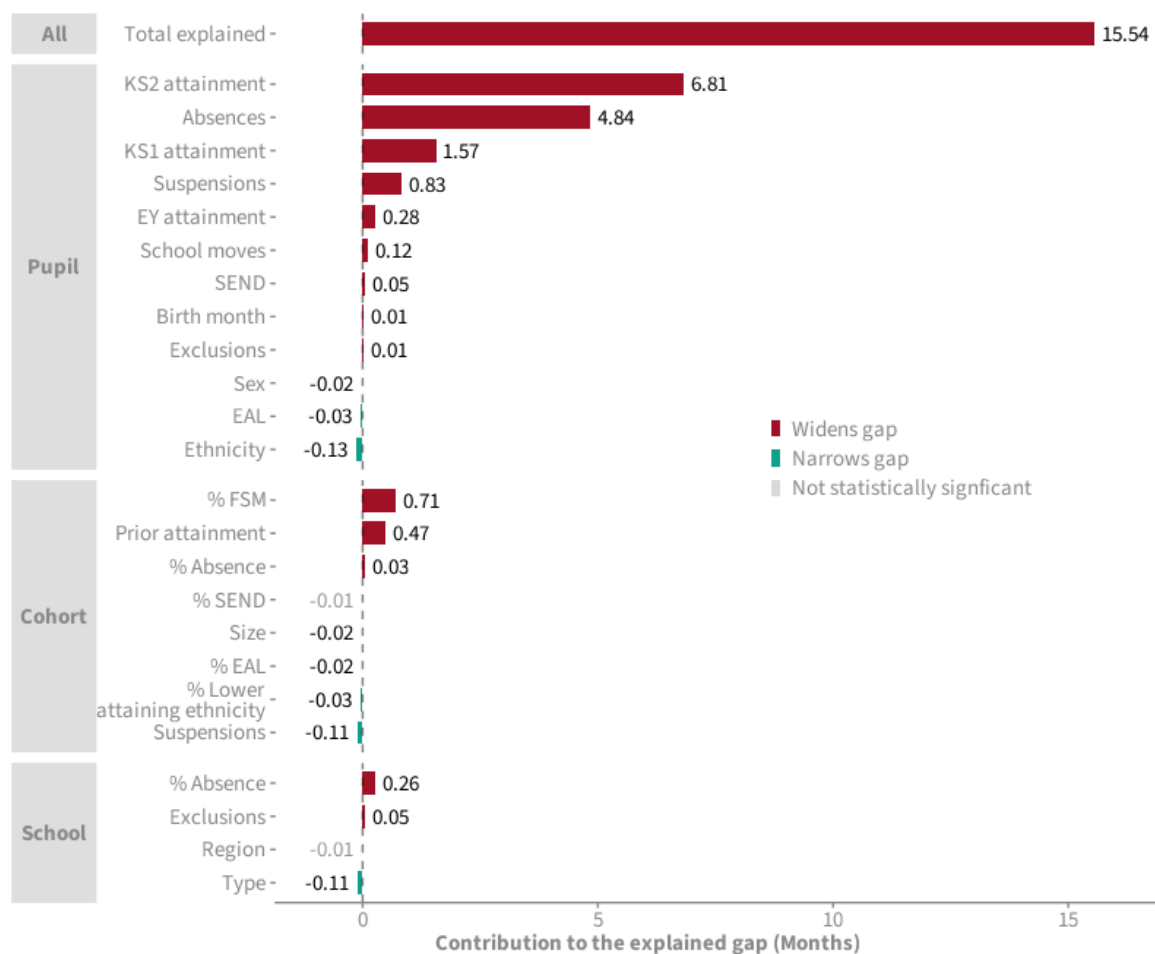
EAL = English as a first language.

School type = Local authority-maintained school.

Region = London.



Figure A5: Factors contributing to the ‘explained’ disadvantage gap (in months) at KS4 in 2023/24



N = 509,308.

Our reference categories are:

Sex = Female.

Birth month = September.

SEND = No identified need.

Ethnicity = White British.

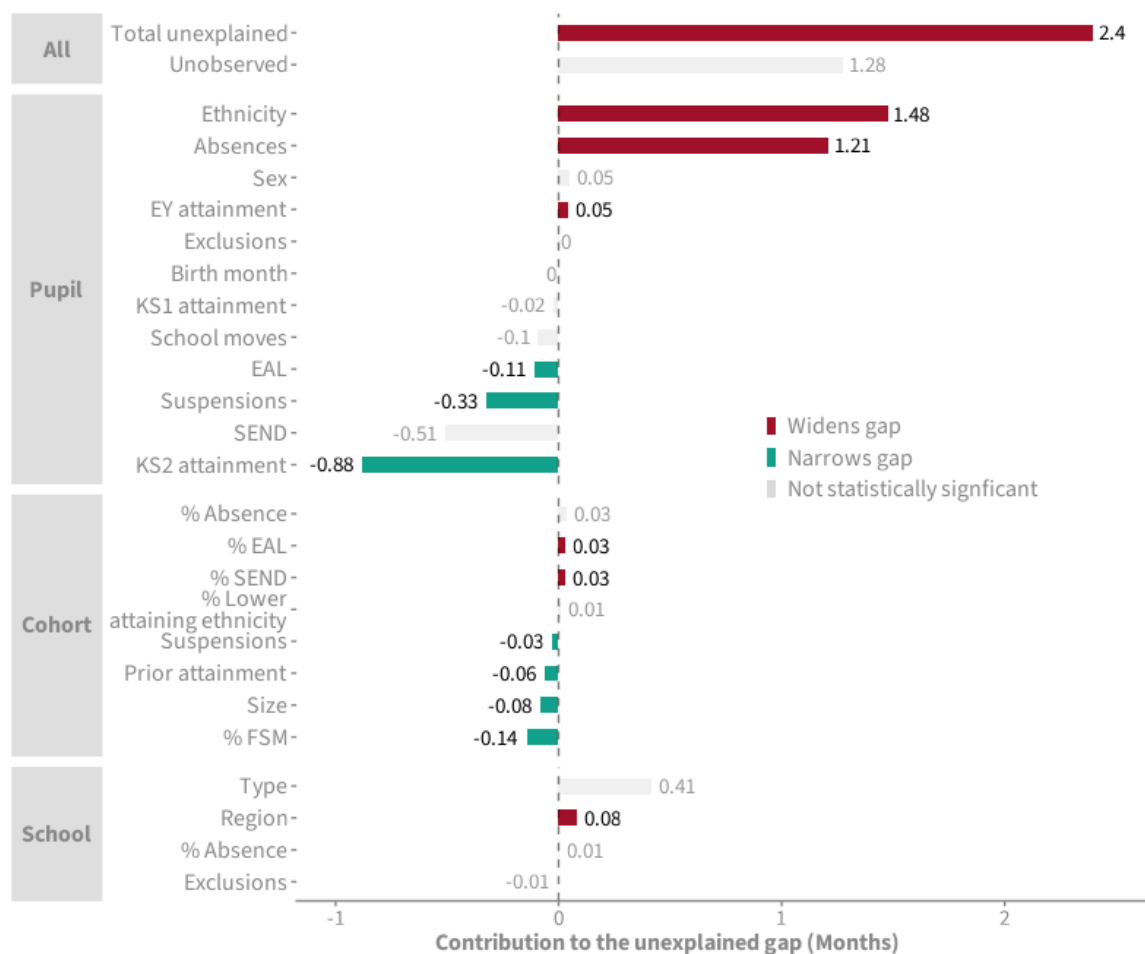
EAL = English as a first language.

School type = Local authority-maintained school.

Region = London.



Figure A6: Factors contributing to the ‘unexplained’ disadvantage gap (in months) at KS4 in 2023/24



N = 509,308.

Our reference categories are:

Sex = Female.

Birth month = September.

SEND = No identified need.

Ethnicity = White British.

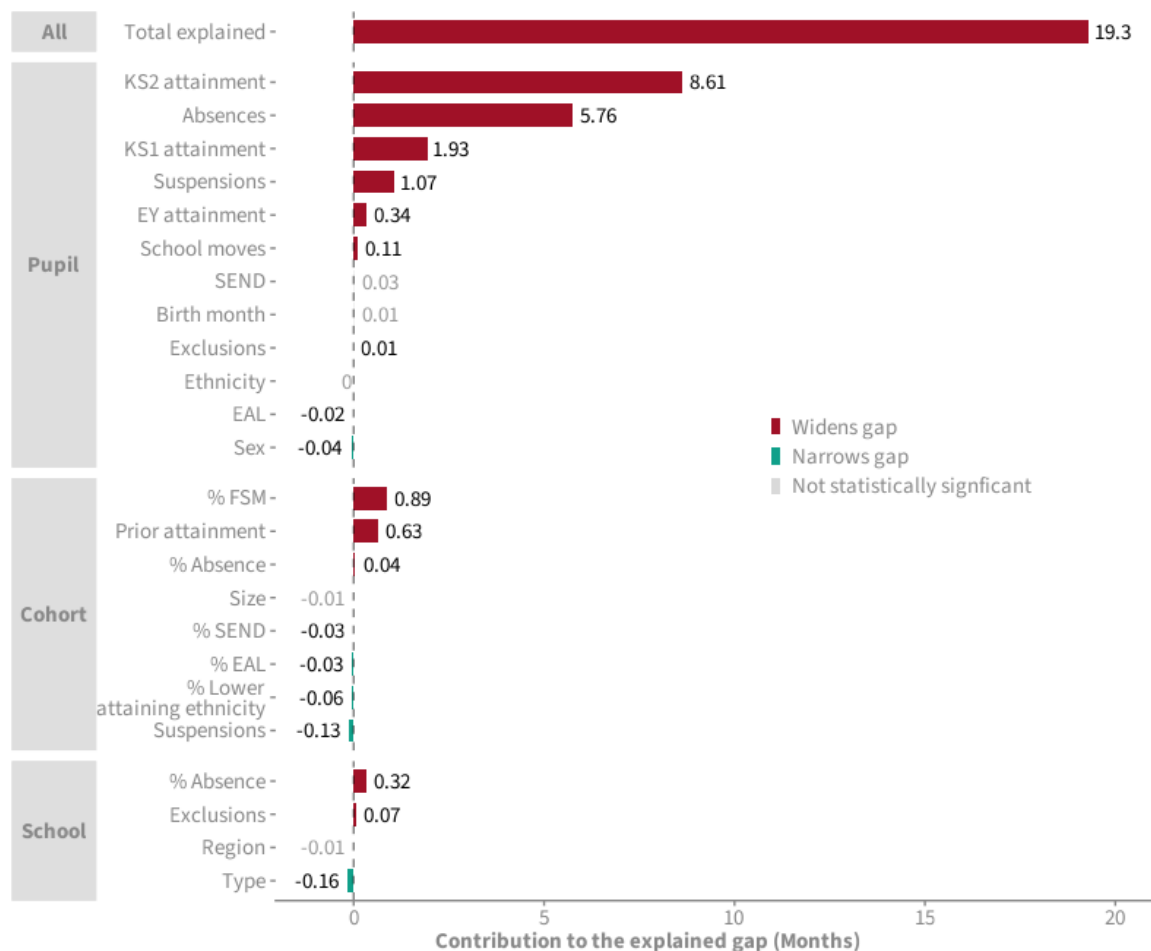
EAL = English as a first language.

School type = Local authority-maintained school.

Region = London.



Figure A7: Factors contributing to the ‘explained’ persistent disadvantage gap (in months) at KS4 in 2023/24



N = 441,855.

Our reference categories are:

Sex = Female.

Birth month = September.

SEND = No identified need.

Ethnicity = White British.

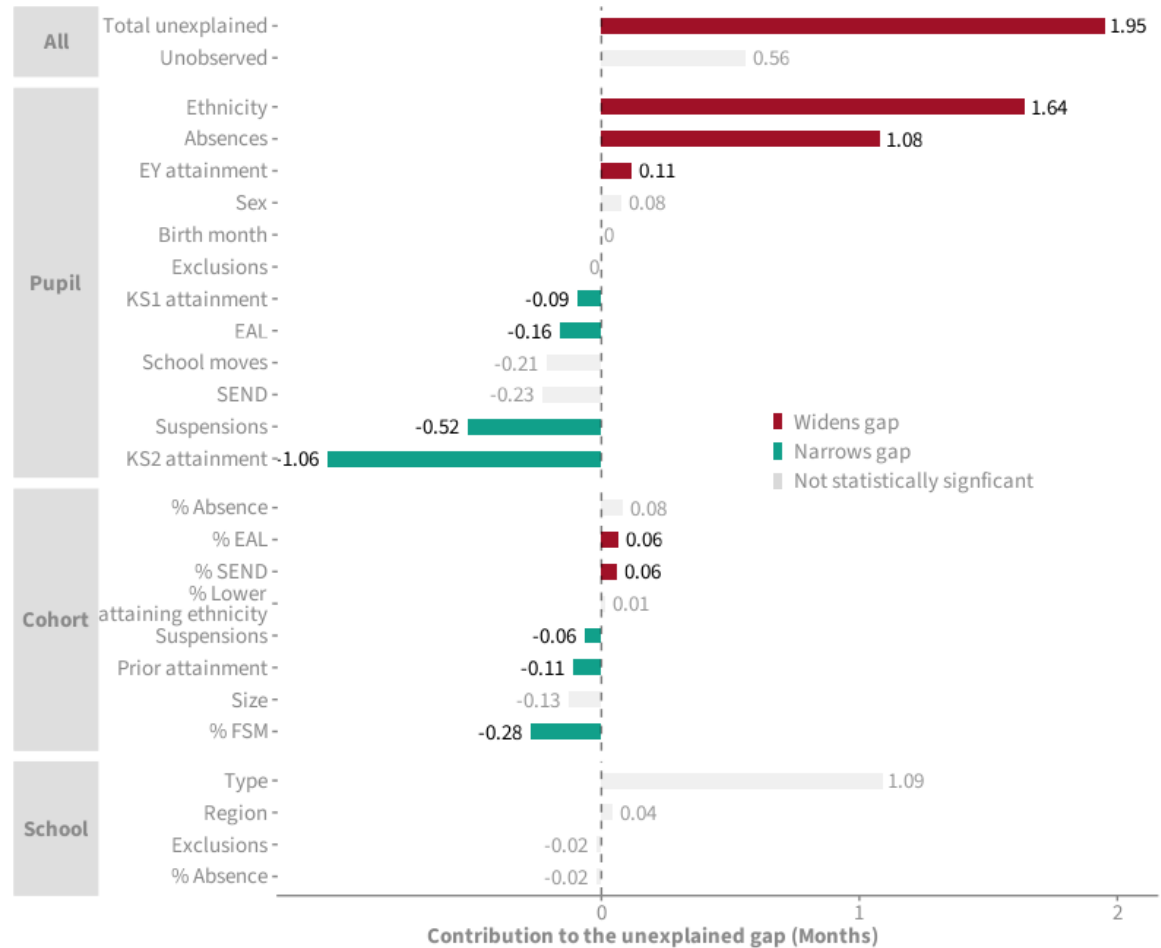
EAL = English as a first language.

School type = Local authority-maintained school.

Region = London.



Figure A8: Factors contributing to the ‘unexplained’ persistent disadvantage gap (in months) at KS4 in 2023/24



N = 441,855.

Our reference categories are:

Sex = Female.

Birth month = September.

SEND = No identified need.

Ethnicity = White British.

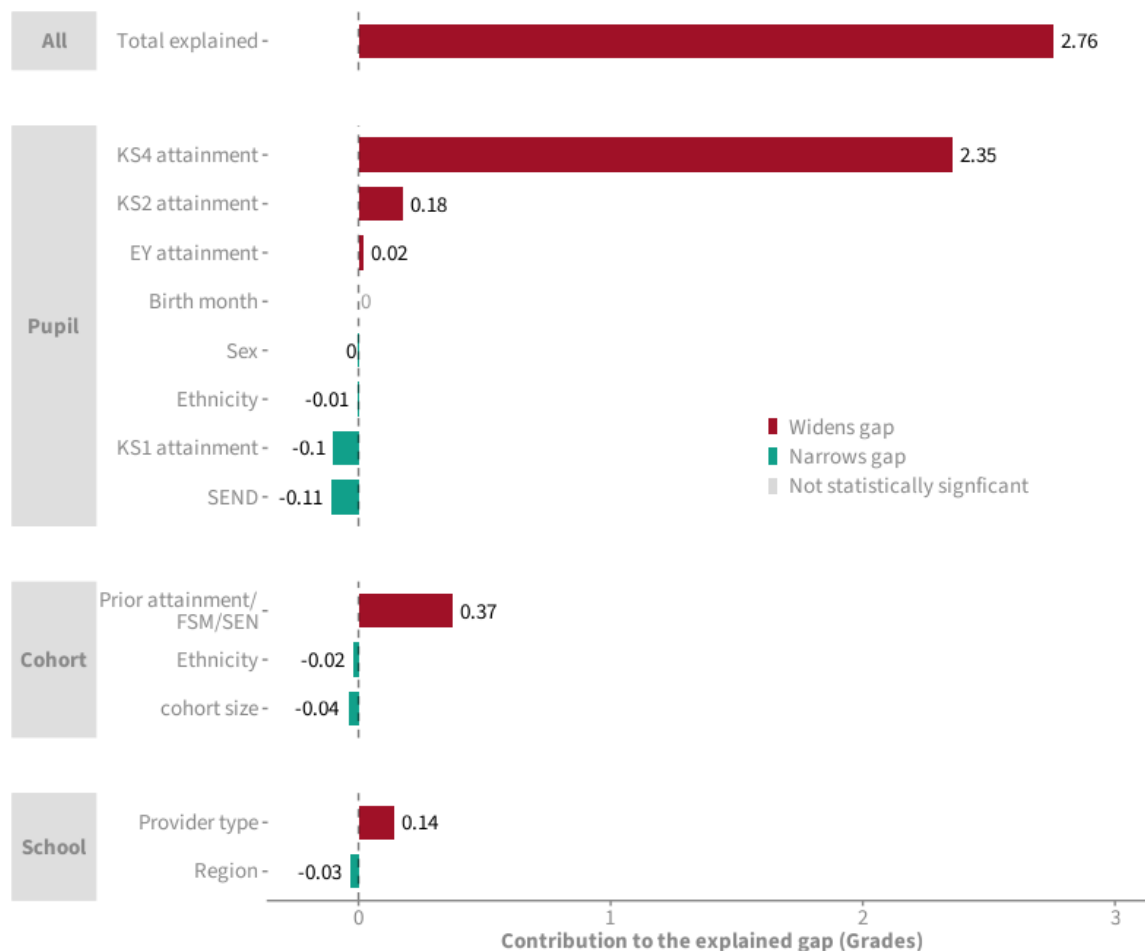
EAL = English as a first language.

School type = Local authority-maintained school.

Region = London.



Figure A9: Factors contributing to the ‘explained’ disadvantage gap (in grades) in the 16-19 phase in 2023/24



N = 433,580.

Our reference categories are:

Sex = Female.

Birth month = September.

SEND = No identified need.

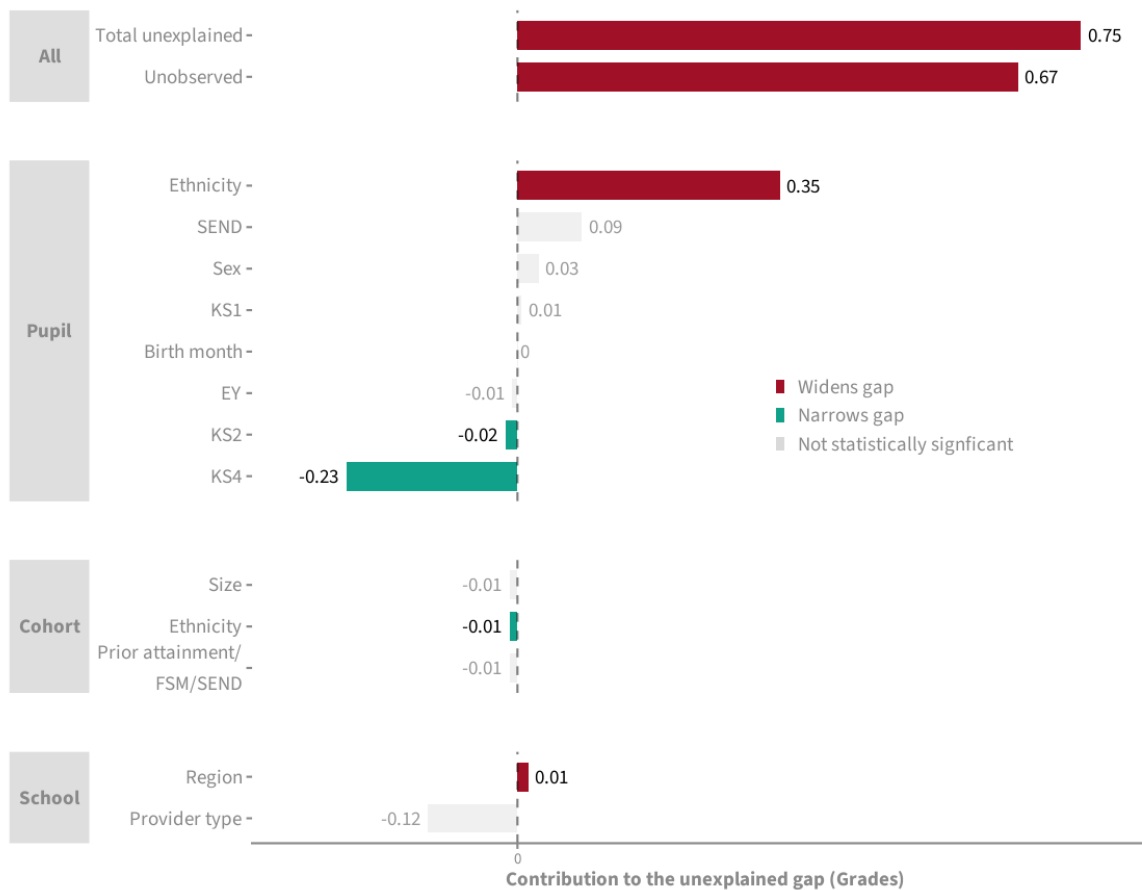
Ethnicity = White British.

Provider type = Local authority-maintained school sixth forms

Region = London.



Figure A10: Factors contributing to the ‘unexplained’ disadvantage gap (in grades) in the 16-19 phase in 2023/24



N = 433,580.

Our reference categories are:

Sex = Female.

Birth month = September.

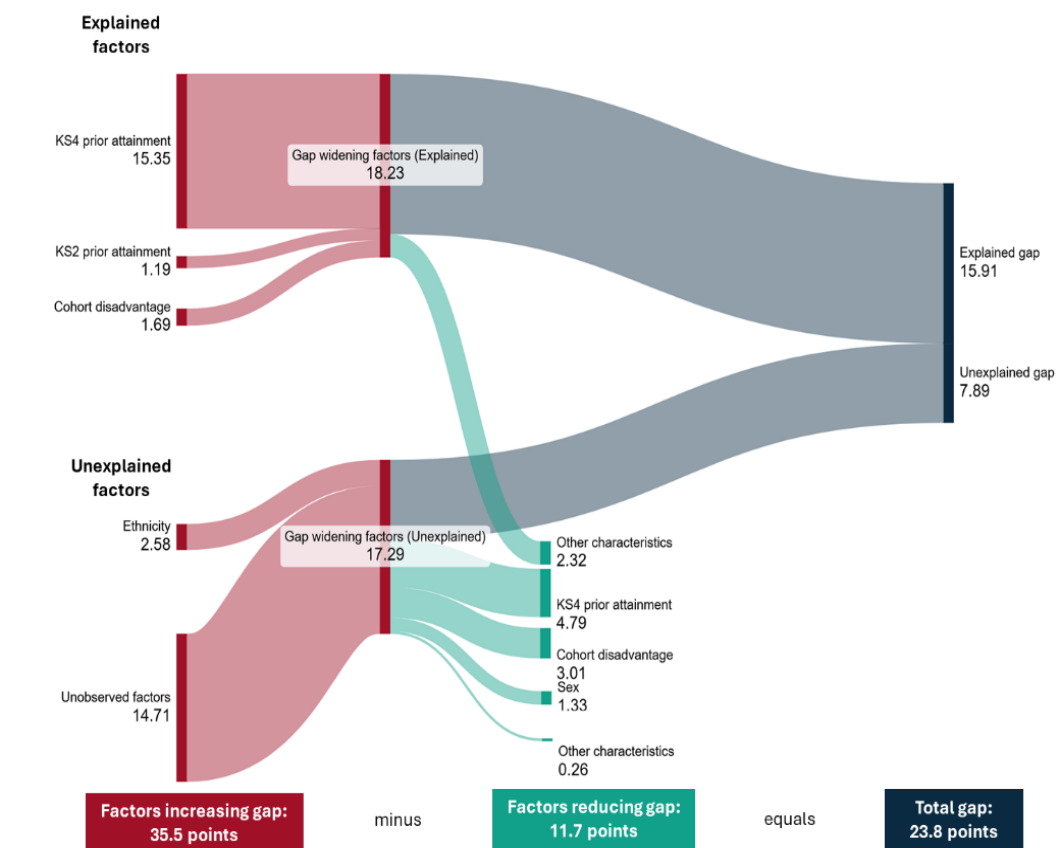
SEND = No identified need.

Ethnicity = White British.

Provider type = Local authority-maintained school sixth forms

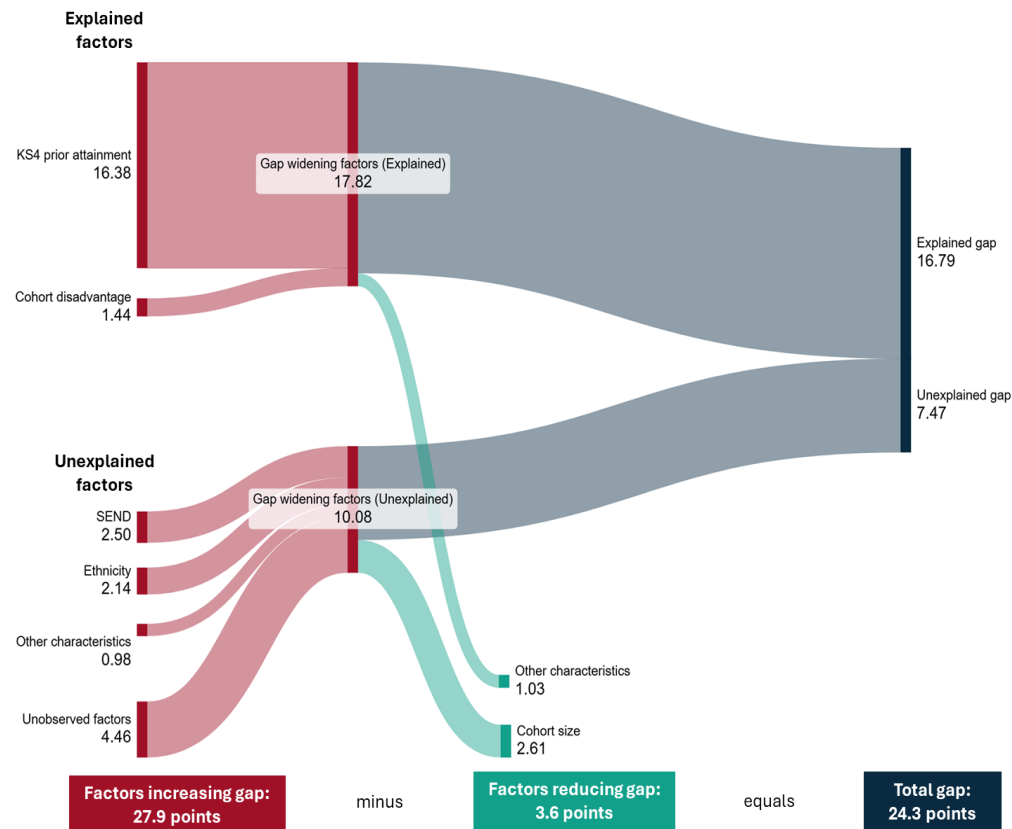
Region = London.

Figure A11: Decomposition of the 16-19 disadvantage gap (in points) in 2023/24, FE colleges, statistically significant factors contributing $\geq 5\%$ of the gap



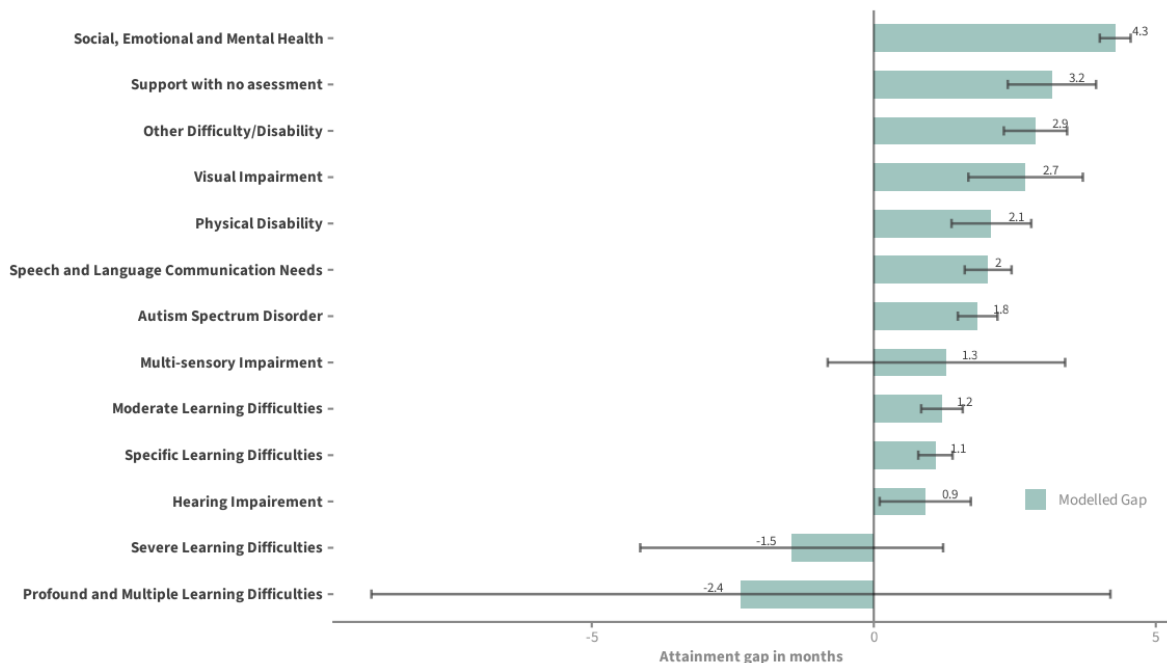
N = 207,796

Figure A12: Decomposition of the 16-19 disadvantage gap (in points) in 2023/24, school sixth forms, statistically significant factors contributing $\geq 5\%$ of the gap



$N = 183,093$.

Figure A13: KS4 attainment gap in 2023/24 by SEND primary need type in Year 11 compared to pupils with no identified SEND



Notes: The modelled gap includes the full set of pupil, cohort, school and regional controls. Within the pupil controls, this includes the duration and complexity of SEND since reception year. Bars show 95 per cent confidence intervals. N = 509,308.

As we found for KS2 in figure 7, GCSE outcomes differ *within* the group of pupils with identified SEND according to primary need type in Year 11.

After accounting for all other factors, figure A13 shows that pupils with hearing impairments, specific learning difficulties, moderate learning difficulties or multi-sensory impairments all have small GCSE gaps (of under one month) compared to their peers with no identified needs and otherwise similar characteristics, including prior attainment. By contrast, pupils with social, emotional and mental health needs are around four months behind their peers, on average.

As with our KS2 SEND findings, important caveats apply here: the analysis is necessarily subject to selection bias as it only includes those pupils remaining on state school rolls by Year 11 with complete data, including on prior attainment. We cannot account for pupils with SEND who may have left state education whose outcomes in some cases may be less favourable than the ones reflected here. Therefore, while our findings show that some pupils with SEND achieve well at the end of KS4, when accounting for all other factors, we must be careful not to interpret these results as generalisable to all SEND students.

Annex B: Model variables

Early years:

- Attainment is based on the statutory teacher assessments of children's development at the end of the EYFS profile that takes place towards the end of reception year. Specifically, we use a pupil's total point score across 12 (out of the 17) early learning goals, corresponding to the Department for Education's 'good level of development' measure.⁶⁸ For each goal, children are assessed as either meeting the expected level of development at the end of reception year (score = 2) or not yet reaching this level (score = 1). Our total points score aggregates scores across these goals, giving a maximum possible score of 24 for pupils meeting the expected level in each goal.
- In addition to sex, month of birth, and ethnicity, our pupil characteristics include:
 - SEND status i.e. whether a pupil is on SEND Support in reception year, has an EHCP, or no identified need; and
 - first language i.e. whether English is a pupil's first or additional language.
- As this is the first year of compulsory education, it is not possible to account for prior attainment or the number of times a pupil has moved school.
- We do not include absence data because four-year-olds in reception year are not yet legally required to attend school.
- We do not include data on pupil-level suspensions or permanent exclusions because too few pupils are affected in reception year.
- Our cohort-level variables are the:
 - percentage of a pupil's peers (i.e. in reception year) who are disadvantaged;
 - percentage of a pupil's peers who have identified SEND;
 - percentage of a pupil's peers who have EAL;
 - percentage of a pupil's peers who are in a lower-attaining ethnic group;⁶⁹ and
 - size of the year group.⁷⁰
- Our school-level variables are the:
 - overall school absence rate (as a percentage);
 - percentage of pupils in the school who are suspended; and
 - school type based on whether the school is local authority-maintained, sponsored academy, converter academy, free school or special school.
- we include regional controls based on where a pupil lives.

⁶⁸ These goals relate to five areas of learning comprising: communication and language; personal, social and emotional development; physical development; literacy; and mathematics.

⁶⁹ At each key stage, we identify these (minor ethnic) groups based on their average attainment being below the national average in each of the last three years (2021/22, 2022/23 and 2023/24) using the same attainment measure as our outcome variable. For the EYFS, lower-attaining ethnic groups comprise: Any other Asian background; Any other Black background; Any other ethnic group; Any other White background; Bangladeshi; Black African; Black Caribbean; Gypsy, Roma and Traveller of Irish Heritage; Pakistani; and White and Black Caribbean.

⁷⁰ At all key stages, we also include a squared-term to account for possible non-linearities between pupil attainment and the number of pupils in a year group.

KS2:

- At the end of primary school, pupil attainment is measured by statutory KS2 assessments using pupils' average scaled scores in reading and mathematics.⁷¹ These are derived from national test results and can take values between 80 and 120. Pupils with missing data in both subjects receive a score of zero and are ranked at the bottom of the attainment distribution. Pupils missing data include those who sat the KS2 assessment but had a raw score below the minimum threshold to be awarded a scaled score, as well as those for whom a score is not applicable.
- In addition to sex, month of birth, and ethnicity, our KS2 pupil characteristics include:
 - SEND duration and severity over the primary school phase – specifically, we use the annual Spring Census to create a categorical variable based on whether a pupil has:
 - never been identified with SEND in the previous seven years;
 - ever been had an EHCP in the previous seven years; and
 - plus additional categories for the number of years on SEND Support for those never having an EHCP (ranging from one to seven).
 - SEND type, based on a pupil's primary type of need in Year 6; and
 - first language i.e. whether English is a pupil's first or additional language.
- KS2 prior attainment is based on a pupil's average (standardised) score on the EYFS profile in each of the seven areas of learning.⁷² There is no KS1 data available for the KS2 cohort completing primary school in 2023/24 due to the pandemic.
- We include a range of measures to capture pupils' disruption to schooling over the final three years of primary school, specifically:
 - Pupil absence using the average absence rate across Years 4 to 6).⁷³ We do not attempt to track pupils' absence histories before Year 4 due to the pandemic. We additionally include a control for the average percentage of absences (over the previous three years) that are unauthorised. This is to test whether the *type* of pupil absence has an effect, over and above the overall level.
 - The number of times a pupil has been suspended during Years 4 to 6.
 - The number of times a pupil has moved schools over the previous three years.
- Our KS2 cohort-level variables are the:
 - average (standardised) prior attainment for Year 6 based on the cohort's mean EYFS profile score across the seven areas of learning;
 - percentage of the Year 6 cohort who are disadvantaged;
 - percentage of the Year 6 cohort who have identified SEND;
 - percentage of the Year 6 cohort who arrive late to the state school system (i.e. in Years 5 of 6) with EAL;

⁷¹ If either subject's result is missing at KS2, the average takes the value of the subject the pupil does have a score for.

⁷² These are: communication and language; personal, social and emotional development; physical development; literacy; mathematics; expressive arts and design; and understanding the world.

⁷³ At KS2 and KS4, our measure captures absence across different settings for pupils who are enrolled in more than one school, though we filter out pupils who appear to have an implausibly high number of sessions. We set the cap on the overall number of sessions in a year at 418, to reflect the standard 380 sessions that a pupil typically has in any given year, plus 10 per cent.



- percentage of the Year 6 cohort who are in a lower-attaining ethnic group;⁷⁴
- number of Year 6 pupils;
- average absence rate of the Year 6 cohort;⁷⁵ and
- average suspension rate of the Year 6 cohort.
- Our KS2 school-level variables are the:
 - average school absence rate;
 - average suspension rate of the school; and
 - school type based on whether the school is local authority-maintained, sponsored academy, converter academy, free school or special school.
- We include regional controls based on where a pupil lives.

KS4:

- To assess attainment at the end of secondary school, we measure pupils' average GCSE grades across English and mathematics. Pupils who fail or do not enter a relevant English or mathematics qualification by the end of KS4 receive a score of zero for that component within the attainment distribution.
- In addition to sex, month of birth, and ethnicity, our KS4 pupil characteristics include:
 - SEND duration and severity over the previous 11 – specifically, we use the annual Spring Census to create a categorical variable based on whether a pupil has:
 - never been identified with SEND in the previous 11 years;
 - ever been had an EHCP in the previous 11 years; and
 - plus additional categories for the number of years on SEND Support for those never having an EHCP (ranging from 1 to 11 years).
 - SEND type, based on a pupil's primary type of need in Year 11; and
 - first language i.e. whether English is a pupil's first or additional language.
- KS4 prior attainment is based on (standardised) measures of a pupil's attainment on a range of statutory assessments at ages 5, 7 and 11 – specifically:
 - EYFS profile: Average score in each of the seven areas of learning;
 - KS1: Average score in each of mathematics, reading and writing; and
 - KS2: Average score in each of mathematics, reading, and grammar, punctuation and spelling.⁷⁶
- We include a range of measures to capture pupils' disruption to schooling over the final three years of secondary school, specifically:
 - Pupil absence using the average absence rate across Years 9 to 11. We do not attempt to track pupils' absence histories before Year 9 due to the pandemic. We additionally include a control for the average percentage of absences (over the previous three years)

⁷⁴ For KS2, lower-attaining ethnic groups comprise: Any other Black background; Any other ethnic group; Black Caribbean; Gypsy, Roma and Traveller of Irish Heritage; Pakistani; White British; and White and Black Caribbean. Our results are not sensitive to dropping White British pupils from this cohort measure.

⁷⁵ As for pupil absence, we convert the continuous absence variable into a categorical one to aide interpretation. We do the same for pupil- and cohort-level absence variables at KS4.

⁷⁶ This includes pupils working below the standard of the test at KS2 who are assigned point scores for the purposes of including them in prior attainment measures.



- that are unauthorised. This is to test whether the *type* of pupil absence has an effect, over and above the overall level.
- The number of times a pupil is suspended during Years 9 to 11.
 - The number of times a pupil is permanently excluded during Years 9 to 11.
 - The number of times a pupil moves schools over the previous three years.
 - Our KS4 cohort-level variables are the:
 - average (standardised) prior attainment for Year 6 based on the cohort's mean EYFS profile score across the seven areas of learning;
 - percentage of the Year 11 cohort who are disadvantaged;
 - percentage of the Year 11 cohort who have identified SEND;
 - percentage of the Year 11 cohort who arrive late to the state school system (i.e. in Years 10 or 11) with EAL;
 - percentage of the Year 11 cohort who are in a lower-attaining ethnic group;⁷⁷
 - number of Year 11 pupils;
 - average absence rate of the Year 11 cohort;
 - average suspensions rate of the Year 11 cohort;⁷⁸ and
 - average permanent exclusions rate of the Year 11 cohort.
 - Our KS4 school-level variables are the:
 - average school absence rate;
 - average permanent exclusions rate of the school; and
 - school type based on whether the school is local authority-maintained, sponsored academy, converter academy, free school, special school, city technology college, university technical college or studio school.
 - We include regional controls based on where a pupil lives.

⁷⁷ For KS4, lower-attaining ethnic groups comprise: Any other Black background; Any other ethnic group; Black Caribbean; Gypsy, Roma and Traveller of Irish Heritage; White British; and White and Black Caribbean. As at KS2, our results are not sensitive to dropping White British pupils from this cohort measure.

⁷⁸ We originally included the school's suspension rate in our school-level controls, alongside our Year 11 suspension rate. However, these two variables are highly correlated, so we dropped the former to satisfy the GVIF threshold and avoid multicollinearity.

16-19:

- At the end of 16-19 study, we measure attainment as a total point score over the students' best three qualifications, based on qualification size rather than number. We map all Level 1-3 qualifications onto the same scale, expressed as a total point score. A point score of 1 is equivalent to one A level grade, so in places we refer to the gap in (A-level equivalent) grades.⁷⁹
- In addition to sex, month of birth, and ethnicity our 16-19 pupil characteristics include:
 - SEND duration and severity over the primary and secondary school phase – specifically, we use the annual Spring Census to create a categorical variable based on whether, at the end of KS4, a pupil had:
 - never been identified with SEND in the previous six years;
 - ever been had an EHCP in the previous six years; and
 - plus additional categories for the number of years on SEND Support for those never having an EHCP (ranging from 1 to 11 years).
 - SEND type, based on a pupil's primary type of need in Year 11.
- Prior attainment is based on (standardised) measures of a pupil's attainment on a range of statutory assessments at ages 5, 7, 11 and 16 – specifically:
 - EYFS profile: Average score in each of the seven areas of learning;
 - KS1: Average score in each of mathematics, reading and writing;
 - KS2: Average score in each of mathematics, reading and grammar, punctuation and spelling; and
 - KS4: Average score in GCSE mathematics, English, and the open component of Progress 8. The EBacc component was not included due to concerns over multicollinearity.
- There are no available 16-19 absence data to use as a model control. We considered using secondary school absence records as a proxy. However, as our 2024 16-19 cohort was drawn from those completing KS4 in either 2021, 2022 or 2023, the impact of the pandemic on absence would not be comparable for students completing 16-19 education in one, two or three years. We excluded suspensions and exclusions for the same reason.
- We also tested the inclusion of an indicator of whether students moved institution during their 16-19 education. This accounted for a large amount of variation in attainment. However, institution mobility in the 16-19 phase is likely to be both a cause AND effect of low attainment. To avoid providing misleading results we decided to exclude this indicator from our analysis.
- We do not include a control for EAL in the 16-19 phase, as this data is not available for post 16 students. We do not take this information from their previous school census records as this data operates with too much of a lag and would miss out those who arrived most recently.
- Cohort KS4 prior attainment, the proportion of the cohort eligible for FSM (pre-16), and the proportion of pupils with SEND were highly correlated with each other. This is likely due to the more selective nature of post-16 education. To address multicollinearity, principal component analysis was applied to these variables. The first principal component explained over 80% of the total variance and loaded strongly on all three measures. This component was retained as a composite indicator of cohort intake characteristics and used in subsequent regression models, identified as 'cohort disadvantage'.

⁷⁹ Full details of our approach can be found in Annex C of '[Measuring the 16-19 disadvantage gap](#)'.

