**School absences, exclusions and criminal sentences amongst high-achieving children from disadvantaged socio-economic backgrounds**

John Jerrim[[1]](#footnote-1) (UCL Social Research Institute)

**Abstract**

A small but growing literature is exploring the later lifetime outcomes of initially high achieving young people from disadvantaged socio-economic backgrounds. These individuals have the potential to break through the glass ceiling and climb up the socio-economic ladder, though unfortunately many fail to achieve this goal. This paper presents new evidence on a selection of behavioural outcomes for this group, focusing on their attendance at and exclusions from school, along with cautions/sentences received for involvement in criminal activity. We find substantial differences in absence rates throughout secondary school relative to their equally able but more socio-economically advantaged peers, with this a particular issue for those of White and Mixed ethnicity. On the other hand, exclusions from school and cautions/sentences are particularly elevated amongst high achieving disadvantaged boys – most prominently those from Black and Mixed-race backgrounds – and peak during Key Stage 4. We also find that differences in attendance, exclusions and cautions/sentences while at school can only partially explain socio-economic differences in the propensity to be cautioned or sentenced as an adult. Our findings thus highlight the broad range of negative outcomes that hinder initially high achieving disadvantaged children’s prospects of achieving upwards social mobility.

**Key Words:** High-ability children, socio-economic disadvantaged, school absences, school exclusions, crime.

Acknowledgment: The Nuffield Foundation is an independent charitable trust with a mission to advance social well-being. It funds research that informs social policy, primarily in Education, Welfare and Justice. The Nuffield Foundation is the founder and co-funder of the Nuffield Council on Bioethics, the Ada Lovelace Institute and the Nuffield Family Justice Observatory. The Foundation has funded this project, but the views expressed are those of the authors and not necessarily the Foundation. Visit [www.nuffieldfoundation.org](http://www.nuffieldfoundation.org). Twitter: [@NuffieldFound](https://twitter.com/nuffieldfound)

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

1. **Introduction**

In many Western countries, social mobility has become a key policy issue. Governments across the world are seeking to ensure that young people have equal opportunities to succeed regardless of their socio-economic background, thereby building a fairer and more economically efficient society. Unfortunately, many countries are struggling to achieve this goal, with socio-economic differences in education, health and labour market outcomes proving difficult to meaningfully shift (Blanden et al., 2022; Machenbach et al, 2018; OECD, 2018).

Improving the later lifetime outcomes of disadvantaged children demonstrating high levels of early academic potential is likely to be pivotal to ensuring this changes in the future (Holt-White & Cullinane, 2023). Despite the socio-economic adversity this group faced early in life, they have managed to build a firm academic platform from which they can further develop. As such, initially high-achieving young people from disadvantaged socio-economic backgrounds have perhaps the best opportunity to break through the glass ceiling and increase diversity within top professional jobs (Social Mobility Commission, 2023). Indeed, if these children are unable to succeed, then this represents a waste of talent, with significant numbers of young people failing to fulfil their early academic potential.

A relatively small number of previous studies have focused on the development of initially high-achieving young people from poor backgrounds, both in England and internationally. Most focus on the academic progress they make at school and their later levels of educational achievement. For instance, a widely cited paper by Feinstein (2003) suggested that developmentally advanced two-year-olds from poor backgrounds were overtaken by developmentally delayed two-year-olds from rich backgrounds before they started school. Blanden and Machin (2010) replicated this finding, showing steep declines in the cognitive skills of initially high-achieving children from disadvantaged backgrounds between ages 3 and 5. This was debated, however, by Jerrim and Vignoles (2013) who suggested that this result was at least partially being driven by regression to the mean[[2]](#footnote-2). Washbrook and Lee (2015) studied this issue in the United States after attempting to address the issue of regression to the mean. They found that disadvantaged children with strong cognitive test scores at age 6 were caught up by socio-economically advantaged children with weak test scores by the early teenage years. Building on this work, Crawford et al. (2017) investigated how secondary school test scores and university entry rates differed across high-achieving primary school children from socio-economically advantaged and disadvantaged homes. They reported between ages 11 and 14 to be a key period, when high-achieving young people from poor backgrounds were particularly likely to fall behind their more advantaged peers academically. This is consistent with the work of Holt-White & Cullinane (2023), who found that children achieving top scores in national assessments at the end of primary school achieved significantly worse GCSE grades if they were from a socio-economically disadvantaged background. Interestingly, they also found disadvantaged young people who were high academic achievers at primary school were more likely to report feeling that they don’t have much of a chance in life than the most socio-economically advantaged group. A recent study by Jerrim and Carvajal (2024) investigated how a range of outcomes varied across 5-year-olds with strong test scores who were from different socio-economic backgrounds. They found that high-achieving children from poor backgrounds were more likely to report becoming disengaged in school, to engage in risky behaviours (e.g. smoking, drinking, taking drugs) and having brushes with the law than their more socio-economically advantaged peers.

While insightful, a notable limitation with this existing literature is the lack of evidence on disadvantaged high-achieving children’s broader behavioural outcomes. For instance, to what extent do initially high-achieving children from poor backgrounds start to display serious behavioural issues during secondary education, such as being persistently absent or excluded from school, and at what point do such issues emerge? This is important as failing to attend school – or consistently breaching school rules – is a marker of academic disengagement, and thus likely to coincide with a tail-off in their academic performance. We however also view these outcomes as important indicators in their own right. Likewise, prior research into “gifted” children more generally (not focused on those from poor backgrounds) has noted how such behavioural issues can emerge, either due to factors such as boredom in the classroom (Diezmann & Watters, 1997) or from their academic prowess making them stand out from their school peers, including their attempts to “fit-in” (Schall et al., 2014) or from being bullied (Martínez-Monteagud et al., 2023). Yet engaging in such behaviours is a known risk factor for several poor later lifetime outcomes, such as lower levels of educational achievement (Klein & Sosu, 2024) and an increased likelihood of engaging in criminal activity (Mazerolle et al., 2019). Indeed, the last thing that society needs is for its highest-achieving young people from socio-economically vulnerable homes to become engaged in serious crime.

The overarching aim of this paper is to contribute new evidence on such matters to the existing literature. Specifically, it documents how serious behavioural issues during adolescence and into early adulthood differ across children from different socio-economic backgrounds, who were amongst the highest academic achievers at the end of primary school. The only previous study to consider this issue in detail was Jerrim and Carvajal (2024), who found that initially high-achieving young people from poor backgrounds were more likely to report engaging in antisocial behaviour, fraternising with a more troublesome peer group and having brushes with the law as teenagers than their equally able but more socio-economically advantaged peers. That work was, however, based on survey data including only around 300 disadvantaged high-achieving young people, surveyed just twice during their teenage years (ages 14 and 17). We advance on this work by using recently linked administrative data from England, tracking school behaviour and criminal cautions/sentences for three entire state school cohorts. These data have some significant advantages, including allowing us to track young people’s absences, exclusions and cautions/sentences throughout their time at secondary school (thus potentially identifying key ages / time points when the behaviour of high-achieving disadvantaged young people starts to tail off), with the large sample size facilitating analysis by key sub-groups (e.g. gender and ethnicity). The latter is likely to be particularly insightful, with there likely to be intersections between socio-economic background, gender and ethnicity when it comes to absence from school (Sosu et al., 2021), school exclusions (Graham et al., 2019) and crime (Burgess-Proctor, 2006).

Thus, in summary, this paper seeks to address the following research questions:

* Research question 1. How do school absences differ between high-achieving young people from different socio-economic backgrounds during secondary school? How does this vary by gender and ethnicity?
* Research question 2. How do school exclusions differ between high-achieving young people from different socio-economic backgrounds during secondary school? How does this vary by gender and ethnicity?
* Research question 3. How do rates of criminal cautions/sentences differ across high-achieving young people from advantaged and disadvantaged socio-economic backgrounds during adolescence and early adulthood? How does this vary by gender and ethnicity?
* Research question 4. To what extent can socio-economic differences in criminal cautions/sentences amongst initially high-achieving young people be explained by differences in their school attendance, exclusions and grades at secondary school?
1. **Data**

Cohorts

Our analysis draws on two administrative databases that have recently been linked in England – the National Pupil Database (NPD) to data from the Police National Computer (PNC). The former includes school records for all children attending state schools. The latter provides data on all criminal cautions and sentences in England. The data owners (Department for Education and Ministry of Justice) have linked these databases using fuzzy matching, with details of the process used provided by Ministry of Justice (2022). Consequently, for state school pupils in England, the linked database includes information on each time they have been cautioned or sentenced (if ever). We have access to these data for three school cohorts:

* Cohort A. All children born September 1990-August 1991. These individuals were in Year 6 in 2001/2002 and completed compulsory schooling (Year 11) in 2006/2007. Caution and sentencing data are available for this cohort between ages 10 and 30.
* Cohort B. All children born September 1994-August 1995. These individuals were in Year 6 in 2005/2006 and completed compulsory schooling (Year 11) in 2010/2011. Caution and sentencing data are available for this cohort between ages 10 and 26.
* Cohort C. All children born September 2000-August 2001. These individuals were in Year 6 in 2011/2012 and completed compulsory schooling (Year 11) in 2016/2017. Caution and sentencing data are available for this cohort between ages 10 and 20.

The data include details on young people’s performance in national examinations, demographic characteristics (e.g. gender, ethnicity), measures of socio-economic background (e.g. eligibility for Free School Meals) and termly information on absences and exclusions from school.

Measurement of socio-economic background

We measure young people’s socio-economic background by combining information across the following two indicators:

* A set of dummy variables capturing whether the child was eligible for Free School Meals (FSM) for each year they were at school.
* Income Deprivation Affecting Children Index (IDACI) decile. This is a measure of the proportion of families within the child’s local area that have low levels of income.

These indicators are combined into a scale using a one-parameter Item Response Model. For each child, we produce Expected A Posteriori (EAP) estimates from this model, reflecting their position on this latent socio-economic background scale. Those young people within the bottom quartile of this scale within their cohort are used to define our “socio-economically disadvantaged” group. In contrast, young people in the top quartile form our “socio-economically advantaged” group.

Measurement of achievement

In England, state school children sit national examinations (Key Stage 2 tests) at ages 10/11, coinciding with then they are finishing primary school. These are relatively long tests, totalling over three hours of test time, and are externally marked. We use children’s scores in the reading and mathematics tests to define high levels of early achievement. Within each cohort, scores on these tests are first standardised to mean zero and standard deviation one, with the average then taken. Those with scores in the top quartile of the distribution within their cohort are then defined as “high early achievers”. Given that exclusions from school and criminal cautions/sentences are a major focus of this paper – instances of which do not start to rise until secondary school – our use of these tests balances the need to define high-achievement using high quality tests, while still capturing children’s academic abilities at a sufficiently early age. The first stages of secondary school have also been highlighted in prior research as a pivotal period for high-achieving disadvantaged children (Crawford et al., 2017; Jerrim & Carvajal, 2024); a time when they start to disengage in school, participate in troublesome behaviour and their academic progress starts to tail off.

Definition of high-achieving young people from different socio-economic backgrounds

Our operationalisation of high initial achievers from disadvantaged socio-economic backgrounds combines information from the two measures described above. Specifically, it is formed of children from the bottom 25% of the socio-economic distribution who are within the top 25% of the Key Stage 2 test distribution within each school cohort. Within our analysis, we will compare the outcomes of this group to equally high-achieving young people from the most advantaged socio-economic backgrounds. This comparison group is defined as young people in the top 25% of the socio-economic distribution who also had Key Stage 2 test scores in the top 25% of the national distribution, within their respective cohort[[3]](#footnote-3).

School absences

The dataset includes detailed information on school absences. For each half-term[[4]](#footnote-4), the data records the percent of “sessions” (half-days) that each pupil missed. This information encompasses Years 10-to-11 (age 14-16) for Cohort A, Years 6-to-11 (age 10-16) for Cohort B and Years 1-to-11 (age 6 – 16) for Cohort C. Within our analysis we consider how absence rates compare across young people with strong Key Stage 2 test scores during their time at school (for Cohort C, this covers almost their entire time in compulsory education).

School exclusions

Within the NPD-MOJ data, information is provided on exclusions from school. This includes fixed-term exclusions (i.e. suspensions – removal from school for a limited period) and permanent exclusions (i.e. expulsion – where the pupil does not return to the school). The precise date is provided for both exclusion types. For fixed-term exclusions, we derive a binary measure for each child for each academic year during secondary school. This is coded as one if the child was excluded at any point, and zero otherwise. We focus only on exclusions during secondary school, as exclusions from primary school are rare (FFT Datalab, 2024). Given the rarity of permanent exclusions (resulting in small sample sizes) we only report on this outcome in the supplementary material (see Online Appendix B)[[5]](#footnote-5).

Criminal cautions and sentences

The data contains information on all criminal cautions and sentences each young person received (if any), including offense date, the age of the individual when the offense was committed, offense type (e.g. motoring, violent crime) and a “disposal code” (i.e. what the caution/sentence led to – e.g. custodial sentence, fine). For each child in the dataset, we derive a set of binary indicators. These are coded as one if they were cautioned or sentenced for any reason at a given age (e.g. ever cautioned or sentenced at age 17) and zero otherwise. We also derive further summary measures for whether the young person ever committed an offense at any age, and whether they committed any offense as an adult (defined as after turning age 18). These variables are used to address our final research question, where we examine the extent that socio-economic differences in criminal caution/sentences amongst high early achievers can be explained by variation in behaviours and outcomes during their time at school.

1. **Methodology**

Research questions 1-3. Differences in school absences, school exclusions and criminal cautions/sentences.

For each of our first three research questions we begin by presenting a series of descriptive statistics. For research question 1 (school absences) this compares the percent of school sessions attended by high-achieving children from socio-economically advantaged and disadvantaged backgrounds during each term they were at school. A similar analysis is presented to address research question 2 (exclusions from school) though now only focusing on occurrences during secondary school (given that few children receive exclusions from primary school). For research question 3 (incidence of criminal cautions/sentences) we present annual figures from secondary school into adulthood (e.g. age 11 through to – where possible - age 30)[[6]](#footnote-6). These results together provide a detailed overview of how the behaviour of high-achieving young people from disadvantaged backgrounds changes as they progress through secondary school.

We then turn to differences across high-achieving disadvantaged children of different genders and ethnicities. Specifically, outcomes are compared across the following sub-groups:

* High-achieving, socio-economically disadvantaged White boys.
* High-achieving, socio-economically disadvantaged White girls.
* High-achieving, socio-economically disadvantaged Asian boys.
* High-achieving, socio-economically disadvantaged Asian girls.
* High-achieving, socio-economically disadvantaged Black boys.
* High-achieving, socio-economically disadvantaged Black girls.
* High-achieving, socio-economically disadvantaged Mixed-race boys.
* High-achieving, socio-economically disadvantaged Mixed-race girls.

This will be done in two ways. First, analogous descriptive statistics will be presented for some of these sub-groups (as sample sizes allow). Second, we will estimate regression models of the form:

$O\_{i}= α+ β.Grp\_{i}+φ.Ach\_{i}+∅.Cohort\_{i}+θ.ABS\\_Prior\_{i}+ε\_{i}$ (1)

Where:

$O\_{i}$ = One of our outcomes of interest.

$Grp\_{i}$ = A vector of dummy variables capturing each of the high-achieving disadvantaged sub-groups listed above.

$Ach\_{i}$ = A vector of academic achievement measures at primary school, including Key Stage 1 and Key Stage 2 reading and mathematics test scores.

$Cohort\_{i}$ = A vector of dummy variables reflecting the cohort.

$ABS\\_Prior\_{i}$ = Attendance during primary school or at the start of secondary school.

$ε\_{i}$ = Random error term.

i = Individual i.

Linear regression will be used for continuous outcomes (school absences) and logistic regression used for binary outcomes (exclusions; criminal cautions/sentences). When estimating these models, the respective outcomes ($O\_{i}$) will be defined as the average (absences) or any incidence (exclusion; criminal cautions/sentences) recorded within a given secondary school year. The analysis will thus capture the extent that high-achieving disadvantaged children of different genders and ethnicities have a record of a serious behaviour incident during their time in secondary education.

When presenting these results, we note the methodological issue of “Kelley’s paradox”, discussed in detail by Jerrim and Carvajal (2024). In particular, when the test scores used to define “high achieving” groups are measured with error, one may overestimate the difference in future outcomes amongst high-achieving children from different socio-economic backgrounds. We interrogate this issue in Appendix G, applying the sensitivity analyses proposed by Jerrim and Carvajal (2024). This approach essentially illustrates how our results change under different assumptions about the measurement error present in the Key Stage 2 tests. We find the magnitude of the difference in attendance, exclusion and cautions/sentences across high-achieving children from different socio-economic backgrounds declines as the assumed measurement error in the Key Stage 2 tests increases. However, even when one assumes the measurement error in the Key Stage 2 tests is reasonably large, most of the socio-economic gaps reported on in this paper remain non-trivial.

Research question 4. To what extent can socio-economic differences in criminal cautions/sentences during adulthood be explained by differences in absences, exclusions and attainment at school, amongst children with high levels of prior achievement?

Our final research question considers how engagement in criminal activity during adulthood differs across initially high-achieving young people from different socio-economic backgrounds, and the extent that such differences can be explained by variation in their attendance, exclusion, academic and criminal caution/sentence outcomes during their time at school. This part of our analysis restricts the sample to Cohorts A and B, given that we can observe their criminal cautions/sentences through to age 30 (cohort A) and 26 (cohort B)[[7]](#footnote-7). For these cohorts we restrict the sample to initially high-achieving young people (top Key Stage 2 quartile) and estimate a series of logistic regression models of the form:

$C18\_{i}=α+β.SES\_{i}+φ.Ach\_{i}+∅.Cohort\_{i}+π. Ab\_{i}+σ.Ex\_{i}+τ.C16\_{i}+ω.GCSE\_{i}$ (2)

Where:

$C18\_{i}$ = A binary outcome coded 1 if the individual was recorded as having been cautioned/sentenced as an adult, and 0 otherwise.

$SES\_{i}$ = A vector of dummy variables capturing socio-economic status quartiles.

$Ach\_{i}$ = A vector of academic achievement measures at primary school, including Key Stage 1 and Key Stage 2 reading and mathematics test scores.

$Cohort\_{i}$ = A vector of dummy variables reflecting the cohort.

$Ab\_{i}$ = The average rate of pupil attendance during secondary school.

$Ex\_{i}$ = The number of times the individual was excluded during secondary school.

$C16\_{i}$ = The number of times the individual was cautioned or sentenced during secondary school.

$GCSE\_{i}$ = Grades achieved in age 16 national GCSE examinations, as captured by overall GCSE point scores and mathematics/English Language grades.

i = Individual i.

Five specifications of this model are estimated, each including additional sets of controls. The baseline model (M0) includes socio-economic group ($Grp\_{i}$), background controls ($D\_{i}$), academic achievement during primary school ($Ach\_{i}$) and cohort ($Cohort\_{i}$) as the only covariates. The $β$ parameter thus captures differences across high-achieving young people with similar levels of academic skills and backgrounds at the end of primary education. Model M1 then includes a control for secondary school absences ($Ab\_{i}$), with secondary school exclusions ($Ex\_{i}$) and secondary school cautions/sentences ($C16\_{i}$) added in specifications M2 and M3. Our primary interest is how the $β$ parameter changes across these model specifications; to what extent can differences in these markers of troublesome behaviour during secondary school jointly explain the socio-economic gap in initially high-achieving children’s receipt of criminal cautions/sentences as adults? The final model specification (M4) additionally controls for GCSE grades to investigate whether academic achievement at the end of secondary school can further explain the socio-economic gap in initially high-achieving children’s propensity to be cautioned/sentenced as adults.

A similar logistic regression modelling approach is then used to investigate differences in criminal caution/sentences as adults amongst initially high-achieving disadvantaged young people of different genders and ethnicities. Our interest is again the extent that differences across these sub-groups can be traced back to variation in their behavioural and academic outcomes during secondary school, relative to initially high-achieving, socio-economically disadvantaged White boys as the reference group.

1. **Results**

Research question 1. School absences

Figure 1 presents the percentage of school sessions missed each academic term between Year 1 and Year 11, pooling data across the three cohorts. (Appendix A – Figure A3 – provides results for each cohort separately). The solid grey line refers to high-achieving children from socio-economically advantaged backgrounds, with the dashed black line providing analogous results for their equally high-achieving but socio-economically disadvantaged peers.

**Figure 1. Average absence rates of high-achieving children from socio-economically advantaged and disadvantaged backgrounds.**

**Notes:** Figures refer to the percent of sessions missed. Vertical dashed green line illustrates the transition between the end of primary (Year 6) and start of secondary (Year) school. Data pooled across the three cohorts. See Appendix A for further detailed results.

For the most advantaged young people, the absence rate is broadly stable during their time at school, typically sitting somewhere between three and four percent of sessions missed each term[[8]](#footnote-8). In contrast, the absence rate for initially high-achieving disadvantaged pupils is higher and has more variable. For instance, in Year 7 the absence rate for the high-achieving disadvantaged group is around 5.5%, relative to around 3.5% for their equally able but more socio-economically advantaged peers. However, by Year 10, the absence rate has reached approximately 7% for the former, but only 4% for the latter. This points towards an increasing socio-economic disparity amongst children with high levels of prior achievement during secondary school.

Table 1 panel (a) reiterates this result, focusing upon annual secondary school absences amongst Cohort C (the group we can observe throughout their time at school). This presents estimates from our regression model, with figures referring to the percentage point difference in school sessions missed relative to the most disadvantaged socio-economic group. Note how, in Year 7, socio-economically advantaged young people with strong Key Stage 2 scores miss 0.9 percentage point fewer school sessions than socio-economically disadvantaged pupils with equally strong Key Stage 2 scores. The magnitude of this gap gradually increases, however, reaching 1.76 percentage point fewer sessions missed in Year 9 and up to 2.5 percentage points in Year 11. Moreover, Appendix G (Table G1) demonstrates how this gap cannot solely be attributed to Kelley’s paradox (i.e. potential measurement error in the Key Stage 2 tests). Together, this illustrates how higher rates of absence from school may serve to undermine the progress and development of initially high-achieving young people from disadvantaged socio-economic backgrounds.

**Table 1. Regression estimates of the difference in secondary school attendance rates across high-achieving children from different socio-economic backgrounds.**

1. Differences across socio-economic groups

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Year 7** | **Year 8** | **Year 9** | **Year 10** | **Year 11** |
|  | **% diff** | **SE** | **% diff** | **SE** | **% diff** | **SE** | **% diff** | **SE** | **% diff** | **SE** |
| Q2 SES | -0.59\* | 0.05 | -0.76\* | 0.05 | -1.14\* | 0.06 | -1.45\* | 0.07 | -1.67\* | 0.08 |
| Q3 SES | -0.74\* | 0.05 | -1.08\* | 0.05 | -1.54\* | 0.06 | -2.00\* | 0.07 | -2.12\* | 0.08 |
| Highest SES group | -0.91\* | 0.05 | -1.25\* | 0.05 | -1.76\* | 0.06 | -2.34\* | 0.07 | -2.53\* | 0.08 |
| **N** | **112,811** | **112,418** | **111,826** | **111,429** | **110,798** |

**Notes:** Figures refer to the percentage point difference in school absence rates in comparison to children from disadvantaged socio-economic backgrounds. SE refers to the standard error. Sample restricted to high achieving (young people in the top Key Stage 2 quartile) in Cohort C. Models control for gender, ethnicity, Key Stage 1 and Key Stage 2 test scores, cohort and annual school absence rates during primary school. \* indicates statistical significance at the 5% level. See Appendix A for further detailed results.

1. Differences between high-achieving disadvantaged children of different genders and ethnicities.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Year 7** | **Year 8** | **Year 9** | **Year 10** | **Year 11** |
|  | **% diff** | **SE** | **% diff** | **SE** | **% diff** | **SE** | **% diff** | **SE** | **% diff** | **SE** |
| White girl | -0.47\* | 0.10 | -0.08 | 0.11 | 0.72\* | 0.13 | 0.43\* | 0.15 | 0.59\* | 0.16 |
| Asian boy | -1.26\* | 0.15 | -1.53\* | 0.17 | -2.00\* | 0.20 | -2.34\* | 0.23 | -2.81\* | 0.25 |
| Asian girl | -1.53\* | 0.14 | -1.88\* | 0.16 | -2.08\* | 0.19 | -2.47\* | 0.22 | -2.90\* | 0.24 |
| Black boy | -0.99\* | 0.18 | -1.25\* | 0.20 | -1.43\* | 0.23 | -1.49\* | 0.27 | -2.02\* | 0.30 |
| Black girl | -1.51\* | 0.16 | -1.56\* | 0.18 | -2.00\* | 0.21 | -2.06\* | 0.25 | -1.98\* | 0.27 |
| Mixed race boy | -0.15 | 0.23 | 0.24 | 0.26 | -0.35 | 0.31 | -0.11 | 0.37 | 0.54 | 0.40 |
| Mixed race girl | -1.38\* | 0.20 | -0.91\* | 0.23 | -0.15 | 0.27 | 0.03 | 0.32 | -0.29 | 0.35 |
| Other ethnicity boy | -1.18\* | 0.24 | -0.86\* | 0.26 | -1.61\* | 0.31 | -1.06\* | 0.37 | -2.06\* | 0.40 |
| Other ethnicity girl | -1.60\* | 0.22 | -1.33\* | 0.24 | -1.34\* | 0.28 | -1.00\* | 0.33 | -1.31\* | 0.36 |
| High achieving, above average SES | -1.33\* | 0.07 | -1.54\* | 0.08 | -1.74\* | 0.10 | -2.40\* | 0.12 | -2.59\* | 0.13 |
| High achieving, Q3 SES | -1.21\* | 0.08 | -1.26\* | 0.09 | -1.44\* | 0.10 | -1.89\* | 0.12 | -2.16\* | 0.13 |
| **N** | **112811** | **112418** | **111826** | **111429** | **110798** |

**Notes:** Figures refer to the percentage point difference in school absence rates in comparison to children from high-achieving, disadvantaged, white boys. SE refers to the standard error. Sample restricted to high achieving (young people in the top Key Stage 2 quartile) in Cohort C. Models control for Key Stage 1 and Key Stage 2 test scores, cohort and annual school absence rates during primary school. See Appendix A for further detailed results.

Figure 2 builds on these results by investigating differences across genders and ethnic groups. Findings for boys are presented on the left, with those for girls on the right.

The broad pattern for initially high-achieving, disadvantaged boys and girls is similar. The starkest result being the higher rates of absence amongst White and mixed-race ethnic groups in comparison to their Asian and Black peers. For instance, high achieving, disadvantaged White and Asian girls have broadly similar absence rates during primary school (see Figure 2b, left of the vertical green line). Yet, during secondary school (right of the dashed green line in Figure 2b) their respective school absence rates clearly diverge. As a result, by Year 10, initially high-achieving disadvantaged White girls are (on average) missing around 8.5% of school sessions compared to only around 5% of their equally able Asian peers. This equates to a difference of around seven extra days of lost learning over the course of an academic year.

Table 1b formalises these descriptive results by presenting estimates from our regression models, focusing on pupils in Cohort C. Figures again refer to percentage point differences in the absence rate compared to initially high-achieving disadvantaged White boys as the reference group. Consistent with Figure 2, there is a clear gap in school absences between White boys and Asian/Black pupils (regardless of their gender) already apparent in Year 7, but grows further still as high-achieving but socio-economically disadvantaged young people progress through secondary school.

**Figure 2. Absence rates amongst high-achieving disadvantaged children. Results by gender and ethnicity.**

1. **Boys (b) Girls**

**Notes:** Figures refer to the percent of sessions missed. Vertical dashed green line illustrates the transition between the end of primary (Year 6) and start of secondary (Year) school. See Appendix A for further detailed results.

Research question 2. School exclusions.

Figure 3 and Table 2b present socio-economic differences in fixed-term exclusions during secondary school amongst pupils within the top Key Stage 2 quartile. Both illustrate how fixed-term exclusions are substantially higher amongst disadvantaged high-achieving young people than their more advantaged peers, even during the first stages of secondary school. Take Year 7, for example. Just 0.2% of high-achieving pupils from the most advantaged socio-economic backgrounds receive a fixed-term exclusion in their first year of secondary school, compared to 1.5% of those from disadvantaged backgrounds (odds ratio in Table 2a of 0.16)[[9]](#footnote-9). Appendix B illustrates how similar results are obtained for each of the three school cohorts (see Appendix Figures B1 and B3).

**Figure 3. Average fixed exclusion rates of high-achieving children from socio-economically advantaged and disadvantaged backgrounds.**

Notes: Figures refer to the percent of pupils receiving a fixed-term exclusion during secondary school. Data pooled across the three cohorts. See Appendix B for further detailed results.

**Table 2. Logistic regression estimates of the difference in secondary school fixed exclusion rates across high-achieving children from different socio-economic backgrounds. Odds ratios.**

1. Differences across socio-economic groups

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|   | **Year 7** | **Year 8** | **Year 9** | **Year 10** | **Year 11** |
| Q2 SES | 0.47\* | 0.50\* | 0.49\* | 0.50\* | 0.53\* |
| Q3 SES | 0.29\* | 0.30\* | 0.29\* | 0.31\* | 0.36\* |
| Highest SES group | 0.16\* | 0.17\* | 0.19\* | 0.19\* | 0.22\* |
| **N** | **241361** | **241361** | **241361** | **241361** | **241361** |

**Notes:** Figures refer to odds ratio in comparison to children from disadvantaged socio-economic backgrounds as the reference group. Sample restricted to high achieving (young people in the top Key Stage 2 quartile) in Cohorts B and C. Models control for gender, ethnicity, Key Stage 1 and Key Stage 2 test scores, cohort and school absence rate in Year 7. \* indicates odds ratio significantly different from one at the 5% level. See Appendix B for further detailed results.

1. Differences between high-achieving disadvantaged children of different genders and ethnicities.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|   | **Year 7** | **Year 8** | **Year 9** | **Year 10** | **Year 11** |
| White girl | 0.35\* | 0.44\* | 0.66\* | 0.58\* | 0.50\* |
| Asian boy | 0.63\* | 0.58\* | 0.83 | 0.88 | 0.72\* |
| Asian girl | 0.25\* | 0.18\* | 0.31\* | 0.24\* | 0.18\* |
| Black boy | 1.87\* | 1.42\* | 1.91\* | 1.57\* | 1.57\* |
| Black girl | 0.79 | 0.85 | 0.91 | 0.83 | 0.61\* |
| Mixed race boy | 1.82\* | 1.42\* | 1.39\* | 1.35\* | 1.23 |
| Mixed race girl | 0.72 | 0.94 | 1.18 | 1.11 | 0.80 |
| Other ethnicity boy | 1.03 | 0.88 | 0.78 | 0.94 | 1.00 |
| Other ethnicity girl | - | 0.42\* | 0.34\* | 0.45\* | 0.33\* |
| High achieving, above average SES | 0.15\* | 0.16\* | 0.19\* | 0.19\* | 0.21\* |
| High achieving, Q3 SES | 0.32\* | 0.34\* | 0.40\* | 0.39\* | 0.38\* |
| **N** | **241361** | **241361** | **241361** | **241361** | **241361** |

**Notes:** Figures refer to odds ratio in comparison to children from high-achieving, disadvantaged, white boys. SE refers to the standard error. Sample restricted to high achieving (young people in the top Key Stage 2 quartile) in Cohorts B and C. Models control for Key Stage 1 and Key Stage 2 test scores, cohort and annual school absence rates during Year 7. Figures for “Other ethnicity girl” in Year 7 suppressed due to small sample size. \* indicates odds ratio significantly different from one at the 5% level. See Appendix B for further detailed results

The absolute difference between socio-economic groups then increases through to Year 10 when it reaches 4.2 percentage points; 5.4% of high-achieving disadvantaged pupils receive a fixed exclusion during Year 10 versus 1.2% of their equally able but more socio-economically advantaged peers. In relative terms, however, Table 2a suggests the socio-economic gap stays broadly the same (the odds ratio equals 0.16 in Year 7 versus 0.19 in Year 10). Appendix G (Table G2) indicates that Kelley’s paradox is only partially able to explain these results. Even when one assumes the measurement error in the Key Stage 2 tests is quite substantial (i.e. that around half the variation in the Key Stage 2 scores can be attributed to measurement error) the socio-economic gap in fixed-term exclusions amongst high-achievers remains at 0.5 percentage points in Year 7 and 2.4 percentage points in Year 10.

Further detail to these results is provided in Figure 4 and Table 2b, where the sample has been restricted to high-achieving young people from disadvantaged backgrounds, and we explore differences across genders and ethnicities.

The first point to note is that fixed-term exclusions are much higher amongst high achieving disadvantaged young people from Black and Mixed-race backgrounds than their peers of White and Asian ethnicity. For boys in Year 10, around 10% of the former receive at least one fixed-term exclusion, compared to 6% to 7% amongst the latter. It is also notable how fixed-term exclusions are much more prevalent amongst boys than girls within this high-achieving, socio-economically disadvantaged sample. Indeed, fixed-term exclusions for the group of girls with the highest exclusion rate (Mixed) is equal to that of the group of boys with the lowest exclusion rate (Asian). The logistic regression results presented in Table 2b further emphasise these findings, with the odds ratio sitting significantly above one for only Black and Mixed-race boys (indicating that their exclusion rate is higher than for high achieving, disadvantaged White boys). Likewise, it is only White and Asian girls where the odds ratio is consistently below one and statistically significant at the 5% level. Thus, together, these results illustrate how there are important intersections between socio-economic status, ethnicity and gender amongst initially high-achieving pupils in terms of school exclusions.

**Figure 4. Fixed exclusion rates amongst high-achieving disadvantaged children. Results by gender and ethnicity.**

1. Boys (b) Girls

Notes: Figures refer to the percent of pupils receiving a fixed-term exclusion during secondary school. Data pooled across the three cohorts. See Appendix B for further detailed results.

Research question 3. Criminal cautions/sentences.

Figure 5 turns to the socio-economic gap amongst early high achievers in the chances they receive a criminal caution or sentence between the ages of 11 and 30. This is supplemented by Appendix C, where we present results separately by cohort, along with estimates from a series of logistic regression models (which produce qualitatively similar results). Appendix D, E and F also provide additional results, where we focus on the chances of a high-achieving young person being cautioned or sentenced for three specific types of crime (drugs, violent crime, theft).

**Figure 5. The percent of high-achieving children being cautioned/sentenced at least once by age**

Notes: Figures refer to the percentage of individuals receiving a caution or sentence by age. Data pooled across cohorts. See Appendix C for further details.

While a difference can already be observed at age 11, the socio-economic gap in criminal cautions/sentences start to rapidly grow after age 12, before peaking around age 16. At this point, there is a five-fold difference in the rate that high-achieving advantaged and disadvantaged young people receive criminal cautions/sentences – standing at 2.5% amongst the highest socio-economic group versus 0.5% for the bottom socio-economic group. The percentage receiving a criminal caution/sentence then declines – particularly amongst the most disadvantaged socio-economic group – in the late teenage years and into their early 20s. Appendix G (Figure G4) illustrates how the measurement error in Key Stage 2 scores would have to be substantial (i.e. for measurement error to account for 75% of the variation in the Key Stage 2 scores) for Kelley’s paradox to fully account for the socio-economic gap observed in these results.

Figure 6 then turns to results for high-achieving disadvantaged young people of different genders and ethnicities (where sample size allows). Note that we now only consider the period between ages 13 and 25 to meet sample size requirements to comply with statistical disclosure control.

**Figure 6. The percent of high-achieving children being cautioned/sentenced at least once by age. Differences by age and ethnicity.**

Notes: Figures refer to the percentage of individuals receiving a caution or sentence by age. Data pooled across cohorts. See Appendix C for further details. Results for Asian girls cannot be reported due to the insufficient sample size for this group.

There are three key points to note. First, high-achieving disadvantaged boys are much more likely to be cautioned or sentenced than high-achieving disadvantaged girls. Second, the peak in the caution/sentencing rate is later for boys (around ages 16 to 19) than for girls (around ages 13 to 15). Finally, Asian boys stand out as being less likely to be cautioned or sentenced than boys from other ethnic groups. For instance, 2.8% of high-achieving disadvantaged Asian boys were cautioned or sentenced for a crime they committed at age 18, compared to around 5.4% of Black and Mixed-race boys (and 3.9% of White boys). As a point of reference, the rate for girls for each ethnic group is just under 1%. Appendix C (Table C5) complements these results with estimates from logistic regression models, including where differences are significantly different from high achieving, disadvantaged White boys (the reference group).

Research question 4. Being cautioned/sentenced as an adult

To conclude, Table 3 panel (a) investigates socio-economic differences in being cautioned or sentenced for a crime as an adult amongst initially high-achieving pupils, conditional on their background characteristics and various outcomes (e.g. attendance, exclusions, test scores) through to the end of secondary school. Estimates refer to odds ratios, with values less than one indicating lower odds of being cautioned or sentenced compared to high-achieving disadvantaged pupils as the reference group.

Model M0 refers to the bivariate relationship between being cautioned/sentenced and socio-economic status amongst pupils with strong Key Stage 2 scores. These unconditional results indicate that the odds of a high-achieving socio-economically advantaged young person receiving a criminal caution or sentence as an adult is around one-third of their equally high-achieving, but socio-economically disadvantaged, peers (odds ratio = 0.32). The inclusion of background demographic and early prior achievement measures in model M1 do not change this result (odds ratio = 0.35). When school absences are included in model M2, the odds ratio does move somewhat closer to one (odds ratio = 0.44). This suggests that – to some extent – the socio-economic gap in being cautioned/sentenced as an adult can be traced back to differences in their attendance at secondary school. While the addition of a control for exclusions from school leads to little further change to the odds ratio in model M3 (odds ratio = 0.47) controlling for whether they received a caution or sentence while they were at school does (odds ratio = 0.55). In other words, the greater propensity for high-achieving disadvantaged pupils to be cautioned/sentenced during their time at school can partly explain – at least in a statistical sense – their greater propensity to be cautioned/sentenced as an adult (relative to their more advantaged peers). The final specification – where GCSE grades are also added (M5) – nudges the estimates odds ratio slightly further towards one (standing at 0.6).

**Table 3. Logistic regression estimates of high-achieving primary school pupils being cautioned/sentenced as an adult. Differences across socio-economic groups. Odds ratio.**

1. Differences across socio-economic groups

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **M0** | **M1** | **M2** | **M3** | **M4** | **M5** |
| Q2 SES | 0.57\* | 0.62\* | 0.70\* | 0.73\* | 0.78\* | 0.82\* |
| Q3 SES | 0.44\* | 0.49\* | 0.57\* | 0.60\* | 0.67\* | 0.70\* |
| Top SES quartile | 0.32\* | 0.35\* | 0.44\* | 0.47\* | 0.55\* | 0.60\* |
| **N** | **281413** | **279693** | **279693** | **279693** | **279693** | **276833** |
| Controls |  |  |  |  |  |  |
| Gender | - | Y | Y | Y | Y | Y |
| Ethnic group | - | Y | Y | Y | Y | Y |
| KS1 maths scores | - | Y | Y | Y | Y | Y |
| Cohort | - | Y | Y | Y | Y | Y |
| KS2 scores | - | Y | Y | Y | Y | Y |
| School absences | - | - | Y | Y | Y | Y |
| School exclusions | - | - | - | Y | Y | Y |
| School cautions/sentences | - | - | - | - | Y | Y |
| GCSE grades | - | - | - | - | - | Y |

**Notes:** Figures refer to odds ratio in comparison to children from disadvantaged socio-economic backgrounds as the reference group. \* indicates odds ratio significantly different from one at the 5% level. See Appendix C for the full set of estimates including standard errors.

1. Differences between high-achieving disadvantaged children of different genders and ethnicities.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **M0** | **M1** | **M2** | **M3** | **M4** | **M5** |
| White girl | 0.25\* | 0.27\* | 0.24\* | 0.25\* | 0.27\* | 0.29\* |
| Asian boy | 0.81\* | 0.88 | 0.99 | 1.00 | 1.13 | 1.27\* |
| Asian girl | 0.08\* | 0.09\* | 0.10\* | 0.12\* | 0.15\* | 0.19\* |
| Black boy | 1.33\* | 1.50\* | 1.73\* | 1.56\* | 1.67\* | 1.70\* |
| Black girl | 0.35\* | 0.42\* | 0.47\* | 0.46\* | 0.54\* | 0.59\* |
| Mixed race boy | 1.13 | 1.26\* | 1.26\* | 1.17 | 1.18 | 1.16 |
| Mixed race girl | 0.32\* | 0.37\* | 0.33\* | 0.32\* | 0.33\* | 0.36\* |
| **N** | **281413** | **279693** | **279693** | **279693** | **279693** | **276833** |
| Controls |  |  |  |  |  |  |
| KS1 maths scores | - | Y | Y | Y | Y | Y |
| Cohort | - | Y | Y | Y | Y | Y |
| KS2 scores | - | Y | Y | Y | Y | Y |
| School absences | - | - | Y | Y | Y | Y |
| School exclusions | - | - | - | Y | Y | Y |
| School cautions/sentences | - | - | - | - | Y | Y |
| GCSE grades | - | - | - | - | - | Y |

**Notes:** Figures refer to odds ratio in comparison to children from high-achieving, disadvantaged, white boys. SE refers to the standard error. \* indicates odds ratio significantly different from one at the 5% level. See Appendix C for further detailed results.

Together, the results presented in Table 3a suggest that socio-economic differences in being cautioned/sentenced as an adult can – to some extent – be predicted by variation in their attendance, behaviour and achievement outcomes during their time at school. However, even after accounting for these factors, initially high-achieving young people from disadvantaged backgrounds are still much more likely to be cautioned or sentenced for a crime as an adult than more advantaged socio-economic groups.

Finally, Table 3 panel (b) focuses on differences amongst disadvantaged high achievers of different genders and ethnicities. There are four key findings to report.

First, for White and Mixed-race girls the odds ratio in the unconditional model (M0) sits closer to zero than one; they are much less likely to be cautioned or sentenced as adults than equally high-achieving, equally disadvantaged White boys. There is little consistent change to the odds ratio across the various model specifications. This indicates that differences in attendance, behaviour, prior cautions/sentences and achievement at school are unrelated to the magnitude of this gap.

Second, the estimated odds ratio for Black and Asian girls in model M0 also falls well below one. It, however, moves closer towards parity as the controls are added. For instance, the odds ratio for Black girls moves up from 0.35 (M0) to 0.59 (M5), with the inclusion of prior achievement scores and cautions/sentences while at school being particularly important. Hence, while there is also a clear gap in being cautioned/sentenced as an adult between high-achieving disadvantaged White boys and equally able, disadvantaged Black/Asian girls, this to some extent reflects patterns in their behaviour and academic outcomes during their time at school.

Third, high achieving disadvantaged Black boys are the only group where the estimated odds ratio is significantly above one in all model specifications. In other words, they are the only group more likely to be sentenced or cautioned as adults than high achieving, disadvantaged White boys. Moreover, the odds ratio tends to increase as we add controls capturing their various outcomes during secondary school, rising from 1.33 in M0 to 1.70 in M5. This means that initially high achieving disadvantaged Black boys are 70% more likely to receive a criminal caution/sentence as an adult than high achieving disadvantaged White boys, even when they had the same level of absences and exclusions from school, prior occurrences of criminal caution/sentences and GCSE grades.

Finally, in the unconditional model, Asian boys are slightly less likely to be cautioned or sentenced as adults than their White peers. However, this difference is quickly explained by differences in background characteristics, prior attainment and absences during secondary school (odds ratio ~ 1 in M2 and M3). Then, after accounting for differences in prior occurrences of criminal cautions/sentences and GCSE grades, the odds ratio is significantly above one, reaching 1.27. Thus, high achieving disadvantaged Asian boys are around 27% more likely to be cautioned or sentenced during early adulthood than equally able, equally disadvantaged White boys with similar attendance, exclusions, prior cautions/sentences and achievement during their time at secondary school.

1. **Conclusions**

High-achieving young people from disadvantaged backgrounds are a key group for enhancing social mobility. They have built solid foundations upon which one would hope they are able to build, though unfortunately many are unable to fulfil their early academic potential. Most previous research on this group has focused on their future academic outcomes, such as achievement in national examinations and entry into university (e.g. Crawford et al., 2017). Fewer studies have focused on broader outcomes, including indicators of serious issues surrounding their behaviour. This is despite the potential for issues such as attendance and exclusions from school – and getting into trouble with the law – being potential drivers of why the academic potential of some high achieving disadvantaged young people goes unfulfilled.

The present paper has sought to contribute insight on such matters to the existing evidence base. Using administrative data covering three school cohorts in England, we have shown how socio-economic gaps in attendance and exclusions amongst previously high achieving pupils emerge during the early stages of secondary school before peaking around Year 10 (age 14/15). These differences are unlikely to be explained by Kelley’s paradox, with non-trivial variation in absences, exclusions and criminal cautions/sentences observed even when one assumes a reasonably high degree of measurement error is present in the Key Stage 2 tests. There appears to be important intersectionality with gender and ethnicity, with increasing levels of secondary school absence being a particular problem amongst teenagers of White and Mixed ethnicity. Likewise, high rates of school exclusions and criminal cautions/sentences while at school are observed for Black and Mixed-race boys. While this variation in attendance, exclusions and cautions/sentences during secondary school can partially explain socio-economic differences amongst initially high achieving pupils in their propensity to receive cautions/sentences as adults, a substantial proportion of the gap also remains unexplained.

These findings have many consistencies with results from previous studies comparing other outcomes across initially high achieving young people from advantaged and disadvantaged backgrounds. For instance, prior research has found a divergence in academic outcomes across early high achievers from different socio-economic groups between the ages of 11 and 14 (Crawford et al., 2017; Jerrim & Carvajal, 2024). This is consistent with when we find the gap in school absences and exclusions to increase. Jerrim (2024) also finds this to be a period when risky health behaviours amongst high achieving disadvantaged pupils start to emerge, including hospital admissions due to alcohol/drug abuse and self-harm. Our findings also serve to strengthen the only prior analysis we are aware of to consider criminal outcomes amongst initially high achieving disadvantaged pupils (Jerrim & Carvajal, 2024), which was based on relatively small scale, self-reported data. In particular, we have been able to demonstrate how important differences exist across initially high achieving disadvantaged young people of different genders and ethnicities, with cautions/sentences being particularly elevated for initially high achieving disadvantaged Black and Mixed race boys. This, in-turn, has helped to build understanding of the important intersectionality that exists when considering the outcomes of young people with high levels of prior achievement.

There are, however, some limitations to our work which we now note. First, our measure of crime is based on cautions/sentences, and thus represents quite an extreme outcome. It will not capture all criminal activity, such as those not caught or reported, nor where no further action was taken. Second, despite pooling data across three school cohorts, sample sizes in parts of our analysis (e.g. cautions/sentences occurring at specific ages for some gender and ethnic sub-groups) becomes relatively small. Third, our measure of disadvantage is based on the proxy measures available within England’s administrative data. While these measures are likely to pick up the most disadvantaged pupils relatively well via the inclusion of Free School Meals indicators across numerous years, it may only capture those from the most socio-economically advantaged homes with a degree of error (Jerrim, 2023). Fourth, relatedly, England’s administrative data – and the Key Stage 2 tests we use to define high achieving pupils – mainly captures those studying in state schools. Many of the roughly 7% of independent school pupils – who disproportionately come from more affluent families – are excluded from our analysis. Finally, all estimates presented throughout the paper are correlational, and do not attempt to capture cause and effect.

What, then, are the implications of our findings for policy? The emergence of socio-economic gaps in attendance and exclusions amongst high-achieving pupils during early secondary school highlights the need for targeted interventions and initiatives at this point in young people’s lives. These might focus on maintaining their engagement and behaviour at school to ensure they do not lose focus, fall into trouble and maximise their early academic potential. The intersectionality with gender and ethnicity also indicates that interventions and policies must be tailored to address the specific needs of different demographic groups. For instance, strategies to reduce school absences should particularly consider the challenges faced by teenagers of White and Mixed ethnicity, while efforts to decrease exclusions and criminal cautions should focus on Black and Mixed-race boys. Additionally, the unexplained part of the socio-economic gap in adult cautions/sentences points towards a need to go beyond school-based interventions, potentially involving community and family support services. It is thus likely that a mix of early, targeted, and intersectional initiatives are needed to ensure high-achieving pupils from disadvantaged backgrounds do not start falling into trouble at school and with the law during secondary school.

**References**

Blanden, J. & Machin, S. (2010). Intergenerational inequality in Early Years assessments. pp. 153-168 in Hansen, K.; Joshi, H. & Dex, S. Children of the 21st century (Volume 2): The first five years <https://doi.org/10.2307/j.ctt9qgrfx.17>

Blanden, J., Doepke, M., & Stuhler, J. (2022). Educational inequality (IZA Discussion Paper No. 15225). Institute of Labor Economics (IZA). <https://docs.iza.org/dp15225.pdf>

Burgess-Proctor, A. (2006). Intersections of Race, Class, Gender, and Crime: Future Directions for Feminist Criminology. *Feminist Criminology*, 1(1), 27-47. <https://doi.org/10.1177/1557085105282899>

Crawford, C., Macmillan, L., & Vignoles, A. (2017). When and why do initially high-achieving poor children fall behind? *Oxford Review of Education*, *43*(1), 88–108. <https://doi.org/10.1080/03054985.2016.1240672>

Cummings, P. (2009). The relative merits of risk ratios and odds ratios. *Archives of Pediatrics & Adolescent Medicine, 163*(5), 438–445. <https://doi.org/10.1001/archpediatrics.2009.31>

Diezmann, C. M., & Watters, J. J. (1997). Bright but bored: Optimising the environment for gifted children. *Australian Journal of Early Childhood, 22(2), 17-21.*

Feinstein, L. (2003). Inequality in the Early Cognitive Development of British Children in the 1970 Cohort, *Economica* 70: 73-97.

FFT Datalab. (2024). The secondary school outcomes of pupils excluded (or suspended) in primary school.<https://ffteducationdatalab.org.uk/2024/04/the-consequences-of-primary-school-exclusions/>

Graham, B., White, C., Edwards, A., Potter, S., & Street, C. (2019). *School exclusion: A literature review on the continued disproportionate exclusion of certain children*. Department for Education. <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/800028/Timpson_review_of_school_exclusion_literature_review.pdf>

Holt-White, E. & Cullinane, C. (2023). *Social Mobility: The Next Generation Lost potential at age 16*. London: Sutton Trust. <https://www.suttontrust.com/our-research/social-mobility-the-next-generation-lost-potential-at-age-16/#:~:text=This%20report%20looks%20at%20the,the%20end%20of%20primary%20school>

Jerrim, J., & Vignoles, A. (2013). Social mobility, regression to the mean and the cognitive development of high ability children from disadvantaged homes. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 176: 887–906.

Jerrim, J. (2023). Measuring parental income using administrative data. What is the best proxy available? *Research Papers in Education*, 1–25. <https://doi.org/10.1080/02671522.2023.2271930>

Jerrim, J. & Carvajal, M. (2024). What happens to bright 5-year-olds from poor backgrounds? Longitudinal evidence from the Millennium Cohort Study. <https://doi.org/10.35542/osf.io/64vg9>

Klein, M., & Sosu, E. (2024). School absences, academic achievement, and adolescents’ post-school destinations. *Oxford Review of Education, 1–18*. <https://doi.org/10.1080/03054985.2024.2308520>

Mackenbach, J. P., Valverde, J. R., Artnik, B., Bopp, M., Brønnum-Hansen, H., Deboosere, P., Kalediene, R., Kovács, K., Leinsalu, M., Martikainen, P., Menvielle, G., Regidor, E., Rychtaříková, J., Rodriguez-Sanz, M., Vineis, P., White, C., Wojtyniak, B., Hu, Y., & Nusselder, W. J. (2018). Trends in health inequalities in 27 European countries. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(25), 6440–6445. <https://doi.org/10.1073/pnas.1800028115>

Martínez-Monteagudo, Á., Martínez-Monteagudo, M. C., & Delgado, B. (2023). School bullying and cyberbullying in academically gifted students: A systematic review. *Aggression and Violent Behavior, 71*, 101842. <https://doi.org/10.1016/j.avb.2023.101842>

Mazerolle, L., Bennett, S., Antrobus, E. *et al.* Disrupting the Pathway from Truancy to Delinquency: A Randomized Field Trial Test of the Longitudinal Impact of a School Engagement Program. *J Quant Criminol* **35**, 663–689 (2019). <https://doi.org/10.1007/s10940-018-9395-8>

Ministry of Justice. (2022). Data First: An Introductory User Guide. <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1056236/data-first-user-guide-version-7.0.pdf>

OECD (2018), *A Broken Social Elevator? How to Promote Social Mobility*, OECD Publishing, Paris, <https://doi.org/10.1787/9789264301085-en>.

Schall, J., Wallace, T. L., & Chhuon, V. (2014). ‘Fitting in’ in high school: how adolescent belonging is influenced by locus of control beliefs. *International Journal of Adolescence and Youth*, *21*(4), 462–475. <https://doi.org/10.1080/02673843.2013.866148>

Social Mobility Commission. (2023). State of the Nation 2023: People and Places. <https://www.gov.uk/government/publications/state-of-the-nation-2023-people-and-places>

Sosu, E. M., Dare, S., Goodfellow, C., & Klein, M. (2021). Socioeconomic status and school absenteeism: A systematic review and narrative synthesis. *Review of Education, 9*(e3291). <https://doi.org/10.1002/rev3.3291>

Wainer, H., & Brown, L. M. (2006). Three Statistical Paradoxes in the Interpretation of Group Differences: Illustrated with Medical School Admission and Licensing Data. In C. R. Rao & S. Sinharay (Eds.), Handbook of Statistics (Vol. 26, pp. 893-918). Elsevier. ISSN 0169-7161. ISBN 9780444521033. [https://doi.org/10.1016/S0169-7161(06)26028-0](https://doi.org/10.1016/S0169-7161%2806%2926028-0).

**Washbrook, E.** & Lee, R. (2015). Social class differences in early cognitive development

1. Social Research Institute, University College London, 20 Bedford Way London, WC1H 0AL. E-mail: j.jerrim@ucl.ac.uk (John Jerrim). [↑](#footnote-ref-1)
2. This is where measurement error in the test used to divide pupils into different achievement groups leads to a statistical artifact, inflating the fall(rise) in initially high-achieving (low-achieving) pupils test scores over time. [↑](#footnote-ref-2)
3. The same threshold for attainment is used for both socio-economically advantaged and disadvantaged pupils (top 25% of the Key Stage 2 distribution nationally). [↑](#footnote-ref-3)
4. In England, the school year is divided into six half-terms, each of which are roughly six to seven weeks long. [↑](#footnote-ref-4)
5. Just 0.16% of secondary pupils are ever permanently excluded (FFT, 2024).When investigating permanent exclusions, we again create binary indicators for whether the child was permanently excluded at any point during the academic year. [↑](#footnote-ref-5)
6. Short time intervals – such as termly figures – would result in insufficient numbers of observations and would not meet statistical disclosure protocols. [↑](#footnote-ref-6)
7. For Cohort C, the latest age we can observe whether they were cautioned/sentenced is age 20. [↑](#footnote-ref-7)
8. The summer term of Year 11 is a notable exception, where the absence rate increases sharply to around 10%. This coincides with the timing of GCSE examinations in England and may be due to pupils deciding to revise at home or not returning to school once their examinations have finished. [↑](#footnote-ref-8)
9. As the binary outcomes investigated in this paper are rare, then the odds ratio will closely approximate the risk ratio (Cummings, 2009). We will thus on occasion interpret odds ratios as if they were risk ratios to facilitate communication of our results. [↑](#footnote-ref-9)