University access, choices and outcomes for high-achieving children from disadvantaged socioeconomic backgrounds.

John Jerrim[1]

Maria Palma Carvajal

Abstract

Young people from disadvantaged backgrounds with high levels of academic achievement at the end of primary school are a key group for enhancing social mobility. Yet many barriers stand in the way of this group converting their early potential into a top professional job, including gaining access to – and graduating from – university. This paper presents new evidence on this issue, providing novel insight into socio-economic differences in university entry and graduation amongst young people with high levels of achievement at the end of primary school. Using National Pupil Database (NPD) linked to Higher Education Statistics Authority (HESA) records from England, we find substantial socio-economic differences in the extent that high achievers at the end of primary school go on to attend a Russell Group or Oxbridge university. There are some indicative signs, however, that these gaps may have narrowed over time. Important differences across genders and ethnic groups are observed. Most gaps in university outcomes can be explained by differences in the extent that early high achievers from different socio-economic backgrounds convert their initial potential into higher school grades.

Key Words: High achieving; disadvantage; Oxbridge; Russell Group; higher education.

Acknowledgments: The Nuffield Foundation is an independent charitable trust with a mission to advance social well-being. It funds and undertakes rigorous research, encourages innovation and supports the use of sound evidence to inform social and economic policy, and improve people's lives. 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. This project has been funded by the Nuffield Foundation, but the views expressed are those of the authors and not necessarily the Foundation. Find out more at: nuffieldfoundation.org.

[1] Social Research Institute, University College London, 20 Bedford Way London, WC1H 0AL. E-mail: j.jerrim@ucl.ac.uk (John Jerrim). +44 7590 761 755

1. Introduction

It is widely known that individuals born into socio-economically disadvantaged homes experience worse outcomes in later life than their more socio-economically advantaged peers. There is widespread concern that this is being driven by inequality of opportunities, including the quantity and quality of education different socio-economic groups experience (Breen & Karlson, 2014). This then has important consequences for the economy and society, with previous research linking lower levels of equality of opportunity to slower rates of economic growth (Neidhöfer et al., 2023). It is thus little wonder why this issue has become a preoccupation of public policymakers, with sustained attempts to build more socially fluid societies across the Western world (OECD, 2018).

One group of particular interest to those looking to enhance social mobility are young people from disadvantaged backgrounds who show signs of early academic potential. According to the theoretical model of lifetime skill formation (Cunha et al., 2006), once a young person has fallen behind academically, it is difficult (and expensive) for them to then catch up. In contrast, children from disadvantaged backgrounds with high levels of academic achievement have managed to overcome socio-economic adversity during the early years. They have thus laid the foundations needed to excel at school and become upwardly mobile. In other words, this group has the best chance of smashing through the glass ceiling and increasing socio-economic diversity amongst high-status, professional jobs.

Yet many obstacles stand in their way of fulfilling their early academic potential. They must, for instance, go on to achieve strong end-of-school grades. They may then also need to secure a place at a high-status university, given how these institutions act as a gateway into many well-paid, professional careers (Sutton Trust, 2019). It is therefore perhaps surprising that relatively little research has focused on higher education access and outcomes of this specific group; for instance, to what extent do young people from disadvantaged backgrounds with high levels of academic achievement at the end of primary school go on to obtain an undergraduate (and then postgraduate) degree?

This paper presents such novel evidence on the higher education access, choices and outcomes for socio-economically disadvantaged young people who were amongst the top academic achievers nationally at the end of primary school. The analysis includes the extent that this group manages to obtain a place at Oxford or Cambridge University, whether they continue into postgraduate study and – when attending university – whether they decide to move out of the family home. The paper hence provides the most comprehensive analysis of the higher

education outcomes for initially high-achieving young people from disadvantaged socio-economic backgrounds to date.

Theoretical framing

The theoretical underpinning of the paper draws upon sociological research into primary and secondary effects (Jackson et al., 2007). This postulates that socio-economic inequalities in the final level of education a young person attains depends on two broad factors.

The first of these are primary effects. These reflect differences in attainment at school that determine the educational options a young person has available. In other words, to what extent do different groups obtain the grades that make – for instance – attending Oxford or Cambridge (henceforth Oxbridge) a feasible choice? In our application, such primary effects can be conceived as the extent that initially high-achieving young people from disadvantaged socio-economic backgrounds go on to obtain the requisite academic credentials that make university entry (including into an academically selective institution) a realistic possibility.

Secondary effects are, on the other hand, factors over and above prior academic attainment that are related to high achieving disadvantaged young people's educational choices. Take entry into a Russell Group or Oxbridge university, for example. Some initially high-achieving young people from poor backgrounds will get the grades needed to enter such an institution - but choose to study elsewhere. This may be due to cultural reasons (e.g. concerns whether they will fit in, self-belief), finance (e.g. being able to afford the costs associated with studying at such an institution), access to information/advice or familial/social factors meaning they are unable or unwilling to migrate. Within this literature, such secondary effects are typically captured as the remaining association between socio-economic background and university entry/outcomes once prior academic attainment has been controlled.

Prior literature

Several studies – both in England and internationally – have studied socio-economic differences in access to higher education, and explored the extent that this can be explained by differences in prior achievement. For instance, Bukodi et al. (2021) used sample survey data to study five educational transitions in England, including access to and graduation from university. While they found that most of the link between social origin and university access and outcomes can be explained by differences in achievement while at school, they also argue that secondary effects – the association that remains after prior achievement has been controlled – is "still of real consequence" (p. 645). This is consistent with the findings of Sullivan et al.

(2014), who reported non-trivial associations in England between parental education, private schooling and higher education destination, even after both childhood cognition and school qualifications had been controlled. Likewise, Campbell et al. (2019) report that socio-economically disadvantaged pupils are more likely to "undermatch" when selecting universities than their more advantaged peers. Findings from these studies are in contrast with the earlier work of Chowdry et al. (2013) who found that – conditional on prior achievement – there was virtually no association between young people's socio-economic background and their entry patterns into university. Jerrim et al (2015) compared access to high-status universities in England to Australia and the United States. They found that access to elite universities was of similar magnitude across the three countries, with non-trivial associations with family background remaining in each country after prior achievement had been controlled. Jackson (2013a) investigated trends over time in access to higher education in England, finding that the relationship with family background remained broadly stable over a 30-year period (individuals born between 1958 and 1986).

A distinct literature has emerged into the educational progress and outcomes of young people from disadvantaged socio-economic backgrounds with strong levels of early development and/or cognitive skills. This work has, however, largely focused on the educational progress this group makes during their time at school. Feinstein (2003) received a lot of public policy attention, with his findings interpreted as showing children form disadvantaged backgrounds with strong levels of early development falling behind their more advantaged peers with lower levels of early development. Jerrim and Vignoles (2013) argued however that this finding was likely a statistical artifact due to regression to the mean. Recent additions have been made to this literature by Jerrim & Carvajal (2024), who tracked the outcomes of initially high-achieving children at age 5 through to age 17. They illustrate how ages 11 to 14 seem to be a key period, when this group starts to fall behind their more affluent peers at school, as well as being more likely to experience behavioural and socio-emotional problems. This built upon the work of Holt-White and Cullinane (2023) who found that 62% of high-achieving children at the end of primary school from the most advantaged socio-economic backgrounds go on to obtain top grades in England's GCSE examinations (taken at age 16), compared to only around 40% of their equally high-achieving disadvantaged peers. Finally, the work of Crawford et al. (2017) is of particular relevance, given it is one of the few existing studies to consider the extent that disadvantaged children who were high academic achievers during primary school go on to attend university. For one school cohort, they illustrate how this group are less likely

to attend research-intensive universities than their more socio-economically advantaged peers, though this inequality largely mirrors differences in the achievement of these respective groups during the latter stages of secondary school.

Research questions

The aforementioned research has made valuable contributions to our understanding of socio-economic inequalities in access to higher education and in understanding educational achievement trajectories amongst initially high-achieving children from different family backgrounds during their time at school. Yet there are also some key gaps in this existing evidence base. For instance, no previous research has investigated the extent that disadvantaged children with high levels of achievement at the end of primary school go on to study at one of England's high-status universities (e.g. Oxbridge or a member of the Russell Group). There has been little consideration of how such entry patterns may vary across disadvantaged high-achieving children of different genders and ethnicities, despite several authors noting the importance of considering intersectionality between socio-economic disadvantage and other background characteristics (Codiroli-Mcmaster & Cook, 2019; Campbell et al., 2019). Several potential outcomes of interest have also not been explored, such as the propensity for high achieving disadvantaged young people to move away from home during their undergraduate degree, whether they obtain good grades while at university (e.g. obtain a 2:1 or 1st class classification) and entry into postgraduate study. We also know little about how the higher education outcomes of this group have changed over time; for instance, are high-achieving children from disadvantaged backgrounds entering Russell Group and Oxbridge universities at a greater rate now than previously?

We attempt to fill these gaps in the literature by addressing four research questions. Our analysis will begin by exploring the extent that high achieving children at the end of primary school convert their early potential into better grades in high-stakes national examinations taken in the latter stages of secondary school, before the transition into higher education takes place. This will provide insight into the extent that initially high achieving young people from different socio-economic backgrounds go on to "credentialise" their early skills which – in turn – has implications for the magnitude of primary effects (i.e. the extent that differences in university entrance rates can be traced back to differences in school grades). Research question 1 is therefore:

- Research question 1. To what extent do young people from disadvantaged socio-economic backgrounds with high levels of achievement at the end of primary school achieve "good" grades at the end of secondary school?

Next, we turn to differences in university entrance rates across initially high-achieving young people from different socio-economic groups. This will include a wide array of outcomes – including entry into Britain's most prestigious institutions – while also considering differences across genders and ethnicities, along with changes over time. We will also explore the extent that any differences across initially high-achieving children from different socio-economic backgrounds can be explained by differences in the grades they achieve in high-stakes examinations at school. Our second research question is thus:

- Research question 2. What proportion of initially high-achieving children from disadvantaged backgrounds go on to study at Oxford, Cambridge or another Russell Group university? How does this compare to their more advantaged peers, and to what extent is this due to differences in prior academic achievement?

Third, our analysis considers whether high-achieving children from disadvantaged backgrounds are more likely to choose to move out of the family home while an undergraduate than their equally able but more socio-economically advantaged peers. We are interested in this outcome for a variety of reasons. First, previous research has discussed the "consumption value" of university (Gong et al., 2021), and that young people are more likely to build new friendships, participate in extra-curricula activities and make use of the broader opportunities' universities offer if they move out of home (Holdsworth, 2006; Davey, 2025). This outcome will hence provide some insight into differences across socio-economic groups in their university experiences. Second, moving out of the family home helps young people to build broader skills such as independence (Mulder & Clark, 2002; Kassenboehmer et al., 2018) which are likely to be valued in the workplace. Third, prior research has suggested that young people that move out of home to go to university are more willing to migrate again in the future (Swinney, 2016). This may hence be linked to the willingness of young people to move away to pursue labour market opportunities once they graduate. Moreover, if young people want to commute from home, their choices of university will be more constrained and they may be more likely to 'undermatch', limiting future prospects (Campbell et al., 2019). Finally, there are good reasons to expect there to be differences across high-achieving young people from different family backgrounds, who have less capacity to afford the financial costs of living away from home. Our third research question is therefore:

- Research question 3. Are high-achieving children from disadvantaged socio-economic backgrounds more likely to live at home as an undergraduate than their more socio-economically advantaged peers?

Finally, it is not only university entry that matters, but also whether young people go on to complete their degree and the classification they obtain. Indeed, dropping out from university – or graduating with poor grades – is likely to reduce the likelihood of initially high-achieving disadvantaged young people entering a well-paid, professional job. At the same time, no previous research has investigated the extent that initially high achieving children from poor backgrounds go on to obtain at least a 2:1 degree, or their propensity to continue into postgraduate study. Our analysis thus concludes by asking:

- Research question 4. How likely are high-achieving children from disadvantaged backgrounds to complete postgraduate study? How does this compare to their more socio-economically advantaged peers?

2. Data

The data we use are drawn from England's National Pupil Database (NPD) linked to Higher Education Statistics Authority (HESA) records. These are administrative data capturing school linked to university records for all state school pupils in England. We have access to data for the following school cohorts:

- Cohort A. All children born September 1990-August 1991. These individuals were in Year 6 in 2001/2002, Year 11 in 2006/2007, entered university in 2009/10 and completed their undergraduate degree in 2012/13.
- Cohort B. All children born September 1994-August 1995. These individuals were in Year 6 in 2005/2006, Year 11 in 2010/2011, entered university in 2013/14 and completed their undergraduate degree in 2016/17.
- Cohort C. All children born September 2000-August 2001. These individuals were in Year 6 in 2011/2012, Year 11 in 2016/2017, entered university in 2019/20 and completed their undergraduate degree in 2021/22. Note that university entry patterns and experiences would to some extent have been impacted by the COVID-19 pandemic.

For each cohort we can observe young people's scores in national examinations throughout their time at school, indicators of their family background, whether they progressed into university (including the institution attended), degree outcomes and – for Cohorts A and B – whether they completed postgraduate study.

Our primary group of interest is young people from disadvantaged socio-economic backgrounds with high levels of early achievement. We operationalise high early achievement as obtaining a score in the top quartile of England's Key Stage 2 reading and mathematics tests. These are national examinations taken by 10/11-year-olds during their final year at primary school. Specifically, for each cohort, we standardise Key Stage 2 mathematics and reading scores to mean zero and standard deviation one. The average standardised score is then taken across these two subjects and the cohort split into four equally sized groups. Those achieving a score in the top quartile for their cohort are defined as "high early achievers". By using Key Stage 2 scores, this measure has the advantage of being based upon lengthy[2] national examinations that are externally set and marked, while also continuing to capture children's academic skills at a relatively young age (before they enter secondary school). While using test scores at a younger age may have some advantages (e.g. by being able to capture disadvantaged high-achieving children who fall behind their more advantaged peers during primary school) they would also come with significant disadvantages (e.g. tests of children at younger ages tend to be shorter and more susceptible to measurement errors). Moreover, previous research has illustrated how the early stages of secondary school is a key period for initially high achieving young people from disadvantaged backgrounds (Crawford et al., 2017).

The other key measure used to define our group of interest is young people's socio-economic background. Following previous research using England's administrative datasets (Jerrim, 2023) we combine information across various indicators to create a socio-economic background scale. A one-parameter Item Response Model is estimated drawing upon the following pieces of information:

- 12 dummy variables indicating – for each year they were at school - whether the young person was eligible for Free School Meals (FSM).
- Income Deprivation Affecting Children Index (IDACI) decile – this captures the number of families in the local area that have low levels of income.
- A binary indicator of parental education (whether either parent holds a university level qualification or not). Note this information is only available for those young people that progressed into university.

[2] Children sit tests totalling around three hours of time over a four day period.

- Parental social class, based upon The National Statistics - Socio-Economic Classification (NS-SEC) schema. This information is only available for those that entered university.

Expected A Posteriori (EAP) estimates are then created for each individual, capturing their position along the latent socio-economic background scale. Descriptive information about this scale is provided in Table 1 and Appendix B. This scale is then divided into quartiles within each cohort, with the least advantaged 25% used to define young people coming from socio-economically disadvantaged backgrounds.

Table 1. Descriptive statistics for initially high achieving children from socio-economically advantaged and disadvantaged backgrounds

	Cohort A		Cohort B		Cohort C	
	Disadvantaged	**Advantaged**	**Disadvantaged**	**Advantaged**	**Disadvantaged**	**Advantaged**
Ethnicity						
White	75%	92%	65%	91%	55%	86%
Asian	10%	2%	15%	2%	16%	3%
Black	5%	<1%	9%	<1%	11%	<1%
Mixed	4%	2%	6%	2%	6%	3%
Other	5%	34%	6%	5%	12%	8%
Gender						
Male	49%	49%	47%	49%	45%	48%
EAL						
Yes	2%	14%	3%	25%	3%	31%
Percent of time FSM eligible						
% time at school	<1%	30%	<1%	45%	<1%	52%
IDACI score						
Mean	0.43	0.05	0.43	0.05	0.42	0.05
Percent of cohort	**3%**	**8%**	**2%**	**8%**	**3%**	**8%**
N	**16,180**	**47,680**	**12,935**	**48,515**	**13,870**	**42,820**

Notes: Percent of cohort refers to percentage of all children in the cohort who are classified as a high achiever from an advantaged/disadvantaged background.

Our analysis begins by focusing on a set of pre-university outcomes, capturing attainment in the latter stages of secondary education. The measure we focus on are:

- Average GCSE total points score. At age 16, young people in England sit high-stakes national examinations across around nine subjects. They are then awarded a grade for each of these subjects, which can be combined and converted into a total points score. We standardise these scores to mean zero and standard deviation one within each cohort.
- Key Stage 5 total point scores. At age 18, young people sit a further set of national examinations in usually three chosen subjects. They receive a grade in each, which can be combined, and a total points score derived. We standardise these scores within each cohort, using this as a measure of their academic achievement at age 18.

We also present results for two additional school outcomes (whether they continued in education and completed exams through to age 18 and whether they obtained at least three B grades at A-Level) in the Appendix (see Appendix C for further details).

When answering Research Question 2 we focus on the following:

- Studying for an undergraduate degree. This variable is coded one if the young person has any HESA record of studying for an undergraduate degree and zero otherwise.
- Studying an undergraduate degree at a Russell Group university. The Russell Group is a self-selecting set of 24 universities across the United Kingdom. These tend to be the most research-intensive institutions that are also generally the most academically selective. Previous research has focused on entry into such institutions (e.g. Hemsley-Brown, 2015). We thus derive a binary variable, coded one if the young person entered a Russell Group university, and zero otherwise.
- Studying an undergraduate degree at Oxbridge. Oxford and Cambridge are the United Kingdom's most well-known universities and receive much public policy interest. They are amongst the most competitive to enter and have long been associated with entry into the most prestigious jobs (e.g. since the second world war, only one British Prime Minster has studied at a university other than Oxford). We hence focus on access to these two institutions, deriving a variable coded one if the young person studied as an undergraduate at either Oxford or Cambridge, and zero otherwise.

We create two versions of the aforementioned variables. The first captures whether the young person was ever recorded to have studied for an undergraduate degree using all timepoints available. The second restricts this to whether they were recorded as studying for an undergraduate degree by age 21 – the latest age we can observe across all three cohorts. Hence an advantage of the second measure is that it has a greater degree of cross-cohort comparability.

Our third research question focuses on whether the young person chose to move out of the family home during their undergraduate studies. This information was recorded each year the young person was enrolled in higher education. We draw upon this information to derive a variable coded as one if the young person was recorded as living with their parents for most of their time as an undergraduate student, and zero otherwise[3].

[3] Those coded as zero will include young people living in university halls, private sector halls and other rented accommodation. This is captured in the HESA variable TTACCOM.

Finally, we consider a set of undergraduate outcomes and entry into postgraduate study. These are operationalised as follows:

- Degree completion. A binary indicator coded as one if the young person was recorded as ever having obtained an undergraduate degree.
- Obtaining a 2:1 or 1st class classification. In the British higher education system, students are awarded a degree classification. We focus on whether young people achieved either a 1st or 2:1 degree. These are the highest two classifications which are often a prerequisite for graduate employers.
- Complete an MSc. A binary variable coded as one if the young person was ever recorded studying for a master's level qualification.
- Enter a PhD. A binary variable coded as one if the young person was ever recorded studying for a PhD.

3. Methodology

Research question 1. School outcomes

Our analysis begins by considering a set of school outcomes prior to university entry, providing insight into the extent that disadvantaged children with high levels of achievement at the end of primary school convert their early academic potential into good school grades. The first of these are GCSEs taken by young people at age 16. We begin by restricting the sample to young people with Key Stage 2 scores in the top quartile. Raw, unconditional average GCSE point scores are then presented across socio-economic groups. A set of regression models are then estimated of the form:

$$O_i = \alpha + \beta.SES_i + \gamma.D_i + \delta.Prior_i + \varepsilon_i \quad (1)$$

Where:

O_i = GCSE total points score.

SES_i = A vector of dummy variables capturing quartiles of the socio-economic background scale (reference group = bottom quartile).

D_i = A vector of demographic background controls, including gender and ethnicity.

$Prior_i$ = A vector of background controls for prior achievement. This includes Key Stage 1 (age 6/7) mathematics and reading levels and Key Stage 2 scores[4].

ε_i = A random error term.

i = Young person i.

The parameter of interest is β. This captures the difference in average GCSE points across high achieving young people from different socio-economic backgrounds. As we standardise the outcome measure, estimates can be interpreted in terms of an effect size. When estimating this model separately by cohort, the difference in the β parameter will reveal whether socio-economic differences across initially high-achieving young people has changed over time.

We then move on to explore – amongst disadvantaged young people within the top Key Stage 2 test quartile – differences in GCSE outcomes across different combinations of gender and ethnicity. This is based upon the following regression model:

$$O_i = \alpha + \beta.SES\_By\_Gender\_Eth_i + \delta.Prior_i + \varepsilon_i \quad (2)$$

Where:

$SES\_By\_Gender\_Eth_i$ = A vector of dummy variables capturing combinations of gender and ethnicity (reference group = high achieving, disadvantaged White boys).

With all other variables as defined under equation (1) above.

The β parameter from (2) captures the difference in GCSE total points amongst high achieving disadvantaged young people of different genders and ethnicities, relative to high achieving disadvantaged White boys. As the sample for several gender-ethnic combinations is relatively small, these results will only be presented after pooling the data across the three cohorts.

We then turn to educational outcomes at age 18 in the form of Key Stage 5 total points scores. This employs models similar to those presented in equations (1) and (2), but with the sample further restricted to young people with this information available – i.e. continued to take educational qualifications through to age 18 - and with GCSE point scores included as a control. These results will thus capture socio-economic differences across initially high

[4] The intuition behind including this control is that there remains some variable in the academic abilities of young people, even amongst those with Key Stage 2 scores in the top quartile. Including controls for Key Stage 1 and Key Stage 2 scores will to some extent control for this variability in academic abilities amongst the "high initial achievement" group.

achieving children in attaining the grades they need to attend an academically selective university.

Research question 2. University entry

Unconditional percentages are first presented by cohort for each of our university entry measures. The sample is then restricted to only those young people with high Key Stage 2 test scores (i.e. in the top quartile) who continued their education through to age 18. Logistic regressions are then estimated of the form:

$$\log(\frac{O_i}{1-O_i}) == \alpha + \beta.SES_i + \gamma.D_i + \delta.Prior_i \quad (3)$$

Where:

O_i = A binary variable coded 1 if the young person entered university and 0 if they did not.

SES_i = A vector of dummy variables capturing quartiles of the socio-economic background scale (reference group = bottom quartile).

D_i = A vector of demographic background controls, including gender and ethnicity.

$Prior_i$ = A vector of background controls for prior achievement, including GCSE total points, Key Stage 5 total points and A-Level grades.

This model is estimated separately for each of our university entry outcomes (entered any university, entered a Russell Group university, entered Oxbridge). The β parameters will be presented in terms of odds ratios, and thus reflect the difference in the odds of going to university across initially high achieving young people from different socio-economic backgrounds. As under research question 1, differences across disadvantaged high achievers of different genders and ethnicities will also be explored.

Research question 3. Moving out of the family home.

Our third research question addresses the issue of whether the young person chooses to move out of the family home. We begin by restricting the sample to only those young people with Key Stage 2 scores in the top quartile and that entered university. Three specifications of the following logistic regression model are then estimated:

$$\log(\frac{O_{ij}}{1-O_{ij}}) == \alpha + \beta.SES_{ij} + \gamma.D_{ij} + \delta.Prior_{ij} + u_j \quad (4)$$

Where:

O_{ij} = A binary variable capturing whether the young person chose to live with their parents when an undergraduate (coded 1) or away from home (coded 0).

u_j = University fixed effects.

With all remaining variables as defined above.

In our first specifications we do not include any controls for demographic characteristics or prior achievement. This hence provides the unconditional difference in the odds of living at home as an undergraduate across high achieving young people from different socio-economic backgrounds. The second specification then adds controls for demographic characteristics and prior achievement. These estimates will reveal the extent that the differences in the unconditional model can be explained by differences in grades achieved in GCSEs and A-Levels. Finally, in our last specification, university fixed effects are added to the model. These results will thus illustrate whether any of the remaining difference across socio-economic groups can be explained by conditioning on the precise university that the young person attends. We then go on to explore differences across high achieving disadvantaged young people of different genders and ethnicities.

Research question 4. University completion, degree classification and postgraduate study

The final research question addresses socio-economic differences in completing university, obtaining at least a 2:1 degree, and continuation into postgraduate study. When doing so, the same broad modelling strategy is followed as outlined under research questions 1-3 above. For the analysis of postgraduate outcomes, the sample is restricted to young people with Key Stage 2 scores in the top quartile that went on to complete an undergraduate degree, with rich controls included for prior achievement (GCSE and A-Level grades, whether they attended a Russell Group or Oxbridge university and degree classification). The results thus reveal whether there are socio-economic differences across initially high achieving young people in accessing postgraduate study, amongst those who obtained similar grades at school and during their undergraduate studies.

Kelley's Paradox / Regression to the mean

As noted by Jerrim and Vignoles (2013), one of the empirical challenges with studying the outcomes of initially high achieving young people is the potential for estimates to be affected by Kelley's Paradox (a form of regression to the mean). Put succinctly, whenever the test used

to classify young people into different achievement groups is measured with error, one will tend to overestimate the magnitude of socio-economic differences in their future outcomes. We discuss this issue in further detail in Appendix A, and follow the approach suggested by Jerrim and Carvajal (2024) to test the sensitivity of our results. We find that while estimated differences across initially high-achieving children from different socio-economic backgrounds are to some extent reduced when making different assumptions of Key Stage 2 test reliability, the general pattern of the results described in the following section continues to hold.

4. Results

Research question 1. School outcomes.

Table 2 begins by presenting the difference in two school outcomes (GCSE and Key Stage 5 total point scores) for initially high achieving young people from the most and least advantaged socio-economic backgrounds. This focuses on results from our regression models, with estimates presented as effect sizes (standard deviation differences).

Table 2. Regression estimates of differences in school outcomes across high achieving young people from socio-economically advantaged and disadvantaged backgrounds

(a) GCSE total points (standardised)

	Q4 SES			
	Effect size	**Lower CI**	**Upper CI**	**N**
Cohort_A	0.29*	0.28	0.30	133120
Cohort_B	0.16*	0.15	0.17	132150
Cohort_C	0.42*	0.41	0.44	118790

(b) Key Stage 5 total points (standardised)

	Q4 SES			
	Effect size	**Lower CI**	**Upper CI**	**N**
Cohort_A	0.32*	0.30	0.34	107995
Cohort_B	0.33*	0.31	0.35	114245
Cohort_C	0.43*	0.41	0.45	98475

Notes: Figures refer to differences in total point scores between young people from the most and least advantaged socio-economic backgrounds in terms of effect sizes. Lower and Upper CI provides the 95% confidence interval. * indicates a statistically significant difference at the 5% level.

There are non-trivial differences across each of the outcomes. Take GCSE total point scores, for instance. High-achieving young people from the most advantaged socio-economic backgrounds obtain scores around 0.3 standard deviations higher than their equally high achieving but socio-economically disadvantaged peers. Although some of this difference can be explained by Kelley's paradox as discussed in Appendix A (e.g. the difference shrinks to around 0.10 to 0.15 standard deviations for GCSE total points under realistic assumptions of the degree of measurement error in the Key Stage 2 tests), it is reasonable to conclude that – during secondary school – initially high achieving children to some extent fall behind their more advantaged peers academically. With respect to Key Stage 5 scores, even after controlling for performance in GCSE examinations, we find that initially high achieving young people from the most advantaged backgrounds achieve scores around 0.3 to 0.4 standard deviations higher than their equally high-achieving, disadvantaged peers.

With respect to changes over time, the results are somewhat inconclusive. For both outcomes, the estimates for Cohort C tend to be slightly larger than for Cohorts A and B. There were however substantial changes to England's curriculum and assessment regime over this period, many of which had implications for high-achieving students (e.g. the GCSE and A-Level grading structure was changed, in part to better distinguish the most able students). Thus, based on these results, we do not believe that there is strong evidence that the school outcomes of initially high achieving young people from advantaged and disadvantaged backgrounds have either narrowed or widened over time.

In Table 3 we consider differences across initially high achieving disadvantaged young people from different gender and ethnic backgrounds. All figures are reported in comparison to initially high achieving disadvantaged White boys as the reference group.

Table 3. Logistic regression estimates for school outcomes of high achieving young people from disadvantaged backgrounds. Differences by gender and ethnicity.

(a) GCSE total points (standardised)

	Effect size	Lower CI	Upper CI
White girl	0.13*	0.11	0.14
Asian boy	0.33*	0.30	0.36
Asian girl	0.53*	0.50	0.56
Black boy	0.16*	0.12	0.20
Black girl	0.32*	0.28	0.35
Mixed race boy	-0.01	-0.05	0.04
Mixed race girl	0.16*	0.12	0.20
Other ethnicity boy	0.28*	0.24	0.33
Other ethnicity girl	0.41*	0.37	0.45
High achieving, above average SES	0.40*	0.38	0.41
Q3 SES, high achieving	0.36*	0.35	0.37
N		**390855**	

(b) Key Stage 5 total point score

	Effect size	Lower CI	Upper CI
White girl	0.04*	0.01	0.07
Asian boy	0.18*	0.14	0.22
Asian girl	0.18*	0.14	0.22
Black boy	0.02	-0.03	0.08
Black girl	0.15*	0.10	0.20
Mixed race boy	0.05	-0.02	0.12
Mixed race girl	0.08*	0.02	0.14
Other ethnicity boy	0.25*	0.18	0.32
Other ethnicity girl	0.22*	0.16	0.28
High achieving, above average SES	0.38*	0.36	0.40
Q3 SES, high achieving	0.23*	0.21	0.25
N		**326795**	

Notes: Panel (a) restricts sample to young people with Key Stage 2 scores in the top quartile, with estimates capturing differences in total GCSE points in comparison to White boys. The sample in panel (b) is restricted to young people with Key Stage 2 scores in the top quartile and who remained in school post-16. Figures are reported as effect sizes. * indicates a statistically significant difference at the 5% level.

These results provide clear evidence that it is initially high achieving disadvantaged White boys who fall behind other groups. Take the results for equally high achieving, disadvantaged Asian boys, for instance. They obtain GCSE total points 0.33 standard deviations higher than their White peers, and then a further 0.18 standard deviations higher at Key Stage 5. Indeed, at least at GCSE, initially high achieving disadvantaged Asian boys, Asian girls and Black girls broadly keep pace with initially high achieving children from the most affluent homes. Table 3 therefore suggests that – during their time at secondary school – it is the early potential of

academically able disadvantaged young people of White ethnicity (particularly boys) that is particularly likely to go unfulfilled.

Research question 2. University entry.

Our second research question turns to differences in university entry rates. In the main text below we focus on entry into university by age 21, with alternatives based on any record of university entry at any age provided in Appendix E. Table 4 begins by presenting the raw, unconditional percentage of initially high achieving young people that go on to attend university by socio-economic background. Results are presented separately for entry into any university to study a bachelor's degree (panel a), entry into a Russell Group university (panel b) and entry into Oxbridge (panel c).

Table 4. The percentage of high-achieving young people that enter university by age 21. Unconditional estimates by socio-economic background for each school cohort.

(a) Any university

	Cohort_A	Cohort_B	Cohort_C
1. High achieving, high SES	57%	60%	75%
2. High achieving, Q3 SES	62%	64%	77%
3. High achieving, Q2 SES	53%	52%	76%
4. High achieving, low SES	32%	40%	57%
5. Missing data	18%	21%	42%

(b) Russell Group

	Cohort_A	Cohort_B	Cohort_C
1. High achieving, high SES	23%	28%	37%
2. High achieving, Q3 SES	21%	25%	33%
3. High achieving, Q2 SES	15%	16%	27%
4. High achieving, low SES	7%	11%	17%
5. Missing data	4%	5%	13%

(c) Oxbridge

	Cohort_A	Cohort_B	Cohort_C
1. High achieving, high SES	2%	3%	3%
2. High achieving, Q3 SES	2%	2%	3%
3. High achieving, Q2 SES	1%	1%	2%
4. High achieving, low SES	<1%	<1%	1%
5. Missing data	<1%	<1%	1%

Notes: Figures refer to the percent of the group that attend university. For instance, in Cohort C, 1% of young people from disadvantaged socio-economic backgrounds with Key Stage 2

test scores in the top quartile go on to attend Oxford or Cambridge university. Figures rounded to nearest whole percent to comply with statistical disclosure control.

There are three key points to note. First, more initially high achieving young people are going to university over time across all socio-economic backgrounds, including to higher status institutions. For instance, only 7% of initially high achieving young people from disadvantaged backgrounds in Cohort A entered a Russell Group university, compared to 11% in Cohort B and 17% in Cohort C. Second, there are nevertheless substantial differences in university attendance amongst initially high achieving students. For instance, within Cohort C, those from the most advantaged backgrounds were still twice as likely as those from the least advantaged backgrounds to enter a Russell Group university (17% versus 37%). Finally, there is some suggestion that the relative difference across initial high achievers from different socio-economic backgrounds may have narrowed over time. Using entry into Oxbridge as an example, in cohort A, early high achievers from the most affluent backgrounds were approximately seven times more likely to attend Oxford or Cambridge than their peers with equally high Key Stage 2 scores from the most disadvantaged homes (entry rates of 2% versus 0%). This difference has shrunk, however, to around a five-fold difference in Cohort and a three-fold difference for Cohort C. While the results for Cohort C should be interpreted with a degree of caution – given how the time they were starting university coincided with the onset of the COVID-19 pandemic – these unconditional estimates nevertheless provide some indication of increasing entry rates into high-status universities amongst initially high achieving young people from the poorest backgrounds. This result is also consistent with increasing levels of widening participation initiatives by universities in recent years.

Table 5 builds on these results by presenting estimates from our logistic regression models. Recall that these models are restricted to only those young people with educational achievement measures through to age 18 and controls included for GCSE and A-Level performance. Odds ratios over (under) one illustrate where – conditional on these factors – high-achieving young people from the most advantaged backgrounds are more (less) likely to enter university than their equally able, socio-economically disadvantaged peers.

Table 5. Logistic regression estimates of high-achieving young people from different socio-economic backgrounds entering university by age 21.

(a) Any university

	Q4 SES			N
	OR	Lower CI	Upper CI	
Cohort_A	1.26*	1.20	1.32	113645
Cohort_B	1.10*	1.05	1.15	117125
Cohort_C	0.67*	0.63	0.72	99420

(b) Russell Group

	Q4 SES			N
	OR	Lower CI	Upper CI	
Cohort_A	1.39*	1.28	1.51	113645
Cohort_B	1.22*	1.13	1.31	117125
Cohort_C	0.93	0.87	1.00	99420

(c) Oxbridge

	Q4 SES			N
	OR	Lower CI	Upper CI	
Cohort_A	1.56*	1.12	2.15	113645
Cohort_B	1.39*	1.06	1.83	117125
Cohort_C	0.59*	0.48	0.73	99420

Notes: Figures refer to odds ratios capturing the difference in the outcome between the most and least disadvantaged group. Sample restricted to young people with Key Stage 2 scores in the top quartile and with academic achievement measures available through to age 18. Lower and Upper CI provides the 95% confidence interval. Controls included for gender, Key Stage 2, GCSE and Key Stage 5 point scores and best three A-Level grades achieved. * indicates a statistically significant difference at the 5% level.

In Cohorts A and B, there is a relatively small difference in favour of high-achieving young people from the most advantaged backgrounds. For instance, in Cohort A, the odds of them attending a Russell Group university by age 21 was around 40% higher than the most disadvantaged group (with around a 50% difference in the odds of entering Oxbridge). Thus, for these two cohorts, while differences in school achievement could explain most of the gap in high status university attendance rates across initially high achieving children from different socio-economic backgrounds, a non-trivial difference remained. However, even between cohorts A and B, the magnitude of this residual association (or "secondary effect") declined

(e.g. with respect to entering a Russell Group university, the odds ratio declined from 1.39 in Cohort A to 1.22 in Cohort B).

The results for Cohort C are somewhat different. The odds ratio now sits below one suggesting that – conditional on prior achievement – initially high achieving young people from disadvantaged backgrounds are *more* likely to enter university than their more affluent peers. While the magnitude of the difference for entry into the Russell Group is small (odds ratio = 0.93) and sits on the boundary of statistical significance, the estimate for entry into Oxbridge is more sizeable (odds ratio = 0.59). We again advise readers to exercise a degree of caution when interpreting this result, given how – to some extent – the pattern of university entry for Cohort C has been impacted by the COVID-19 pandemic. Nevertheless, Table 5 provides some signs that "secondary effects" – i.e. factors other than achievement at school that lead to differences in university entry rates across socio-economic groups – may have declined in England over time.

Table 6 concludes our analysis of research question 2 by considering – amongst high achievers from disadvantaged socio-economic backgrounds – differences in Russell Group entry rates across genders and ethnicities. Note that we focus on the Russell Group here as the sample size for Oxbridge as an outcome becomes small for each group, leading to very wide confidence intervals (these results – along with those for entry into any university – are however provided in Appendix D for reference). These estimates are again based on logistic regression models controlling for rich measures of prior achievement at school, with high-achieving White boys acting as the reference group.

Table 6. Logistic regression estimates of high-achieving young people from disadvantaged socio-economic backgrounds entering a Russell Group university by age 21. Differences across genders and ethnicities.

	OR	Lower CI	Upper CI
White girl	0.81*	0.72	0.91
Asian boy	1.46*	1.26	1.69
Asian girl	1.61*	1.41	1.85
Black boy	1.39*	1.14	1.71
Black girl	1.61*	1.36	1.89
Mixed race boy	1.50*	1.17	1.94
Mixed race girl	1.35*	1.09	1.68
Other ethnicity boy	1.49*	1.19	1.88
Other ethnicity girl	1.46*	1.19	1.79
High achieving, above average SES	1.32*	1.21	1.44
Q3 SES, high achieving	1.20*	1.10	1.31

N	**340440**

Notes: Sample restricted to young people from disadvantaged socio-economic backgrounds with Key Stage 2 scores in the top quartile and with academic achievement measures available through to age 18. Figures are odds ratios, with values greater (less) than one indicating how much more (less) likely the group is to enter a Russell Group university than high achieving disadvantaged White boys. Controls included for Key Stage 2, GCSE and Key Stage 5 point scores and best three A-Level grades achieved. * indicates a statistically significant difference at the 5% level.

The key finding from this analysis is that – after controlling for achievement at school – most groups are more likely to enter a Russell Group university than high achieving disadvantaged White boys. For instance, the odds of high-achieving disadvantaged Asian and Black boys and girls attending a Russell Group university are around 40% to 60% higher than equally able, equally disadvantaged White boys. Indeed, there is only one group with less chance of entering a Russell Group university – high achieving disadvantaged White girls (odds ratio = 0.81). Table 6 hence reveals that it is primarily high achieving disadvantaged young people of White ethnicity that choose to not enter a high-status university, even when they obtain the grades to do so.

Research question 3. Living at home.

Table 7 addresses our third research question, exploring differences across socio-economic groups in living at home while studying for an undergraduate degree. The unconditional percentages presented in panel (a) illustrate a clear pattern that has remained broadly the same across the three cohorts. That is, high achieving young people from disadvantaged backgrounds are much more likely to continue living with their parents while an undergraduate than their peers from the most advantaged backgrounds. Specifically, around half of disadvantaged high achievers live at home while at university, compared to only a fifth of the most advantaged groups.

Table 7. The propensity for high-achieving young people from different socio-economic backgrounds living at home as an undergraduate.

(a) Unconditional percentages

	Cohort_A	Cohort_B	Cohort_C
1. High achieving, high SES	22%	19%	20%
2. High achieving, Q3 SES	28%	26%	26%
3. High achieving, Q2 SES	39%	38%	42%
4. High achieving, low SES	54%	50%	55%

5. Missing data	48%	46%	40%

(b) Logistic regression model estimates

	Unconditional	Conditional	University Fixed Effects
2nd SES quartile	0.58*	0.67*	0.72*
3rd SES quartile	0.32*	0.46*	0.53*
top SES quartile	0.23*	0.37*	0.45*
N	**242845**	**223870**	**223870**

Notes: Figures in panel (a) refers to the percent that lived at home as a undergraduate. Figures refer to odds ratios from logistic regression models where the sample is restricted to young people with Key Stage 2 scores in the top quartile, has information on academic achievement through to age 18 available and that started an undergraduate degree. The unconditional model does not include any controls. The conditional model includes controls for gender, ethnicity, Key Stage 2, GCSE and Key Stage 5 total points scores and best three grades achieved at A-Levels. The final model additionally controls for university fixed effects. Values below 1 indicate that the group in question are less likely to live at home as an undergraduate than the reference group (most disadvantaged quartile). Standard errors are reported in Appendix Table F2. * indicates odds ratio significantly different from one at the 5% level.

Panel (b) investigates whether this group can be explained by differences in the grades these young people achieve at school (middle column) and whether it continues to hold true when they attend the same higher education institution (right-hand column). These factors can only explain part of the difference in the propensity to continue living at home. In the unconditional model, the odds ratio sits at 0.23. This increases to 0.37 once school achievement controls are added to the model, and 0.45 when university fixed effects are included as well. Hence, amongst high achievers at the end of primary school that go on to achieve similar GCSE and A-Level grades, and who then go to the same university, the odds are almost twice as high that those from the most disadvantaged backgrounds will continue living with their parents.

In Table 8 we illustrate how this is to some extent being driven by the decisions made by certain ethnic groups – most prominently Asian boys and girls. The odds of high-achieving disadvantaged Asian students living at home as an undergraduate are around three to four times higher than their White peers. There is also a notable difference between White and Black boys (odds ratio = 1.45) but not girls (odds ratios of 1.10 versus 1.17). Overall, Tables 7 and 8 illustrate how there remains sizeable differences in university living arrangements – and hence experiences – across high achieving young people from different socio-economic backgrounds, particularly amongst certain ethnic groups.

Table 8. Logistic regression estimates of high-achieving young people from disadvantaged socio-economic backgrounds living at home as an undergraduate. Differences across genders and ethnicities.

	OR	Lower CI	Upper CI
White girl	1.10	1.00	1.22
Asian boy	3.24*	2.81	3.75
Asian girl	4.36*	3.79	5.02
Black boy	1.45*	1.22	1.72
Black girl	1.17	1.00	1.36
Mixed race boy	1.06*	0.83	1.34
Mixed race girl	0.99	0.80	1.21
Other ethnicity boy	1.49*	1.20	1.85
Other ethnicity girl	1.57*	1.29	1.90
High achieving, above average SES	0.61*	0.57	0.66
Q3 SES, high achieving	1.00	0.93	1.09
N		**231,430**	

Notes: Sample restricted to young people with Key Stage 2 scores in the top quartile from disadvantaged socio-economic backgrounds, with academic achievement measures available through to age 18 and who started an undergraduate degree. Estimates based on a logistic regression model controlling for Key Stage 2, GCSE and Key Stage 5 point scores, best three A-Level grades achieved and university fixed effects. Odds ratios greater than one indicate that the group was more likely to live at home as an undergraduate than high achieving disadvantaged White boys as the reference group. * indicates a statistically significant difference at the 5% level.

Research question 4. University outcomes and postgraduate study

Our final research question focuses on university outcomes and postgraduate study. This part of our analysis focuses on the eldest two cohorts (A and B) given that – with the data available – many individuals in Cohort C would not have yet had chance to complete their undergraduate studies and start a postgraduate degree. Table 9 begins by presenting results from our logistic regressions, focusing on the (conditional) difference between high early achievers from the most and least advantaged backgrounds.

Table 9. University outcomes for initially high-achieving young people different socio-economic backgrounds. Logistic regression estimates.

(a) Complete degree (any age/record)

	Q4 SES			N
	OR	Lower CI	Upper CI	
Cohort_A	1.64*	1.51	1.79	95410
Cohort_B	2.09*	1.91	2.28	96990

(b) Complete degree by age 21

	Q4 SES			N
	OR	Lower CI	Upper CI	
Cohort_A	1.15*	1.08	1.21	95410
Cohort_B	1.16*	1.09	1.23	96990

(c) Obtain at least a 2:1

	Q4 SES			N
	OR	Lower CI	Upper CI	
Cohort_A	1.50*	1.41	1.59	95410
Cohort_B	1.66*	1.56	1.76	96990

Notes: Sample restricted to young people with Key Stage 2 scores in the top quartile, with academic achievement measures available through to age 18 and who started an undergraduate degree. Estimates based on a logistic regression model controlling for Key Stage 2, GCSE and Key Stage 5 point scores and best three A-Level grades achieved. Values greater than one indicate that young people from the most advantaged socio-economic backgrounds are more likely to complete an undergraduate degree or obtain a 2:1 than the most disadvantaged group. * indicates a statistically significant difference at the 5% level.

The headline finding is that, when entering a university with similar qualifications, those from more advantaged socio-economic backgrounds are more likely to complete their degree and obtain at least a 2:1. In Appendix G, we illustrate how this finding continues to hold after including university fixed effects in the model as well. For instance, the odds of achieving at least a 2:1 are around 50% higher for young people from the most (as compared to the least) socio-economically advantaged families. In Appendix G we extend this analysis to consider differences across high achieving disadvantaged young people of different genders and ethnicities (see Appendix Table G4). We find consistent evidence that White and Asian girls are slightly more likely to complete university studies and obtain a 2:1 than high achieving disadvantaged White boys. On the other hand, high achieving disadvantaged Black boys are the least likely to graduate with a good degree. In particular, the odds of them obtaining a 2:1

are around half that as their equally high achieving, equally disadvantaged peers of White ethnicity.

Table 10 concludes by examining entry into postgraduate study. The limited number of disadvantaged young people with high levels of achievement during primary school that go on to postgraduate study means that the confidence intervals surrounding our results are now relatively wide. Thus, while the point estimates suggest that initial high achievers from the most advantaged group may be slightly less likely to continue into postgraduate study than those from disadvantaged backgrounds, most of these estimates are statistically insignificant at the 5% level. The only exception is for entry into MSc degrees in Cohort B; here the odds are 15% lower for the most (compared to the least) advantaged group, conditional upon their achievement at school and as an undergraduate.

Table 10. Entry into postgraduate study for initially high-achieving young people different socio-economic backgrounds.

(a) Entry into MSc

	Q4 SES			N
	OR	Lower CI	Upper CI	
Cohort_A	0.99	0.93	1.06	87040
Cohort_B	0.85*	0.80	0.91	89255

(b) Entry into PhD

	Q4 SES			N
	OR	Lower CI	Upper CI	
Cohort_A	0.91	0.77	1.07	87040
Cohort_B	0.86	0.73	1.01	89255

Notes: Sample restricted to young people with Key Stage 2 scores in the top quartile, with academic achievement measures available through to age 18 and who started and completed an undergraduate degree. Estimates based on a logistic regression model controlling for Key Stage 2, GCSE and Key Stage 5 point scores, best three A-Level grades achieved, whether they completed an undergraduate at a Russell Group or Oxbridge University and degree classification obtained. Values less than one indicate that young people from the most advantaged socio-economic backgrounds are less likely to complete a postgraduate degree than the most disadvantaged group. * indicates a statistically significant difference at the 5% level.

While the confidence intervals are now wide, Appendix H reports how this varies across genders and ethnic groups. With respect to entering an MSc, girls across each ethnic group are more likely to continue their education than boys. For instance, the odds of starting an MSc are 35% higher for White girls than White boys, with this increasing to around 50% for

Black/Asian girls and 70% for mixed race girls. Initially high achieving boys are hence less likely to study for an MSc than initially high achieving girls, amongst those with similar levels of achievement at school and during their undergraduate studies, regardless of their ethnicity. In contrast, we find that high achieving disadvantaged White boys are the most likely to study for a PhD in comparison to other gender-by-ethnicity combinations, once their prior achievement at school and as an undergraduate has been controlled (see Appendix Table H3 for the full set of results).

5. Discussion

Young people from disadvantaged socio-economic backgrounds who excel academically during primary school are a key group for enhancing social mobility. They have developed the platform needed to go on to excel at secondary school, enter a high status university and find well-paid, professional employment. If they can't go on to succeed, then who can? Most previous research into this group has focused on the academic progress they make while at school, with comparatively little written about their later educational outcomes, such as access to – and graduation from – university. This paper has sought to contribute these insights to the existing literature, providing the most comprehensive analysis into the higher education outcomes of socio-economically disadvantaged children who excelled during primary school to date.

Our analysis has shown how this group are much less likely to attend and graduate from a Russell Group or Oxbridge university than their equally able but more socio-economically advantaged peers. Much of this gap can be explained by differences in the academic progress these groups make during secondary school. That is, after conditioning upon end-of-school performance, differences in higher education access and outcomes across early high achievers from different socio-economic backgrounds become relatively small (though remain non-trivial). Tentative evidence has also emerged that the socio-economic gap in entering Russell Group and Oxbridge universities amongst early high achievers may have narrowed over time. Yet these groups may have rather different experiences while at university, given how high early achievers from poor backgrounds are much more likely to stay living at home with their parents. We also illustrate for the first-time important differences in higher education access, choices and outcomes across socio-economically disadvantaged high early achievers of different genders and ethnicities. For instance, those of Asian ethnicity are much more likely than their White peers to convert their early academic potential into high grades at the end of secondary school, and thus go on to study at a high-status university.

These findings are broadly consistent with the existing evidence base. For instance, previous research within the higher education literature has generally noted how there are significant differences in university access across socio-economic groups (Budoki et al., 2021), with this to a great extent being explained by variation in performance in end-of-school examinations (Chowdry et al., 2013). Indeed, prior evidence from the sociology of education literature has typically found that "primary effects" (the proportion of the gap in higher education outcomes attributable to differences in prior achievement) are typically much more substantial than "secondary effects" (the gap in higher education entry rates that remains once prior achievement has been controlled) – Jackson (2013b). Our contribution has been to show how this also holds true amongst young people who excelled academically during primary school, how this intersects with other background characteristics such as gender and ethnicity, and how these relationships have changed across three school cohorts.

With respect to previous research into the future educational outcomes of disadvantaged young people with high levels of early achievement, only Crawford et al. (2017) has considered entry into university in England. Like our study, they found that much of the difference in the university entrance rates of this group compared to their more advantaged peers can be traced back to how they progressed during secondary school. We have shown, however, that this broad finding masks important differences by sub-group, with a lack of progress made during secondary school a particular issue amongst initially high achieving disadvantaged young people of White ethnicity. Our analysis has also built on prior work in this field by demonstrating the extent that disadvantaged young people with strong academic skills at the end of primary school go on to enter universities of particular public interest (e.g. Oxbridge), how they differ in associated choices made (e.g. whether to move out from home) and in their eventual graduation rates, degree outcomes and entry into postgraduate courses.

One must of course interpret these findings in the context of the limitations of our research. First, our measure of high early achievement is based on tests taken at age 10/11 that focus on reading and mathematics. One may argue that measures at a younger age might be advantageous (e.g. to also capture progress made during primary school) and that the tests should ideally cover broader aspects of school's curricula (e.g. science, history, geography). While such alternative measures would also come with drawbacks (e.g. increased levels of measurement error), further research into the primary school experiences of disadvantaged children with strong pre-school skills is needed. Second, our analysis is restricted to young people who sat the Key Stage 2 tests and thus studying in the state school sector. Our analysis

has therefore been unable to include around 7% of young people in England that attend independent (private) schools. Third, while we have presented results across three school cohorts, there are several issues that limit their comparability. One is that substantial changes have been made to England's GCSE and A-Level examinations over time, which has led to some changes to the school performance measures available for Cohort C relative to cohorts A and B. Another is that Cohort C were entering and progressing through higher education at the time of the COVID-19 pandemic, and would thus have had rather different university experiences to older cohorts. Unfortunately. It is not possible to firmly establish the impact these issues have on the cross-cohort comparability of our results. Finally, while our use of administrative data has many advantages – including providing the requisite sample size to study differences across genders and ethnicities – it also comes with certain limitations. Most notably, we cannot probe in much detail what may be driving our results. For instance, while we have established high achieving young people from poor backgrounds are much more likely to continue living with their parents while at university, we are unable to establish whether this is being driven by financial, social or familial concerns. Likewise, we are unable to establish the extent to which this is being driven by the different benefits of remaining at home while studying at university, such as maintaining close contact with family bonds, access to local employment opportunities, less graduate debt and greater ability to focus on studies without needing so many hours of paid work.

Our findings do nevertheless have some important implications. One is that more needs to be done to help young people from poor backgrounds who excel during primary school to convert this early potential into strong school grades. This is particularly true for those of White ethnicity, given they are the least likely to keep pace academically during secondary school with their peers from more advantaged backgrounds. Another is that, if one believes that moving out of the family home is a key part of the university experience, further efforts are needed to help high achieving disadvantaged young people to take this step. While further work is needed to establish why this group is more reluctant to leave home, it may be that they need to feel greater financial security to do so. Finally, there may now be relatively little to gain from universities putting further resource into widening access schemes targeting teenagers, given how the link between socio-economic background and university entry is limited once prior achievement has been controlled. Rather, it may be better for universities to invest any available resource – or to reallocate some existing resource – into supporting disadvantaged

high-achievers during their university studies, helping them to make it through to graduation with the best possible grades.

References

Breen, R. & Karlson, K. (2014). Education and social mobility: new analytical approaches. *European Sociological Review* 30(1): 107-118.

Bukodi, E., Goldthorpe, J. H., & Zhao, Y. (2021). Primary and secondary effects of social origins on educational attainment: New findings for England. *British Journal of Sociology, 72*, 627–650. https://doi.org/10.1111/1468-4446.12845

Campbell, S.; Macmillan, L. & Wyness, G. (2019). Mismatch in higher education: prevalence, drivers and outcomes. Accessed from https://www.nuffieldfoundation.org/wp-content/uploads/2019/11/Wyness-42856-MPO-Final-Nov19.pdf

Chowdry, H., Crawford, C., Dearden, L., Goodman, A., & Vignoles, A. (2013). Widening participation in higher education: Analysis using linked administrative data. Journal of the Royal Statistical Society, Series A, 176, 431– 457.

Codiroli McMaster, N., & Cook, R. (2019). The contribution of intersectionality to quantitative research into educational inequalities. *Review of Education*, 7, 271-292. https://doi.org/10.1002/rev3.3116

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

Cunha, F. Heckman, J. Lochner, L and Masterov, D. (2006). 'Interpreting the Evidence on Life Cycle Skill Formation'. In *Handbook of the Economics of Education* (eds E. Hanushek and F. Welch), pp. 697-812. Amsterdam: Holland North.

Davey, K. (2025). Beyond the promise of social mobility? Re-thinking the purpose of higher education in England through high-achieving, working-class girls' reflexive reasons for applying to high-tariff universities. *Research in Post-Compulsory Education*, *30*(1), 184–202. https://doi.org/10.1080/13596748.2025.2469004

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

Gong, Y., Lochner, L., Stinebrickner, R., & Stinebrickner, T. R. (2021). The consumption value of college. National Bureau of Economic Research Working Paper No. 26335. http://www.nber.org/papers/w26335

Hemsley-Brown, J. (2015). Getting into a Russell Group university: high scores and private schooling. *British Educational Research Journal*, *41*(3), 398–422. http://www.jstor.org/stable/24464021

Holdsworth, C. (2006). 'Don't you think you're missing out, living at home?' Student experiences and residential transitions. *The Sociological Review* 54(3): 495-519. https://doi.org/10.1111/j.1467-954X.2006.00627.x

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

Jackson, M., Erikson, R., Goldthorpe, J. H., & Yaish, M. (2007). Primary and secondary effects in class differentials in educational attainment: the transition to A-Level courses in England and Wales. *Acta Sociologica*, *50*(3), 211–229. http://www.jstor.org/stable/20459999

Jackson, M. (Ed.). (2013a). *Determined to Succeed?: Performance versus Choice in Educational Attainment* (1st ed.). Stanford University Press. https://doi.org/10.2307/j.ctvqsdrjr

Jackson, M. (2013b). Social Background and Educational Transitions in England. In M. Jackson (Ed.), *Determined to Succeed?: Performance versus Choice in Educational Attainment* (1st ed., pp. 253–278). Stanford University Press. https://doi.org/10.2307/j.ctvqsdrjr.13

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*.

Jerrim, J., & Carvajal, M. (2024). What happens to bright 5-year-olds from poor backgrounds? Longitudinal evidence from the Millennium Cohort Study.

Kassenboehmer, S., Leung, F. & Schurer, S. (2018). University education and non-cognitive skill development. *Oxford Economic Papers* 70(2): 538–562. https://doi.org/10.1093/oep/gpy002

Mulder, C. H., & Clark, W. A. V. (2002). Leaving home for college and gaining independence. *Environment and Planning A: Economy and Space* 34(6): 981-999. https://doi.org/10.1068/a34149

Neidhöfer, G., Ciaschi, M., Gasparini, L & Serrano, J. (2023). Social mobility and economic development. *Journal of Economic Growth* https://doi.org/10.1007/s10887-023-09234-8

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

Sullivan, A., Parsons, S., Wiggins, R., Heath, A., & Green, F. (2014). Social origins, school type and higher education destinations. *Oxford Review of Education, 40*(6), 739-763. https://doi.org/10.1080/03054985.2014.979015

Sutton Trust. (2019). Elitist Britian 2019. The educational backgrounds of Britian's leading people. Retrieved from https://www.suttontrust.com/wp-content/uploads/2020/01/Elitist-Britain-2019-Summary-Report.pdf

Swinney, P. (2016). The Great British brain drain. Retrieved from https://www.centreforcities.org/reader/great-british-brain-drain/migration-students-graduates/

Appendix A. Robustness tests to explore the impact of Kelley's paradox (regression to the mean)

Jerrim and Carvajal (2024) discuss in detail the issue of Kelley's paradox and the related issue of regression to the mean. This is an issue whenever a test score that is less than perfectly reliable is used to classify individuals into different achievement groups, such as – in our application – "high achievers". In particular, if these individuals were to complete the same test(s) again the very next day, they would – on average – achieve a lower score than on the initial test used to classifying them as high-achieving individuals (i.e. their scores will regress towards the mean).

As explained by Wainer & Brown (2007), the extent of this regression may be group specific. For instance, children from lower socioeconomic status backgrounds achieve scores that are – on average – lower than their peers from higher socioeconomic backgrounds. The regression effect will hence be particularly large for children identified as "high achieving" on the first test(s) if they come from a disadvantaged socioeconomic background. In other words, this may lead one to erroneously conclude that high-achieving children from disadvantaged backgrounds have "fallen behind" their more socioeconomically advantaged peers, when this is really an artefact of the measurement error present in the tests used to classify individuals into different ability groups.

In this appendix we follow the four step approach suggested by Jerrim and Carvajal (2024) to investigate the robustness of our results to this issue. This can be summarised as follows. In the first step we estimate the difference in our outcome measures across high achieving children (top Key Stage 2 quartile) from socioeconomically advantaged and disadvantaged backgrounds. Then, in step 2, we estimate the difference in the academic abilities of high achieving children from different socioeconomic backgrounds under different assumptions of Key Stage 2 reliability. This is based upon the formula provided by Wainer & Brown (2007):

$$\tau_i = \rho(x_i) + (1 - \rho).u_g$$

Where:

τ_i = The young person's "true" academic abilities at the end of primary school.

ρ = The assumption one makes regarding the reliability of the Key Stage 2 tests as a measure of young people's academic abilities.

x_i = The score the young person achieves on the Key Stage 2 test.

u_g = The average Key Stage 2 score within the individual's socioeconomic group (g) (the average Key Stage 2 score of children from disadvantaged backgrounds, for example).

The key input here is ρ – the assumption one makes about Key Stage 2 test reliability as a marker of young people's overall academic abilities at the end of primary school (a point to which we return below).

In the third step we then estimate the link between the academic ability of young people at the end of primary school and each of the relevant outcome measures. The final step is that to adjust downwards the "raw" difference from step 1 by the estimated difference in high achieving advantaged and disadvantaged children's academic abilities (from step 2) multiplied

by how strongly this is related to the outcome in question (from step 3). Jerrim & Carvajal (2024: section 3, Appendix B and Appendix C) provides for further details about this approach.

A key question is what value of ρ to use in this process (i.e. the assumption one makes about Key Stage 2 test reliability). While the reported reliability of these tests are high – and they involve over three hours of test time – they only cover the subjects of English and mathematics (and not other areas of the curriculum). Consequently, following Jerrim and Carvajal (2024), we test the sensitivity of results to using a range of different values of ρ, from a low of 0.5 to a high of 1.0 (essentially equivalent to assuming the Key Stage 2 tests are perfectly reliable).

Our results can be found in Appendix Table A1 to A13 below for each of our outcomes of interest. The results in panel (a) are based upon logistic regressions (with estimates presented as odds ratios) with those in panel (b) based upon linear probability models (with estimates presented as percentage point differences). The value of rho refers to the assumption made regarding the reliability of the Key Stage 2 test as a measure of 10/11-year-olds academic achievement. Lower values of rho refer to lower levels of reliability.

Our findings can be summarised as follows. First, in the unconditional results, the difference between high achieving young people from advantaged and disadvantaged backgrounds is somewhat reduced once one accounts for Kelley's paradox. However, in general, the change in the results is usually fairly modest. Take entry into university (Appendix Table A13), for instance. The unconditional difference between advantaged and disadvantaged young people with high Key Stage 2 maths scores in entering university is 21 percentage points when no adjustment is made for Kelley's paradox (i.e. when one assumes that the Key Stage 2 tests measure primary school children's academic abilities without any error). This difference falls to 14 percentage points when one instead assumes that these tests measure primary school children's academic skills with a reasonable degree of error (rho = 0.7). Thus, while the magnitude of the difference declines, the broad pattern of unconditional results continues to hold for most outcomes.

Second, in general, adjusting for Kelley's paradox has less impact on the conditional results. This is as expected; the controls included for prior achievement in the conditional models will have already soaked up much of the measurement error present in the Key Stage 2 tests. Take the chances of obtaining a 2:1 degree for instance (Appendix Table A3). The conditional difference between initially high achieving young people from advantaged and disadvantaged socio-economic backgrounds is estimated to be 12 percentage points when no adjustment is made for Kelley's paradox. This difference falls only slightly – to 11 percentage points – even when measurement error in the Key Stage 2 tests is assume to be very large (rho = 0.5). This suggests that the condition results – that are the focus in the main body of the paper – are unlikely to be severely affected by Kelley's paradox.

Finally, the results for GCSE achievement (Appendix Table A6) is somewhat of an exception, where both the conditional and unconditional estimates decline after adjusting for Kelley's paradox. This is somewhat expected, given how the conditional estimates only include weak measures of prior achievement (Key Stage 1 levels) which will not soak up much of the residual error in the Key Stage 2 tests. For instance, the estimated difference in GCSE total points between initially high achieving young people from advantaged and disadvantaged backgrounds falls from 0.26 standard deviations (assuming the Key Stage 2 tests are measured without error) down to 0.08 standard deviations when assuming the measurement error is reasonably sizeable. The results for GCSE outcomes presented in the main text may thus to

some extent be affected by Kelley's paradox, leading to a moderate upward bias in the estimated difference across socio-economic groups for this particular outcome.

Appendix Table A1. Obtain BBB at A-Level

(a) Logistic regression

Rho	Unconditional				Conditional			
	OR	Lower CI	Upper CI	N	OR	Lower CI	Upper CI	N
Raw	0.25	0.25	0.26	810980	0.34	0.33	0.35	149135
1.0	0.32	0.31	0.33	810980	0.39	0.38	0.40	149135
0.9	0.39	0.38	0.41	810980	0.40	0.39	0.41	149135
0.8	0.51	0.49	0.52	810980	0.41	0.40	0.42	149135
0.7	0.69	0.67	0.71	810980	0.42	0.41	0.44	149135
0.6	0.98	0.95	1.01	810980	0.44	0.43	0.46	149135
0.5	1.47	1.43	1.51	810980	0.47	0.46	0.49	149135

(b) Linear probability model

Rho	Unconditional				Conditional			
	PP diff	Lower CI	Upper CI	N	PP diff	Lower CI	Upper CI	N
Raw	-27%	-27%	-27%	810980	-21%	-22%	-21%	149135
1.0	-26%	-26%	-26%	810980	-18%	-19%	-17%	149135
0.9	-25%	-25%	-25%	810980	-18%	-18%	-17%	149135
0.8	-24%	-24%	-23%	810980	-17%	-18%	-16%	149135
0.7	-22%	-23%	-22%	810980	-16%	-17%	-16%	149135
0.6	-20%	-21%	-20%	810980	-15%	-16%	-15%	149135
0.5	-18%	-18%	-18%	810980	-14%	-15%	-13%	149135

Appendix Table A2. Obtained a 1st class degree

(a) Logistic regression

Rho	Unconditional				Conditional			
	OR	Lower CI	Upper CI	N	OR	Lower CI	Upper CI	N
Raw	0.38	0.37	0.40	810980	0.64	0.62	0.67	120075
1.0	0.43	0.42	0.45	810980	0.65	0.63	0.68	120075
0.9	0.48	0.46	0.50	810980	0.66	0.63	0.68	120075
0.8	0.55	0.53	0.57	810980	0.66	0.63	0.69	120075
0.7	0.65	0.63	0.68	810980	0.66	0.64	0.69	120075
0.6	0.79	0.76	0.83	810980	0.67	0.64	0.70	120075
0.5	1.01	0.97	1.05	810980	0.68	0.65	0.71	120075

(b) Linear probability model

Rho	Unconditional				Conditional			
	PP diff	Lower CI	Upper CI	N	PP diff	Lower CI	Upper CI	N
Raw	-10%	-10%	-10%	810980	-6%	-7%	-6%	120075
1.0	-10%	-10%	-9%	810980	-6%	-7%	-6%	120075
0.9	-9%	-9%	-9%	810980	-6%	-7%	-5%	120075
0.8	-9%	-9%	-8%	810980	-6%	-7%	-5%	120075
0.7	-8%	-8%	-8%	810980	-6%	-6%	-5%	120075
0.6	-7%	-8%	-7%	810980	-6%	-6%	-5%	120075
0.5	-6%	-7%	-6%	810980	-6%	-6%	-5%	120075

Appendix Table A3. Obtained a 2:1 degree

(a) Logistic regression

Rho	Unconditional				Conditional			
	OR	Lower CI	Upper CI	N	OR	Lower CI	Upper CI	N
Raw	0.37	0.36	0.38	810980	0.60	0.58	0.62	120075
1.0	0.41	0.40	0.42	810980	0.61	0.59	0.63	120075
0.9	0.46	0.45	0.47	810980	0.61	0.59	0.63	120075
0.8	0.52	0.51	0.54	810980	0.61	0.59	0.63	120075
0.7	0.61	0.60	0.63	810980	0.62	0.60	0.64	120075
0.6	0.74	0.73	0.76	810980	0.62	0.60	0.64	120075
0.5	0.93	0.91	0.96	810980	0.63	0.61	0.65	120075

(b) Linear probability model

Rho	Unconditional				Conditional			
	PP diff	Lower CI	Upper CI	N	PP diff	Lower CI	Upper CI	N
Raw	-22%	-23%	-22%	810980	-12%	-13%	-12%	120075
1.0	-21%	-22%	-21%	810980	-12%	-13%	-11%	120075
0.9	-20%	-20%	-20%	810980	-12%	-13%	-11%	120075
0.8	-19%	-19%	-18%	810980	-12%	-13%	-11%	120075
0.7	-17%	-17%	-17%	810980	-12%	-13%	-11%	120075
0.6	-15%	-15%	-15%	810980	-12%	-12%	-11%	120075
0.5	-12%	-13%	-12%	810980	-11%	-12%	-10%	120075

Appendix Table A4. Graduated from university

(a) Logistic regression

Rho	Unconditional				Conditional			
	OR	Lower CI	Upper CI	N	OR	Lower CI	Upper CI	N
Raw	0.41	0.40	0.42	810980	0.71	0.69	0.74	120075
1.0	0.46	0.45	0.47	810980	0.72	0.70	0.75	120075
0.9	0.51	0.49	0.52	810980	0.72	0.70	0.75	120075
0.8	0.57	0.55	0.58	810980	0.73	0.70	0.75	120075
0.7	0.65	0.63	0.66	810980	0.73	0.71	0.76	120075
0.6	0.76	0.75	0.78	810980	0.74	0.72	0.77	120075
0.5	0.92	0.90	0.95	810980	0.75	0.73	0.78	120075

(b) Linear probability model

Rho	Unconditional				Conditional			
	PP diff	Lower CI	Upper CI	N	PP diff	Lower CI	Upper CI	N
Raw	-22%	-22%	-21%	810980	-7%	-7%	-6%	120075
1.0	-20%	-21%	-20%	810980	-6%	-7%	-6%	120075
0.9	-19%	-19%	-18%	810980	-6%	-7%	-6%	120075
0.8	-17%	-18%	-17%	810980	-6%	-7%	-6%	120075
0.7	-15%	-16%	-15%	810980	-6%	-7%	-5%	120075
0.6	-13%	-13%	-12%	810980	-6%	-7%	-5%	120075
0.5	-10%	-10%	-9%	810980	-6%	-6%	-5%	120075

Appendix Table A5. Educational achievement data through to age 18

(a) Logistic regression

Rho	Unconditional				Conditional			
	OR	Lower CI	Upper CI	N	OR	Lower CI	Upper CI	N
Raw	0.26	0.25	0.27	810980	0.36	0.35	0.37	810935
1.0	0.29	0.28	0.30	810980	0.35	0.34	0.36	810935
0.9	0.32	0.31	0.33	810980	0.36	0.35	0.37	810935
0.8	0.36	0.35	0.37	810980	0.38	0.37	0.39	810935
0.7	0.42	0.41	0.43	810980	0.40	0.39	0.41	810935
0.6	0.50	0.49	0.51	810980	0.44	0.43	0.45	810935
0.5	0.63	0.61	0.64	810980	0.50	0.48	0.51	810935

(b) Linear probability model

Rho	Unconditional				Conditional			
	PP diff	Lower CI	Upper CI	N	PP diff	Lower CI	Upper CI	N
Raw	-23%	-24%	-23%	810980	-14%	-15%	-14%	810935
1.0	-21%	-21%	-20%	810980	-15%	-15%	-14%	810935
0.9	-19%	-19%	-18%	810980	-14%	-14%	-14%	810935
0.8	-17%	-17%	-16%	810980	-13%	-14%	-13%	810935
0.7	-14%	-14%	-13%	810980	-12%	-12%	-11%	810935
0.6	-10%	-11%	-10%	810980	-10%	-11%	-10%	810935
0.5	-6%	-6%	-5%	810980	-8%	-8%	-8%	810935

Appendix Table A6. GCSE total point scores (effect sizes)

Rho	Unconditional				Conditional			
	Effect size	Lower CI	Upper CI	N	Effect size	Lower CI	Upper CI	N
Raw	-0.32	-0.33	-0.31	810935	-0.26	-0.26	-0.25	810935
1.0	-0.27	-0.28	-0.26	810935	-0.26	-0.26	-0.25	810935
0.9	-0.22	-0.23	-0.21	810935	-0.21	-0.22	-0.20	810935
0.8	-0.15	-0.16	-0.15	810935	-0.15	-0.16	-0.14	810935
0.7	-0.08	-0.09	-0.07	810935	-0.08	-0.09	-0.07	810935
0.6	0.01	0.00	0.02	810935	0.00	-0.01	0.01	810935
0.5	0.12	0.11	0.13	810935	0.10	0.09	0.11	810935

Appendix Table A7. Key Stage 5 points (effect sizes)

Rho	Unconditional				Conditional			
	Effect size	Lower CI	Upper CI	N	Effect size	Lower CI	Upper CI	N
Raw	-0.44	-0.46	-0.43	397780	-0.38	-0.39	-0.37	149135
1.0	-0.41	-0.42	-0.39	397780	-0.33	-0.34	-0.32	149135
0.9	-0.37	-0.38	-0.36	397780	-0.32	-0.34	-0.31	149135
0.8	-0.32	-0.33	-0.31	397780	-0.32	-0.33	-0.30	149135
0.7	-0.26	-0.28	-0.25	397780	-0.30	-0.31	-0.29	149135
0.6	-0.20	-0.21	-0.18	397780	-0.29	-0.30	-0.28	149135
0.5	-0.11	-0.13	-0.10	397780	-0.26	-0.28	-0.25	149135

Appendix Table A8. Live with parents as an undergraduate

(a) Logistic regression

Rho	Unconditional				Conditional			
	OR	Lower CI	Upper CI	N	OR	Lower CI	Upper CI	N
Raw	4.39	4.24	4.53	232985	3.67	3.53	3.80	98720
1.0	4.05	3.92	4.19	232985	3.55	3.42	3.68	98720
0.9	3.78	3.66	3.91	232985	3.51	3.39	3.64	98720
0.8	3.46	3.35	3.58	232985	3.47	3.35	3.60	98720
0.7	3.09	2.99	3.19	232985	3.41	3.29	3.54	98720
0.6	2.67	2.58	2.76	232985	3.33	3.21	3.45	98720
0.5	2.22	2.14	2.29	232985	3.19	3.08	3.31	98720

(a) Linear probability model

Rho	Unconditional				Conditional			
	PP diff	Lower CI	Upper CI	N	PP diff	Lower CI	Upper CI	N
Raw	33%	32%	33%	232985	28%	28%	29%	98720
1.0	31%	30%	32%	232985	28%	27%	28%	98720
0.9	29%	29%	30%	232985	28%	27%	28%	98720
0.8	27%	27%	28%	232985	27%	27%	28%	98720
0.7	25%	24%	26%	232985	27%	26%	28%	98720
0.6	22%	21%	22%	232985	27%	26%	27%	98720
0.5	18%	17%	18%	232985	26%	25%	27%	98720

Appendix Table A9. Studied for an MSc

(a) Logistic regression

Rho	Unconditional				Conditional			
	OR	Lower CI	Upper CI	N	OR	Lower CI	Upper CI	N
Raw	0.56	0.54	0.59	810980	1.04	1.00	1.09	91370
1.0	0.62	0.60	0.64	810980	1.05	1.01	1.10	91370
0.9	0.67	0.65	0.70	810980	1.06	1.01	1.10	91370
0.8	0.74	0.71	0.77	810980	1.06	1.01	1.11	91370
0.7	0.84	0.81	0.87	810980	1.06	1.01	1.11	91370
0.6	0.97	0.93	1.00	810980	1.07	1.02	1.11	91370
0.5	1.14	1.10	1.19	810980	1.07	1.02	1.12	91370

(b) Linear probability model

Rho	Unconditional				Conditional			
	PP diff	Lower CI	Upper CI	N	PP diff	Lower CI	Upper CI	N
Raw	-5%	-5%	-5%	810980	1%	0%	2%	91370
1.0	-5%	-5%	-5%	810980	1%	0%	2%	91370
0.9	-5%	-5%	-4%	810980	1%	0%	2%	91370
0.8	-4%	-4%	-4%	810980	1%	0%	2%	91370
0.7	-4%	-4%	-4%	810980	1%	0%	2%	91370
0.6	-3%	-3%	-3%	810980	1%	0%	2%	91370
0.5	-3%	-3%	-2%	810980	1%	0%	2%	91370

Appendix Table A10. Entered an Oxbridge university

(a) Logistic regression

Rho	Unconditional				Conditional			
	OR	Lower CI	Upper CI	N	OR	Lower CI	Upper CI	N
Raw	0.25	0.22	0.28	810980	0.82	0.72	0.94	149135
1.0	0.36	0.32	0.41	810980	0.88	0.77	1.01	149135
0.9	0.51	0.45	0.58	810980	0.90	0.78	1.03	149135
0.8	0.76	0.67	0.86	810980	0.91	0.80	1.05	149135
0.7	1.23	1.09	1.39	810980	0.94	0.82	1.08	149135
0.6	2.18	1.93	2.47	810980	0.97	0.85	1.11	149135
0.5	4.41	3.90	4.99	810980	1.02	0.89	1.17	149135

(b) Linear probability model

Rho	Unconditional				Conditional			
	PP diff	Lower CI	Upper CI	N	PP diff	Lower CI	Upper CI	N
Raw	-2%	-2%	-2%	810980	0%	0%	0%	149135
1.0	-2%	-2%	-2%	810980	0%	0%	0%	149135
0.9	-2%	-2%	-2%	810980	0%	0%	0%	149135
0.8	-2%	-2%	-2%	810980	0%	0%	0%	149135
0.7	-2%	-2%	-2%	810980	0%	0%	1%	149135
0.6	-2%	-2%	-1%	810980	0%	0%	1%	149135
0.5	-1%	-1%	-1%	810980	1%	0%	1%	149135

Appendix Table A11. Started a PhD

(a) Logistic regression

Rho	Unconditional				Conditional			
	OR	Lower CI	Upper CI	N	OR	Lower CI	Upper CI	N
Raw	0.35	0.32	0.39	810980	0.91	0.82	1.02	91370
1.0	0.42	0.38	0.46	810980	0.93	0.83	1.04	91370
0.9	0.49	0.44	0.54	810980	0.93	0.84	1.04	91370
0.8	0.58	0.53	0.65	810980	0.94	0.84	1.05	91370
0.7	0.73	0.66	0.81	810980	0.95	0.85	1.06	91370
0.6	0.95	0.86	1.06	810980	0.96	0.86	1.07	91370
0.5	1.31	1.18	1.45	810980	0.97	0.87	1.08	91370

(b) Linear probability model

Rho	Unconditional				Conditional			
	PP diff	Lower CI	Upper CI	N	PP diff	Lower CI	Upper CI	N
Raw	-2%	-2%	-2%	810980	0%	-1%	0%	91370
1.0	-2%	-2%	-2%	810980	0%	-1%	0%	91370
0.9	-2%	-2%	-2%	810980	0%	0%	0%	91370
0.8	-2%	-2%	-2%	810980	0%	0%	0%	91370
0.7	-1%	-2%	-1%	810980	0%	0%	0%	91370
0.6	-1%	-1%	-1%	810980	0%	0%	0%	91370
0.5	-1%	-1%	-1%	810980	0%	0%	1%	91370

Appendix Table A12. Entered a Russell Group university

(a) Logistic regression

Rho	Unconditional				Conditional			
	OR	Lower CI	Upper CI	N	OR	Lower CI	Upper CI	N
Raw	0.31	0.30	0.32	810980	0.79	0.76	0.82	149135
1.0	0.38	0.37	0.40	810980	0.80	0.77	0.84	149135
0.9	0.46	0.45	0.47	810980	0.81	0.77	0.84	149135
0.8	0.57	0.55	0.59	810980	0.81	0.78	0.85	149135
0.7	0.74	0.72	0.77	810980	0.82	0.79	0.85	149135
0.6	1.01	0.98	1.04	810980	0.83	0.79	0.86	149135
0.5	1.43	1.39	1.48	810980	0.84	0.81	0.87	149135

(b) Linear probability model

Rho	Unconditional				Conditional			
	PP diff	Lower CI	Upper CI	N	PP diff	Lower CI	Upper CI	N
Raw	-18%	-18%	-18%	810980	-2%	-3%	-2%	149135
1.0	-17%	-17%	-17%	810980	-2%	-2%	-1%	149135
0.9	-16%	-17%	-16%	810980	-2%	-2%	-1%	149135
0.8	-15%	-16%	-15%	810980	-2%	-2%	-1%	149135
0.7	-14%	-15%	-14%	810980	-2%	-2%	-1%	149135
0.6	-13%	-13%	-13%	810980	-2%	-2%	-1%	149135
0.5	-12%	-12%	-11%	810980	-1%	-2%	-1%	149135

Appendix Table A13. Entered any university

(a) Logistic regression

Rho	Unconditional				Conditional			
	OR	Lower CI	Upper CI	N	OR	Lower CI	Upper CI	N
Raw	0.42	0.42	0.43	810980	0.98	0.95	1.01	149135
1.0	0.47	0.46	0.48	810980	0.98	0.95	1.01	149135
0.9	0.52	0.51	0.53	810980	0.98	0.95	1.01	149135
0.8	0.58	0.57	0.59	810980	0.98	0.95	1.01	149135
0.7	0.67	0.65	0.68	810980	0.98	0.95	1.01	149135
0.6	0.78	0.77	0.80	810980	0.98	0.95	1.01	149135
0.5	0.94	0.92	0.96	810980	0.98	0.95	1.01	149135

(b) Linear probability model

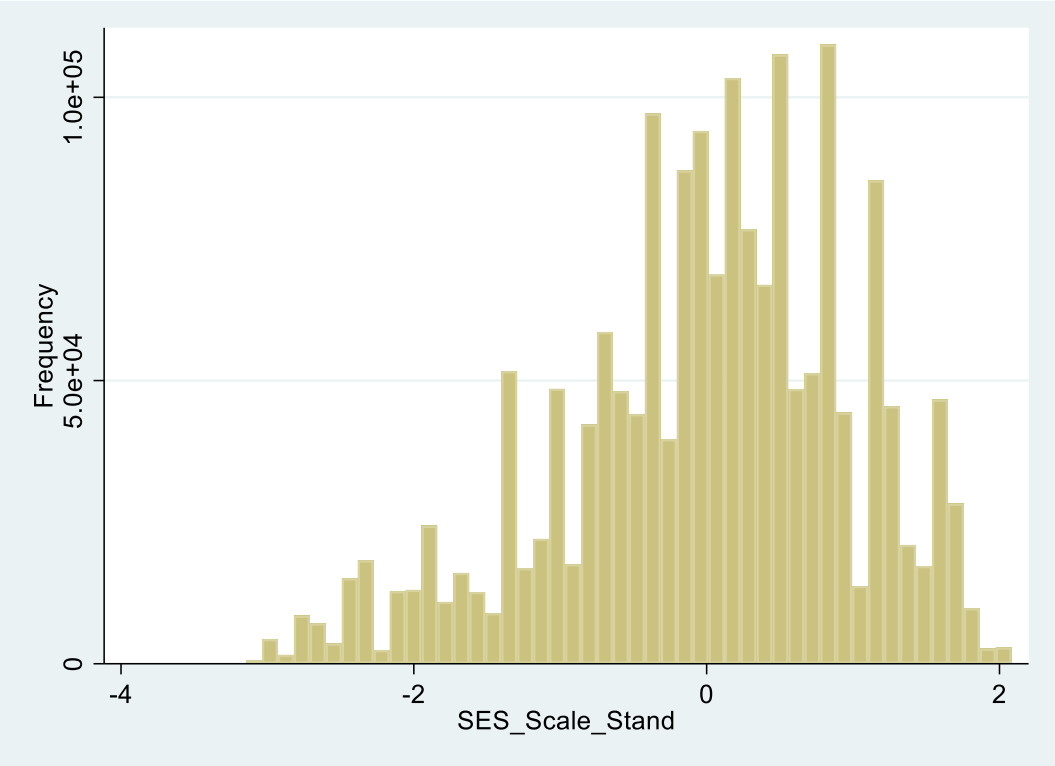
Rho	Unconditional				Conditional			
	PP diff	Lower CI	Upper CI	N	PP diff	Lower CI	Upper CI	N
Raw	-21%	-21%	-21%	810980	-1%	-1%	0%	149135
1.0	-19%	-20%	-19%	810980	-1%	-1%	0%	149135
0.9	-18%	-18%	-17%	810980	-1%	-1%	0%	149135
0.8	-16%	-17%	-16%	810980	-1%	-1%	0%	149135
0.7	-14%	-14%	-13%	810980	-1%	-1%	0%	149135
0.6	-11%	-12%	-11%	810980	-1%	-1%	0%	149135
0.5	-8%	-9%	-8%	810980	-1%	-1%	0%	149135

Appendix B1. Distribution of the socio-economic status scale

Variable	Group	Mean	SD	N
IDACI decile	Decile 1	1.39	0.40	176,575
	Decile 2	1.01	0.34	176,565
	Decile 3	0.70	0.32	176,575
	Decile 4	0.43	0.33	176,575
	Decile 5	0.17	0.37	176,565
	Decile 6	-0.10	0.43	176,565
	Decile 7	-0.38	0.50	176,565
	Decile 8	-0.69	0.58	176,570
	Decile 9	-1.05	0.67	176,585
	Decile 10	-1.49	0.78	176,550
NSSEC	Higher managerial	0.92	0.55	137,350
	Lower managerial	0.54	0.54	177,810
	Intermediate occupations	0.30	0.54	81,620
	Small employers	0.02	0.55	60,015
	Lower supervisory / technical	-0.04	0.52	35,180
	Semi-routine	-0.45	0.63	87,240
	Routine	-0.63	0.67	46,035
	Never worked	-1.51	0.92	2,930
Parental education	Doesn't hold a degree	0.04	0.90	340,975
	Holds a degree	0.43	0.69	326,395
FSM eligible 2006	No	0.26	0.80	1,414,320
	Yes	-1.42	0.86	256,880
FSM eligible 2007	No	0.27	0.78	1,420,690
	Yes	-1.47	0.84	255,660
FSM eligible 2008	No	0.31	0.74	1,101,655
	Yes	-1.57	0.81	186,350
FSM eligible 2009	No	0.31	0.73	1,068,650
	Yes	-1.58	0.80	181,915
FSM eligible 2010	No	0.29	0.73	928,025
	Yes	-1.55	0.82	182,215
FSM eligible 2011	No	0.29	0.73	911,720
	Yes	-1.54	0.83	177,030
FSM eligible 2012	No	0.34	0.71	632,685
	Yes	-1.54	0.82	112,705
FSM eligible 2013	No	0.34	0.71	599,250
	Yes	-1.52	0.83	107,345
FSM eligible 2014	No	0.29	0.72	459,715
	Yes	-1.58	0.81	91,190
FSM eligible 2015	No	0.27	0.73	449,010
	Yes	-1.61	0.81	81,370
FSM eligible 2016	No	0.25	0.75	457,670
	Yes	-1.63	0.82	75,575
FSM eligible 2017	No	0.23	0.77	462,475
	Yes	-1.61	0.84	72,055
FSM eligible 2018	No	0.34	0.73	194,240
	Yes	-1.58	0.88	15,975
Gender	Female	0.00	1.00	817925
	Male	0.01	1.02	853275
Ethnicity	White	0.11	0.98	1372775
	Asian	-0.52	0.95	123620
	Black	-0.83	0.90	67350

	Mixed	-0.33	1.04	51285
	Other	-0.09	0.93	162830
EAL	English	0.08	0.98	1458190
	English as Additional Language	-0.64	0.96	183970

Appendix Figure B1. Distribution of the socio-economic status scale



Appendix C. Full set of results for school level outcomes.

Table C1. School outcomes for initially high-achieving children. Unconditional estimates by socio-economic background for each school cohort.

(a) GCSE total points (standardised)

	Cohort_A	Cohort_B	Cohort_C
1. High achieving, high SES	0.87	0.73	1.07
2. High achieving, Q3 SES	0.88	0.76	1.05
3. High achieving, Q2 SES	0.78	0.68	0.97
4. High achieving, low SES	0.52	0.53	0.64
5. Missing data	-0.78	-0.77	-0.57

(b) In post-16 education

	Cohort_A	Cohort_B	Cohort_C
1. High achieving, high SES	85%	90%	88%
2. High achieving, Q3 SES	86%	90%	85%
3. High achieving, Q2 SES	80%	82%	83%
4. High achieving, low SES	59%	70%	65%
5. Missing data	33%	41%	41%

(c) Key Stage 5 total points (standardised)

	Cohort_A	Cohort_B	Cohort_C
1. High achieving, high SES	0.43	0.46	0.58
2. High achieving, Q3 SES	0.40	0.41	0.48
3. High achieving, Q2 SES	0.25	0.26	0.30
4. High achieving, low SES	0.02	0.10	0.01
5. Missing data	-0.04	0.00	0.21

(d) Achieved at least BBB at A-Level.

	Cohort_A	Cohort_B	Cohort_C
1. High achieving, high SES	40%	46%	44%
2. High achieving, Q3 SES	38%	42%	39%
3. High achieving, Q2 SES	27%	29%	31%
4. High achieving, low SES	13%	18%	18%
5. Missing data	7%	8%	13%

Table C2. Logistic regression estimates of differences in school outcomes across high achieving young people from socio-economically advantaged and disadvantaged backgrounds

(a) GCSE total points (standardised)

	Q2 SES			Q3 SES			Q4 SES		
	Effect size	Lower CI	Upper CI	Effect size	Lower CI	Upper CI	Effect size	Lower CI	Upper CI
Cohort_A	0.22	0.21	0.23	0.30	0.29	0.31	0.29	0.28	0.30
Cohort_B	0.13	0.12	0.15	0.19	0.18	0.20	0.16	0.15	0.17
Cohort_C	0.29	0.28	0.31	0.39	0.38	0.41	0.42	0.41	0.44

(b) In post-16 education

	Q2 SES			Q3 SES			Q4 SES			N
	OR	Lower CI	Upper CI	OR	Lower CI	Upper CI	OR	Lower CI	Upper CI	
Cohort_A	2.34	2.23	2.45	3.39	3.24	3.56	3.00	2.87	3.15	139840
Cohort_B	1.92	1.82	2.03	3.71	3.52	3.92	3.68	3.49	3.89	135400
Cohort_C	2.20	2.08	2.32	2.77	2.62	2.92	3.69	3.49	3.90	119885

(c) Key Stage 5 total points (standardised)

	Q2 SES			Q3 SES			Q4 SES			
	Effect size	Lower CI	Upper CI	Effect size	Lower CI	Upper CI	Effect size	Lower CI	Upper CI	N
Cohort_A	0.17	0.15	0.19	0.29	0.27	0.31	0.32	0.30	0.34	107995
Cohort_B	0.14	0.12	0.16	0.28	0.26	0.30	0.33	0.31	0.35	114245
Cohort_C	0.18	0.16	0.20	0.33	0.31	0.35	0.43	0.41	0.45	98475

(d) Achieved at least BBB at A-Level.

	Q2 SES			Q3 SES			Q4 SES			N
	OR	Lower CI	Upper CI	OR	Lower CI	Upper CI	OR	Lower CI	Upper CI	
Cohort_A	2.04	1.91	2.17	3.21	3.02	3.41	3.66	3.45	3.89	119280
Cohort_B	1.66	1.57	1.77	2.84	2.68	3.00	3.52	3.32	3.72	115730
Cohort_C	1.37	1.29	1.46	1.88	1.77	2.00	2.36	2.22	2.51	96090

Table C3. Logistic regression estimates for school outcomes of high achieving young people from disadvantaged socio-economic backgrounds. Differences by gender and ethnicity.

(a) GCSE total points (standardised)

	Effect size	Lower CI	Upper CI
White girl	0.13	0.11	0.14
Asian boy	0.33	0.30	0.36
Asian girl	0.53	0.50	0.56
Black boy	0.16	0.12	0.20
Black girl	0.32	0.28	0.35
Mixed race boy	-0.01	-0.05	0.04
Mixed race girl	0.16	0.12	0.20
Other ethnicity boy	0.28	0.24	0.33
Other ethnicity girl	0.41	0.37	0.45
High achieving, above average SES	0.40	0.38	0.41
Q3 SES, high achieving	0.36	0.35	0.37
N		**390855**	

(b) In post-16 education

	OR	Lower CI	Upper CI
White girl	1.17	1.11	1.24
Asian boy	3.02	2.71	3.37
Asian girl	4.65	4.10	5.26
Black boy	2.37	2.08	2.70
Black girl	3.70	3.24	4.23
Mixed race boy	1.85	1.59	2.15
Mixed race girl	1.91	1.66	2.20
Other ethnicity boy	1.90	1.62	2.24
Other ethnicity girl	2.44	2.07	2.87
High achieving, above average SES	4.07	3.91	4.24
Q3 SES, high achieving	2.86	2.74	2.99
N		**406565**	

(c) Key Stage 5 total points (standardised)

	Effect size	Lower CI	Upper CI
White girl	0.04	0.01	0.07
Asian boy	0.18	0.14	0.22
Asian girl	0.18	0.14	0.22
Black boy	0.02	-0.03	0.08
Black girl	0.15	0.10	0.20
Mixed race boy	0.05	-0.02	0.12
Mixed race girl	0.08	0.02	0.14
Other ethnicity boy	0.25	0.18	0.32
Other ethnicity girl	0.22	0.16	0.28
High achieving, above average SES	0.38	0.36	0.40
Q3 SES, high achieving	0.23	0.21	0.25
N		**326795**	

(d) Achieved at least BBB at A-Level.

	OR	Lower CI	Upper CI
White girl	1.16	1.07	1.26
Asian boy	2.48	2.21	2.78
Asian girl	2.48	2.23	2.76
Black boy	1.74	1.48	2.04
Black girl	2.31	2.02	2.64
Mixed race boy	1.64	1.34	2.02
Mixed race girl	1.87	1.57	2.23
Other ethnicity boy	2.35	1.96	2.82
Other ethnicity girl	2.35	1.99	2.76
High achieving, above average SES	3.88	3.64	4.14
Q3 SES, high achieving	2.38	2.23	2.54
N		**341225**	

Appendix D. Full set of estimates for university entry by age 21.

Table D1. The percentage of high-achieving young people that enter university by age 21. Unconditional estimates by socio-economic background for each school cohort.

(a) Any university

	Cohort_A	Cohort_B	Cohort_C
1. High achieving, high SES	57	60	75
2. High achieving, Q3 SES	62	64	77
3. High achieving, Q2 SES	53	52	76
4. High achieving, low SES	32	40	57
5. Missing data	18	21	42

(b) Russell Group

	Cohort_A	Cohort_B	Cohort_C
1. High achieving, high SES	23	28	37
2. High achieving, Q3 SES	21	25	33
3. High achieving, Q2 SES	15	16	27
4. High achieving, low SES	7	11	17
5. Missing data	4	5	13

(c) Oxbridge

	Cohort_A	Cohort_B	Cohort_C
1. High achieving, high SES	2	3	3
2. High achieving, Q3 SES	2	2	3
3. High achieving, Q2 SES	1	1	2
4. High achieving, low SES	1%<	1%<	1
5. Missing data	1%<	1%<	1

Notes: Figures refer to the percent of the group that attend university. For instance, in Cohort C, 1% of young people from disadvantaged socio-economic backgrounds with Key Stage 2 test scores in the top quartile go on to attend Oxford or Cambridge university.

Table D2. Logistic regression estimates of high-achieving young people from different socio-economic backgrounds entering university by age 21.

(a) Any university

	Q2 SES			Q3 SES			Q4 SES			N
	OR	Lower CI	Upper CI	OR	Lower CI	Upper CI	OR	Lower CI	Upper CI	
Cohort_A	1.40	1.33	1.48	1.63	1.56	1.71	1.26	1.20	1.32	113645
Cohort_B	1.16	1.11	1.23	1.39	1.32	1.46	1.10	1.05	1.15	117125
Cohort_C	1.30	1.21	1.39	0.98	0.92	1.05	0.67	0.63	0.72	99420

(b) Russell Group

	Q2 SES			Q3 SES			Q4 SES			
	OR	Lower CI	Upper CI	OR	Lower CI	Upper CI	OR	Lower CI	Upper CI	N
Cohort_A	1.15	1.06	1.26	1.30	1.20	1.41	1.39	1.28	1.51	113645
Cohort_B	0.98	0.91	1.06	1.12	1.04	1.21	1.22	1.13	1.31	117125
Cohort_C	0.87	0.81	0.93	0.89	0.83	0.95	0.93	0.87	1.00	99420

(c) Oxbridge

	Q2 SES			Q3 SES			Q4 SES			N
	OR	Lower CI	Upper CI	OR	Lower CI	Upper CI	OR	Lower CI	Upper CI	
Cohort_A	1.44	1.02	2.02	1.46	1.05	2.02	1.56	1.12	2.15	113645
Cohort_B	0.97	0.72	1.31	1.10	0.83	1.44	1.39	1.06	1.83	117125
Cohort_C	0.68	0.54	0.84	0.61	0.49	0.75	0.59	0.48	0.73	99420

Table D3. Logistic regression estimates of high-achieving young people from disadvantaged socio-economic backgrounds entering university by age 21. Differences across genders and ethnicities.

(a) Any university

	OR	Lower CI	Upper CI
White girl	1.17	1.10	1.25
Asian boy	2.06	1.84	2.29
Asian girl	2.60	2.33	2.89
Black boy	2.43	2.10	2.80
Black girl	2.98	2.61	3.40
Mixed race boy	1.69	1.42	2.01
Mixed race girl	1.60	1.37	1.86
Other ethnicity boy	1.98	1.65	2.37
Other ethnicity girl	2.21	1.86	2.61
High achieving, above average SES	1.70	1.61	1.79
Q3 SES, high achieving	1.88	1.78	1.98
N		**340440**	

(b) Russell Group

	OR	Lower CI	Upper CI
White girl	0.81	0.72	0.91
Asian boy	1.46	1.26	1.69
Asian girl	1.61	1.41	1.85
Black boy	1.39	1.14	1.71
Black girl	1.61	1.36	1.89
Mixed race boy	1.50	1.17	1.94
Mixed race girl	1.35	1.09	1.68
Other ethnicity boy	1.49	1.19	1.88
Other ethnicity girl	1.46	1.19	1.79
High achieving, above average SES	1.32	1.21	1.44
Q3 SES, high achieving	1.20	1.10	1.31
N		**340440**	

(c) Oxbridge

	OR	Lower CI	Upper CI
White girl	1.14	0.77	1.69
Asian boy	0.52	0.29	0.94
Asian girl	0.80	0.48	1.33
Black boy	1.77	0.94	3.34
Black girl	1.14	0.61	2.13
Mixed race boy	0.87	0.38	1.99
Mixed race girl	1.16	0.54	2.50
Other ethnicity boy	1.21	0.65	2.24
Other ethnicity girl	0.99	0.52	1.89
High achieving, above average SES	0.98	0.73	1.32
Q3 SES, high achieving	0.93	0.69	1.26
N		**340440**	

Appendix E. Full set of estimates for university entry (any age / record)

Table E1. The percentage of high-achieving young people that enter university. Unconditional estimates by socio-economic background for each school cohort.

(a) Any university

	Cohort_A	Cohort_B	Cohort_C
1. High achieving, high SES	73	76	75
2. High achieving, Q3 SES	8	82	77
3. High achieving, Q2 SES	72	71	76
4. High achieving, low SES	48	57	57
5. Missing data	31	34	42

(b) Russell Group

	Cohort_A	Cohort_B	Cohort_C
1. High achieving, high SES	28	34	37
2. High achieving, Q3 SES	26	31	33
3. High achieving, Q2 SES	18	21	27
4. High achieving, low SES	9	14	17
5. Missing data	5	7	13

(c) Oxbridge

	Cohort_A	Cohort_B	Cohort_C
1. High achieving, high SES	2	3	3
2. High achieving, Q3 SES	2	2	2
3. High achieving, Q2 SES	1	1	2
4. High achieving, low SES	1%<	1	1
5. Missing data	1%<	1%<	1

Table E2. Logistic regression estimates of high-achieving young people from different socio-economic backgrounds entering university.

(a) Any university

	Q2 SES			Q3 SES			Q4 SES			
	OR	Lower CI	Upper CI	OR	Lower CI	Upper CI	OR	Lower CI	Upper CI	N
Cohort_A	1.68	1.58	1.79	2.17	2.05	2.31	1.10	1.04	1.16	113645
Cohort_B	1.21	1.13	1.29	1.53	1.44	1.62	0.82	0.77	0.87	117125
Cohort_C	1.30	1.21	1.39	0.98	0.92	1.05	0.67	0.63	0.72	99420

(b) Russell Group

	Q2 SES			Q3 SES			Q4 SES			N
	OR	Lower CI	Upper CI	OR	Lower CI	Upper CI	OR	Lower CI	Upper CI	
Cohort_A	1.13	1.04	1.23	1.27	1.18	1.37	1.40	1.30	1.51	113645
Cohort_B	0.94	0.87	1.01	1.07	0.99	1.15	1.13	1.05	1.22	117125
Cohort_C	0.87	0.81	0.93	0.89	0.83	0.95	0.93	0.87	1.00	99420

(c) Oxbridge

	Q2 SES			Q3 SES			Q4 SES			
	OR	Lower CI	Upper CI	OR	Lower CI	Upper CI	OR	Lower CI	Upper CI	N
Cohort_A	1.44	1.04	1.98	1.44	1.06	1.95	1.56	1.15	2.12	113645
Cohort_B	0.97	0.74	1.28	1.05	0.82	1.36	1.32	1.03	1.70	117125
Cohort_C	0.68	0.54	0.84	0.61	0.49	0.75	0.59	0.48	0.73	99420

Table E3. Logistic regression estimates of high-achieving young people from disadvantaged socio-economic backgrounds entering university. Differences across genders and ethnicities.

(a) Any university

	OR	Lower CI	Upper CI
White girl	1.14	1.06	1.22
Asian boy	3.32	2.88	3.82
Asian girl	3.88	3.36	4.48
Black boy	3.46	2.90	4.13
Black girl	4.12	3.47	4.89
Mixed race boy	1.86	1.52	2.26
Mixed race girl	2.12	1.76	2.54
Other ethnicity boy	2.90	2.30	3.66
Other ethnicity girl	2.95	2.37	3.67
High achieving, above average SES	1.74	1.65	1.84
Q3 SES, high achieving	2.14	2.03	2.27
N		**340440**	

(b) Russell Group

	OR	Lower CI	Upper CI
White girl	0.77	0.69	0.86
Asian boy	1.62	1.40	1.87
Asian girl	1.65	1.44	1.88
Black boy	1.44	1.18	1.76
Black girl	1.54	1.31	1.82
Mixed race boy	1.49	1.16	1.92
Mixed race girl	1.48	1.20	1.83
Other ethnicity boy	1.52	1.21	1.92
Other ethnicity girl	1.48	1.21	1.81
High achieving, above average SES	1.30	1.19	1.41
Q3 SES, high achieving	1.16	1.06	1.26
N		**340440**	

(c) Oxbridge

	OR	Lower CI	Upper CI
White girl	1.13	0.78	1.65
Asian boy	0.48	0.27	0.84
Asian girl	0.82	0.50	1.32
Black boy	1.79	0.97	3.31
Black girl	1.21	0.67	2.18
Mixed race boy	0.92	0.42	2.00
Mixed race girl	1.35	0.66	2.75
Other ethnicity boy	1.11	0.60	2.03
Other ethnicity girl	0.89	0.47	1.69
High achieving, above average SES	0.98	0.74	1.31
Q3 SES, high achieving	0.93	0.70	1.25
N		**340440**	

Appendix F. All results for living at home while an undergraduate

Table F1. The propensity for high-achieving young people from different socio-economic backgrounds living at home as an undergraduate. Unconditional percentages.

	Cohort_A	Cohort_B	Cohort_C
1. High achieving, high SES	22	19	20
2. High achieving, Q3 SES	28	26	26
3. High achieving, Q2 SES	39	38	42
4. High achieving, low SES	54	50	55
5. Missing data	48	46	40

Table F2. The propensity for high-achieving young people from different socio-economic backgrounds living at home as an undergraduate. Logistic regression model estimates (log-odds)

Log-odds	Unconditional		Conditional		University Fixed Effects	
	Beta	SE	Beta	SE	Beta	SE
2nd SES quartile	-0.547	0.017	-0.400	0.020	-0.329	0.022
3rd SES quartile	-1.129	0.017	-0.777	0.019	-0.636	0.021
top SES quartile	-1.478	0.017	-0.995	0.020	-0.790	0.022
N	**242845**		**223870**		**223870**	

Table F3. The propensity for high-achieving young people from different socio-economic backgrounds living at home as an undergraduate. Logistic regression model estimates across cohorts.

	Q2 SES			Q3 SES			Q4 SES			
	OR	Lower CI	Upper CI	OR	Lower CI	Upper CI	OR	Lower CI	Upper CI	N
Cohort_A	0.62	0.58	0.67	0.46	0.43	0.49	0.38	0.36	0.41	77655
Cohort_B	0.72	0.68	0.78	0.50	0.47	0.53	0.40	0.37	0.42	78685
Cohort_C	0.64	0.60	0.68	0.41	0.38	0.44	0.34	0.32	0.37	67530

Table F4. Logistic regression estimates of high-achieving young people from disadvantaged socio-economic backgrounds living at home as an undergraduate. Differences across genders and ethnicities.

	OR	Lower CI	Upper CI
White girl	1.10	1.00	1.22
Asian boy	3.24	2.81	3.75
Asian girl	4.36	3.79	5.02
Black boy	1.45	1.22	1.72
Black girl	1.17	1.00	1.36
Mixed race boy	1.06	0.83	1.34
Mixed race girl	0.99	0.80	1.21
Other ethnicity boy	1.49	1.20	1.85
Other ethnicity girl	1.57	1.29	1.90
High achieving, above average SES	0.61	0.57	0.66
Q3 SES, high achieving	1.00	0.93	1.09
N		**231430**	

Appendix G. All results for university outcomes.

Table G1. University outcomes for initially high-achieving young people different socio-economic backgrounds. Unconditional percentages.

(a) Complete degree (any age/record)

	Cohort_A	Cohort_B	Cohort_C
1. High achieving, high SES	67	71	30
2. High achieving, Q3 SES	74	75	31
3. High achieving, Q2 SES	63	62	30
4. High achieving, low SES	40	47	20
5. Missing data	26	28	13

(b) Complete degree by age 21

	Cohort_A	Cohort_B	Cohort_C
1. High achieving, high SES	28	30	30
2. High achieving, Q3 SES	31	32	31
3. High achieving, Q2 SES	25	26	30
4. High achieving, low SES	15	18	20
5. Missing data	7	8	13

(c) Obtain at least a 2:1

	Cohort_A	Cohort_B	Cohort_C
1. High achieving, high SES	52	58	27
2. High achieving, Q3 SES	55	61	27
3. High achieving, Q2 SES	45	47	24
4. High achieving, low SES	25	33	15
5. Missing data	14	18	10

(d) Obtain a 1st

	Cohort_A	Cohort_B	Cohort_C
1. High achieving, high SES	17	24	12
2. High achieving, Q3 SES	18	24	11
3. High achieving, Q2 SES	13	17	10
4. High achieving, low SES	7	11	5
5. Missing data	3	6	4

Table G2. University outcomes for initially high-achieving young people different socio-economic backgrounds. Logistic regression estimates.

(a) Complete degree (any age/record)

	Q2 SES			Q3 SES			Q4 SES			N
	OR	Lower CI	Upper CI	OR	Lower CI	Upper CI	OR	Lower CI	Upper CI	
Cohort_A	1.23	1.13	1.34	1.46	1.35	1.59	1.64	1.51	1.79	95410
Cohort_B	1.37	1.25	1.49	1.84	1.69	2.00	2.09	1.91	2.28	96990
Cohort_C	1.13	1.07	1.20	1.11	1.05	1.18	1.10	1.04	1.17	82435

(b) Complete degree by age 21

	Q2 SES			Q3 SES			Q4 SES			N
	OR	Lower CI	Upper CI	OR	Lower CI	Upper CI	OR	Lower CI	Upper CI	
Cohort_A	1.09	1.03	1.16	1.16	1.10	1.23	1.15	1.08	1.21	95410
Cohort_B	1.11	1.05	1.18	1.19	1.12	1.26	1.16	1.09	1.23	96990
Cohort_C	1.13	1.07	1.20	1.11	1.05	1.18	1.10	1.04	1.17	82435

(c) Obtain at least a 2:1

	Q2 SES			Q3 SES			Q4 SES			N
	OR	Lower CI	Upper CI	OR	Lower CI	Upper CI	OR	Lower CI	Upper CI	
Cohort_A	1.22	1.15	1.29	1.39	1.31	1.47	1.50	1.41	1.59	95410
Cohort_B	1.24	1.17	1.32	1.54	1.45	1.63	1.66	1.56	1.76	96990
Cohort_C	1.23	1.15	1.31	1.27	1.19	1.35	1.28	1.20	1.37	82435

(d) Obtain a 1st

	Q2 SES			Q3 SES			Q4 SES			N
	OR	Lower CI	Upper CI	OR	Lower CI	Upper CI	OR	Lower CI	Upper CI	
Cohort_A	1.13	1.05	1.23	1.23	1.14	1.33	1.24	1.15	1.34	95410
Cohort_B	1.13	1.05	1.21	1.29	1.21	1.38	1.33	1.24	1.42	96990
Cohort_C	1.31	1.19	1.44	1.35	1.23	1.48	1.41	1.28	1.55	82435

Notes: Sample restricted to young people with Key Stage 2 scores in the top quartile, with academic achievement measures available through to age 18 and who started an undergraduate degree. Estimates based on a logistic regression model controlling for Key Stage 2, GCSE and Key Stage 5 point scores and best three A-Level grades achieved. Values greater than one indicate that young people from the most advantaged socio-economic backgrounds are more likely to complete an undergraduate degree or obtain a 2:1 than the most disadvantaged group. * indicates a statistically significant difference at the 5% level.

Table G3. University outcomes for initially high-achieving young people different socio-economic backgrounds. Logistic regression estimates – with and without university fixed effects. Log-odds.

(a) Complete degree (any age/record)

Log-odds	Unconditional		Conditional		University Fixed Effects	
	Beta	SE	Beta	SE	Beta	SE
2nd SES quartile	0.665	0.012	0.068	0.019	0.068	0.019
3rd SES quartile	1.061	0.011	0.244	0.018	0.261	0.019
top SES quartile	0.883	0.011	0.182	0.019	0.221	0.019
N	**406570**		**274835**		**274835**	

(b) Complete degree by age 21

Log-odds	Unconditional		Conditional		University Fixed Effects	
	Beta	SE	Beta	SE	Beta	SE
2nd SES quartile	0.559	0.015	0.112	0.018	0.112	0.018
3rd SES quartile	0.771	0.014	0.150	0.017	0.158	0.017
top SES quartile	0.678	0.014	0.139	0.017	0.159	0.018
N	**406570**		**274835**		**274835**	

(c) Obtain at least a 2:1

Log-odds	Unconditional		Conditional		University Fixed Effects	
	Beta	SE	Beta	SE	Beta	SE
2nd SES quartile	0.684	0.013	0.158	0.017	0.160	0.017
3rd SES quartile	1.093	0.012	0.327	0.017	0.338	0.017
top SES quartile	1.005	0.012	0.320	0.017	0.344	0.017
N	**406570**		**274835**		**274835**	

(d) Obtain a 1st

Log-odds	Unconditional		Conditional		University Fixed Effects	
	Beta	SE	Beta	SE	Beta	SE
2nd SES quartile	0.636	0.021	0.143	0.023	0.166	0.023
3rd SES quartile	0.992	0.020	0.259	0.022	0.309	0.023
top SES quartile	0.967	0.020	0.249	0.022	0.324	0.023
N	**406570**		**274835**		**274835**	

Table G4. University outcomes for initially high-achieving young people disadvantaged socio-economic backgrounds. Differences across genders and ethnicity.

(a) Complete degree (any age/record)

	OR	Lower CI	Upper CI
White girl	1.12	1.03	1.23
Asian boy	1.11	0.99	1.26
Asian girl	1.22	1.09	1.37
Black boy	0.60	0.52	0.70
Black girl	0.76	0.66	0.86
Mixed race boy	0.71	0.58	0.86
Mixed race girl	0.97	0.81	1.15
Other ethnicity boy	0.97	0.80	1.18
Other ethnicity girl	1.17	0.98	1.40
High achieving, above average SES	1.37	1.28	1.46
Q3 SES, high achieving	1.07	1.00	1.14
N		**283975**	

(b) Complete degree by age 21

	OR	Lower CI	Upper CI
White girl	1.34	1.23	1.45
Asian boy	1.01	0.89	1.13
Asian girl	1.45	1.31	1.62
Black boy	0.76	0.65	0.90
Black girl	1.16	1.02	1.32
Mixed race boy	0.92	0.74	1.13
Mixed race girl	1.00	0.84	1.19
Other ethnicity boy	0.98	0.81	1.19
Other ethnicity girl	1.48	1.26	1.74
High achieving, above average SES	1.38	1.29	1.47
Q3 SES, high achieving	1.30	1.22	1.39
N		**283975**	

(c) Obtain at least a 2:1

	OR	Lower CI	Upper CI
White girl	1.28	1.18	1.38
Asian boy	0.94	0.84	1.05
Asian girl	1.15	1.04	1.27
Black boy	0.51	0.44	0.59
Black girl	0.85	0.75	0.96
Mixed race boy	0.68	0.56	0.82
Mixed race girl	1.10	0.93	1.29
Other ethnicity boy	0.89	0.74	1.06
Other ethnicity girl	1.16	0.99	1.35
High achieving, above average SES	1.63	1.53	1.73
Q3 SES, high achieving	1.25	1.17	1.33
N		**283975**	

(d) Obtain a 1st

	OR	Lower CI	Upper CI
White girl	1.03	0.92	1.15
Asian boy	0.85	0.73	0.99
Asian girl	0.88	0.76	1.01
Black boy	0.46	0.36	0.59
Black girl	0.62	0.51	0.75
Mixed race boy	0.80	0.61	1.06
Mixed race girl	0.86	0.68	1.08
Other ethnicity boy	0.81	0.63	1.04
Other ethnicity girl	0.91	0.74	1.13
High achieving, above average SES	1.42	1.31	1.54
Q3 SES, high achieving	1.12	1.03	1.22
N		**283975**	

Notes: Sample restricted to young people with Key Stage 2 scores in the top quartile from disadvantaged socio-economic backgrounds, with academic achievement measures available through to age 18 and who started an undergraduate degree. Estimates based on a logistic regression model controlling for Key Stage 2, GCSE and Key Stage 5 point scores and best three A-Level grades achieved. Odds ratios greater than one indicate that the group was more likely to achieve the outcome than high achieving disadvantaged White boys as the reference group. * indicates a statistically significant difference at the 5% level.

Appendix H. All results for postgraduate entry.

Table H1. Postgraduate entry for initially high-achieving young people different socio-economic backgrounds. Unconditional percentages.

(a) Entry into MSc

	Cohort_A	Cohort_B
1. High achieving, high SES	19	19
2. High achieving, Q3 SES	21	20
3. High achieving, Q2 SES	18	17
4. High achieving, low SES	10	13
5. Missing data	7	7

(b) Entry into MSc by age 27

	Cohort_A	Cohort_B
1. High achieving, high SES	14	19
2. High achieving, Q3 SES	16	20
3. High achieving, Q2 SES	13	17
4. High achieving, low SES	7	13
5. Missing data	5	7

(c) Entry into PhD

	Cohort_A	Cohort_B
1. High achieving, high SES	4	4
2. High achieving, Q3 SES	4	4
3. High achieving, Q2 SES	3	3
4. High achieving, low SES	1	2
5. Missing data	1%<	1%<

(d) Entry into PhD by age 27

	Cohort_A	Cohort_B
1. High achieving, high SES	3	4
2. High achieving, Q3 SES	3	4
3. High achieving, Q2 SES	2	3
4. High achieving, low SES	1%<	2
5. Missing data	1%<	1%<

Table H2. Entry into postgraduate study for initially high-achieving young people different socio-economic backgrounds.

(a) Entry into MSc

	Q2 SES			Q3 SES			Q4 SES			N
	OR	Lower CI	Upper CI	OR	Lower CI	Upper CI	OR	Lower CI	Upper CI	
Cohort_A	1.08	1.01	1.16	1.05	0.99	1.13	0.99	0.93	1.06	87040
Cohort_B	0.95	0.89	1.02	0.92	0.86	0.98	0.85	0.80	0.91	89255

(b) Entry into MSc by age 27

	Q2 SES			Q3 SES			Q4 SES			N
	OR	Lower CI	Upper CI	OR	Lower CI	Upper CI	OR	Lower CI	Upper CI	
Cohort_A	1.03	0.95	1.11	1.04	0.96	1.12	1.02	0.95	1.10	87040
Cohort_B	0.95	0.89	1.02	0.92	0.86	0.98	0.85	0.80	0.91	89255

(c) Entry into PhD

	Q2 SES			Q3 SES			Q4 SES			N
	OR	Lower CI	Upper CI	OR	Lower CI	Upper CI	OR	Lower CI	Upper CI	
Cohort_A	1.04	0.88	1.23	0.95	0.81	1.12	0.91	0.77	1.07	87040
Cohort_B	0.93	0.79	1.11	0.90	0.77	1.06	0.86	0.73	1.01	89255

(d) Entry into PhD by age 27

	Q2 SES			Q3 SES			Q4 SES			N
	OR	Lower CI	Upper CI	OR	Lower CI	Upper CI	OR	Lower CI	Upper CI	
Cohort_A	1.11	0.90	1.38	1.05	0.85	1.28	1.03	0.84	1.26	87040
Cohort_B	0.93	0.79	1.11	0.90	0.77	1.06	0.86	0.73	1.01	89255

Notes: Sample restricted to young people with Key Stage 2 scores in the top quartile, with academic achievement measures available through to age 18 and who started and completed an undergraduate degree. Estimates based on a logistic regression model controlling for Key Stage 2, GCSE and Key Stage 5 point scores, best three A-Level grades achieved, whether they completed an undergraduate at a Russell Group or Oxbridge University and degree classification obtained. Values less than one indicate that young people from the most advantaged socio-economic backgrounds are less likely to complete a postgraduate degree than the most disadvantaged group. * indicates a statistically significant difference at the 5% level.

Table H3. Entry into postgraduate study for initially high-achieving young people disadvantaged socio-economic backgrounds. Differences across genders and ethnicity.

(a) Entry into MSc

	OR	Lower CI	Upper CI
White girl	1.35	1.20	1.52
Asian boy	1.17	0.99	1.38
Asian girl	1.47	1.27	1.71
Black boy	0.95	0.73	1.23
Black girl	1.52	1.26	1.83
Mixed race boy	1.29	0.96	1.74
Mixed race girl	1.69	1.34	2.13
Other ethnicity boy	1.17	0.90	1.53
Other ethnicity girl	1.21	0.95	1.54
High achieving, above average SES	1.18	1.07	1.29
Q3 SES, high achieving	1.27	1.15	1.40
N		**179465**	

(b) Entry into MSc by age 27

	OR	Lower CI	Upper CI
White girl	1.33	1.18	1.51
Asian boy	1.20	1.01	1.43
Asian girl	1.55	1.33	1.82
Black boy	1.10	0.84	1.43
Black girl	1.54	1.26	1.87
Mixed race boy	1.20	0.87	1.65
Mixed race girl	1.67	1.31	2.13
Other ethnicity boy	1.22	0.92	1.62
Other ethnicity girl	1.30	1.01	1.67
High achieving, above average SES	1.19	1.08	1.32
Q3 SES, high achieving	1.23	1.11	1.37
N		**179465**	

(c) Entry into PhD

	OR	Lower CI	Upper CI
White girl	0.56	0.43	0.72
Asian boy	0.29	0.18	0.46
Asian girl	0.43	0.30	0.63
Black boy	0.31	0.13	0.70
Black girl	0.42	0.24	0.73
Mixed race boy	0.59	0.29	1.19
Mixed race girl	0.53	0.29	0.98
Other ethnicity boy	0.65	0.36	1.18
Other ethnicity girl	0.55	0.32	0.96
High achieving, above average SES	0.61	0.51	0.73
Q3 SES, high achieving	0.63	0.53	0.76
N		**179465**	

(d) Entry into PhD by age 27

	OR	Lower CI	Upper CI
White girl	0.56	0.42	0.74
Asian boy	0.26	0.15	0.45
Asian girl	0.48	0.32	0.71
Black boy	0.31	0.13	0.78
Black girl	0.41	0.22	0.77
Mixed race boy	0.64	0.30	1.35
Mixed race girl	0.60	0.32	1.15
Other ethnicity boy	0.60	0.31	1.18
Other ethnicity girl	0.49	0.26	0.93
High achieving, above average SES	0.64	0.53	0.77
Q3 SES, high achieving	0.64	0.52	0.78
N		**179465**	

Notes: Sample restricted to young people with Key Stage 2 scores in the top quartile from disadvantaged socio-economic backgrounds, with academic achievement measures available through to age 18 and who started and completed an undergraduate degree. Estimates based on a logistic regression model controlling for Key Stage 2, GCSE and Key Stage 5 point scores, best three A-Level grades achieved, whether they completed an undergraduate at a Russell Group or Oxbridge University and degree classification obtained. Odds ratios greater than one indicate that the group was more likely to enter postgraduate study than high achieving disadvantaged White boys as the reference group. * indicates a statistically significant difference at the 5% level.