When and why do initially high-achieving poor children fall behind?

Crawford, Claire; Macmillan, Lindsey; Vignoles, Anna

DOI:
10.1080/03054985.2016.1240672

Document Version
Peer reviewed version

Citation for published version (Harvard):
Crawford, C, Macmillan, L & Vignoles, A 2016, 'When and why do initially high-achieving poor children fall behind?', *Oxford Review of Education*, vol. 43, no. 1, pp. 88-108.
https://doi.org/10.1080/03054985.2016.1240672

When and why do initially high achieving poor children fall behind?[1]

Claire Crawford*, Lindsey Macmillan** and Anna Vignoles***

** Department of Economics, University of Warwick and Institute for Fiscal Studies*

*** Department of Social Science, UCL Institute of Education*

**** Faculty of Education, University of Cambridge*

June 2016

Abstract

In this paper, we examine the trajectories of initially higher- and lower-achieving children from lower and higher socio-economic status families from primary school through to university in England for the first time. We also explore what explains these trajectories. This enables us to provide new insights into when and why the performance of children with similar initial achievement diverges on the basis of their socio-economic background. Our results indicate that pupils from poor backgrounds who are higher achievers in primary school fall behind their better-off but lower achieving peers during secondary school. This suggests that secondary school may be a critical period to intervene to prevent poor children from falling behind their richer peers. Our analysis suggests that there is less divergence in performance between pupils from different socio-economic backgrounds who attend the same schools. This result is particularly strong for children with low initial achievement. While we remain cautious about the implications of these findings, they provide suggestive evidence that schools (or the sorting of pupils into schools) play an important role in explaining why the test scores of richer and poorer children diverge over time.

Keywords: social mobility, education achievement, regression to the mean

Acknowledgements:

This project is part of the Social Policy in a Cold Climate programme of work which is funded by Joseph Rowntree Foundation, the Nuffield Foundation and Trust for London. It builds on earlier work funded by the Social Mobility and Child Poverty Commission. The authors would like to thank Ruth Lupton and John Hills for useful comments on an earlier draft. All errors remain the responsibility of the authors.

[1] Corresponding author: Dr Lindsey Macmillan, UCL Institute of Education, 20 Bedford Way, London, WC1H OAL (l.macmilllan@ucl.ac.uk)

1. Introduction

The role of education as a potential driver of social mobility has been well established in both the theoretical (Blau and Duncan, 1967; Becker and Tomes, 1986) and empirical literature (Atkinson, 1980; Atkinson and Jenkins, 1984; Breen and Goldthorpe, 2001; Breen and Jonsson, 2007; Blanden et. al, 2007) across disciplines over the past fifty years. Many view reducing educational inequality as a key policy lever for improving levels of social mobility.[2] This is certainly true in the UK, where the Government now actively tracks levels of educational inequality across the life course as a proxy for longer term trends in social mobility (Cabinet Office, 2011).

Of particular importance from a policy perspective is whether educational inequalities increase as children get older, as the existence of substantial inequalities at the end of compulsory schooling may be detrimental to future social mobility. However, much of the existing evidence focuses on cross-sectional trends in achievement gaps across cohorts (Blanden and Gregg 2004; Strand, 2014a), which conflate changes in achievement as children grow older with changes in achievement over time (amongst different cohorts). While there has been a growing literature on the trajectories of different groups of children and the value-added of different schools for different types of student (Strand et al. 2006; Strand, 2014b; Thomas et al. 1997a, 1997b; Strand, 1997, 2010) there is limited evidence to date on the trajectories of educational achievement for the same individuals over time from different family backgrounds.

Within this literature, an important issue is the extent to which initially higher achieving poor children fall behind their better-off peers, and what, if anything, can be done to mitigate these patterns. This issue has received less attention to date, and the existing evidence provides little consensus. For example, a seminal paper by Feinstein (2003) for the UK found that high-achieving children from low income families fell behind low-achieving children from high

[2] Although we note that this claim is disputed by some (e.g. Goldthorpe, 2013).

income families at a very early age.[3] However, Jerrim and Vignoles (2013) demonstrated that these findings could be at least partially driven by measurement error in the initial achievement level, which was also used to plot the trajectories over time. This can lead to a statistical phenomenon known as "regression to the mean", which can arise when some particularly high (or particularly low) initial scores are driven by 'luck' rather than reflecting the individual's underlying potential. When replicating the Feinstein (2003) analysis to account for this issue, Jerrim and Vignoles (2013) found little evidence of a convergence in performance between poor children with high initial development and richer children with lower initial development.

Both studies are limited by only focusing up to age 7, however; and neither explore what factors might drive the differential trajectories that they observe between children from different socio-economic backgrounds. This paper seeks to add to the literature in these two respects. We start by examining the trajectories of initially high- and low-achieving children from lower and higher socio-economic status families from primary school to university, for a cohort of pupils born in England between September 1990 and August 1991. We use tests in different subjects taken at the same age to minimise the impact of regression to the mean on our results. We also try to account for what might be driving the differential trajectories by initial achievement and family background using basic demographic characteristics and information on the schools children attend.

In line with previous literature, we find large differences in educational achievement by socio-economic background at age 7 that increase as pupils get older - by around two thirds by the end of compulsory schooling (age 16). When looking at trajectories by socio-economic background and initial achievement, we find that the performance of initially high achieving children from the most deprived families and that of average-achieving affluent children

[3] Other studies using comparable approaches have found similar results (e.g. Schoon 2006; Feinstein 2003, 2004; Blanden and Machin 2007, 2010; Parsons et al 2011).

converge, but this occurs somewhat later than previous literature has suggested, namely between ages 11 and 14, when most young people in England are at the start of secondary school. We find a similar pattern of convergence between initially low-achieving children from the least deprived families and average-achieving poorer children. These findings are robust to alternative definitions of initial achievement, suggesting that this convergence is not driven by regression to the mean.

When exploring the role of other factors in explaining these trends, we find that the schools attended by pupils from different backgrounds may be important. In particular, we find less divergence in performance between pupils of similar initial achievement but from different socio-economic backgrounds when we compare pupils attending the same schools. (There is also less convergence between richer initially lower-achieving pupils and poorer initially higher-achieving pupils.) However, we are cautious about ascribing a causal interpretation to these findings, as the direction of causality is not obvious and the sorting of pupils into schools, rather than school quality, may partly explain these patterns.

This paper now proceeds as follows: the next section reviews the related literature; Section 3 outlines the data that we use and Section 4 our empirical approach. The main results are discussed in Section 5 and we end in Section 6 with some brief conclusions.

2. Related literature

Large socio-economic gaps in children's cognitive skills at an early age are well documented in the literature (Cunha et al 2006, Goodman et al 2009). These findings are consistent with both theoretical and empirical work which has indicated that the early years are particularly critical in terms of children's cognitive development (Cunha et al 2006). A key policy question is whether these differences are narrowed by the school system or if instead socio-economic gaps widen further as children progress through school.

Theoretically there is reason to believe that socio-economic differences in cognitive achievement might widen through time if the greater levels of investment made by parents of higher socio-economic status (SES) in their children in turn enable high SES children to benefit more from later investments, such as schooling. In other words, if inputs into children's development are complementary, then the cognitive skills gap between richer and poorer children is likely to widen through time – in the words of Heckman and co-authors, skill begets skill (see Cunha et al., 2006). The empirical evidence on this point is somewhat mixed. Some studies have found a widening of the socio-economic gap in cognitive achievement as children get older (Caro, 2009; Feinstein, 2003; Goodman et al, 2009). Other studies (Blanden and Machin, 2010, Reardon, 2011, Duncan and Magnuson 2011, Cunha et al 2006) have found little change in the magnitude of the SES gap across childhood.

This literature has largely focused on average differences in cognitive skill by social background between different cohorts at different ages. But such analysis may conflate changes that occur as children get older with changes that occur over time/across cohort. To better understand how education performance changes for particular children, one must rely on longitudinal or panel data, following the same individuals as they get older, collecting information on achievement at multiple time points. Goodman and Gregg (2010) piece together within-child changes from three different cohorts, representing an initial step in this direction using UK data.[4] They find that the differences in achievement by socio-economic background start large (around 23 percentiles at age 3) and widen up to age 14 (36 percentiles). Similarly, Washbrook and Lee (2015) using the US Early Childhood Longitudinal Survey – Kindergarten cohort and Caro (2009) using data from the Canadian National Longitudinal Study of Children and Youth find evidence of widening socio-economic gaps in achievement as children get

[4] Specifically, they use a cohort of children born in 2000-01 to illustrate changes in achievement by socio-economic background between ages 3 and 7, a cohort born in 1991-92 to show changes between ages 7 and 11, and a cohort born in 1989-90 to show changes between ages 11 and 16.

older. Cunha et al (2006), by contrast, using the US Children of the National Longitudinal Survey of Youth (CNLSY), find large achievement gaps at age 6 that remain broadly stable up to age 12 (see also Goodman et al., 2011).

Another way to consider whether socio-economic gaps have widened during a particular period of schooling is to use a value-added model, in which one would model achievement at time period 2, controlling for achievement at time period 1. There is a large body of work that has taken this approach, though such studies often constrain prior attainment to have the same effect throughout the distribution, and very rarely interact the effect of attainment with socio-economic background. (The set of studies referred to in Goodman and Gregg (2010) are notable exceptions, however; see also Strand (2014b) who focuses on ethnic differences in achievement.) Of particular interest is whether the academic performance of children from poorer backgrounds who start out with higher levels of cognitive skill declines relative to their richer (but lower achieving) counterparts. The few UK papers that have examined this issue largely concur that high achieving children from poorer backgrounds do fall behind relative to their richer lower-achieving peers (Schoon 2006; Feinstein 2003, 2004; Blanden and Machin 2007, 2010; Parsons et al 2011; amongst others).

But, as outlined above, there is a concern that at least part of the story highlighted by these papers might be driven by the problem of regression to the mean. This issue was first identified by Galton (1886), and an increasing number of papers are paying attention to the associated methodological challenges (e.g. Jerrim and Vignoles (2013), Washbrook and Lee (2015)). Using simulations, Jerrim and Vignoles (2013) attempt to estimate the extent of the bias in estimates of the educational achievement trajectories of children from different socio-economic backgrounds and achievement groups that may arise if one does not account for the possibility of regression to the mean, and showed that apparently substantial declines in test

scores for poor high achieving children can occur, even when no real change in achievement is taking place. They attribute this to regression to the mean.

To overcome this issue, Jerrim and Vignoles adopt a standard approach in this literature – namely to define initial achievement using a different test to the one that is used as the baseline from which to measure a child's education trajectory. In doing so, they find little evidence that the cognitive skills of initially high achieving children from lower socio-economic backgrounds suffer a significant decline between the ages of 3 and 7. They are not, however, able to follow children to the end of primary school and beyond. Washbrook and Lee (2015) adopt an analogous approach within a value-added context, using a test score from an earlier time period (age 5 in their case) as an instrument for prior achievement. They find evidence of a significant widening of educational inequality between ages 6 and 14 between those measured to have the same initial achievement, but they do not differentiate between the patterns found for individuals at particular parts of the distribution of prior achievement.

This study builds on the literature in this area, using a census of children attending state schools in England and the methods adopted by Jerrim and Vignoles (2013) to demonstrate how educational trajectories differ by socio-economic background and initial achievement between primary school and university. It also explores the drivers of these trajectories. This has not been done before for the same individuals over time and split by initial achievement as well as family background.

3. Data

Given that our analysis is based on the educational careers of students in England, we start by providing a brief overview of the English education system. Pupils in England generally start school in the academic year (September to August) in which they turn five. For most pupils, this means that they start school in the September after their fourth birthday. The first three

years are spent in Key Stage 1, with a further four years spent working through Key Stage 2. This takes pupils to the end of primary school. In the academic year in which pupils' turn 11 they make the transition from primary to secondary school, where attendance is compulsory up until the end of Key Stage 4 (taken at the end of the academic year in which they turn 16). Key Stage 3 usually runs for the first three years of secondary school and Key Stage 4 for the final two years. The vast majority of students start school at the expected time and progress through the system as expected, with very few held back or advanced a year. The compulsory school leaving age was 16 for our cohort. Students who choose to stay on beyond compulsory this point generally study for a further two years (known as Key Stage 5). Thereafter, students can enter university if they choose to do so.

National achievement tests are taken at the end of each Key Stage by all pupils in state schools. At the end of Key Stage 1 (the academic year in which pupils turn 7), they are tested in reading, writing and maths. These tests were introduced in 1997-98. For the cohort that we analyse, they were externally marked (although performance at the end of Key Stage 1 is now teacher assessed). At the end of Key Stage 2 (the academic year in which pupils turn 11), pupils are tested in English, maths and science. These tests were introduced in 1994-95 and have always been externally marked. Up until 2009 there were compulsory national achievement tests at the end of Key Stage 3 in English, Maths and Science which were externally marked. At the end of Key Stage 4, almost all pupils (including those attending private schools) take public exams, General Certificates in Secondary Education (GCSEs) or equivalent qualifications, which largely determine their participation in post-compulsory schooling. Around two thirds of students reach the benchmark of 5 A*-C grades in GCSEs or equivalents. The majority of pupils stay in education beyond age 16, with approximately 60% achieving two or more Advanced level (A level) qualifications by age 19 in 2013 and approximately 35% entering a higher education institution at age 18 or 19 (Crawford, 2014).

To measure trajectories across individuals' educational careers we require longitudinal data on the educational achievement of children. We use linked individual-level administrative data (the NPD-ILR-HESA data) which enables us to follow students from the start of primary schooling through to the end of university.[5] This data includes a limited set of demographic information – including gender, ethnicity, month of birth, eligibility for free school meals (a proxy for low family income) and home postcode, from which information on the child's local neighbourhood can be merged in – and detailed results from the national achievement tests described above. We use a cohort born between September 1990 and August 1991, enabling us to examine educational trajectories from age 7 through to university entry at age 18 or 19.

Given our focus on educational trajectories throughout primary and secondary school, we restrict attention to individuals for whom we observe English and Maths test scores at each stage from Key Stage 1 to Key Stage 4. Because Key Stage tests are not compulsory in private schools, this means that we focus on pupils attending state schools only. We additionally exclude individuals who attended a special secondary school since many do not access the full curriculum and will not take standardised tests.[6] Our final sample therefore comprises 460,653 pupils from the cohort born in 1990-91.

The premise of our approach – described in more detail below – is to track the education trajectories of children with different initial achievement and from different family backgrounds. To do so, we split pupils into groups on the basis of early measures of achievement and family background. Our analysis measures initial achievement at Key Stage 1, with pupils split into groups on the basis of their performance in maths. Specifically, pupils are classified as "high achieving" if they reach Level 3 (above the government's expected

[5] The NPD-ILR-HESA data links together the National Pupil Database (NPD), the school administrative data set, which is a census of pupils, with data on their characteristics and achievement, the Individual Learner Records (ILR), the Further Education administrative data set, which is a census of students' learning episodes and achievement, and Higher Education Statistics Agency (HESA) data, the higher education administrative data set, with data on students' characteristics and higher education attainment.

[6] Only 314 pupils were omitted because they attended special secondary schools.

level), "average achieving" if they achieve the government's expected level (Level 2) and "low achieving" if they score Level 1 or below. The distributions of initial achievement are summarised in Table 1.

Family circumstance is measured by placing each pupil into a quintile group on the basis of an index of socio-economic status (SES). This index is created using individuals' eligibility for free school meals and a set of local area characteristics linked in on the basis of their home postcode at age 16. These include the Index of Multiple Deprivation score, their ACORN (A Classification of Residential Neighbourhoods) score, and three measures from the 2001 census: the proportion of individuals who work in professional or managerial jobs, the proportion of highly educated individuals and the proportion of individuals who own their own home.[7] This index is of course going to measure the socio-economic circumstances in which pupils were raised with some error, particularly given that it uses information on pupils' circumstances when they were age 16. Crawford et al. (2015) confirms that our results are robust to using richer data that contain better quality individual level information on the socio-economic status of each pupil from the Longitudinal Study of Young People in England. Our analysis focuses on pupils in the top 20% and bottom 20% of this index of socio-economic status (the most and least deprived children). Table 1 documents the percentage of our final sample who fall into these two quintile groups.

We combine information on initial achievement and family circumstance to create six groups whose educational trajectories we track: most deprived (bottom quintile) with low, average and high initial achievement (containing 2.2%, 12.1% and 2.2% respectively of our final sample; and least deprived (top quintile) with low, average and high initial achievement (containing 6.7%, 14.1% and 0.9% respectively of our final sample).

[7] See Chowdry et al. (2013) for further information on this measure including how it compares with various individual measures of socio-economic status from a cohort born at a similar time.

4. Empirical approach

As described above, when tracking the educational performance of pupils over time, regression to the mean (RTM) may be an issue. Any child defined as having 'high' or 'low' achievement on any given day may have over- or under-performed relative to their 'true' underlying potential, meaning that the next time they are tested they will look more like the average individual. When viewed over time as an educational trajectory, this statistical artefact would drive those from high and low initial achievement groups towards the mean value, creating an artificial convergence in educational trajectories over time. Jerrim and Vignoles (2013) emphasise that any measurement error in the test that is used to define initial achievement will be more prominent in the tails of the distribution, where sample sizes are smaller, and hence this is more likely to be the case for the low and high initial achievement groupings, particularly when split by socio-economic status. Finding that high achieving poor children fall behind their lower achieving but better-off peers may therefore arise, at least in part, as a result of RTM. If we do not account for this, then our conclusions regarding the trajectories of children from different socio-economic backgrounds – and, in particular at what point high achieving poor children appear to fall behind their lower achieving more advantaged peers – may be misleading.

The problem of RTM is exacerbated by defining initial achievement groupings based on measures that are then also used to plot educational trajectories over time. To overcome this, we use a method initially proposed by Ederer (1972) (see also Davis (1976) and Marsh and Hau (2002)) and implemented by Jerrim and Vignoles (2013). This involves having two test scores taken at the same time point (t_1) and using one test to determine which pupils are classified as "high achieving" and which "low achieving", and the other test as the baseline observation from which changes in achievement are measured. Defining initial achievement

using a different test, measured at the same time, be it on the same day or within the same short time period, will go some way to reducing any effect of RTM by reducing the correlation (and hence measurement error) between the initial grouping and the first observed achievement measure. We use performance in Key Stage 1 maths to define whether a child has high, average or low achievement at baseline, and Key Stage 1 reading scores as the basis from which changes in achievement are measured.[8]

A second approach would be to use a test score measured at an earlier time point (t_0) to determine whether a child is high, average or low achieving, and then start measuring trajectories of achievement from t_1. Any measurement error in the initial test should not be correlated with error in a test taken at a different time (based on the classical measurement error assumption of no serial correlation). This method should therefore, in principle, be even more robust to the presence of RTM than the first.

Test scores are not available before Key Stage 1, but when describing what happens to the performance of children defined as high, average or low achieving at baseline, we focus on what happens to their performance from the next period (Key Stage 2) onwards. This approach enables us to be as confident as possible that any convergence in test scores between pupils from different socio-economic backgrounds represents a true change in underlying performance rather than a spurious result arising from RTM, though of course subsequent tests are themselves not necessarily immune from measurement error. Although there is remarkably little evidence on the validity and reliability of Key Stage 2 and 3 tests during this period, there are concerns that "teaching to the test" for the high stakes (for schools) exams at the end of primary school may lead to higher than expected test scores at Key Stage 2, with some fall

[8] We also check that our results hold by defining initial achievement based on reading and using maths as the starting point from which change is measured instead in Appendix Figure A1; our results are robust to these choices.

away thereafter. More generally Key Stage 2 results appear not to be as predictive of Key Stage 4 results as one might expect (Strand, 2006; Black and Wiliam, 2006).

From our perspective, these concerns regarding test validity should not matter as long as the biases are random; but if they are systematically different across different types of student, then it is possible that we may not completely purge the effect of RTM. For this to cause problems for our analysis, initially high achieving children from poor backgrounds would have had to be more effectively "taught to the test" than initially average or high achieving children from rich backgrounds, for example, which seems relatively unlikely.

Information from the NPD is used to rank individuals within the distribution of overall achievement from Key Stage 1 through to Key Stage 5. Key Stage 1 achievement is based on pupils' reading level[9]; Key Stage 2 and 3 achievement is based on their fine grade score in English, and Key Stage 4 achievement is based on their GCSE English point score. For those who participate at Key Stage 5, achievement is measured based on their A level point score.[10] At university level, information for those who participate is available on the institution that they attend at age 18/19 from HESA data.[11] We rank individuals on the basis of their institution's average score from the 2001 Research Assessment Exercise. Clearly there are a number of alternative rankings of universities that we could have used here, including rankings by course rather than institution. For our purposes, however, we argue that overall RAE ranking is a reasonable proxy measure for pupils' achievement on entry to university since entry into more selective institutions is closely aligned with pupils' A level scores and these in turn vary

[9] Given that fine grade scores are not available at Key Stage 1, the ranking is based on a measure of six levels meaning that a lot of data is clustered around the three standard levels.

[10] Or their individual Learner Record (ILR) score for those missing A level points in the NPD-ILR-HESA data. Unfortunately this data is not available in the LSYPE.

[11] 163,104 individuals who went to university could be assigned an RAE ranking on the basis of the institution they reported attending in the first year. 13,992 university attendees could not be assigned an RAE ranking. 35% of our final sample attended an identifiable university.

systematically by RAE ranking. This ranking does not necessarily fully capture institution quality, however.

When ranking individuals at Key Stage 5 and university, we encounter issues in assigning a ranking for those who do not participate. We take the following approach: for those who do not participate at Key Stage 5, we assign them a ranking based on their predicted probability of participation from a probit model of participation and prior achievement at Key Stage 4. For those who do not participate at university we similarly assign a ranking based on their predicted probability of participation from a probit model based on their prior achievement at Key Stage 4 and 5. The underlying models are presented in Appendix Table A1 and the predicted percentile rankings by participation are reported for each of our six groups in Appendix Table A2. In both cases, individuals who do not participate are allocated a ranking below those who do participate. By predicting the probability of participation based on prior attainment, this assignment process allows us to distinguish between those who may have had the grades to stay on in education but chose not to do so, who will be allocated a higher probability of participation, and those who were not able to make the choice to stay on as their prior achievement was too low, who will be allocated a lower probability of participation.

We standardise each measure of educational achievement (from Key Stage 1 to university) within our sample and assign each pupil a percentile rank, with an average of around 50.5.

We are of course interested in the factors that may be driving any changes in education trajectories that we observe. To do this we control for characteristics that are likely to drive differences in these trends by analysing the conditional educational distribution of our sample: that is, the distribution after accounting for various other factors. We begin by controlling for individual-level characteristics that are fixed over time including gender, ethnicity and month

of birth, measured when children are age 16. We do not anticipate that these are major drivers of the story.

One factor which we think is more likely to be a key driver of differences in pupils' achievement trajectories by socio-economic background is the school that they attend. It is well known that sorting into schools on the basis of socio-economic background is extensive (Allen and Vignoles, 2007; Allen, 2007) and that sorting is moderately high in England compared to other countries (Jenkins et al. 2008). If children from different socio-economic backgrounds attend different quality schools, this may help to explain the trajectories that we see. This is likely since part of the mechanism through which such social segregation occurs is via the property market, with wealthier parents able to purchase houses nearer to certain sought after state-funded schools (Gibbons and Machin, 2003; Allen et al. 2010).[12] In addition, in some areas, sorting into schools occurs on the basis of test scores rather than socio-economic background (although the two are, of course, highly correlated).[13]

To explore the extent to which sorting into schools can help to explain different academic trajectories we add school-level fixed effects to our model, based on the school in which pupils take their Key Stage 4 exams. This enables us to control for any differences that arise because pupils from different backgrounds with differing initial achievement attend different schools and does not require us to make some of the more stringent assumptions required with random effects or multi-level models (Clarke et al. 2010). Whilst the fixed effects model comes at the cost of a loss in efficiency compared to the random effects model, in this instance since we know there is substantial sorting into schools that we are unlikely to be able

[12] A small percentage of pupils in England (around 7%) also attend private schools at age 16, with sorting occurring more explicitly on parents' ability to pay in this case. But these pupils are not part of our sample.
[13] While the percentage of pupils in England attending academically selective schools is small (less than 5%), it is much higher in areas which operate a grammar school system, such as Buckinghamshire or Kent.

to fully account for using the data at our disposal, the fixed effects model is the more conservative choice.[14]

The method we use to account for differences in demographic characteristics and school effects is to regress the achievement measure at age 16 on our range of pupil characteristics and school fixed effects, and then to consider the average positions of those from our six prior achievement/SES groups *after* controlling for these differences (in the residual distribution of achievement).

5. Results

We begin by looking at how SES gradients change as pupils move through the education system, from age 7 to university entry at age 18/19. Figure 1 plots the average percentile ranking by SES quintile (lines) and the gap between the most and least deprived quintile (bars) from Key Stage 1 through to university participation. (Underlying figures in top panel of Appendix Table A3.) Consistent with previous work in the UK, there are stark SES gradients at each stage of education, which increase as children move through their educational careers and then flatten out during post-16 education. Large socio-economic differences are observed in the earliest achievement tests at age 7 with a 16.1 percentile achievement gap between the most and least deprived pupils at Key Stage 1. At Key Stage 2, Key Stage 3 and Key Stage 4 these gaps widen to 22.1, 27.4 and 28.9 percentiles respectively. This is driven both by the lower SES groups' average percentile ranking falling (from 42 at Key Stage 1 to 35.3 at Key Stage 4) and the higher SES groups' average percentile ranking increasing (from 58.1 to 64.2). The gap remains at almost 30 percentiles at Key Stage 5 and into university: this is perhaps unsurprising given

[14] Note that we are unable to measure school processes, e.g. streaming, which may be an important part of the story.

the close link between achievement at the end of secondary school and A level grades, as well as the type of university a student can access. This means that the achievement gap between the most and least deprived quintile groups increases by around two thirds between age 7 and age 18/19.

To explore how these patterns differ depending on pupils' initial achievement and socio-economic background, Figure 2 plots the average percentiles of those from the most and least deprived families who were high, average and low achievers in their age 7 maths tests. (Underlying figures in second panel of Appendix Table A3.) The groups who experience the largest changes in performance over time are the initially high achieving poor children and the initially low achieving rich children. While some of this convergence occurs between Key Stage 1 and Key Stage 2, the majority happens between Key Stage 2 and Key Stage 4. More specifically, initially lower-achieving affluent children move closer to higher-achieving deprived children between Key Stage 2 and Key Stage 3 and, to a lesser extent, between Key Stage 3 and Key Stage 4. Conversely, initially high-achieving children from the most deprived families move closer to lower-achieving students from the least deprived families by Key Stage 3. Note that while there is some slight further convergence between these groups at Key Stage 5, most of the movement occurs between Key Stage 2 and Key Stage 4. This suggests that the period of compulsory secondary schooling is a crucial time when children from deprived families are at risk of falling behind their similarly achieving more affluent peers. We look into what might be driving these differences later in this section.

While we are using different tests taken at age 7 to define initial achievement and from which to measure educational performance in order to reduce the impact of measurement error, there may still be some correlation in the error between the two tests that drives our findings, if for example the pupil took both tests on a particularly 'bad day' and therefore performed unusually badly on both measures. We argue, however, that our main findings between Key

Stage 2 and Key Stage 4 are unlikely to be driven by measurement error as initial achievement is defined in the period before the main convergence occurs. Furthermore, Crawford et al. (2015) show that the patterns of educational trajectories observed are practically identical whether Key Stage 2 or Key Stage 1 tests are used to define initial achievement. These findings strongly support the idea that the convergence between Key Stage 2 and Key Stage 4 is not being driven by regression to the mean.

Drivers of educational trajectories

Given our finding that compulsory secondary schooling is a potentially critical period during which initially higher achieving poorer children are at risk of falling behind their initially lower achieving more affluent peers, we move on to consider why this might be: what might help to explain the patterns that we see?

To assess the extent to which demographic characteristics and school attended affect educational trajectories, we regress achievement from Key Stage 2 onwards on these measures and consider the remaining achievement distribution once differences in these factors are taken into account. We build our model in stages, focusing first on taking account of individual-level characteristics before moving to a school fixed effects model assessing how much of the pattern is driven by differences between schools that children from different backgrounds attend (see Appendix Table A4 for full estimates from these models). We plot raw achievement at Key Stage 1, followed by the conditional achievement rankings from Key Stage 2 to 4: that is, we show how educational trajectories from Key Stage 1 to Key Stage 4 change, after differences between ethnicity, gender, month of birth and then the school attended are removed (i.e. how achievement changes for children with the same gender, ethnicity and month of birth, and who attend the same school).

Panel A of Figure 3 replicates the unconditional trajectory from Figure 2 (on a different scale) while panel B plots the trajectory conditional on the individual characteristics we consider, namely ethnicity, gender and month of birth. As can be seen, controlling for observed differences in these fixed individual-level characteristics does little to change the picture of educational trajectories over this crucial period, and this is exactly what we might have expected given the relatively limited evidence of substantial differences in these characteristics between children in our different background and initial achievement groupings (see Appendix Table A5).

Panel C of Figure 3 illustrates the educational trajectories between Key Stage 1 and Key Stage 4 after controlling for differences in these individual-level characteristics and differences between schools (using school fixed effects). This specification removes any differences that are driven by children sorting into different schools, and essentially compares the achievement levels of children from each of our six groups who attend the same schools. We might expect this to make more of a difference to our results, as there are clear differences across both SES and initial achievement groups in terms of the types of schools they attend. For example, Appendix Table A5 shows that high achieving affluent children attend the best performing secondary schools with the lowest proportion of children eligible for free school meals, while low achieving deprived children attend the worst performing schools with the highest proportion of children eligible for free school meals.

Panel C of Figure 3 shows that, within initial achievement group, the extent of the divergence in performance between children from higher and lower socio-economic backgrounds seen in Panels A and B is somewhat less once we allow for the school attended. This indicates that some of the reason for the observed divergence is driven by the sorting of pupils of similar initial achievement but different socio-economic backgrounds into different schools. This pattern is particularly strong amongst the low initial achievement group, in which

the trajectories of those from higher and lower socio-economic backgrounds almost entirely coincide once we compare pupils who attend the same school.[15] Moreover, this appears to be driven by improved test score performance amongst low achieving children from poor backgrounds.

While there is a narrowing of the gap within initial achievement groups by SES, there is little change between initial achievement groups when school differences are accounted for. In other words, attending different schools explains why poor children fall further behind their richer counterparts but does not really explain differences between high and low achieving pupils generally.

Although we are cautious about the interpretation of these results given that they do not identify the causal effect of school variation on pupil achievement, they are nevertheless indicative of a potentially important role for schools, teachers or peers influencing the attainment of low achieving children from poor backgrounds to ensure that they do not fall behind relative to their richer low achieving peers within the same school.

6. Conclusions

This paper has described the education trajectories of pupils from different socio-economic backgrounds, focusing on how they differ between pupils of initially high, average and low achievement, after accounting as far as possible for the possibility of regression to the mean. We find clear evidence of a convergence of achievement during secondary school, particularly from ages 11 to 14, between affluent children of lower initial achievement and

[15] Around 40% of schools in our data have at least one rich and one poor high-achieving pupil; 65% have at least one rich and one poor pupil with average achievement; 27% have at least one rich and one poor low-achieving pupil. Around 50% of schools have at least one deprived high-achieving and one affluent average achieving pupil.

poorer children of higher initial achievement. This finding is robust to our attempts to minimise the problem of regression to the mean.

A lot of policy attention has been focused on trying to improve the achievement of poor children in England in recent years. Amongst pupils who attend the same secondary school, initially low-achieving poor children seem to be keeping up with their richer initially low-achieving peers, which may suggest that policies aimed at reducing the achievement gap between rich and poor children – such as the provision of additional financial resources in the form of the pupil premium – have been focused on those poor children with low initial achievement (although more research is needed to understand how pupil premium resources have been targeted within schools).

However, there is still evidence of some divergence in the performance of initially average and high achieving children by socio-economic background even within schools, which needs to be better understood. Work by the Education Endowment Foundation and others has identified some policies which may help to reduce socio-economic gaps in attainment across the distribution in secondary school – such as more one-to-one tuition and summer schools – and outreach efforts by universities may also be successful at targeting higher achieving children from poorer backgrounds.

The bigger issue highlighted by our research, however, is the fact that at least part of the reason why poorer children fall behind their richer peers seems to be because they go to different secondary schools. While our findings are not causal, they do seem to suggest that policymakers interested in improving the educational progress of children from lower socio-economic backgrounds need to adopt policies that ensure all pupils have access to good schools and perhaps to consider the role of school choice and admissions policies in this.

We have identified early secondary as a key transition period. Given that Key Stage 3 tests have not been set or marked externally since 2009, there is currently no nationally

consistent way to identify those falling behind at this crucial time before entering into Key Stage 4. Providing better information on students' performance during the transition to secondary school may help both families and schools support students in making the transition from primary to secondary and on to GCSE.

References

Allen, R. (2007). 'Allocating pupils to their nearest secondary school: The consequences for social and ability stratification'. *Urban Studies* Vol. 44 (4), pp. 751-770.

Allen, R., Burgess, S., and Key, T. (2010). 'Choosing Secondary School by Moving House: School Quality and the Formation of Neighbourhoods.' CMPO Working Paper No. 10/238.

Allen, R., and Vignoles, A (2007). 'What should an index of school segregation measure?' *Oxford Review of Education* Vol. 33 (5), pp. 643-668.

Atkinson A.B. (1980). 'Inheritance and the redistribution of wealth', in *Public Policy and the Tax System*, (G.A. Hughes and G.M. Heal, eds.), Allen and Unwin: London.

Atkinson A.B., Jenkins, S.P. (1984). 'The steady state assumption and the estimation of distributional and related models', *Journal of Human Resources*, Vol. 19, pp.358-376.

Becker, G. and Tomes, N. (1986). 'Human Capital and the Rise and Fall of Families' *The Journal of Labor Economics* Vol. 4(3).

Black, P. and Wiliam, D., (2006), 'The reliability of assessments'. *Assessment and Learning*, pp.119-131.

Blanden, J., Gregg, P. and Macmillan, L. (2007). 'Accounting for Intergenerational Income Persistence: Noncognitive Skills, Ability and Education'. *Economic Journal*, Vol. 117, pp. C43-C60.

Blanden, J., and Gregg, P. (2004). 'Family income and educational attainment: a review of approaches and evidence for Britain'. *Oxford Review of Economic Policy*, Vol. 20(2), pp. 245-263.

Blanden, J. and Machin, S. (2007). 'Recent Changes in Intergenerational Mobility in Britain'. Sutton Trust, London.

Blanden, J. and Machin, S. (2010). 'Intergenerational Inequality in Early Years Assessments'. In *Children of the 21st century: The first five years* (eds K. Hansen, H. Joshi**,** and S**.**Dex**),** pp. 153-168. Bristol: The Policy Press.

Blau, P.M. and Duncan, O.D. (1967). *The American Occupational Structure*. New York: Wiley.

Breen, R. and Goldthorpe, J.. (2001). 'Class, mobility and merit: The experience of two British birth cohorts.' *European Sociological Review* Vol. 17(2), pp. 81-101.

Breen, R. and Jonsson, J.O. (2007). 'Explaining change in social fluidity: educational equalization and educational expansion in twentieth-century Sweden.' *American Journal of Sociology* Vol. 112(6), pp. 1775-1810.

Cabinet Office (2011). 'Opening Doors, Breaking Barriers: A Strategy for Social Mobility', HM Government, London.

Caro, D. (2009). 'Socio-economic status and academic achievement trajectories from childhood to adolescence', *Canadian Journal of Education*, Vol. 32, pp. 558-590.

Chowdry, H., Crawford, C., Dearden, L., Goodman, A., and Vignoles, A. (2013). 'Widening participation in higher education: Analysis using linked administrative data'. *Journal*

of the Royal Statistical Society: Series A (Statistics in Society), Vol. 176, pp. 431–457.

Clarke, P., Crawford, C., Steele, F., and Vignoles, A., (2010). 'The choice between fixed and random effects models: some considerations for educational research'. CMPO working paper: University of Bristol.

Crawford, C. (2014). 'The Link between Secondary School Characteristics and University Participation and Outcomes'. Report to the Department for Education.

Crawford, C., Macmillan, L. and Vignoles, A. (2015). **'**When and Why do Initially High Attaining Poor Children Fall Behind?' SPSS WP 20, CASE, London School of Economics

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.

Davis, C. (1976). 'The Effect of Regression to the Mean in Epidemiologic and Clinical Studies', *American Journal of Epidemiology*, Vol. 104, pp. 493-498.

Duncan, G. and Magnuson, K. (2011). 'The Nature and Impact of Early Achievement Skills, Attention Skills and Behavior Problems'. In *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances* (eds G. Duncan and R. Murnane), pp. 47-69. New York: Russell Sage Foundation.

Ederer, F. (1972). 'Serum Cholesterol: Effects of Diet and Regression Toward the Mean', *Journal of Chronic Disorders*, Vol. 25, pp. 277-289.

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

Feinstein, L. (2004). 'Mobility in Pupils' Cognitive Attainment During School Life', *Oxford Review of Economic Policy*, Vol. 20, pp. 213-229.

Galton, F. (1986). 'Regression Towards Mediocrity in Hereditary Stature', *The Journal of the Anthropological Institute of Great Britain and Ireland*, Vol. 15, pp. 246–263.

Gibbons, S, and Machin, S. (2003). 'Valuing English primary schools'. *Journal of Urban Economics*, Vol. 53(2), pp. 197-219.

Goodman, A., & Gregg, P. (2010). 'Poorer Children's Educational Attainment: How Important Are Attitudes and Behaviour?' Report to the Joseph Rowntree Foundation.

Goodman, A., Gregg, P., & Washbrook, E. (2011). 'Children's educational attainment and the aspirations, attitudes and behaviours of parents and children through childhood.' *Longitudinal and Life Course Studies*, Vol. 2(1), pp. 1-18.

Goodman A. Sibieta L. and Washbook E. (2009). 'Inequalities in Educational Outcomes Among Children Aged 3 to 16', Final report for the National Equality Panel, Institute for Fiscal Studies, London.

Jenkins, S P., Micklewright, J. and Schnepf. S.V. (2008). 'Social segregation in secondary schools: how does England compare with other countries?' *Oxford Review of Education* Vol. 34(1), pp. 21-37.

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)*, Vol. 176, pp. 887–906.

Marsh, H. and Hau, K. (2002). 'Multilevel Modelling of Longitudinal Growth and Change: Substantive Effects or Regression Toward the Mean Artefacts?', *Multivariate Behavioral Research*, Vol. 37, pp. 245-282.

Parsons, S. Schoon, I. Rush, R. and Law, J. (2011). 'Long-term Outcomes for Children with Early Language Problems: Beating the Odds', *Children and Society*, Vol. 25, pp. 202 – 214.

Reardon, S. (2011). 'The Widening Academic Achievement Gap Between the Rich and the Poor: New Evidence and Possible Explanations'. In *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*, (eds G. Duncan and R. Murnane), pp.91-116. New York: Russell Sage Foundation.

Schoon, I. (2006). *Risk and Resilience: Adaptations in Changing Times*, Cambridge: Cambridge University Press.

Strand, S., (2006). 'Comparing the predictive validity of reasoning tests and national end of Key Stage 2 tests: which tests are the 'best'?'. *British Educational Research Journal*, Vol. 32(2), pp.209-225.

Strand, S., (1997). 'Pupil progress during Key Stage 1: a value added analysis of school effects'. *British Educational Research Journal*, Vol. 23(4), pp.471-487.

Strand, S., (2010). 'Do some schools narrow the gap? Differential school effectiveness by ethnicity, gender, poverty, and prior achievement'. *School Effectiveness and School Improvement*, Vol. 21(3), pp.289-314.

Strand, S. (2014a). 'Ethnicity, gender, social class and achievement gaps at age 16: Intersectionality and 'Getting it' for the white working class'. *Research Papers in Education*, Vol. 29(2), pp.131-171.

Strand, S. (2014b). 'School effects and ethnic, gender and socio-economic gaps in educational achievement at age 11'. *Oxford Review of Education*, Vol. 40(2), pp. 223-245.

Strand, S., Deary, I.J. and Smith, P., (2006). 'Sex differences in cognitive abilities test scores: A UK national picture'. *British Journal of Educational Psychology*, Vol. 76(3), pp. 463-480.

Thomas, S., Sammons, P., Mortimore, P. and Smees, R., (1997a). 'Stability and Consistency in Secondary Schools' Effects on Students' GCSE Outcomes over Three Years'. *School effectiveness and school improvement*, Vol. 8(2), pp.169-197.

Thomas, S., Sammons, P., Mortimore, P. and Smees, R., (1997b). 'Differential secondary school effectiveness: comparing the performance of different pupil groups'. *British Educational Research Journal*, Vol. 23(4), pp.451-469.

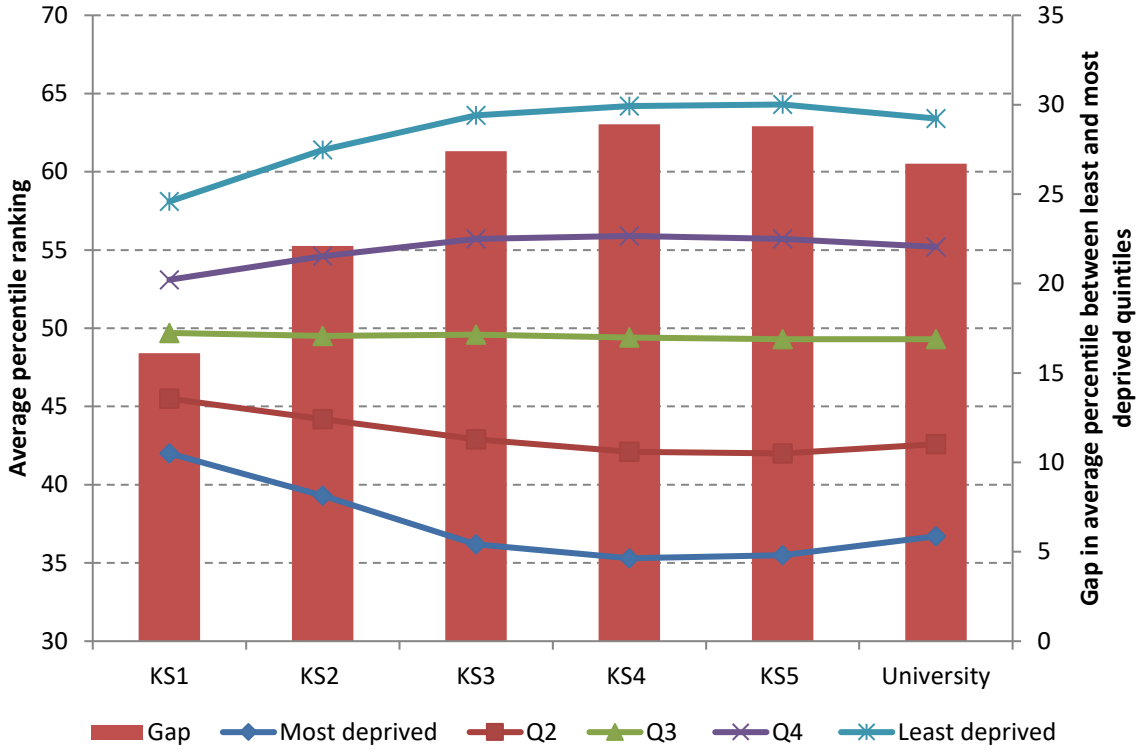
Washbrook, L. and Lee, R. (2015). 'Beyond the Feinstein chart: investigating differential achievement trajectories in a US cohort', *Longitudinal and Life Course Studies*, Vol. 6, pp. 359-368.

Table 1 Descriptive statistics of final sample

Most deprived	16.5%
Least deprived	21.7%
Low initial achievement	21.4%
Average initial achievement	70.3%
High initial achievement	8.3%
Most deprived high initial achievement	2.2% (10,146)
Most deprived average initial achievement	12.1% (55,841)
Most deprived low initial achievement	2.2% (10,173)
Least deprived high initial achievement	6.7% (30,744)
Least deprived average initial achievement	14.1% (64,766)
Least deprived low initial achievement	0.9% (4,325)
Total N	460,653

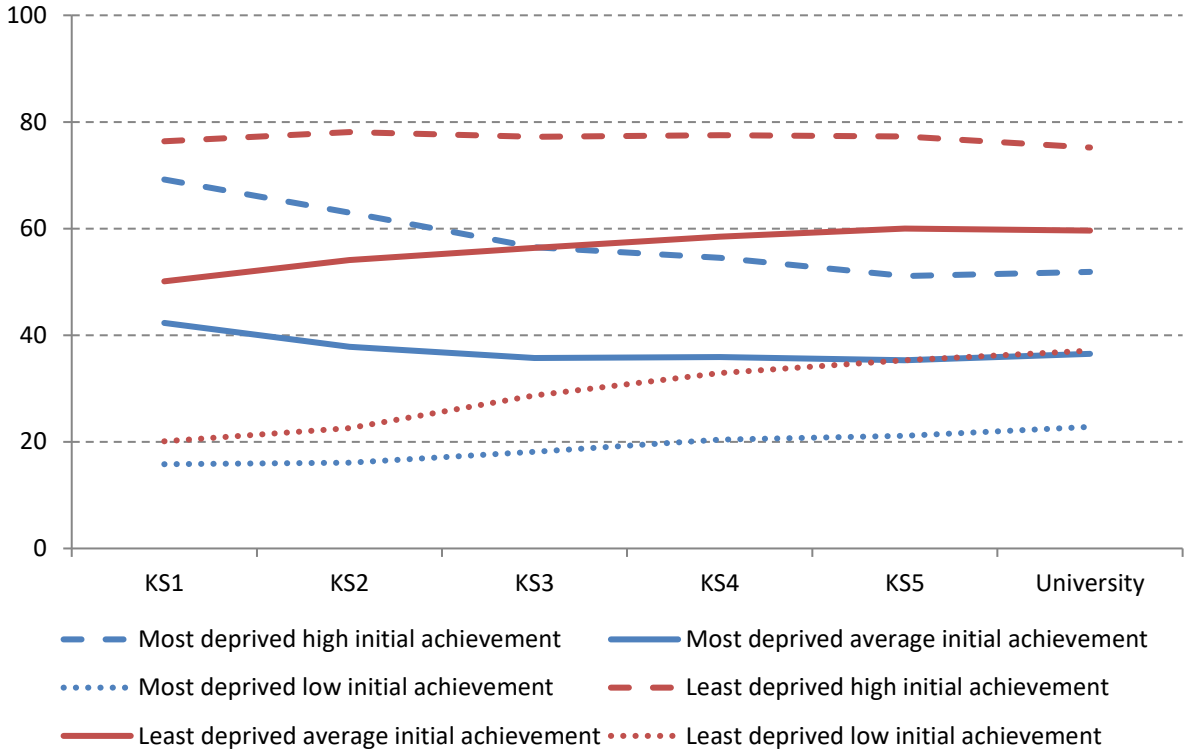
Notes: Initial achievement defined based on performance in Key Stage 1 maths tests: low = level 1, average = level 2, high = level 3 or above.

Figure 1 Average percentile rankings at each stage of the educational trajectory by SES and the achievement gap between the most deprived and least deprived quintiles of SES



Notes: Assessments occur at each point in time with lines used to connect the points. Achievement is measured using standardised percentile rankings based on average reading and maths scores at Key Stage 1, average English and maths scores from Key Stage 2 to Key Stage 4, average scores across all subjects combined with predicted probabilities of attendance at Key Stage 5, and university rankings combined with predicted probabilities of attendance at university.

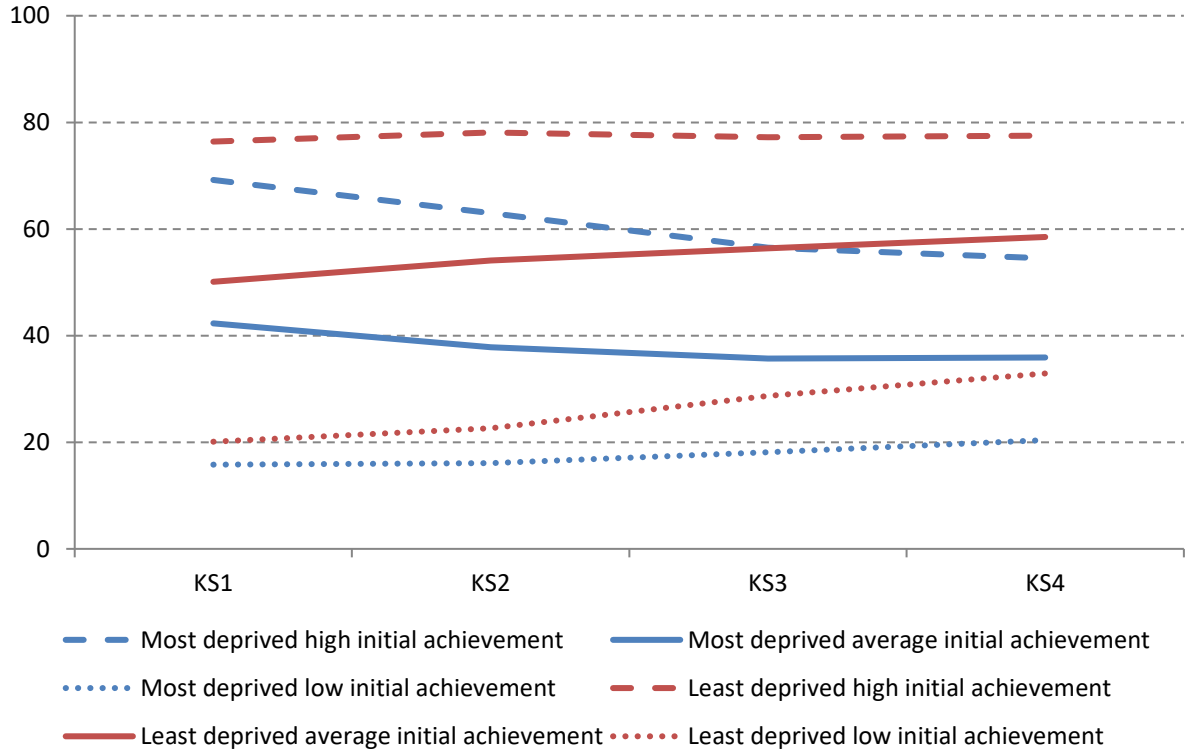
Figure 2 Trajectories from Key Stage 1 to university by initial achievement (defined using Key Stage 1 maths) for the most deprived and least deprived quintiles of SES



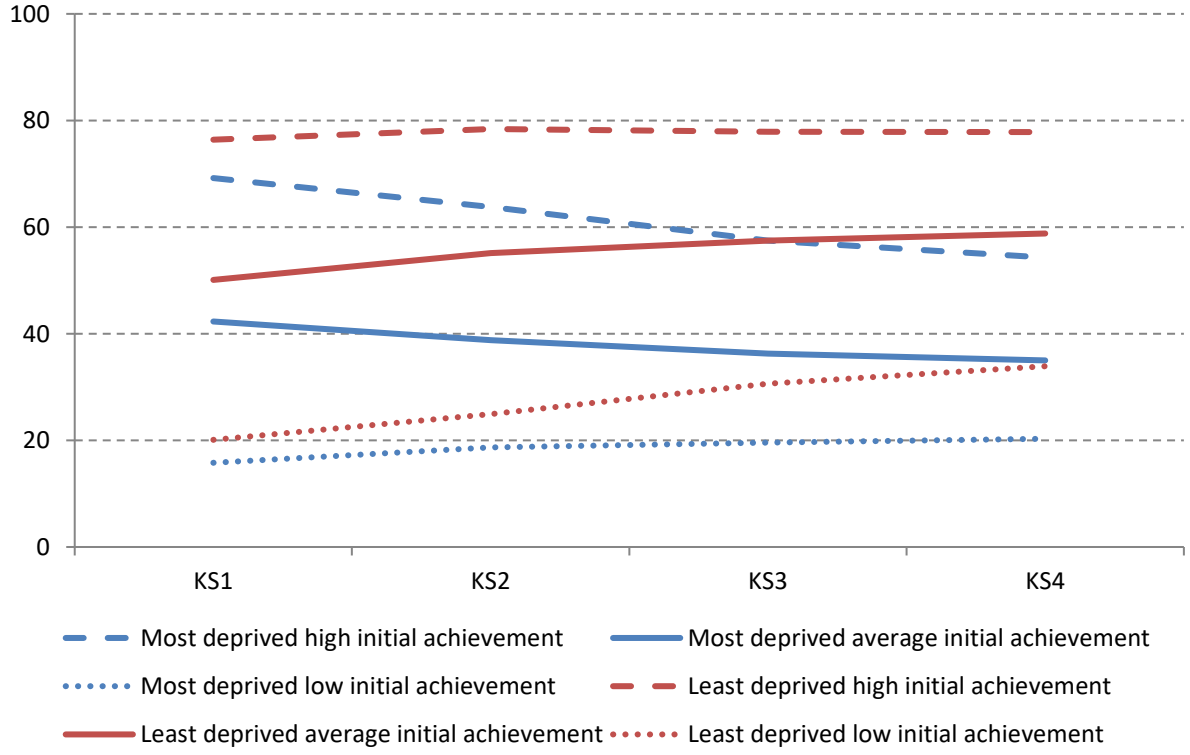
Notes: Assessments occur at each point in time with lines used to connect the points. Achievement is measured using standardised percentile rankings based on reading scores at Key Stage 1, English scores at Key Stages 2 to 4, average scores across all subjects combined with predicted probabilities of attendance at Key Stage 5, and university rankings combined with predicted probabilities of attendance at university.

Figure 3 Trajectories from Key Stage 1 to Key Stage 4 by initial achievement (defined using Key Stage 1 maths) for the most deprived and least deprived SES quintiles, conditional on demographics and school fixed effects

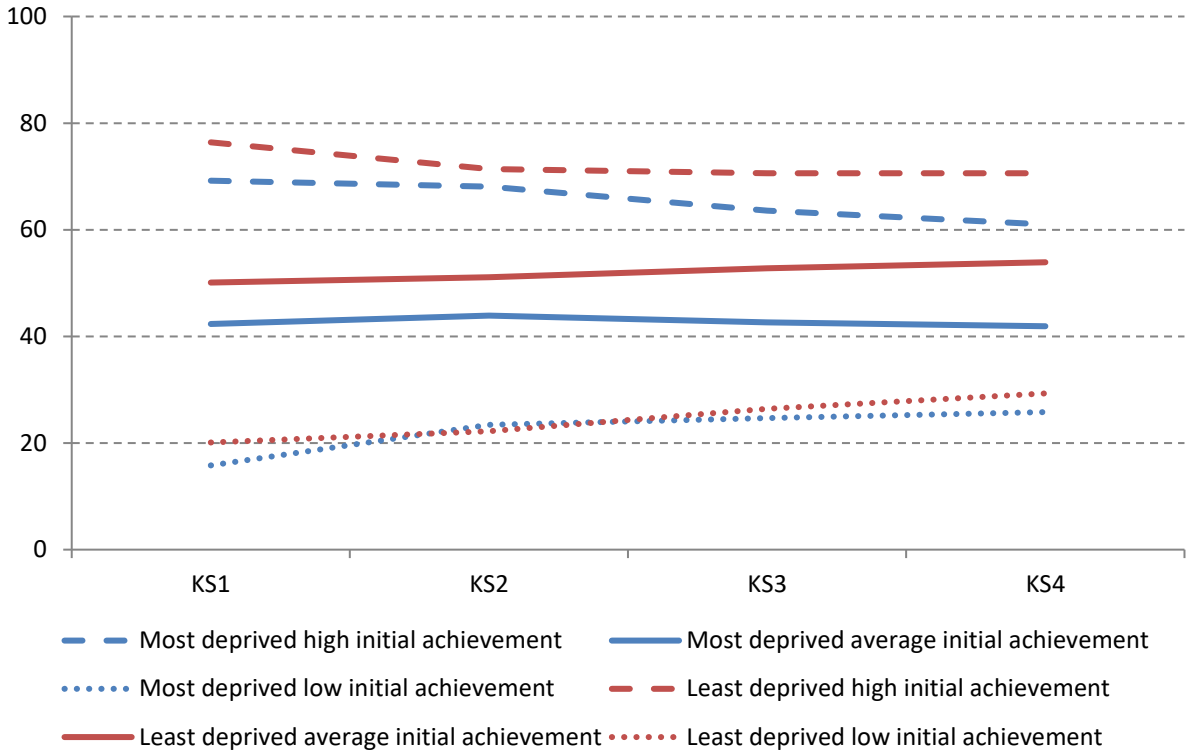
Panel A: Unconditional



Panel B: Conditional on demographics



Panel C: Conditional on demographics and school fixed effects



Notes: Assessments occur at each point in time with lines used to connect the points. Achievement is measured using standardised percentile rankings based on reading scores at Key Stage 1 and English scores at Key Stages 2 to 4. Demographic controls are gender, ethnicity and month of birth. School fixed effects are based on the secondary school in which students sit their Key Stage 4 tests.

Appendix

Table A1 Marginal effects from probit models predicting participation at Key Stage 5 and university based on prior achievement at Key Stage 4

	Participation at Key Stage 5	**Participation at university**
Key Stage 4 points score	0.012*** (0.0001)	0.003*** (0.0001)
Key Stage 4 English A*-C	0.245*** (0.002)	0.072*** (0.003)
Key Stage 4 maths A*-C	0.124*** (0.002)	0.054*** (0.003)
Number of A*s in EBacc subjects	0.130*** (0.004)	0.016*** (0.002)
Number of As in EBacc subjects	0.136*** (0.002)	0.023*** (0.001)
Number of Bs in EBacc subjects	0.099*** (0.001)	0.022*** (0.001)
Number of Cs in EBacc subjects	0.063*** (0.001)	0.011*** (0.001)
Participation at Key Stage 5		0.231*** (0.003)
Key Stage 5 point score		0.0003*** (0.000005)
Number of As at A-level		0.117*** (0.002)
Number of Bs at A-level		0.105*** (0.002)
Number of Cs at A-level		0.078*** (0.002)
Number of Ds at A-level		0.036*** (0.002)
Number of Es at A-level		-0.012*** (0.002)
Pseudo R2	0.432	0.463
N	460,653	460,653

Table A2 Predicted percentile rankings at Key Stage 5 and University by participation status, initial achievement and socio-economic status

	High SES			Low SES		
	High initial achievement	Average initial achievement	Low initial achievement	High initial achievement	Average initial achievement	Low initial achievement
Participated at Key Stage 5	82.2	74.0	66.0	69.8	65.5	61.3
Did not participate at Key Stage 5	35.3	26.9	17.7	28.0	19.9	12.9
Participated at University	84.7	80.9	78.0	80.3	78.6	77.7
Did not participate at University	47.4	37.5	23.5	35.6	25.2	15.8

Table A3 Numbers underlying Figures 1 to 3

	KS1	KS2	Cumulative change KS1 to KS2	KS3	KS4	Cumulative change KS2 to KS4	KS5	University
Figure 1								
Most deprived (Q1)	42	39.3		36.2	35.3		35.5	36.7
Q2	45.5	44.2		42.9	42.1		42	42.6
Q3	49.7	49.5		49.6	49.4		49.3	49.3
Q4	53.1	54.6		55.7	55.9		55.7	55.2
Least deprived (Q5)	58.1	61.4		63.6	64.2		64.3	63.4
Gap (Q5-Q1)	16.1	22.1		27.4	28.9		28.8	26.7
Figure 2								
Low SES, high initial achievement	69.2	63.0	-6.2	56.5	54.5	-8.5	51.1	51.9
Low SES, average initial achievement	42.3	37.8	-4.5	35.7	35.9	-1.9	35.3	36.5
Low SES, low initial achievement	15.8	16.1	0.3	18.1	20.4	4.3	21.1	22.8
High SES, high initial achievement	76.4	78.1	1.7	77.2	77.5	-0.6	77.3	75.2
High SES, average initial achievement	50.1	54.1	4.0	56.4	58.5	4.4	60.0	59.6
High SES, low initial achievement	20.1	22.6	2.5	28.7	32.9	10.3	35.3	37.1
Figure 3								
Panel A: unconditional (see Figure 2)								
Panel B: conditional on demographics								
Low SES, high initial achievement	69.2	63.8		57.5	54.3			
Low SES, average initial achievement	42.3	38.8		36.3	35			
Low SES, low initial achievement	15.8	18.7		19.6	20.3			
High SES, high initial achievement	76.4	78.4		77.9	77.8			
High SES, average initial achievement	50.1	55.1		57.5	58.8			
High SES, low initial achievement	20.1	24.9		30.6	33.9			
Panel C: demographics and school FE								
Low SES, high initial achievement	69.2	68.1		63.6	61			
Low SES, average initial achievement	42.3	43.9		42.6	41.9			
Low SES, low initial achievement	15.8	23.4		24.7	25.8			
High SES, high initial achievement	76.4	71.4		70.6	70.6			
High SES, average initial achievement	50.1	51.1		52.8	53.9			
High SES, low initial achievement	20.1	22.2		26.4	29.3			

Table A4 Regressions coefficients predicting outcomes at Key Stage 4 based on demographics and school fixed effects

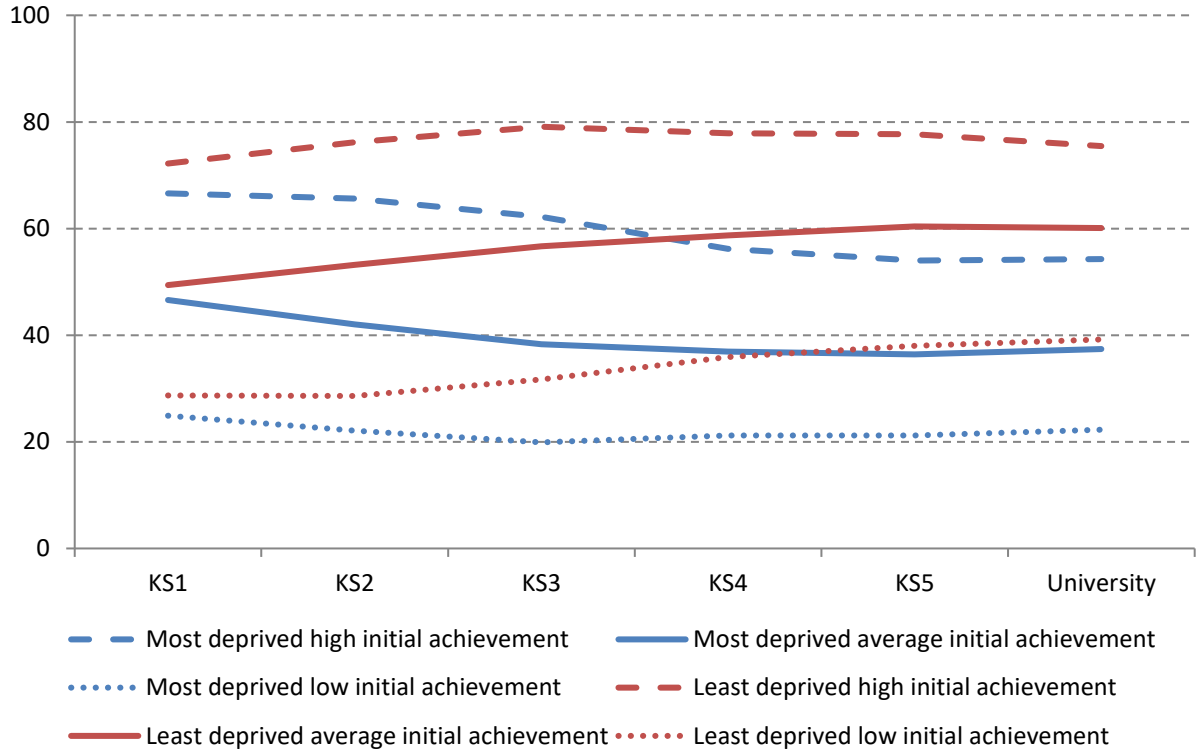
	Controlling for demographic characteristics	**Controlling for demographic characteristics and school fixed effects**
Male	-9.521*** (0.082)	-9.550*** (0.079)
White	(omitted category)	(omitted category)
Other White	6.365*** (0.310)	2.047*** (0.296)
Black African	2.894*** (0.427)	2.668*** (0.425)
Black Caribbean	-5.903*** (0.373)	-3.452*** (0.367)
Other Black	-6.577*** (0.672)	-3.573*** (0.629)
Indian	8.101*** (0.279)	6.954*** (0.293)
Pakistani	-4.318*** (0.282)	-0.530*** (0.310)
Bangladeshi	-1.274*** (0.452)	3.706*** (0.489)
Chinese	14.681*** (0.772)	9.905*** (0.566)
Other Asian	9.437*** (0.603)	4.029*** (0.566)
Mixed	1.395*** (0.277)	1.138*** (0.259)
Other	4.102*** (0.549)	1.786*** (0.516)
Month of birth		
September	(omitted category)	(omitted category)
October	-0.599*** (0.196)	-0.571*** (0.179)
November	-0.949*** (0.199)	-0.949*** (0.181)
December	-2.170*** (0.200)	-1.857*** (0.182)
January	-2.337*** (0.198)	-2.258*** (0.180)
February	-2.780*** (0.202)	-2.664*** (0.184)
March	-2.967*** (0.198)	-2.868*** (0.181)
April	-3.240*** (0.200)	-3.140*** (0.182)
May	-3.555*** (0.197)	-3.458*** (0.179)
June	-4.262*** (0.198)	-4.120*** (0.181)
July	-4.708*** (0.196)	-4.465*** (0.179)
August	-5.330*** (0.198)	-4.963*** (0.180)
School FEs		Yes
Group N		3,767
N	460,653	460,653

Table A5 Descriptive statistics of key measures of demographics and school characteristics by SES status and initial achievement

	All	High SES			Low SES	
		High initial achievement	Average initial achievement	Low initial achievement	High initial achievement	Average initial achievement	Low initial achievement
Male	53.3	48.3	55.7	54.9	45.3	46.1
Ethnicity						
White	92.3	92.1	88.0	79.4	81.3	76.8
Other White	2.2	2.0	2.5	1.9	1.5	1.8
Black African	0.2	0.2	0.4	2.6	2.0	2.0
Black Caribbean	0.3	0.4	0.7	2.5	2.1	2.1
Other Black	0.1	0.1	0.3	0.7	0.8	0.8
Indian	1.5	1.6	2.3	1.5	1.4	1.5
Pakistani	0.4	0.7	1.8	3.2	3.9	6.6
Bangladeshi	0.1	0.1	0.6	2.0	2.3	3.8
Chinese	0.4	0.3	0.4	0.5	0.3	0.2
Other Asian	0.3	0.3	0.3	0.5	0.4	0.5
Mixed	1.9	1.6	2.2	4.3	3.0	3.0
Other	0.3	0.4	0.6	0.9	0.9	0.9
Month of birth						
September	12.9	6.9	4.0	13.9	8.5	4.1
October	11.8	7.2	3.8	13.9	8.5	5.0
November	10.5	7.3	4.4	11.2	8.4	5.4
December	9.4	7.0	4.8	10.8	8.6	6.5
January	9.3	7.9	5.8	9.5	8.7	6.9
February	7.7	7.7	7.1	7.5	7.7	7.0
March	7.9	8.6	7.9	7.4	8.4	8.6
April	7.0	8.7	8.8	6.4	8.0	8.7
May	7.0	9.5	11.2	5.6	8.2	10.1
June	6.0	9.5	12.1	5.0	8.1	10.9
July	5.8	9.8	13.9	4.8	8.7	12.7
August	4.7	9.7	16.1	4.0	8.1	14.1

School characteristics						
School 5 A*-C at KS4	73.6	68.7	64.7	57.1	53.9	52.3
School % FSM	6.3	7.3	8.6	23.2	23.9	25.5

Figure A1 Trajectories from Key Stage 1 to university by initial achievement (defined using Key Stage 1 reading) for the most deprived and least deprived quintiles of SES



Notes: achievement is measured using standardised percentile rankings based on maths scores at Key Stages 1 to 4, average scores across all subjects combined with predicted probabilities of attendance at Key Stage 5, and university rankings combined with predicted probabilities of attendance at university.