Centre for Economic Performance

School qualifications and youth custody

Stephen Machin
Sandra McNally
Jenifer Ruiz-Valenzuela

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

UKRI

Economic and Social Research Council

Abstract
A very small number of young people enter youth custody between age 16 and 18 (about 4 in 1000 males), yet the consequences are severe. They spend an average of 7 months in youth custody and such incarceration has been related to negative outcomes in the longer term even if they can establish themselves in the labour market. In this paper, we evaluate whether there is a relationship between GCSE qualifications in English and maths and the probability of youth custody using administrative data in England. We are hindered in this because the majority of young people who end up in youth custody are not entered or fail their GCSEs in these subjects. Although regression results are consistent with educational achievement being a factor in why people end up in youth custody, they strongly suggest that both non-entry/low achievement and youth custody are correlated with severe vulnerabilities which are partially picked up by the explanatory variables available in administrative data (in particular indicators for special needs, disadvantage and being from some ethnic minority groups). Another interesting insight is that for many, problems only emerge (or at least become evident) in early or middle adolescence.

Keywords: youth custody; school qualifications; crime; educational attainment; special needs
JEL: I2; J1

Summary

A very small number of young people enter youth custody between age 16 and 18 (about 4 in 1000 males), yet the consequences are severe. They spend an average of 7 months in youth custody and such incarceration has been related to negative outcomes in the longer term even if they can establish themselves in the labour market. There is also a high unit cost of keeping individuals in youth custody. It would be in the private and social interest to prevent individuals from having to enter in the first place.

Previous research has shown a causal relationship between educational achievement and the probability of youth crime. In this paper, we evaluate whether there is a relationship between GCSE qualifications in English and maths and the probability of youth custody. We are hindered in this because the majority of young people who end up in youth custody are not entered or fail their GCSEs in these subjects. Although regression results are consistent with educational achievement being a factor in why people end up in youth custody, they strongly suggest that both non-entry/low achievement and youth custody are correlated with severe vulnerabilities which are partially picked up by the explanatory variables available in administrative data.

The vulnerable population are much more likely to have a serious special needs designation or be in a 'special type' of institution (such as a Pupil Referral Unit) than in the population as a whole. In fact, 75 per cent of students ending up in youth custody had been identified in school as having Special Educational Needs (most often with a Statement or School Action Plus). They are also much more likely to come from a disadvantaged background as indicated by their eligibility to receive free school meals. Those from Black Caribbean or Black African backgrounds are greatly over-represented amongst those who end up in youth custody. Although there is a negative correlation between achievement at age 11 and ending up in youth custody, about half of such boys had met the government 'expected level' in English or maths at that time, many of whom did very badly in exams at age 16 (if they were entered at all). This suggests that for many, problems either emerge or become evident in early or middle adolescence.

Given the large number of boys in youth custody with either a serious special needs designation when at school and/or being in a special type of educational institution for their secondary education, a useful direction for future research would be to investigate the efficacy of the types of intervention that occur within these settings.

1. Introduction

There is a known causal relationship between education and crime, which is often explored using changes in compulsory school laws to establish the direction of causality from education to crime (e.g Bell et al, 2022; Bell et al., 2016; Lochner and Moretti, 2004). Much of this work in the economics of crime field focusses on crime measures that do not result in incarceration (like arrests or convictions), although some work does study the likelihood of ending up in prison as an adult. In this paper, we investigate the link between education, specifically GCSE grades, and the probability of entering youth custody in England.

Only a very small proportion of young people enter youth custody (and we restrict our attention here to those aged between 16 and 18), but the consequences are severe – with those in our data spending an average of 7 months in secure centres. Such a penalty is imposed by courts because the crime is so serious there is no other suitable option, or the young person has committed crimes before, or the judge or magistrate thinks the young person is a risk to the public.[1] Apart from being a very negative outcome for young people themselves, this is expensive for the public purse, with the unit cost for youth custody alone being estimated as almost £5,000 per month served in prison back in 2011 (NAO, 2011).

The question is whether education can act as a mediating mechanism to prevent young people ending up in youth custody in the first place. While existing evidence suggests that education is a relevant factor, in this paper we show that young people entering youth custody are often not even entered for GCSE exams. Unsurprisingly, but empirically validated here with administrative data, the correlation between GCSE grades and the probability of ending up in youth custody is driven by variation at the lower end of the distribution (when other factors are accounted for). Enabling a student to get a pass grade might plausibly have some causal effect on their probability of ending up in youth custody. Our results point to the

---

[1] https://www.gov.uk/young-people-in-custody

importance of investigating the efficacy of existing interventions that are already in place for these young people. This is especially the case for interventions that take place in those institutions concentrating the majority of young people that will subsequently enter youth custody, such as Pupil Referral Units.

In this paper, we use administrative data to analyse this issue. We first summarise briefly why crime and education are related (Section 2); then explain the data and methodology used in this analysis (Section 3); discuss descriptive statistics (Section 4); report on our results (Section 5) before concluding (Section 6).

2. Why might education affect crime?

As outlined by Lochner and Moretti (2004), there are several mechanisms through which educational achievement may affect young people's propensity to engage in crime. These are as follows:

(a) Higher educational achievement increases the probability of receiving higher wages in the labour market. This increases the opportunity cost of crime.

(b) Incarceration involves time out of the labour market and this is more costly for the more highly educated (who are likely to get better jobs).

(c) Schooling may change an individual's time preference or risk aversion; more risk-averse or patient individuals will place more weight on the possibility of punishment in the future.

(d) Schooling might directly affect the psychic costs of breaking the law.

(e) Schooling may help keep young people occupied and therefore reduce the time for criminal participation – and this may have larger effects if crime at a point in time is correlated with the future propensity to commit crime.

By now a quite sizable literature establishes that education does have a crime-reducing impact, in a range of settings and time periods (see Lochner, 2011, and Draca and Machin, 2015, for reviews of some of this work). It is also clear that the channels listed in (a) to (e) play a role in explaining why education reduces crime. Evidence has been garnered at times for education have an income/earnings effect that underpins the causal crime-education relationship (for example, in the case of compulsory school leaving reforms in the US in the 1960s, 1970s and 1980s in Lochner and Moretti, 2004, and for the England and Wales compulsory school leaving age reform in the 1972/73 school year in Machin et al, 2011). Incapacitation from being kept in school, and therefore "off the streets", has also been identified as important (for example, see US evidence from Jacob and Lefgren, 2003, and Luallen, 2006, for direct tests of incapacitation and or Bell et al, 2022, who study dynamic incapacitation effects due to shifts in crime-age profiles resulting from compulsory school leaving age law changes between 1980 and 2010).

As achieving higher qualifications at school leaving exams (GCSEs) enables people to move on to the next stage with a broader range of options and therefore become more employable, all of the above mechanisms are potentially relevant in the present context. Machin et al. (2020) show that even those who marginally fail to get a good grade at GCSE increase the risk of ending up 'not in education, employment or training' by age 18. It is plausible that faced with such prospects, young people may turn to crime. In fact, there is evidence to show that if young people are 'not in education, employment or training', this puts them at a high risk of wage scarring effects and crime participation resulting from youth unemployment in the longer term (Gregg and Tominey, 2005; Bell, Bindler and Machin, 2018).

3. Data and methods

We use administrative data from the National Pupil Database (NPD) linked to the National Client Caseload Information System (NCCIS). The former is a census of school students in state schools, where we have information on achievement in primary school (Key Stage 2 tests at age 11), some demographics (ethnicity, gender, whether eligible to receive free school meals), school characteristics, and achievement in national exams at age 16 (GCSE). We use the NCCIS to identify whether the young person ends up in youth custody, together with producing relevant descriptive statistics for that sub-population.

We use the cohorts who did their GCSE exams in the academic years 2011-12, 2012-13 and 2013-14.[2] We link this with exam board data on marks for GCSE English, GCSE English Language (both counting towards the requirement of getting a C in GCSE English) and GCSE mathematics that are available for all years for the AQA exam board for English specifications and available in 2014 for both subjects for all four exam boards in England. The AQA exam board accounts for well over half of all exam entries in GCSE English (see Machin et al. 2020 for further details). As it turns out, the match between individuals who end up in youth custody and their mark in these subjects is very limited - available for only 419 out of 3128 individuals in youth custody for English (2012-2014) and 225 out of 865 individuals for maths (2014).[3] Much of this is due to these individuals not being entered for GCSE exams in the first place. We will explore this further in the next section.

We first use this data to explore descriptive statistics of the sample, understanding how those who enter youth custody differ from the general student population in their observable

---

[2] We denote each academic year by the year in which the exams were taken (i.e. 2011-12 is referred to as 2012).
[3] As the vast majority of those in youth custody that can be linked to marks are entered for exams run by Pearson, we use this awarding body alone for the analysis. Using exam boards together is complicated for maths because of different marking schemes in relation to grade thresholds. The number 3128 refers to the number of males in youth custody. The total number of males and females in youth custody in our sample is 3331.

characteristics. As the vast majority of those in youth custody are male (94%), our analysis is for boys only.

We estimate OLS regressions of the following form (suppressing the subscript for individuals, i)[4]:

$$Y \ = \ \alpha + \ \beta Q + \ \gamma X + \ \rho S + \ \varepsilon \qquad (1)$$

Where Y is a binary variable for whether the individual ends up in youth custody from age 16-18 (i.e. in the two years after students typically complete their GCSE exams); Q is the qualification achieved (if any) at GCSE. Indicators for whether the qualification is entered are also included. We run different regressions where Q is a dummy variable for (a) whether the student achieves 5 good grades at GCSE (A*-C); (b) the grade and entry status for English GCSE; (c) the grade and entry status for maths GCSE. We progressively include controls – X is a vector of individual-level controls; S represents school type; and ε is an error term. In regressions where multiple years are included, we always include a set of year dummies. Then we progressively include the following controls: demographics (ethnicity and whether eligible to receive free school meals); primary school test scores (KS2); dummies for whether the individual has been identified as having 'special needs' or attends a non-conventional institution type (community special school; alternative provision; secure unit; pupil referral unit) and finally, school fixed effects.

Then for a very restricted sample of students – where it is possible to match individuals to marks – we implement a sharp regression discontinuity approach of the following form[5]:

---

[4] Even though there are very few people who are in youth custody, estimates are very similar whether we use the linear probability model (LPM), the Logit or the Probit. Use of the LPM (i.e. OLS where the dependent variable is binary) simplifies our analysis where we want to control for school fixed effects in some specifications and hence we adapt this strategy.

[55] Note that in previous work, Machin et al (2020) have used fuzzy regression discontinuity design strategies to study the impact of failing to get a C grade in GCSE English on educational outcomes three years after GCSEs. We cannot use fuzzy RD here because we are missing the so-called original marks (that is, pre-remarking of exams for some students that ask for this option). However, Machin et al (2020) explain that estimates are very similar between fuzzy and sharp RD due to a very strong (close to 1) first stage.

$$Y = \alpha + \beta CF + \varphi M + \gamma X + \rho S + \varepsilon \quad (2)$$

Where *CF* is an indicator variable for the grade achieved by the student and *M* represents his underlying marks. This regression is estimated for groups of students whose grade falls within narrow categories (i.e. C to D; D-E; E-F; F-G; G-U). There are hardly any students in youth custody that obtain any higher grade than a C and hence grades above this are not relevant. Whereas 'C' is seen as a 'good' grade within the GCSE system, 'G' is another important threshold, especially for low attaining students, because it represents a pass.[6]

4. Descriptive statistics

4.1 Young people in youth custody – activities before and after

Only a very small percentage of students ends in youth custody at the age of 16 or 17. As noted above they are predominately male (94%) and this comes to 3128 individuals out of around 884,000 males who complete Year 11 at (state) secondary school at the same time. The probability of ending up in youth custody is approximately 0.004 or 4 in 1000. Figure 1 shows their activities before and after entering youth custody. For many this is not known, but where this is known, there is no indication of better outcomes after a spell in youth custody relative to before in terms of education, training or employment (to the extent there is much difference). The biggest known category is NEET (not in education, employment or training). Roughly, 35 per cent of the sample are categorised as NEET before a spell of youth custody and 36 per cent after.

---

[6] In more recent years, the GCSE grading system has changed to a numeric grading system, but this does not affect the cohorts of interest here.

**Figure 1: Activities before and after spell in youth custody**
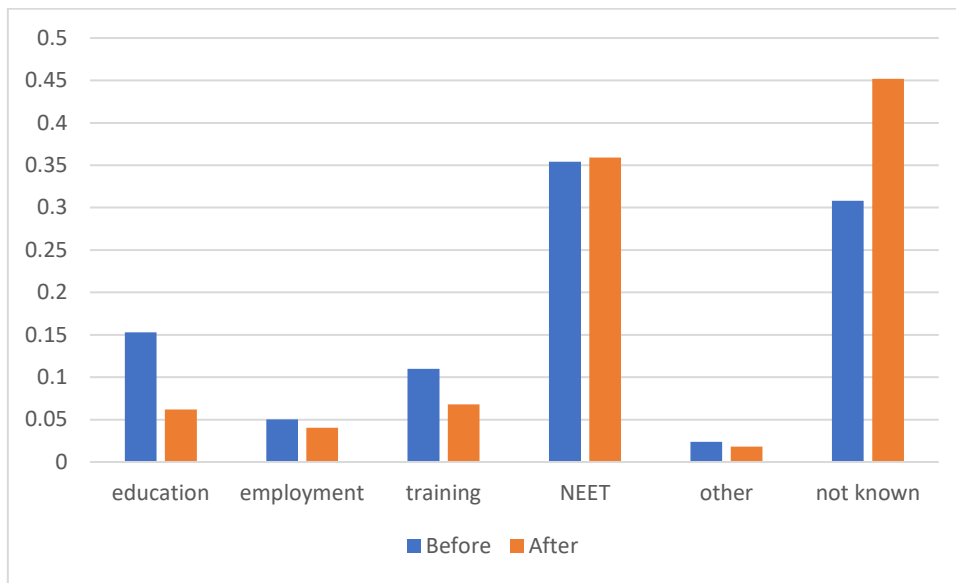
Note. Own calculations based on NCCIS data. The Y axis shows the fraction in each category both before (blue) and after (orange) youth custody.

## 4.2 Educational qualifications in English and maths

We are interested in whether qualifications at age 16 affect the probability of entering youth custody (restricting the sample to males only). Many of those entering youth custody do not even have an entry for GCSE English or maths. This is shown in Figure 2 and 3.

Figure 2: GCSE grades in English or English language (males only)
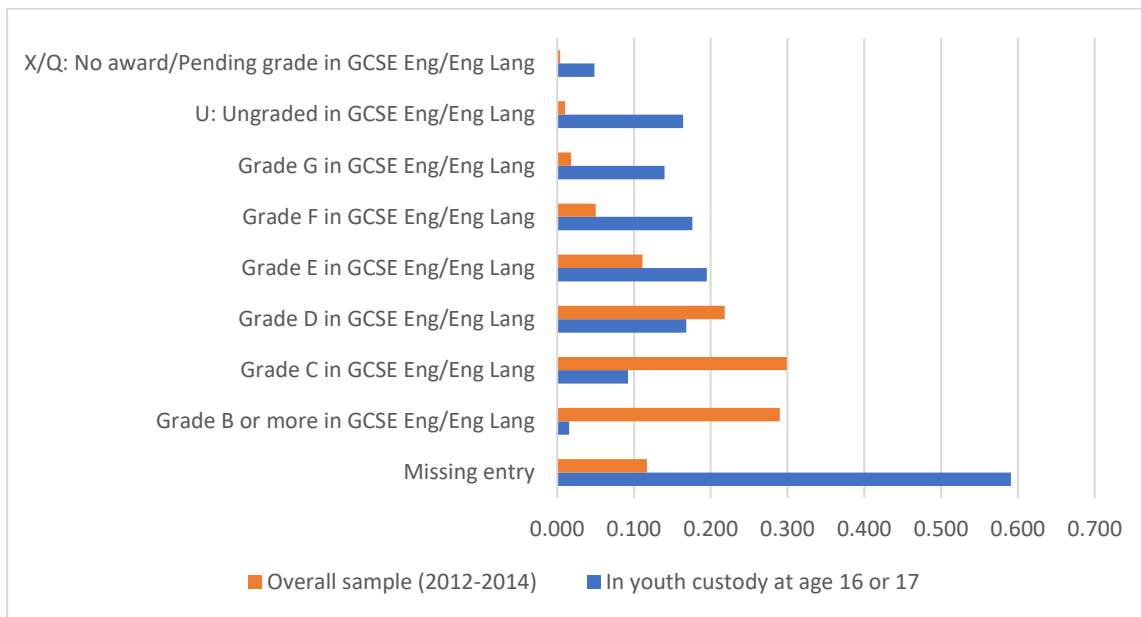
Note: Own calculations based on NCCIS data and KS4 data. The X axis shows the fraction of students falling in each qualification grade for GCSE English/GCSE English Language. Note that.a further 15% of those in youth custody have other types of English entry. This still leaves 45% with no entry at all.

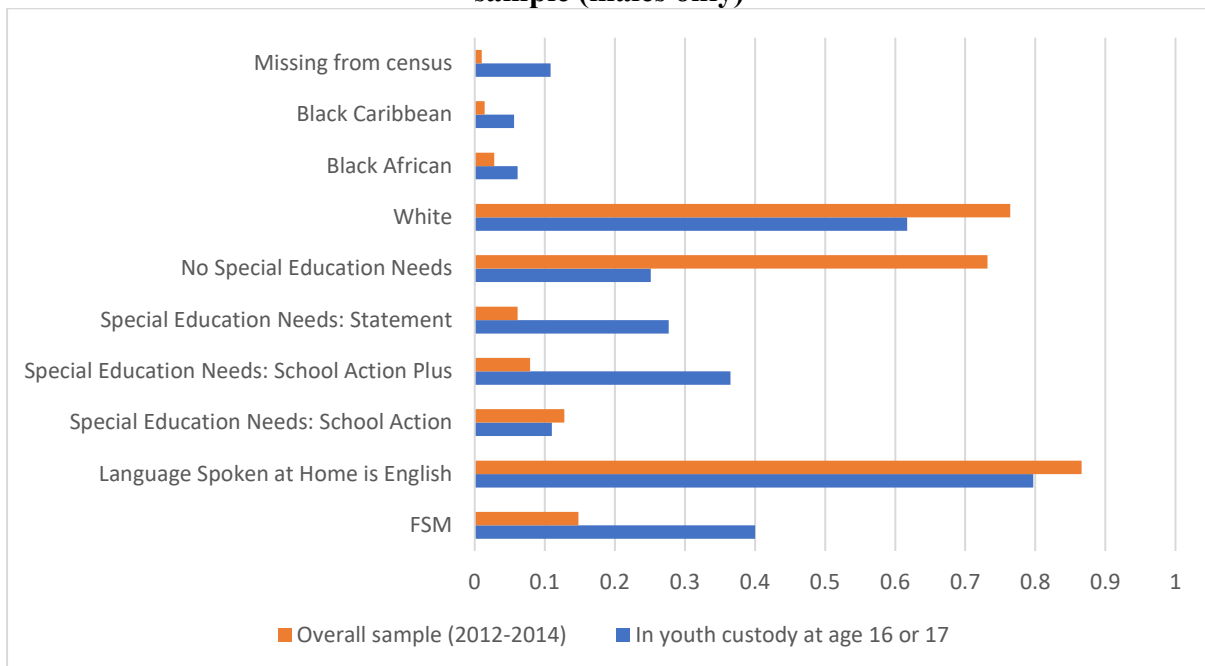**Figure 3: GCSE Grades in Maths (males only)**
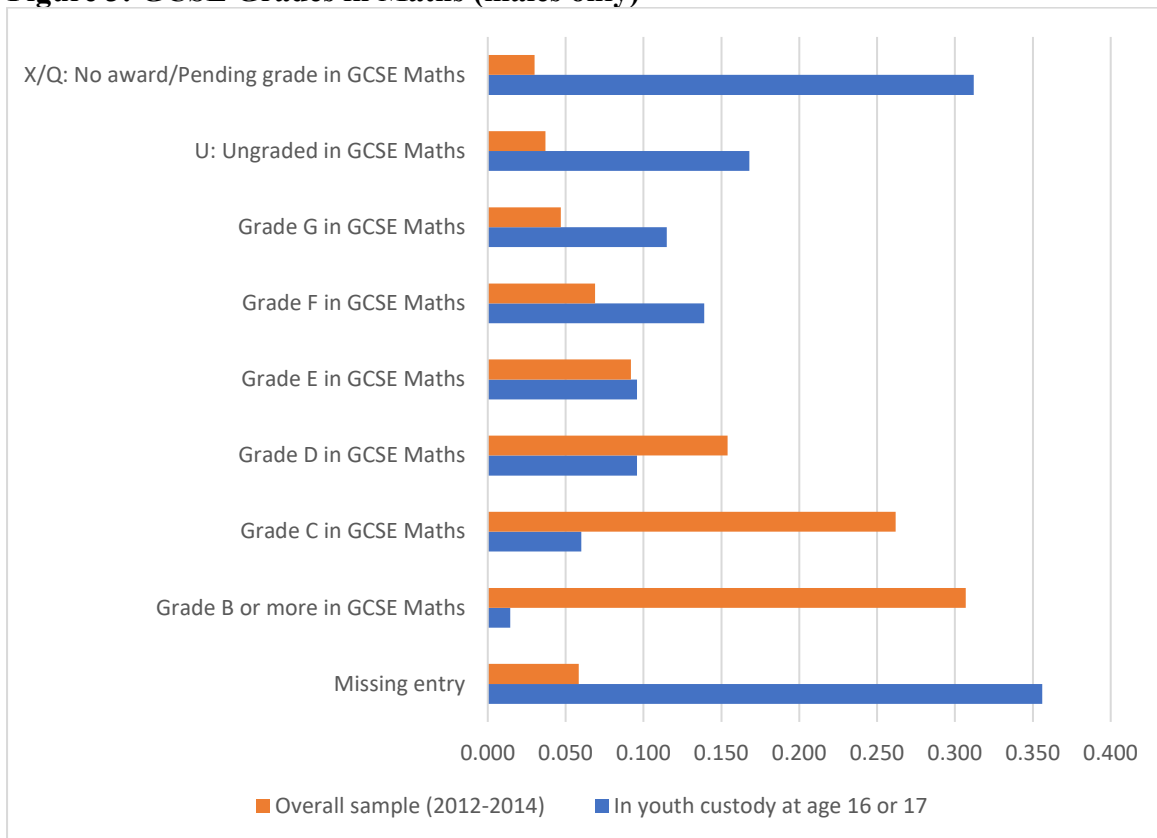
Note: Own calculations based on NCCIS data and KS4 data. The X axis shows the fraction of students falling in each qualification grade for GCSE Mathematics. Note that a further 5% of those in youth custody have other types of maths entry. This still leaves 30% with no entry

As many as 45 per cent of those who subsequently enter youth custody have a missing entry for English (including those who take a non-GCSE English subject in Year 11) and this number is 30 per cent for maths. Where there is an entry, those who enter youth custody at age 16/17 are very likely to have received very low grades relative to other males attending state schools. If we sum those who either have a missing entry (including non-GCSE subjects), no award or a fail in GCSEs, this comes to 66 per cent for English and 78 per cent for maths. Where there is a definite fail of U (as opposed to 'no award' or missing entry), this is 16 per cent for English and 17 per cent for maths, and compares to only 1 and 3.7 per cent for English and maths, respectively, among the rest of the male population. At the other end of the distribution, it is rare for a person who subsequently ends up in youth custody to have done well at GCSE: 9 percent and 6 per cent receive a grade C in English and maths, respectively, and it is very rare to have got a grade B or more.

Thus, there is most certainly a correlation between entry to and grades in GCSE English and maths and the probability of ending up in youth custody at age 16 or 17. However, as so many people destined for youth custody are concentrated at the bottom of the distribution or not entered at all, there are limits to what we can say about the effect of improving skills or human capital at age 16 on the probability of entering youth custody. It is very likely that the reasons why people enter youth custody are similar to reasons why they are not even entered for exams (or fail the exams) and that the correlation between these two things represent the influence of these other factors. In a quantitative study relying on administrative data alone, we cannot say much about what these other factors might be. Nonetheless there are some interesting insights coming from some simple descriptive analysis.

4.3 Demographics and school type

Figure 4: Personal characteristics of those in youth custody compared to the rest of the sample (males only)

Note: Own calculations based on NCCIS data and Census data. The X axis shows the fraction of students falling in each category as described by the Y-axis.

Figure 4 shows personal characteristics of males in youth custody at age 16 or 17 compared to the rest of the male population who were in Year 11 at the same time. There are very striking differences. Those who enter youth custody are much more likely to have been eligible to receive free school meals when at school (i.e. 40% relative to 15% in the rest of the sample). They are much more likely to be Black African or Black Caribbean compared to the rest of the sample (i.e. 6% and 5% of those observed in youth custody at ages 16-17, respectively, are Black African or Black Caribbean relative to 2.8% and 1.4% among other males).[7] They are less likely to speak English as a first language – though the differences with the rest of the sample are not as striking. And they are much more likely to be classified as either having a statement of Special Educational Needs (SEN) or having 'School Action Plus' as a SEN category. There is much less difference in the more basic 'School Action' SEN category. Taken

---

[7] The percentage of those in youth custody from the other individual categories of ethnicity are too small to report.

as a whole, 75 per cent of those males ending up in youth custody at age 16 or 17 were designated under a 'special needs' category while at school (usually a more serious one) whereas this is only 25 per cent among other males (of which about half received the more serious designation of School Action Plus or a Statement).

This very big correlation between SEN and the probability of ending up in youth custody is also reflected in the institution attended when in secondary school. Figure 5 shows the school type attended by males who subsequently went to youth custody compared to all other males attending state schools in Year 11.

Figure 5: School type when in secondary school
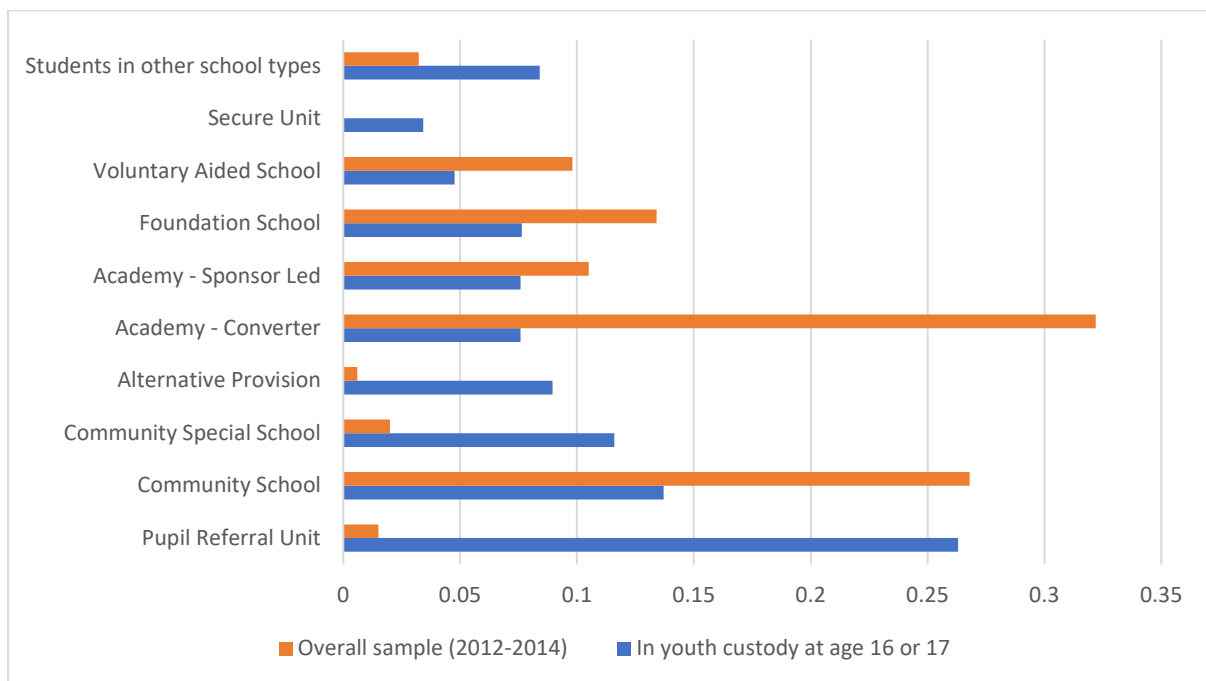
Note: Own calculations based on NCCIS data and KS4 data. The X axis shows the fraction of students falling in each category (school type) as described by the Y-axis.

Figure 5 shows that a high percentage of males who subsequently went to youth custody attended either Pupil Referral Units (26 per cent), Community Special Schools (11.6 per cent), 'Alternative Provision' (9 per cent) or Secure Units (3.4 per cent) compared to a very small

proportion attending these school types in the rest of the sample (4 per cent for all these types of institution collectively).

It seems likely the type of complex needs causing students to enrol in such institutions or receive a serious type of SEN designation are also behind non-entry or very poor grades at GCSE and the eventual outcome of youth custody at age 16 or 17. Of course, there are many more students classified as SEN or in a special type of institution than those who end up in youth custody.[8] Thus, even though the correlations are high, there is no simple, mechanical relationship. Appendix A shows that the prevalence of SEN or attendance at 'special institutions' is even greater among those students who end up in youth custody and have no English entry.

Most students do not start out as having SEN or attending some type of special institution. We next look at their test scores in primary school compared to the general population.

4.4 Educational attainment in primary school

Figures 6 and 7 show that attainment in primary school at age 11 in English and maths tests (at Key Stage 2) were already very much worse for those entering youth custody later. But the representation of the sample who go to youth custody is much better here than it was when looking at KS4 results. For example, Key Stage 2 data for English is only missing for about 7 per cent of people who subsequently enter youth custody.

---

[8] The percentages might lead one to believe that there is hardly anyone in, for example, secure units, that do not subsequently end up in youth custody. But this is deceptive as the number of those not in youth custody is such a big number (888774). Numerically, about 38 per cent of males attending secure units end up in youth custody (and hence, most do not).

Figure 6: English Key Stage 2 test results
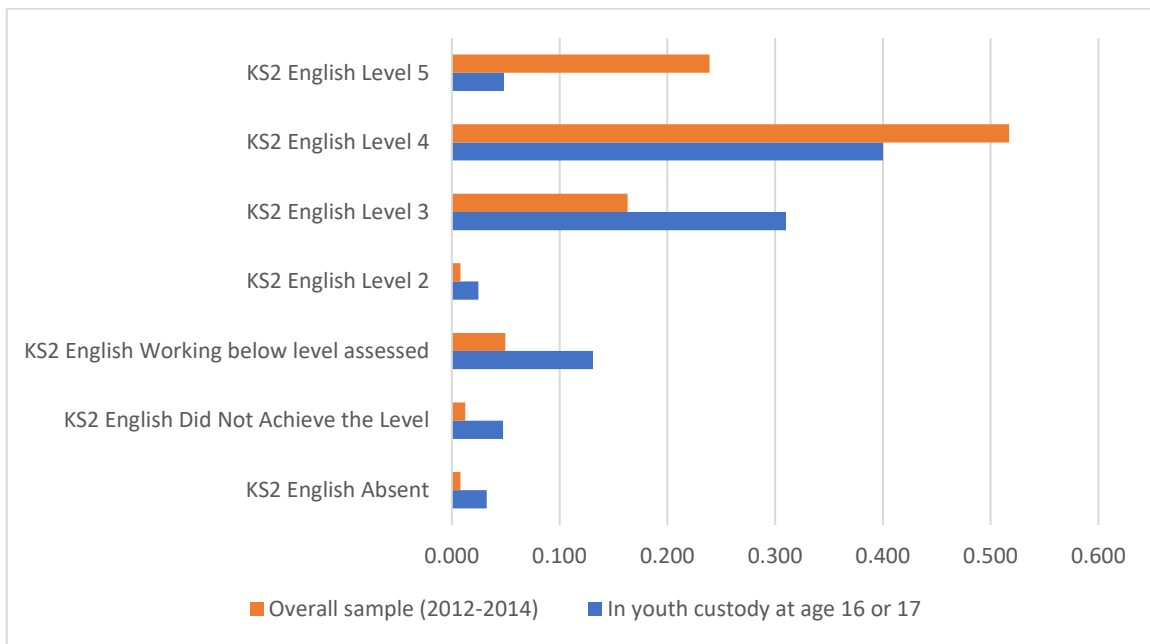
Note: Own calculations based on NCCIS data and KS2 data. The X axis shows the fraction of students falling in each category (KS2 English results) as described by the Y-axis.

Figure 7: Maths Key Stage 2 test results
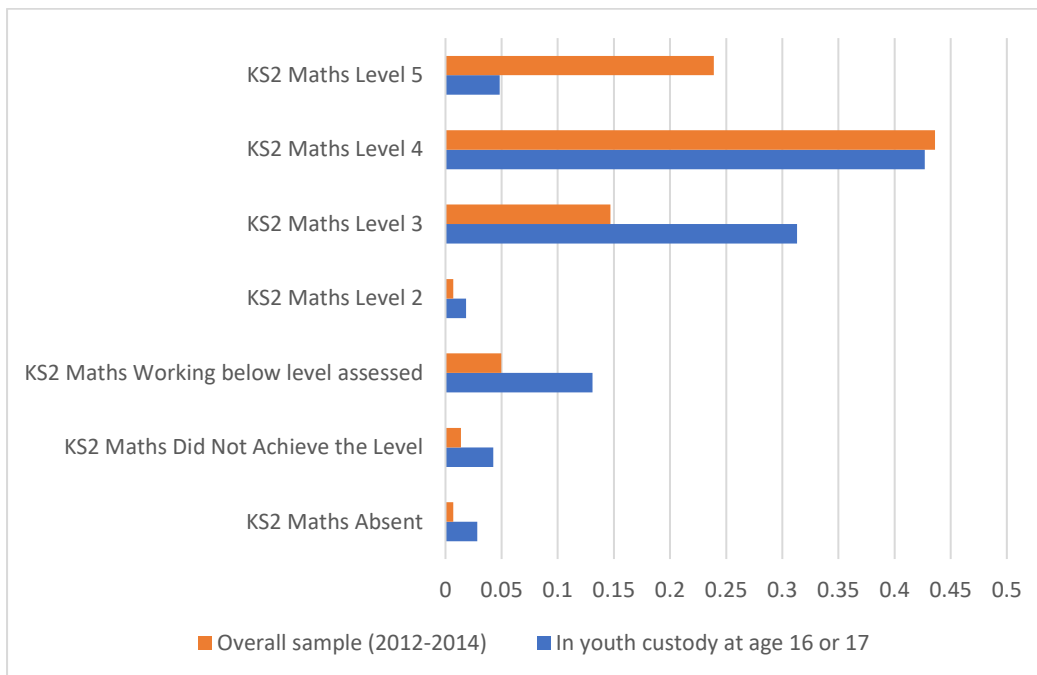
Note: Own calculations based on NCCIS data and KS2 data. The X axis shows the fraction of students falling in each category (KS2 maths results) as described by the Y-axis.

The negative correlation between educational achievement and the probability of ending up in youth custody is evident from an early age. Yet, many students who subsequently ended up in

youth custody received an acceptable grade at age 11 while failing completely to do so at age 16. Among those with some form of entry for Key Stage 2 (over 90% of those who entered youth custody), 45 per cent and 48 per cent met the thresholds of Level 4 or above for English and maths respectively compared to 76 per cent and 68 per cent for English and maths, respectively, among other male students.[9] The relatively better average performance in primary school than at secondary school for students ending up in youth custody suggests that underlying problems become evident (in terms of education) in early/mid adolescence for many of these individuals.

5. Regression results

The above discussion suggests that the correlation between an individual's educational achievement and their probability of later entering youth custody is driven by many factors that are not well captured in administrative data. Many who enter youth custody have no qualifications at GCSE and hence we cannot conclude very much about the relationship between underlying human capital or skills and crime in this context. Nonetheless, it is of interest to analyse how the correlation between educational qualifications and the probability of youth custody changes as we try to account for contextual factors – even though these factors (such as SEN or institutional type) reflect the influence of variables not captured in administrative data rather than these designations in themselves. It also remains extremely likely that we are not fully capturing (even indirectly) those factors that are both correlated with non-entry/poor exam grades and the probability of entering youth custody. The more rigorous method – the Regression Discontinuity approach – described above, can be applied only to a very small number of students who end up in youth custody. Unfortunately, the results

---

[9] These are thresholds deemed at the 'expected level' according to National Curriculum guidelines.

are for the most part under-powered to find statistically significant results, though they provide interesting suggestive evidence on the potential influence (or lack thereof) of GCSE grades.

5.1 The association between good grades overall at GCSE and youth custody

Table 1 shows the association between whether a male student achieves 5 good grades at GCSE (including English and maths) and the probability of ending up in youth custody at age 16 or 17. As discussed in Section 3, we successively include various controls so as to observe how this association is mediated by various individual and institutional characteristics described above. Results are reported in Table 1. Column 1 shows the correlation while only controlling for year dummies; column 2 controls for whether the individual is eligible for free school meals and his ethnicity; column 3 augments the specification by including Key Stage 2 scores in English and maths; column 4 adds dummies for school type and whether the student has been identified as having special educational needs. Finally, column 5 adds school fixed effects. As demographics and Key Stage 2 results are included (in columns 2 and 3), the association between good grades and going to youth custody declines about equally (by roughly 14 per cent in each specification). This reduces more dramatically when SEN status and school type are included (i.e. dummies for the four 'special institutions' described above) and is unchanged by additionally controlling for school fixed effects. If the final estimate in column (4) or (5) were to be taken at face value, this would suggest that enabling students to achieve 5 good grades at GCSE (including English and Maths) would half the probability of entering youth custody from 0.004 to 0.002 or from 4 in 1000 to 2 in 1000. This would be a very strong 'effect' and very unlikely to be causal for the reasons discussed above. Most of the other estimates (reported below) are at least as high. The probability of entering youth custody in the sample may be extremely small, but there are some high associations with individual and institutional characteristics contained in administrative data. This is helpful for identifying the vulnerable population.

Table 1: Relationship between youth custody and whether student achieves 5 or more good grades at GCSE including English and maths.

	(1) Acad Year	(2) +FSM ETH	(3) +KS2	(4) +SpecSch SEN	(5) +KS4 school Fes
5+GCSEs (incl E & M)	-0.007***	-0.006***	-0.005***	-0.002***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
FSM		0.004***	0.004***	0.003***	0.001***
		(0.000)	(0.000)	(0.000)	(0.000)
black_african		0.001**	0.001**	0.002***	0.002***
		(0.001)	(0.001)	(0.001)	(0.001)
black_caribbean		0.007***	0.007***	0.005***	0.005***
		(0.001)	(0.001)	(0.001)	(0.001)
white_british		-0.002***	-0.002***	-0.002***	-0.002***
		(0.000)	(0.000)	(0.000)	(0.000)
KS2 points score			-0.001***	-0.000	0.000
			(0.000)	(0.000)	(0.000)
SEN Action				0.000**	0.001***
				(0.000)	(0.000)
SEN Action Plus				0.006***	0.006***
				(0.000)	(0.000)
SEN Statement				0.006***	0.003***
				(0.001)	(0.001)
Special School				0.013***	
				(0.001)	
Alternative provision				0.051***	
				(0.003)	
Secure Unit				0.670***	
				(0.037)	
Pupil Referral Unit				0.054***	
				(0.002)	

Notes: robust standard errors. Controls for year dummies and missing variable dummies also included.
N=888,774

5.2 The association between good grades in English and maths and youth custody

Figures 8 and 9 show results from regressions where the explanatory variable is in the form of dummy variables for the grade achieved in English and maths respectively (in different sets of regressions). The baseline or omitted category is whether the student achieves Grade C or above. The graphs show how the coefficients change as various controls are added (in the same way as described above). The equivalent regression results are shown in Appendix B.

These results show that most of the correlation between specific grades and the probability of entering youth custody is driven by variation at the bottom of the distribution, from Grade F or lower (or non-entry). This is unsurprising given the descriptive statistics discussed above. When full controls are added, there is no association between getting either grade D or E (relative to a grade C) and the probability of later entering youth custody. But there is an association at levels below that, even with controls. For grades F and G (relative to grade C) this is clearer for English than for maths. Failing either subject (grade U) or not being entered is clearly strongly correlated with entering youth custody – though it is difficult to interpret differences between these residual categories. In English, not being entered is a stronger correlate of youth custody compared to failing but the opposite is true for maths.

Overall, the association between GCSE grades and the probability of entering youth custody is consistent with educational achievement having a role to play for influencing this outcome. But this analysis has mainly served to highlight that there are likely to be common unobserved factors influencing both non-entry/failing GCSE and entering youth custody.

Figure 8. The association between English GCSE grade and youth custody
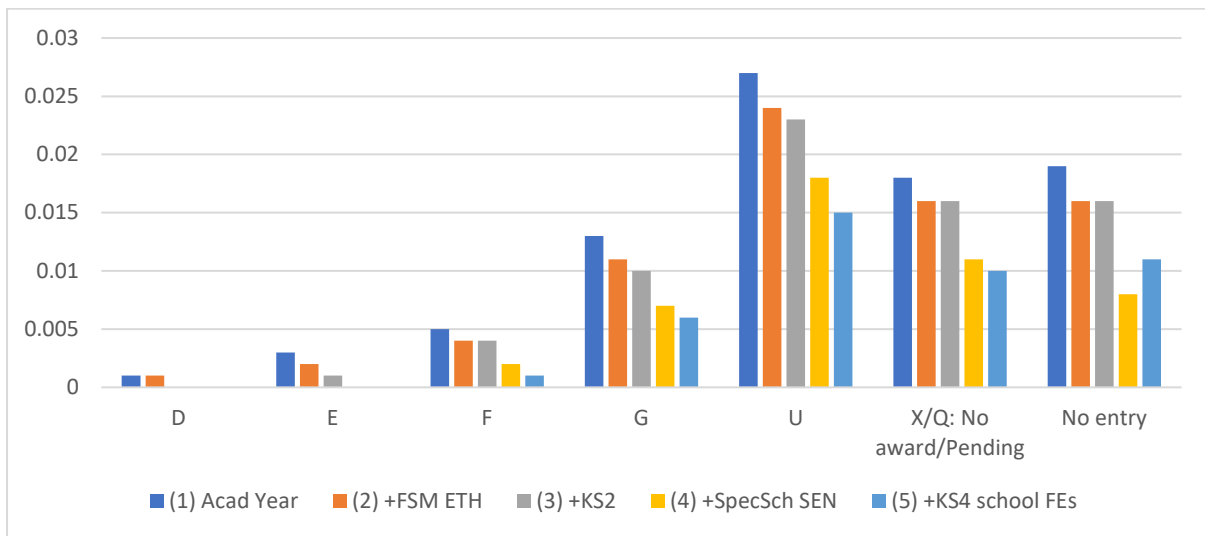
Note. Coefficient on grade (relative to C and above) for GCSE English as controls are added. See Appendix Table B1.

Figure 9. The association between maths GCSE grade and youth custody
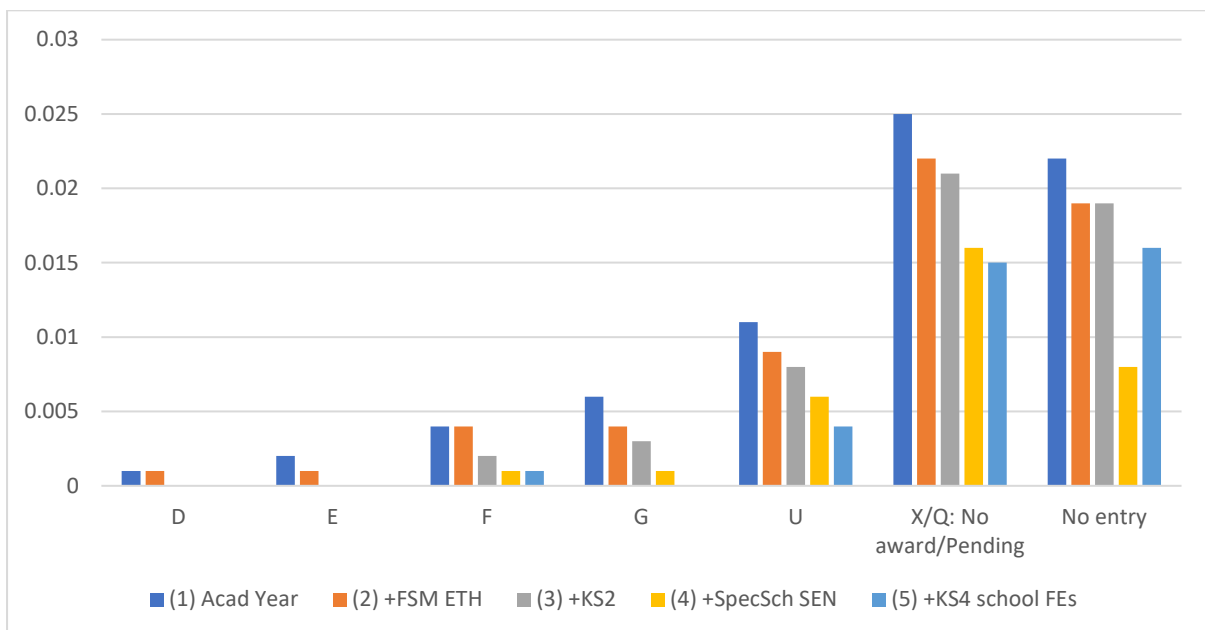
Note. Coefficient on grade (relative to C and above) for GCSE English as controls are added. See Appendix Table B2.

5.3 A Regression Discontinuity Approach

As described in Section 3, we implement a sharp regression discontinuity approach within the small sample of students who can be linked to a GCSE mark within the following grade boundaries: C-D; D-E; E-F; F-G and G-U. We do this from 2012-2014 for those students who can be linked to AQA for English marks and in 2014 for those students who were entered to the Pearson awarding body (i.e. the main provider) for maths scores. The number of individuals who have a mark and end up in youth custody is very low. This has two consequences for our analysis: first, any results here do not extrapolate to most individuals who enter youth custody, for whom we have no mark – usually because they are ungraded or not entered for the qualification; secondly, the analysis is underpowered to find results that are statistically different from zero – making it difficult to interpret the results.

Table 2 presents results where columns (1) to (3) show results for English and (4) to (6) for maths. In columns (1) and (4), only marks to the left and right of the relevant grade threshold are included (i.e. the forcing or running variable). In colums (2) and (5) controls are additionally included for year where relevant (i.e. for English only), number of GCSE entries and GCSE English (or maths) entries, KS2 results and individual level controls for free school meal eligibility, ethnicity and SEN status. In columns (3) and (6), school fixed effects are also included.

For the most part, coefficients are close to zero and not statistically significant. As the number of students both in the sample and in youth custody is very small in each case, it would be difficult to distinguish between a meaningful 'zero' and a very small causal effect that cannot be detected. To the extent there is some hope of the latter, it is most plausibly around the G-U boundary as this represents the pass-fail distinction and hence may have some positive correlation with future opportunities for the young person (whether in education or the labour market). In the G-U boundary, in every specification the estimated effect of passing has a

negative relationship with the probability of entering youth custody (and thus it has a sign in line with the hypothesis one would have according to previous literature).However, it is only significant in one of six specifications (the most detailed specification for maths – column 6). This may reflect causality but it could also have occurred by chance. The size of the coefficient is sizeable in relation to the mean of the dependent variable and thus is consistent with results from OLS regressions. It is certainly plausible that enabling people to achieve a better education would help keep them out of youth custody – though this analysis suggests it might be very challenging to do so given the signifcant vulnerabilities faced by this group of young people and reflected in their characteristics.

Table 2: Regression Discontinuity results

	English			Maths		
Sample C-D						
	(1)	(2)	(3)	(4)	(5)	(6)
Grade C	-0.00019	-0.00027	-0.00033	0.00040	0.00042	0.00032
	(0.00025)	(0.00025)	(0.00025)	(0.00037)	(0.00037)	(0.00036)
ymean	0.00074	0.00074	0.00074	0.00068	0.00068	0.00068
N	194251	194251	194251	87185	87185	87185
Sample D-E Grade						
D	0.00064	0.00067	0.00068	0.00048	0.00052	0.00007
	(0.00047)	(0.00047)	(0.00047)	(0.00095)	(0.00095)	(0.00095)
ymean	0.00140	0.00140	0.00140	0.00140	0.00140	0.00140
N	117810	117810	117810	39191	39191	39191
Sample E-F						
Grade E	0.00015	0.00009	0.00053	-0.00217	-0.00218	-0.00208
	(0.00108)	(0.00107)	(0.00102)	(0.00161)	(0.00163)	(0.00169)
ymean	0.00259	0.00259	0.00259	0.00212	0.00212	0.00212
N	54815	54815	54815	20710	20710	20710
Sample F-G						
Grade F	0.00228	0.00255	0.00410*	-0.00100	-0.00089	-0.00119
	(0.00213)	(0.00213)	(0.00212)	(0.00155)	(0.00155)	(0.00152)
ymean	0.00469	0.00469	0.00469	0.00236	0.00236	0.00236
N	20690	20690	20690	16107	16107	16107
Sample G-U Grade						
G	-0.00856	-0.00778	-0.00913	-0.00206	-0.00226	-0.00364*
	(0.00588)	(0.00585)	(0.00587)	(0.00182)	(0.00184)	(0.00197)
ymean	0.00927	0.00927	0.00927	0.00429	0.00429	0.00429
N	6690	6690	6690	15609	15609	15609

Note: the dependent variable is whether the individual is in youth custody at age 16 or 17. Each panel represents a separate set of regressions. Column 1 controls for the running variable (i.e. final marks) only; column 2 additionally controls for individual level characteristics (see text); and Column 3 also controls for school fixed effects. ymean is the mean of the dependent variable in each specification.

6. Conclusion

Individuals entering youth custody at a young age are typically male and are much more likely to come from disadvantaged backgrounds than the population as a whole. They are also much more likely to have a serious special needs designation (Statement or School Action Plus) and/or more likely to be in a special type of institution for secondary school. Although like the rest of the population, they are more likely to be White British than any other ethnicity, those from Black Caribbean or Black African backgrounds are greatly over-represented amongst those who end up in youth custody.

Against this background, it is not surprising to see that there is a high negative correlation between educational achievement at GCSE and the probability of ending up in youth custody. In fact, the majority of those in youth custody have a missing entry, no award or a fail for English (66 per cent) and/or maths (78 per cent), even if we include non-GCSE subjects in English and maths for this age group. This means, for the majority of individuals in youth custody, we cannot tell from these data the extent to which incremental changes in skill/human capital would make a difference to their probability of entering youth custody. But the results are consistent with some relationship between getting a qualification (even if low-level) and this probability.
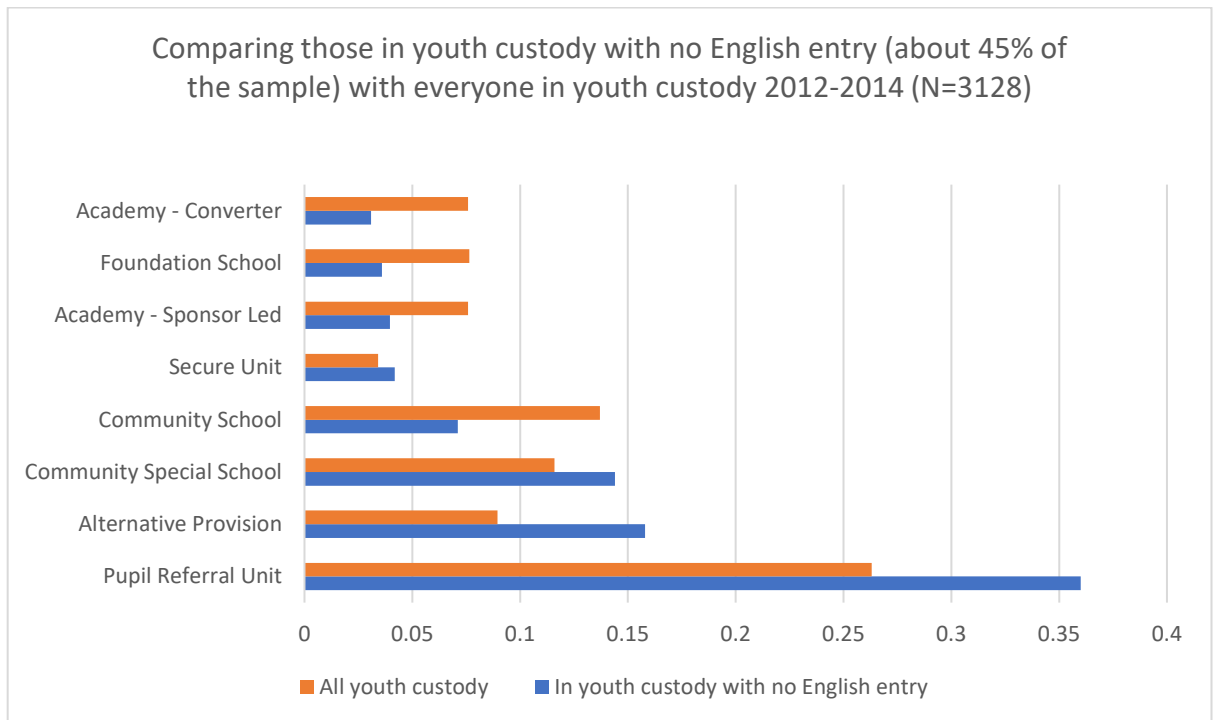
Many individuals ending up in youth custody have problems that become evident in early adolescence (as primary school test scores were typically better). Also, many students vulnerable to ending up in youth custody have previously been designated as special needs and/or gone to a special type of institution. This points to the need to understand what types of intervention are being undertaken within these SEN or institutional categories and to discover the extent to which they are effective. In future work, it would also be interesting to try to think of research designs that would enable analysis of the causal influence of education and skills

on youth custody – though clearly qualifications alone are not the most useful measures of educational achievement in this context.

References

Bell, B., R. Costa, and S. Machin. (2016). Crime, Compulsory Schooling Laws and Education, *Economics of Education Review*, 54, 214-26.

Bell, B., R. Costa, and S. Machin. (2022). Why Does Education Reduce Crime, *Journal of Political Economy*, 130, 732-65.

Bell, B., A. Bindler and S. Machin (2018). Crime Scars: Recessions and the Making of Career Criminals. *Review of Economics and Statistics*, 100, 392-404.

Draca, M. and S. Machin (2015). Crime and Economic Incentives. *Annual Review of Economics,* 7, 389-408.

Gregg, P. and E. Tominey (2005). The Wage Scar from Male Youth Unemployment. *Labour Economics*, 12, 487-509.

Jacob, B. and L. Lefgren (2003). Are Idle Hands the Devil's Workshop? Incapacitation, Concentration and Juvenile Crime. *American Economic Review*, 93, 1560-77.

Lochner, L. (2011). Non-Production Benefits of Education: Crime, Health, and Good Citizenship. In Hanushek E., S. Machin and L. Woessmann (eds.) Handbook of the Economics of Education, Volume 4, Amsterdam: Elsevier.

Lochner, L. and E. Moretti (2004). The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports. *American Economic Review*. 94, 155-189.

Luallen, J. (2006). School's Out… Forever: A Study of Juvenile Crime, At-Risk Youths and Teacher Strikes. *Journal of Urban Economics*, 59, 75-103.

Machin, S., O. Marie, and S. Vujic (2011). The Crime Reducing Effect of Education. *Economic Journal*, 121, 463-84.

Machin, S., McNally, S., and Ruiz-Valenzuela, J. (2020). Entry Through the Narrow Door: The Costs of Just Failing High Stakes Exams. *Journal of Public Economics*. 190.

NAO (2011). The cost of a cohort of young offenders to the criminal justice system. Technical Paper. National Audit Office. https://www.nao.org.uk/wp-content/uploads/2010/12/1011663_technical_paper.pdf

Appendix A: Characteristics of boys in youth custody with no English GCSE entry compared to the full sample of boys in youth custody.

Comparing those in youth custody with no English entry (about 45% of the sample) with everyone in youth custody 2012-2014 (N=3128)

SEN status of those in youth custody overall In youth custody with no English entry All youth custody and without an English GCSE entry (SEN info missing for about 10%)
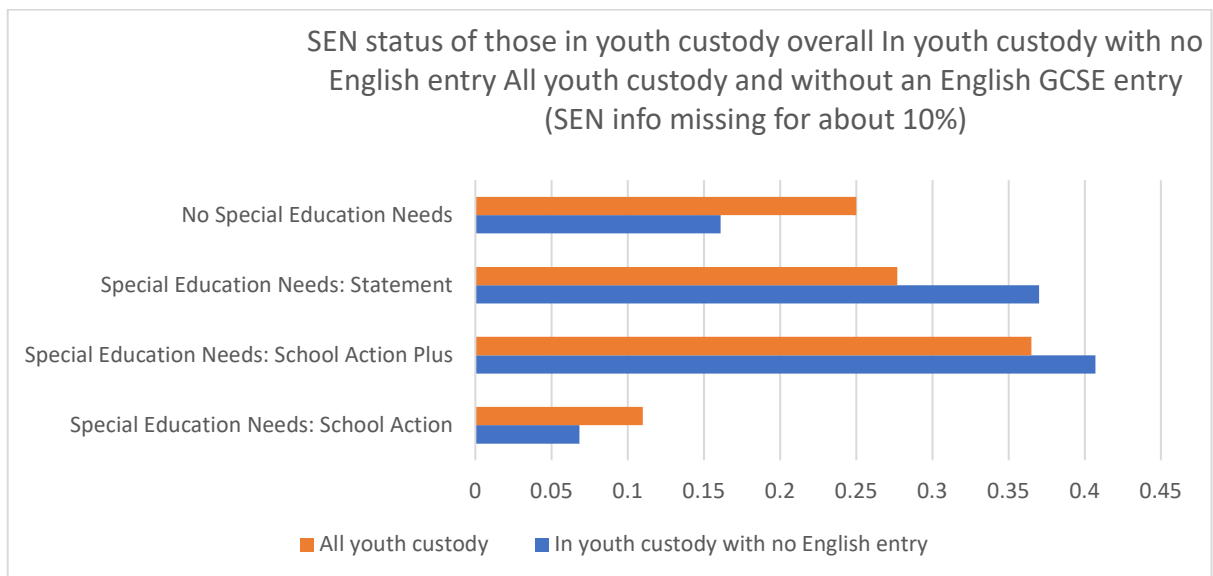
Appendix B: Regressions for English and maths qualifications

B.1. Youth custody and grade in GCSE English.

	(1) Acad Year	(2) +FSM ETH	(3) +KS2	(4) +SpecSch SEN	(5) +KS4 school FEs
D	0.001***	0.001***	-0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
E	0.003***	0.002***	0.001***	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
F	0.005***	0.004***	0.004***	0.002***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
G	0.013***	0.011***	0.010***	0.007***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
U	0.027***	0.024***	0.023***	0.018***	0.015***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
X/Q: No award/Pending	0.018***	0.016***	0.016***	0.011***	0.010***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
No entry	0.019***	0.016***	0.016***	0.008***	0.011***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)

Notes: N=888,774. Robust standard errors in parenthesis. Regression specifications as described in text for Table 1. Coefficient estimates reported for variable of interest only. Full details available on request. Base category is whether young person achieves a grade C or more.

B.2. Youth custody and grade in GCSE maths

	(1) Acad Year	(2) +FSM ETH	(3) +KS2	(4) +SpecSch SEN	(5) +KS4 school FEs
D	0.001***	0.001***	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
E	0.002***	0.001***	0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
F	0.004***	0.004***	0.002***	0.001***	0.001**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
G	0.006***	0.004***	0.003***	0.001**	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
U	0.011***	0.009***	0.008***	0.006***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
X/Q: No award/Pending	0.025***	0.022***	0.021***	0.016***	0.015***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
No entry	0.022***	0.019***	0.019***	0.008***	0.016***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

Notes: N=888,774. Robust standard errors in parenthesis. Regression specifications as described in text for Table 1. Coefficient estimates reported for variable of interest only. Full details available on request. Base category is whether young person achieves a grade C or more.

CENTRE FOR ECONOMIC PERFORMANCE
Occasional Papers