Working Paper No. 22-05

Walking the line: Does crossing a high stakes exam threshold matter for labour market outcomes?

Oliver Anderson
University College London

This paper offers new insight into the link between success in high stakes exams and subsequent education and labour market outcomes. It is the first study to look holistically at the impact of crossing an important high stakes threshold on both academic *and* vocational education choices and ultimately labour market outcomes. It does so by comparing those either side of a formerly important threshold in the English education system at the end of compulsory schooling (achieving five general certificate of secondary education A* to C passes) which was commonly regarded as the minimum benchmark for continuing into post-compulsory education. I find that crossing this threshold led to an 6.3-6.7 percentage point increase in the proportion of men and women (respectively) going on to take academic qualifications, with little change in the proportion taking vocational qualifications, leading to a net increase in those staying on after compulsory schooling. Women's daily earnings in 2017-18 (11-13 years after leaving compulsory schooling) were 3.1 percentage points higher for those just crossing the threshold, but men's early labour market outcomes were unchanged. The results for men can be explained by low returns to academic qualifications for marginal learners. The findings for women do not disappear after accounting for subsequent education choices, suggesting that crossing the threshold may play a signalling role for employers as well as education institutions.

VERSION: April 2022

CENTRE FOR EDUCATION POLICY & EQUALISING OPPORTUNITIES

Disclaimer

Any opinions expressed here are those of the author(s) and not those of the UCL Institute of Education. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

CEPEO Working Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

Highlights

- There is a small amount of literature on the effect of crossing important compulsory high stakes exams thresholds on labour market outcomes. This literature shows mixed results.
- Using English administrative data, I examine crossing an important threshold in the English education system at the end of compulsory education on early career labour market outcomes. I am able to link crossing the threshold with post compulsory education, including both academic and vocational pathways, and ultimately labour market outcomes.
- I show that crossing the five GCSEs A* to C threshold (the one I focus on) leads to an increase in academic education for both men and women, and an increase in women's early career earnings, but not for men.
- The results are explained by returns to higher levels of academic education being positive and sizeable for the marginal female learner, but returns for the male marginal learner being smaller.
- These results show for women show that whilst crossing the threshold leads to higher early career earnings due to the increased uptake of academic education, this only explains some of the differences. Therefore, crossing the threshold may send some sort of signal to employers (in addition to education institutions).

Why does this matter?

Understanding the long-term consequences of just passing or achieving a certain grade/mark, just failing to pass or not achieving certain grade/mark in high stakes tests, is crucial for understanding the economic and social consequences of how we design education systems.

Walking the line: Does crossing a high stakes exam threshold matter for labour market outcomes?

Oliver Anderson

Abstract

This paper offers new insight into the link between success in high stakes exams and subsequent education and labour market outcomes. It is the first study to look holistically at the impact of crossing an important high stakes threshold on both academic *and* vocational education choices and ultimately labour market outcomes. It does so by comparing those either side of a formerly important threshold in the English education system at the end of compulsory schooling (achieving five general certificate of secondary education A* to C passes) which was commonly regarded as the minimum benchmark for continuing into post-compulsory education. I find that crossing this threshold led to an 6.3-6.7 percentage point increase in the proportion of men and women (respectively) going on to take academic qualifications, with little change in the proportion taking vocational qualifications, leading to a net increase in those staying on after compulsory schooling. Women's daily earnings in 2017-18 (11-13 years after leaving compulsory schooling) were 3.1 percentage points higher for those just crossing the threshold, but men's early labour market outcomes were unchanged. The results for men can be explained by low returns to academic qualifications for marginal learners. The findings for women do not disappear after accounting for subsequent education choices, suggesting that crossing the threshold may play a signalling role for employers as well as education institutions.

Acknowledgements

I would like to thank Dr Claire Crawford, Dr Gill Wyness and Professor Lindsey Macmillan for helpful comments.

1. Introduction

It is well established that the results of high stakes exams, such as the awarding of a high school diploma, have a significant effect on subsequent education outcomes (Machin et al 2020, Canaan and Mouganie 2018, Dee at al 2016, Diamond and Persson 2016, De Philippis 2016, Ebenstein et al 2016). Given the strong causal impact of education on labour market outcomes, this could mean that just passing a high stakes exam, or just attaining a certain grade, has a long-term impact on labour market outcomes. However, the evidence here is more mixed. For example, Clark and Martorell (2014) found there was no impact of just achieving a high school diploma in the United States on earnings, whilst Canaan and Mouganie (2018) found there was a 12.5 percent increase in earnings associated with just passing the Baccalaureate in France. Importantly, no previous study has been able to explore the links between just passing or just failing important exams, and both labour market outcomes and the full range of educational pathways, that individuals falling just either side of the threshold may go on to take.

This paper offers new insight into the relationship between just crossing important exam thresholds and future earnings, and in particular to what extent any effects on labour market outcomes are mediated by education choices. It does this by using detailed administrative records on compulsory exams taken at age 16 in England linked to subsequent education and tax records up to age 29. I focus on a formerly important threshold in the English exam system: achieving five General Certificate of Secondary Education (GCSE) passes A* to C. This threshold still applied to the cohorts under consideration in this paper, relating to attainment at age 16 which was concentrated around the middle of the overall results distribution. This threshold was seen as a secondary school pass and was typically regarded as the benchmark required in order to continue on the academic track in post-compulsory education. It is comparable to a high school diploma in the United States in stature (though the level and age taken is different). Crucially, I am able to explore whether post 16 education choices and outcomes mediate any observed differences in earnings. I focus on labour market outcomes in the latest tax year I have data available for, 2017-18, when my cohorts were aged 27 to 29, i.e., 11 to 13 years after the exams took place, and link this to education outcomes at age 25 (nine years after exams took place).

I do not have access to exam marks, but do have a rich set of socio-economic and demographic characteristics, as well as the full basket of qualifications, subjects and grades taken by young people at the end of compulsory education (and earlier in the education system). I use overall attainment at age 16 as a kind of running variable and estimate the effect of crossing the threshold using regression analysis with a rich set of controls. The key assumption I make is that those either side of the threshold are otherwise identical and thus crossing the threshold is due to arguably random factors which are not fully under the control of the individual or other stakeholders. For example, a number of studies have shown that external factors such as temperature (Park 2020) and air pollution (Ebenstein et al 2016) can impact upon exam performance and there can be measurement error when exams are marked (Rhead et al 2016, Rhead et al 2018). This would lead to allocative efficiency and fairness issues, if future education and labour market outcomes were determined strongly by such factors.

I find that crossing the threshold in high stakes exams at age 16, despite leading to higher post 16 education levels for both men and women, has little impact on labour market outcomes for men, but some for women. I go on to show that this is because men who just cross the threshold do not appear to benefit from taking advanced and

higher-level academic qualifications, whilst women do. This finding is similar to other studies that focus on the marginal learner (Dearden et al 2004), i.e., those at the margins of an educational investment and also chimes with the wider literature focusing on average returns, which shows that women tend to have stronger returns to academic qualifications than men (Belfield et al 2018b, Walker and Zhu 2003, Dearden et al 2004). The finding of higher earnings for women is partially, but not fully, explained by the fact that those just crossing the threshold are more likely to stay on to take academic post-compulsory qualifications, suggesting that the threshold may act as a signal to employers of the underlying productivity of these women compared to their peers who do not cross the threshold.

This paper makes two important contributions to the literature and also has some important policy implications. Firstly, I offer a holistic picture of the impact of marginally attaining or not attaining a grade/mark in high stakes qualifications by exploring both education and labour market outcomes, including the educational route taken, for men and women separately. My analysis highlights the importance of splitting by gender and not only focusing on short term outcomes, but also on longer term outcomes to obtain a full picture. Machin et al (2020) examine the impact of crossing an important English language exam threshold but are only able to consider short term education outcomes (at age 19). Clark and Martorell (2014) compare the earnings of those barely passing or barely failing exit exams that lead to high school diplomas in Texas and find that just crossing the threshold does not impact upon college enrolment or subsequent earnings, but are not able to consider a wider range of education or training routes. Canaan and Mouganie (2018) focus on high stakes exams at the end of secondary school in France, finding that those just passing go on to earn higher early career salaries, which can be explained by the fact that they are more likely to access higher 'quality' higher education (but not higher quantities), but again they focus solely on subsequent higher education choices.[1] By focusing on both academic and vocational pathways I am able to offer new insight.

Secondly, it speaks to the human capital versus signalling debate. Human capital theory suggests that investments in human capital lead to better labour market outcomes by increasing productivity (Smith 1776, Marshall 1890, Becker 1964, Mincer 1974). The signalling model, on the other hand, suggests that education has no effect on productivity but is simply a way for the market to overcome asymmetry of information, with employers using education as a 'signal' of an individual's underlying productivity (Spence 1973, Weiss 1995, Riley 2001, Caplan 2018). Differentiating between human capital theory and signalling is extremely difficult because both theories imply a positive relationship between education and earnings.

Comparing the earnings of individuals just crossing or just failing to cross important exam thresholds has been argued to be an appropriate way to test between the models, as those just passing have a piece of paper signalling their underlying abilities to other stakeholders, while those just failing, while having a similar amount of knowledge, do not. Clark and Martorell (2014), for example, interpret their results as providing no evidence in favour of the signalling model, because those just

[1] In addition, there are several papers that focus on admission exams to colleges and universities to establish causal links between education and labour market outcomes (Hoekstra 2009 in the USA, Saavedra 2014 in Colombia, Hastings et al 2013 and Zimmerman in Chile, Ebenstein et al 2016 in Israel, Anelli 2020 in Italy, Jia and Li 2021 in China). Collectively they find that crossing the threshold of high stakes exams needed to access university has an impact on labour market outcomes. These papers make an important contribution to the overall high-stakes exams debate, but are less relevant to this paper, as they focus on non-compulsory exams for those at the higher end of the attainment spectrum.

achieving a high school diploma do not earn more than those just failing. They are able to do this because they find no effect of just achieving a high school diploma on subsequent education outcomes. (If individuals just above the threshold achieve more education than those just below, then we can no longer interpret an effect of the cut-off on earnings as due to signalling alone.)

In my study, I can check for the possibility of signalling at different stages: at the point at which individuals are deciding whether or not to continue in compulsory education, just achieving or just missing out on the benchmark five GCSEs at grades A* to C may send a signal to education institutions that an individual is more likely to be successful in completing an academic education (or may signal this to the individuals themselves). Moreover, if over and above any differences in subsequent educational attainment, there remain differences in earnings as a result of just crossing the threshold, then it could still be the case that crossing the threshold is sending a signal to employers about an individual's unobserved productivity. I find strong evidence in favour of the threshold being important, possibly indicating a role for signalling, in terms of subsequent education choices. I also find some tentative evidence of labour market signalling for women, but no evidence of signalling for men.

From a policy perspective, my findings suggest that, in England, something is stopping those that just miss out on crossing the threshold from continuing in academic education. This could be down to education institutions using this cut off when making admission decisions, the individuals themselves making different choices or potentially other factors. Policymakers and education institutions may want to reflect on this and consider what interventions may help to overcome this. This could, for example, be education institutions offering more flexibility to those with very similar profiles but one grade difference or teacher and careers advice to support these individuals. Currently it seems those just missing out have fewer options than those just passing. However, the fact that their early career outcomes, their education profiles in compulsory education and their socioeconomic and demographic characteristics, are very similar suggests that those just missing out may have been just as capable of completing an academic route as those just passing.

Findings in this paper also have some important policy implications for the future of high stakes exams. The Covid-19 pandemic led to the cancellation of in-person exams in 2020 in many countries (e.g., United Kingdom, France, Belgium, Ireland, the Netherlands, Norway, Mexico, Pakistani, Saudi Arabia and India). This has led a number of countries to question the existence and purpose of qualifications at important progression points, which have high stakes for students and schools[2]. For men, at least, my findings do not provide any evidence that high stakes exams are likely to lead to substantial allocative inefficiency: at least for this marginal group of male learners, just missing the threshold does not appear to have significantly harmed their future labour market prospects. There is some evidence of an effect for women, but the effects are relatively small. Therefore, the findings in this paper do not present a strong case for the abolition or major revision of high stakes exams, although of course my results only apply to a marginal sample and there are many

[2] Will GCSEs be scrapped? Why England's exam system could face major overhaul after Covid results fiasco (inews.co.uk)
SQA exams reform: 'rushed policy is bad policy' | Tes
Committee hears calls for reform of Leaving Cert exams
IBCC recommends exams, assessment reforms in intermediate, secondary education - Pakistan - Aaj.tv
Govt to scrap exams up to class 3, no public exams before class 10 | Dhaka Tribune

other reasons why high stakes exams may or may not be the right assessment method[3].

This paper now proceeds as follows. Section 2 covers background and some information on the English education system. Section 3 gives details of the data used and section 4 covers the empirical strategy. This is followed by the results (section 5) and discussion (section 6).

2. Background

This section gives details of the English education system and specifically GCSEs, post 16 education and the former importance of the five A* to C threshold. My sample includes those that took their GCSEs between 2004 and 2006 and thus when I refer to the English education system, I am referring to how it was at this time.

GCSEs are a series of compulsory qualifications that are taken by all pupils at the end of compulsory education, when pupils are around 16 years old. On average pupils entered around 9-10 GCSE or equivalent exams. Maths, English and science are compulsory, but a range of other subjects can be undertaken from the humanities, arts, languages and design and technology groups. Grades were awarded from A* to G, highest to lowest, with C or above seen as a 'good pass'.[4]

GCSEs are high stakes qualifications and there is a considerable amount of pressure on pupils to perform well in them. Better GCSE grades are associated with better education and labour market outcomes (Machin et al 2020, Gayle 2002, Crawford 2014, De Philippis 2016, Galindo-Rueda et al 2004, Chowdry et al 2013). GCSEs essentially act as a facilitator for continuing in education (and accessing higher levels) and certain grades are expected to access higher levels. Achieving what were formerly known as five 'good passes' (A* to C) was a salient 'cut off' and was expected by many institutions' entry policies for individuals to be able to progress to Advanced level qualifications (A-levels) and equivalents, qualifications often needed to go to university. Education providers (school sixth forms, sixth form colleges and further education colleges) use GCSE results to screen potential candidates. The proportion achieving five GCSEs A* to C was 56.8 percent of the whole population (of those doing GCSEs in 2004 to 2006).

GCSEs are externally set and awarded by one of four exam boards in England. The format of GCSE examinations differs for each subject, though there has always been a final exam, i.e., an externally set and marked assessment. The model for marking the final exam varies depending on subject and exam board, but it is commonly a standardised exam that is marked by external examiners. Grade boundaries are not pre-set and are decided by an external committee, using statistics and evidence (from previous years) and inspecting exam papers. The rigorous, independent process means there are limited concerns around school or teacher manipulation of results. However, there exists the possibility of remarking exams, which can be requested by the school (only). It is therefore possible that the grade awarded can change after remarking. I explore the implications of this in validity of design (section 4.2).

Before GCSEs were reformed post 2013 to remove tiering, most subjects had a two-

[3] cepeobn14.pdf (ucl.ac.uk)

[4] GCSEs were reformed in 2015 and the grade system is now 9-1. This is no longer the case since the reform of the school accountability system and GCSEs.

tier system, with a higher or foundation tier. There were some exceptions which did not have a tiers (such as history, physical education, art and drama) and notably maths, which had a three-tier system The tier entered by pupils for maths was therefore an important factor when considering grades A* to C, or thresholds, as it limited the possibility of grades achieved.

At 16 years old, after completing GCSEs, individuals had the choice to continue in education or do something else[5] (for example finding a job). For those staying in education, there are two broad options: 1) an academic route or 2) a vocational route. The academic route involves studying A-levels and usually university afterwards, whilst a vocational route includes apprenticeships and other technical qualifications. Having better GCSE grades gives more choice in terms of options, i.e., those with poorer GCSE grades may have fewer options. Overall, on completing GCSEs, various different pathways are possible.

Post 16 education in England is generally seen in levels, based on a qualifications framework[6]. These range from entry (0) to level 8. Achieving five GCSEs A* to Cs is a full level 2 qualification, whilst two A-level passes (or equivalent) is a full level 3. Higher education is level 4 and above and within that completing a (first) degree is a level 6.

GCSEs were, and still are, also high stakes exams for schools, i.e., teachers and leaders, as well as pupils. Since the Education Act in 1992 school league tables were published based on GCSE and equivalent results. Thus, prospective parents were able to see schools' overall 'performance' in exam results and make choices based on this. For a long period achieving five GCSEs A* to C was the headline measure used in these performance tables and though the accountability system changed, the 'five good passes' cut off continued to be reported on. In addition to being considered like passing secondary school, in a way a high school diploma in the US might be seen, the aforementioned factors left quite a legacy of the achieving five GCSEs A* to C threshold.

3. Data and sample selection

3.1 Data

This paper examines the impact of crossing the five GCESs A* to C threshold on labour market outcomes. The analysis draws upon a range of government administrative data, crucially data on exams taken at 16 and subsequent education and tax records. This has the advantage of very large numbers and completeness, i.e., it includes almost everyone who went to school in England (including those attending private, selective, and special schools) and is not based on sampling. This gives the opportunity for highly robust analysis (reducing risks of bias and small sample sizes).

I start with administrative records on GCSEs and equivalent qualifications. This includes information on entries, grades, qualifications, subject, and tier. The detailed GCSE data enables me to construct a continuous measure of overall attainment. I do not, however, observe continuous marks within each grade boundary or whether a grade was changed due to remarking (just the final grade). I discuss my identification

[5] Since 2015 it is now compulsory to stay in some form of education or training until 18.
[6] What qualification levels mean: England, Wales and Northern Ireland - GOV.UK (www.gov.uk)

strategy in more detail in Section 4.

I use three cohorts of individuals, those taking their GCSEs in England between 2004 and 2006. I combine cohorts to increase sample size. The records of these individuals that did their GCSEs (and equivalents) in England during these three years are then matched to other education administrative records which include test scores from exams taken at 11 years old and socioeconomic and demographic characteristics, including proxies for socioeconomic status (Free School Meals (FSM) eligibility and Income Deprivation Affecting Children Index (IDACI)[7]), special educational need status (SEN), first language spoken, ethnicity and region of school attended. I also add a school fixed effects variable to control for the school attended and address any unobservable factors that could influence labour market outcomes. This is subsequently matched to post 16 education data on qualifications undertaken and completed. This includes the level of the qualification, the type of qualification, the subject or sector area, the institution or provider and the grades or classifications.

Finally, this is matched to administrative tax and benefit records, which include employment spells, earnings from employment and out of work benefits claims. Collectively, these matched datasets are known as the Longitudinal Education Outcomes (LEO) dataset. The labour market variables are available up to 2017-18 tax year and hence individuals can be tracked for 11 to 13 (full tax) years after completing their GCSEs (dependent on the cohort, i.e., those taking GCSEs in 2004 can be tracked 13 years, whilst those taking exams in 2006 11 years).

I focus on employment, earnings and out of work benefits claims in the latest (tax) year available in the data: 2017-18. This approach is common in the literature (Clark and Martorell 2014, Angrist and Krugeur 1991, Lemieux and Card 2001, Anelli 2020, Aakvik et al 2010). I also show that results hold if I select a given age, i.e., ages 27, 28 or 29. This is around 11 to 13 full years after GCSEs (depending on the GCSE cohort), i.e., around 27 to 29 years old. Three labour market variables are constructed for this paper: 1) proportion of year in employment, 2) proportion of year claiming out of work benefits and 3) logarithm of earnings at age 27 to 29. The employment variable captures the proportion of the tax year that someone has been employed. This is a continuous variable (from 0 to 1) calculated by dividing the number of days employed, by the days in the tax year. Being out of work is also defined as the proportion of the year claiming out of work benefits, in the way employment is defined. I use the definition used in Department for Education official statistics[8]. This includes more 'classic' unemployment benefits and disability out of work benefits[9]. Claiming out of work benefits and being in employment are not mutually exclusive, i.e. it is still possible to claim certain out of work benefits whilst working. This however comes with restrictions and is usually those on low incomes, working few hours or with a disability. For example, jobseeker's allowance (the main unemployment benefit) can be claimed if someone is working fewer than 16 hours per week and actively seeking other employment.

I calculate daily earnings by dividing annual earnings by the number of days in

[7] FSM is a proxy for lower socioeconomic status. To be able to claim FSM parents/guardians need to be claiming certain benefits and are effectively on very low incomes: Apply for free school meals - GOV.UK (www.gov.uk). IDACI measures the proportion of children in a local neighbourhood that are in poverty.

[8] Further education: outcome-based success measures, Methodology – Explore education statistics – GOV.UK (explore-education-statistics.service.gov.uk)

[9] For more details see page 8 (paragraphs 12 to 14) of this report: Technical Report for Education and Labour Market Pathways of Individuals (LEO) (publishing.service.gov.uk)

employment in the tax year. Earnings are only included for those in employment for the whole tax year[10] and are adjusted to the latest tax year of data available (2017-18) using the Consumer Prices Index including owner occupiers' housing costs (CPIH). 47,105 (57 percent) of the 83,128 individuals in my sample meet this criterion. There are 17,834 individuals (around 22 percent of the sample) who were not employed at all, leaving 17,199 (21 percent of the sample) who were employed for at least one day but less than 365 days. See the appendix (Table A1 and Figure A1) for more details, including a gender split (the patterns are very similar). I opt to focus on earnings of those in employment for the whole tax year to focus on those in 'stable' or continuous employment, thus removing casual or temporary work. Checks are taken to see if different employment and earnings definitions make a difference to results (shown in section 5: results). I remove the first and 99th percentiles of earnings to remove outliers separately for men and women. The main limitation of the earnings data is that it does not contain hours worked. Because women are more likely to work part time than men,[11] I make comparisons of crossing the threshold separately for men and women. This is explored when interpreting results in section 5.

3.2 Sample selection

I start with 1.9m individuals who did their GCSEs in 2004 to 2006. 95 percent are successfully matched to labour market data, leaving me with 1.8m individuals. From here I restrict my sample to those who 'just achieve' or 'just miss out' on the threshold of achieving five full[12] GCSEs at grades A* to C. This is defined using a 'one grade away' approach. Ideally, I would use exam marks as a running variable and carry out a RDD approach around the C/D cut off of the marginal exam. However, exam marks are unavailable. Instead, I use total GCSE points score as a kind of running variable, which enables me to account for an individual's total attainment across all qualifications and subjects, as well as the marginal exam.

The treatment group is defined as those that achieve five full GCSEs A* to C and have at least one C grade (i.e., have at least one subject in which they achieved a 'marginal' good pass). The control group are those that have four full GCSEs A* to C and at least one D grade amongst their remaining grades (i.e., they 'just' missed out on meeting the benchmark five A* to C grades). Thus, the threshold is the C versus D grade, and the treatment and control groups are only separated by that one GCSE grade. The treatment group and control group can have a range of grades, as long as they meet the criteria outlined above. This means that the treatment group could include CCCCC, and the control group could include CCCCD, or they could include A*A*A*A*C and A*A*A*A*D (respectively). As shown in Section 4.2 below, the overall basket of GCSE results – as well as a range of other characteristics – are very similar across the two groups.

Restricting the sample in this way produces a sample of 83,460 individuals. I also restrict attention to individuals who, in any given exam they took, had the possibility of attaining either a C grade or a D grade, to maximise the possibility that random

[10] 365 days or 366 in a leap year

[11] https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/datasets/allemployeesashetable1

[12] Equivalent qualifications such as vocational GCSEs, short GCSEs, general national vocational qualifications (GNVQs) and key skills were taken at the time. I restrict my sample to full GCSEs only with the exception of allowing one short GCSE (the value of half a GCSE). entry/pass. I ran a robustness check to show this has not impacted results.

factors could have contributed to them 'just passing' or 'just missing out'.[13] This is a particular problem for maths and reduces the sample slightly further, to 82,138 individuals: 42,799 in the treatment group and 39,339 in the control group. For men, there are 22,460 and 20,574 in the treatment and control groups respectively, whilst for women it is 20,339 and 18,765.

3.3 Representativeness of my sample versus wider population

There are 82,128 individuals in my sample, which is almost four percent of the overall population (of 1.8m). This implies that my findings are for a very marginal group of students with attainment around the threshold (i.e., it does not include very high or very low achievers). This is not to say that focusing on my sample is not important, as this is a crucial threshold in the English education system and hence these marginal learners are of key interest. However, it is unlikely that my findings can be extrapolated to the population as a whole. This section illustrates some of the differences between my sample and the rest of the population taking GCSEs between 2004 and 2006 who have been matched to labour market data and provides important context for my findings.

Table 1 shows that my sample is slightly less likely to be from a lower socioeconomic status, to have a special educational need, be from a minority ethnic background or have a language other than English as their first language. Average test results at 11 vary for men and women, with my sample being higher for men and the opposite for women. At 16, the patterns remain the same for women but the converse for men.

Table 1: Comparison of my sample and wider population - socioeconomic and demographic characteristics

	Men in sample	All other men	Differences	Women in sample	All other women	Differences
Socioeconomic and demographic characteristics						
IDACI	0.185	0.205	-0.019***	0.202	0.206	-0.005***
FSM eligible	0.089	0.128	-0.039***	0.110	0.128	-0.018***
SEN	0.108	0.212	-0.104***	0.066	0.120	-0.054***
Minority ethnic	0.121	0.131	-0.010***	0.143	0.135	0.008***
First language English	0.081	0.089	-0.007***	0.094	0.090	0.004**
Education factors						
KS2 English and maths marks	104.6	103.9	0.745***	101.6	106.4	-4.776***
KS4 points	40.3	40.5	-0.167	41.4	46.5	-5.087***
Sample size	43,034	903,345		39,104	854,715	
** significant at 5% level, *** significant at 1% level.						

[13] As introduced in the background section, maths previously had a three-tier system. The higher tier awarded grades A* to C and foundation tier awarded grades D to G. Comparing someone with a C in the higher tier with a D in the foundation tier would be invalid as there is essentially no threshold, i.e., those with a C in higher tier never had any chance of getting a D nor did those achieving a D in foundation tier have the chance of achieving a C. Thus, I only include those entering the intermediate tier. This awarded grades B to E and hence there was a natural threshold. I only have information on exam tier from the AQA exam board and therefore I only include those individuals that crossed the (C/D) threshold in maths in an exam administered by the AQA exam board.

My sample includes those that achieved five full GCSEs A* to C, including at least one C and those who just missed out, i.e., achieving 4 full GCSEs A* to C and at least one D.
All other men and women are all those taking their GCSEs in the same years (2004 to 2006) who achieved other results (i.e., not either side of the threshold).
Columns 2, 3, 5 and 6 show means. Columns 4 and 7 show results from two-sided t-test.

Whilst the sample and overall population are similar in compulsory schooling, some differences show at higher levels of education (see Table 2). The sample becomes a marginal group at higher levels of education, with generally poorer outcomes than those from the overall population.

Table 2: Comparison of my sample and wider population - post 16 education variables

	Men in sample	All other men	Differences	Women in sample	All other women	Differences
Level 3 participation by age 20	0.430	0.448	-0.018***	0.471	0.544	-0.073***
Level 3 participation by age 20 - points	514.67	726.96	-212.30***	542.18	765.17	-223.00***
Degree participation	0.286	0.372	-0.086***	0.322	0.464	0.142***
Degree completion	0.184	0.298	-0.114***	0.220	0.392	-0.172***
Degree participation but not completion ('drop out')[14]	0.360	0.199	0.169***	0.318	0.155	0.159***
Degree classification – 2:1 or above	0.447	0.626	-0.179***	0.445	0.672	-0.225***
University 'status'[15] 1	0.001	0.028	-0.027***	0.000	0.020	-0.020***
University 'status' 2	0.090	0.386	-0.296***	0.069	0.356	-0.287***
University 'status' 3	0.726	0.496	0.230***	0.724	0.513	0.211***
University 'status' 4	0.147	0.066	0.082***	0.174	0.084	0.090***
Sample size	43,034	903,345		39,104	854,715	

** significant at 5% level, *** significant at 1% level.
My sample includes those that achieved five full GCSEs A* to C, including at least one C and those who just missed out, i.e., achieving 4 full GCSEs A* to C and at least one D.
All other men and women are all those taking their GCSEs in the same years (2004 to 2006) who achieved other results (i.e., not either side of the threshold).
Columns 2, 3, 5 and 6 show means. Columns 4 and 7 show results from two-sided t-test.

4. Empirical strategy

4.1 Research design

The previous section defined the groups on either side of the five GCESs A* to C threshold. Here I explain how I estimate the impact of crossing the threshold on

[14] This is the proportion of individuals that participated in a degree but had not completed it (by the latest year's data). This is therefore not exactly 'drop out' as individuals could still be doing their degree, and this is a risk for late starters.
[15] Using clusters from Boliver (2015)

subsequent outcomes.

I observe the impact of crossing the threshold on labour market outcomes and subsequent short- and medium-term post compulsory education outcomes, before finally looking at how they all fit together. I use mediation analysis to decompose the link between crossing the threshold, post 16 education outcomes and labour market outcomes (and carry out a robustness check using instrumental variables (IV) to verify results – see below). This is essentially a framework that allows me to carry out several components of analysis and then put them together, but with each of the components able to stand alone and provide insight in their own right. I refer to them as steps, in line with the theoretical framework first outlined by Baron and Kenny (1986).

In step 1, I use regression analysis to estimate the impact of achieving five full GCSEs at A* to C on labour market outcomes versus just missing out (i.e., achieving four full GCSEs at A* to C, and one full GCSE at D). The definition of these two groups is discussed in detail below. Formally, the model estimated is the following:

$$(1) \qquad Y_i = \beta_0 + \beta_1 T_i + \beta_2 X_i + e_i$$

Where Y is the labour market variable of interest, for example (logarithm of) earnings. T is a dummy variable that indicates if someone crossed the threshold. X is a vector of control variables and includes proxies for socioeconomic status (FSM and IDACI), SEN status, ethnicity, first language, region (where they went to school), prior attainment (English and maths marks from test scores at 11), academic year cohort, the school attended (using school fixed effects) and variables reflecting the overall basket of GCSE results (exam entries, a points measure for overall results, subjects taken and exam tier). e is the error term.

The range of these variables allows me to compare those with similar performance at school, similar personal characteristics, growing up in similar parts of the country, going to similar types of schools, and taking a similar set of exams in which they obtain a similar overall basket of grades in similar subjects but, crucially, in which they obtained a C vs. D grade in one particular subject. I do not have exam marks and therefore a regression discontinuity design (RDD) is not possible. However, the richness of the variables I use helps to overcome this. This is explored in sub section 4.2 validity of design.

Ordinary Least Squares (OLS) regressions are used to estimate the impact on the three main labour market outcome variables at age 27 to 29, i.e., years 11 to 13 after GCSEs.: 1) proportion of year in employment, 2) proportion of year claiming out of work benefits and 3) logarithm of earnings.

In Step 2, Probit models are used to estimate the same equation, calculating marginal effects, but where the outcome variables are several short-term and medium-term (post 16) education variables. This is because the dependent variables are binary, i.e., achieved the education level (=1) or not (=0), and an appropriate model is needed. Formally this can be expressed as:

$$(2) \qquad M_i = \beta_0 + \beta_1 T_i + \beta_2 X_i + e_i$$

Where M_i is the mediator, i.e., the short- or medium-term education variable. These are: achieved any level 3 qualification by age 20, achieved level 3 academic by 20 and achieved level 3 vocational by 20. And similarly, for medium term: achieved level 3 or above by 25, achieved level 4 or above by 25, achieved level 3 academic or

above by 25 and achieved level 3 vocational or above by 25. I choose to consider level 3 by age 20 as this is the next level after completing GCSEs (five 'good passes' is a level 2) and the route (academic or vocational) is important. Level 3 is the highest possible qualification that could have been completed by this age. At age 25 I focus on the same categories, but consider whether individuals have completed level 4 or above by this age. T and X are the same as in step 1 (above).

For step 3, I add the mediator variable to the regression model outlined in step 1. This is expressed as:

$$(3) \qquad Y_i = \beta_0 + \beta_1 T_i + \beta_2 M_i + \beta_3 X_i + e_i$$

Y is logarithm of earnings, in this case, for those employed in the latest tax year and M is the mediator variable: post 16 education. I start by including those achieving level 3 or above by age 25 as my mediator. This is then replaced with two mediator variables representing whether an individual achieved level 3 or above by 25 via an academic route (1) or a vocational route (2). I subsequently add level 4 or above achievement by 25 to the first model and finally add controls for degree classification and university 'status' to this model. I choose to focus primarily on level 3 because this is not only the next level above achieving five GCSEs A* to C; the route achieving level 3 is also crucial, as it is expected that returns would be different for those doing academic and vocational qualifications. Level 4 and above captures all levels of education above this (and is generally known as higher education).

I also carry out instrumental variables (IV) analysis as an alternative to step 3, i.e., this is not my main approach. Here, T is used as an instrument for post 16 education choices, M_i, in this instance whether an individual achieved level 3 or above. This is an alternative to step 3 in the sense that step 3 seeks to understand descriptively how much of the relationship between crossing the threshold and earnings can be explained by the findings highlighted in step 2, i.e. the fact that there is a relationship between crossing the threshold and subsequent education choices, and the IV approach is another way of doing this, although it requires some stronger assumptions (discussed in more detail below). Formally this can be expressed as:

$$(4) \qquad Y_i = \beta_0 + \beta_1 \hat{M}_i + \beta_2 Z_i \beta_3 X_i + e_i$$

$$(5) \qquad \hat{\hat{M}}_i = \alpha Z_i + u_i$$

Where Z is a vector of covariates of observed attributes (between T and M). The IV approach rests on the exclusion restriction that the instrument only affects the outcome variable indirectly through its effect on the endogenous variable of interest and does not affect the outcome through any other means. More formally this can be expressed as $E(\hat{M}_{I,} e_i) = E(Z_i, u_i) = 0$. Under a human capital framework, it would be assumed that crossing the threshold does not lead directly to higher productivity and thus higher wages. However, crossing the threshold may influence education activity and thus productivity and earnings, i.e., indirectly. Therefore, theoretically it could be argued that 'just' achieving or 'just' missing out on achieving five GCSEs A* to C is a good instrument. It should be noted that signalling theory would dispute this assumption and suggest that just passing may signal to employers higher levels of ability and directly command a higher salary.

4.2 Validity of design

The OLS models will give unbiased estimates if T is exogenous, i.e., not correlated with the error term. The main threat to this assumption is endogeneity in the form of

there being a difference in characteristics, observed or unobserved, of those just falling either side of the cut off. I have a rich set of variables I can draw on from the administrative data to try to ensure that there are no inherent differences between the treatment and control groups. This includes background characteristics (such as ethnicity, socioeconomic status, first language and SEN status), the region where they went to school and the school they attended. I also account for English and maths test scores at 11 and the overall GCSE results and entries. By comparing similar individuals, from the same place, attending the same (type of) school, surrounded by similar types of pupils and with the same test scores at 11 and (crucially) almost exactly the same results at 16 (except for the C/D difference in question), I aim to account as far as possible for underlying unobserved factors like motivation and ability.

The existence of no or only small differences in observed characteristics between the treatment and control groups would further support the assumption that there are unlikely to be large differences in unobservable characteristics between the groups that could potentially bias my results. Differences in GCSE attainment and other characteristics between the treatment and control group are presented in Tables 3 and 4 respectively. Nevertheless, I cannot rule out some form of bias. It could be that there are differences between the treatment and control group that I have not accounted for, which are not available in the data. This would mean that the error term is correlated with T and results are biased. I discuss potential bias throughout this section and return to this explicitly at the end of this sub-section.

Table 3 presents average differences between the treatment and control group in terms of the number of full GCSEs entered, and two different measures of total GCSE points. One measure is for the total 'basket' of GCSE results, whilst the other is for the best five results, allowing me to look at overall results and hone in on results that make up just achieving or just missing out. It shows that the number of GCSEs entered does not differ significantly between the two groups. GCSE points is a continuous variable representing results for all GCSEs with A* being 8 points and a G being 1 point. The two measures for GCSE points included in Table 3, for all results and for the best five, are slightly different (and statistically significant). The number of qualifications entered by those in the treatment group is not statistically significantly different to the control group. Removing the one point for crossing the threshold (a C is 5 points and a D four points) the mean is 40.4 for the treatment and 39.2 for the control group for men, and 41.5 for the treatment and 40.3 for the control group for women. This difference of 1.1 and 1.2 points for men and women equates to just over one grade. Considering they took on average nine exams this one grade difference could be argued to be small. It should also be stressed that this is the overall combination of results and number of exams entered ranges from five to seventeen. Focusing on the best five results is perhaps a more salient measure as the other D/E-G grades could be considered as noise. The adjusted KS4 points score for the 'best five' GCSE grades is 25.3 for the treatment group and 24.8 for the control group for men, and 25.6 for the treatment group and 25.1 for the control group for women. One point is explained by the threshold (i.e., C/D difference being one point). This leaves a difference of around half a mark for both men and women, thus making the remaining four grades very similar (though a t-test shows the distributions to be statistically significant even after the adjustment). The average the treatment group would have achieved grades BCCCC and the control group grades BCCCD (from their 'best five GCSEs). The distribution is shown in the appendix (Figure A1). Table A2 in the appendix shows that the proportion of pupils entering different subjects is very similar for the treatment and control groups. It should be noted, as detailed below, that I include a series of controls in the model for number of GCSEs entered, subjects entered and crucially a measure for overall results.

Table 3: descriptive statistics (means): GCSE results

Variable	Men			Women		
	Treatment (T)	Control (C)	T – C	Treatment (T)	Control (C)	T – C
Full GCSEs entered	9.39	9.38	0.010	9.41	9.42	0.010
GCSE points (adjusted[16]	40.4	39.2	-1.128***	41.5	40.3	-1.236***
GCSE points (best five) adjusted[11]	25.3	24.9	-0.407***	25.6	25.1	-0.514***
Sample size	22,460	20,574		20,339	18,765	

*** significant at 1% level.
This is descriptive analysis of those that achieved five full GCSEs A* to C, including at least one C and those who just missed out, i.e., achieving 4 full GCSEs A* to C and at least one D.

The crucial element to my analysis is the range of controls I am able to apply to the model. I take into account socioeconomic and demographic variables, attainment at age 11 and the overall basket of GCSE results. Despite these strong controls, ideally the treatment and control group should be similar in all other factors that affect productivity. This is examined in the remainder of this sub-section.

As mentioned above, whilst the treatment and control group are 'one grade away' from each other, there is a wide range of potential exam results. This variance may lead to some small differences in socioeconomic and demographic factors and hence a simple comparison of descriptive statistics might not be relevant. For this reason, I compare differences in socioeconomic and demographic (and other education) factors after my GCSE controls. I carry out a series of regression models with different socioeconomic, demographic and education factors as the dependent variable, T as the explanatory variable, and controls for KS4 points, number GCSEs entered, exam tier and subjects entered. The results are presented in Table 4. Table A3 in the appendix shows the descriptive statistics for the treatment and control group based on sample means for completeness and transparency. Table 4 shows that, after accounting for GCSE performance, the background characteristics of the two groups are very similar. There are no statistically significant differences between the treatment and control group for the socioeconomic and demographic factors I can account for, except for some small differences in region where GCSEs undertaken. There was a difference in education performance at 11 between those that crossed the threshold and those that did not. There is around a 2 marks difference for men women, out of 200, in combined English and maths scores(national standard attainment tests). This shows that before I control for the basket of GCSEs there are some differences between the treatment and control group. However, the crucial aspect is whether there are inherent differences between the treatment and control group after the (GCSE) controls in my model are applied.

Table 4: Comparison of treatment and control group: socioeconomic, demographic and education factors

Variable	Men	Women
Socioeconomic and demographic characteristics		
IDACI	-0.002	-0.003

[16] Removing the one point (i.e., one grade) for crossing the threshold, i.e., the C versus D difference. A* is eight points and a G one point.

	(0.002)	(0.002)
FSM eligible	-0.005 (0.003)	-0.005 (0.003)
SEN	-0.001 (0.003)	-0.005 (0.003)
Minority ethnic	0.000 (0.003)	-0.001 (0.004)
First language English	0.000 (0.003)	-0.001 (0.003)
Region at GCSEs: London	-0.005 (0.003)	-0.001 (0.004)
Region at GCSEs: North East	0.005** (0.002)	0.003 (0.002)
Region at GCSEs: North West	0.005 (0.004)	-0.001 (0.004)
Region at GCSEs: Yorkshire and the Humber	0.004 (0.003)	0.008*** (0.003)
Region at GCSEs: East Midlands	-0.003 (0.003)	-0.006** (0.003)
Region at GCSEs: West Midlands	-0.003 (0.003)	0.004 (0.003)
Region at GCSEs: South East	0.001 (0.004)	-0.003 (0.004)
Region at GCSEs: South West	-0.003 (0.003)	0.004 (0.003)
Region at GCSEs: East of England	-0.001 (0.003)	-0.007** (0.003)
Education factors		
National test age 11 total marks (English and maths)	1.908*** (0.170)	2.259*** (0.189)
Sample size	43,034	39,104

** significant at 5% level, *** significant at 1% level.
These models have the socioeconomic, demographic and education variables as the dependent variable.
The analysis presented includes those that achieved five full GCSEs A* to C, including at least one C and those who just missed out, i.e., achieving 4 full GCSEs A* to C and at least one D.
The coefficients shown are for the dummy variable indicating if someone just achieved five GCSEs A* to C, i.e., the main explanatory variable of interest (T).
Control variables are included for number of GCSEs taken, subjects entered, GCSE cohort, exam tier and GCSE points (adjusted).

As noted above, the empirical strategy will deliver valid estimates for β_1 only if crossing the threshold is as good as randomly assigned. The key factor here is whether pupils have control over GCSE grades achieved, similar to the way Clark and Martorell (2014) show with test scores. If pupils do not have full control of their test scores, then crossing a threshold could be deemed to be random. It seems perfectly reasonable that two (groups of) individuals with similar ability levels achieve slightly different grades, i.e., C or D, based on whether they had a good or bad day, revised the right or wrong topic(s), had a harsh or lenient examiner or some other random factors. This is explored more in the discussion. However, schools can request GCSE grades to be remarked and this can result in them being regraded. This may introduce a form of bias and mean achieving a grade C or D is not fully randomised. I am unable to control for this as I do not have data on remarking

requests or regrading. However, Machin et al (2020) found that socioeconomic and demographic characteristics and national test results at 11 years old were not statistically significant factors in predicting an individual's paper being regraded. Whilst these are only proxies for unobsevables, this suggests that regrading may not necessarily have introduced bias and may not be a major concern for my study. The scale of requests and regrading may be important too. In 2013 for example, two percent of all GCSE exams were appealed and around one in six of these appeals led to a change in grade (Machin et al 2020). For those in the sample used by Machin et al (2020), i.e., sitting the English language higher tier that year with AQA and achieving a D (or C), the average probability of requesting a remark was 10 percent and 12 percent of these cases were upgraded. This infers that the overall proportion of the sample that was actually upgraded was around one percent. In the absence of exam marks and regrading data, I use the data available to test this key assumption, i.e., crossing the threshold is conditionally as good as random. But on the whole the results from Machin et al (2020) and the relatively small scale in regrading gives me confidence this is unlikely to be a major risk for my study.

We have already seen (in Table 4) that the treatment and control groups have similar socioeconomic and demographic characteristics (after factoring in GCSE attainment controls). The difference in test marks at 11 could be seen as cause of concern. Even though these are controlled for in the model, the concern is that they could signal some unobservable differences between individuals in the treatment and control group. If these unobservables are linked with higher productivity, it could be something in the error term that drives any differences in labour market outcomes. However overall, the similar profiles give confidence that background characteristics are unlikely to drive whether someone crosses the threshold or not.

I have already shown the overall 'basket' of GCSE exams taken and the results. I also examine the prior attainment of individuals in the treatment and control groups. As mentioned, the sample includes exam marks in English and maths from tests taken at the end of primary school (i.e., around 11 years old). These are important in examining this issue and in my empirical approach, i.e., whether there may be unobservable factors such as innate ability and motivations explaining crossing the threshold. Table 4 shows there is a difference of 2 marks for men and women (out of 200). These differences are statistically significant, suggesting that that the treatment group were performing slightly better in English and maths at 11 years old than the control group. The key question is whether a difference of 1.5-2 marks out of 200 means crossing the threshold is a sign of a difference in ability levels. I argue such a small difference in performance levels at 11 years old is not enough to suggest that those in the treatment group have a significantly better cognitive abilities than those in the control group. It should be noted that these are attainment tests, not ability tests.

The prior attainment results examined above, as well as socioeconomic and background characteristics and other GCSE results explored earlier, are used as controls, and are therefore aimed to ensure there is no endogeneity in the model. This gives confidence that the OLS models are unbiased. Overall, the socioeconomic and demographic profiles, attainment at school and other GCSE results are very similar. This all supports the case that crossing the threshold is conditionally as good as random. Nevertheless, some of the analysis in this section casts doubt on this key assumption and could suggest other (unobservable) factors are driving whether those in my sample achieve a C or D grade. For example, if the slight differences in test marks at age 11 indicate a difference in innate ability, this could be driving who crosses the threshold (and ultimately the labour market outcomes). Whilst I argue that this is not the case, and such small differences in test scores five years earlier

are not indicative of ability levels, I cannot rule it out.

Should T be correlated with e, i.e., crossing the threshold be correlated with the error term and bias be introduced, it would likely to be a 'positive' bias. By this I mean that there are some unobservables in the treatment group that are slightly more favourable than those in the control group. These could be higher levels of ability, motivation, parental aspirations or other socio-cultural factors that are likely to be related to productivity. Observing Table 4 and linking back to the discussion above, the difference in attainment at 11 stands out for both men and women and FSM eligibility for men. The attainment results at 11 years old imply there was a difference in English and maths levels five years before the exams, for example. This could imply that crossing the threshold was non-random and those just passing the exam had slightly higher levels of ability or other non-observed factors. I argue that I account for this, but it is important to bear this in mind when reviewing results. Should there be any positive differences between the labour market outcomes of the treatment and control group, it could potentially be down to the 'positive' bias discussed here.

5. Results

In this section I present the estimates of achieving five full GCSEs A* to C (my primary model). I approach this in the way I describe the mediation analysis in the previous section. I begin by reporting the estimates of achieving five GCSEs A* to C on labour market outcomes (step 1), next on post 16 education outcomes (step 2) and then linking them together (step 3). Thus, there is an overarching mediation analysis framework, but each sub-section has findings in their own right. I then present some IV estimates to verify the results at step 3, i.e., the returns to post 16 education conditional on crossing the threshold. Subsequently, I present results on heterogeneity. Finally, I present robustness checks.

5.1 Labour market outcomes (step 1)

Table 5 shows estimates of crossing the threshold on three labour market variables in 2017-18, when my cohorts are age 27 to 29, i.e. around 11 to 13 full tax years after GCSEs were taken: 1) proportion of year in employment, 2) proportion of year claiming out of work benefits and 3) logarithm of daily earnings amongst those who work for the whole year. This is equation (1) from section 4.1. These are reported separately for males and females. As shown in row 1, for men there are no statistically significant differences in being in employment or claiming out of work benefits between those just passing and just missing out. Equally, for those men in employment, achieving five full GCSEs A* to C is not associated with statistically significantly higher earnings.

For women, as with men we do not observe any employment increase to achieving five full GCSEs A* to C, but we do see a reduction in claiming out of work benefits. This may seem contradictory, however there are a few factors that could be explaining this. Firstly, employment and being out of work (claiming benefits) are not mutually exclusive. As explained in the background section, some out of work benefits can still be claimed whilst working, however there are often restrictions on the activity (e.g., a maximum income or hours worked). Secondly, labour force participation could be different. Not all those not in employment are actively seeking employment (a prerequisite for claiming out of work benefits) and there may be differences in choice between the treatment and control group. The coefficient of -0.007 implies that those in the treatment group are likely to be claiming out of work benefits for 0.7 percentage points less of the year than the control group. This implies

that just achieving five GCSEs A* to C leads just over two and a half days fewer claiming out of work benefit on average. Crossing the threshold is also associated with a statistically significant increase in log of daily earnings for women, around 3.1 percent (i.e., 1.031, derived from the exponential of 0.031). On average, this equates to around £1.54 per day or around £561 gross annual earnings (i.e., before tax).

Table 5 – Effect of achieving five GCSEs A* to C on early career labour market outcomes – regression output

	Model 1) Employed years 11 to 13 - OLS	**Model 2) Claiming out of work benefits years 11 to 13 - OLS**	**Model 3) Log (earnings) years 11 to 13 - OLS**
Men	0.000 (0.004)	0.000 (0.002)	0.001 (0.006)
Sample size	43,034		23,798
Women	0.001 (0.002)	-0.007*** (0.003)	0.031*** (0.008)
Sample size	39,104		22,367

** significant at 5% level
This is the primary model including those that achieved five full GCSEs A* to C, including at least one C and those who just missed out, i.e., achieving 4 full GCSEs A* to C and at least one D. Models 1 and 2 include the whole sample, whilst model 3 only includes those that were employed for 365 days in the (tax) year.
The coefficients shown are for the dummy variable indicating if someone just achieved five GCSEs A* to C, i.e., the main explanatory variable of interest (T).
All regressions include a set of controls described in section 4.1, research design.

5.2 Education outcomes (step 2)

I next move on to consider what might be driving these results, and in particular, whether just crossing the threshold leads to differences in subsequent education choices, which could plausibly affect subsequent labour market outcomes. Table 6 shows that crossing the threshold has quite a significant impact on short term education outcomes, reporting models based on equation (2) from section 4.1. Just passing five GCSEs A* to C leads to higher levels of (post compulsory) education and a higher likelihood of taking an academic route for both men and women. Models 1, 2 and 4 show that by 20 years old, those in the treatment group are more likely to have achieved a level 3 qualification, to have achieved it via an academic route, and to have started a level 4 or above course. (This is primarily driven by those just crossing the threshold being more likely to go on to do a degree (a level 6 qualification) than those just missing out.) A crucial factor here is that crossing the threshold does not impact (negatively) on achieving level 3 vocational (model 3). This means that it leads to genuine increase in level 3 academic achievement and not a switching between vocational and academic. The displacement effect actually seems

to be taking place between level 2 vocational (as shown in Table A4 in the appendix) and level 3 academic, i.e., crossing the threshold leads to a decrease in level 2 vocational achievement by age 20 and an increase in level 3 academic achievement by age 20. I also extend this to age 25 and still observe differences in highest level of education achieved at this age (also shown in Table A4 in the appendix).

These results may suggest that education institutions are 'screening' candidates based on GCSE results, although it could be that the individuals themselves decide not to even apply to continue their academic education. The difference in level 3 achievement and academic split (plus the switch from level 2 vocational to level 3 academic) show that these institutions value the achievement of five GCSEs A* to C. Put another way, crossing the threshold sends a signal to education institutions that candidates are more likely to be successful continuing to higher levels of education and doing so via an academic route. Or alternatively individuals receive a signal themselves about their ability to continue.

Table 6: Effect of achieving five GCSEs A* to C on short term education outcomes - regression output

	Model 1) Achieved level 3 by 20 - Probit	**Model 2) Achieved level 3 academic by 20 - Probit**	**Model 3) Achieved level 3 vocational by 20 - Probit**	**Model 4) Started level 4 by 20 - Probit**
Men	0.063*** (0.005)	0.063*** (0.004)	0.000 (0.004)	0.035*** (0.004)
Sample size	43,034			
Women	0.058*** (0.005)	0.067*** (0.004)	-0.008 (0.005)	0.044*** (0.004)
Sample size	39,014			
** significant at 5% level, *** significant at 1% level This is the primary model including those that achieved five full GCSEs A* to C, including at least one C and those who just missed out, i.e., achieving 4 full GCSEs A* to C and at least one D. Models 1 to 4 include the whole sample. The coefficients shown are for the dummy variable indicating if someone just achieved five GCSEs A* to C, i.e., the main explanatory variable of interest (T). Estimated using Probit regression, calculating marginal effects. All regressions include a set of controls described in section 4.1, research design.				

5.3 Combining education and labour market outcomes (step 3)

Putting together the education and labour market findings (in sections 5.1. and 5.2 above) in step 3 of the mediation analysis (based on equation (3) from section 4.1), Table 7 examines the extent to which the differences in labour market outcomes observed in Section 5.1 can be explained by the differences in education outcomes reported in Section 5.2. Baron and Kenny (1986) stipulated that the explanatory variable in the first step (section 5.1) needed to be significant for mediation analysis to work, though this has been disputed by other authors (Shrout and Bolger 2002). Thus, despite step 1 being insignificant for men, I complete the mediation analysis for

both men and women.

For men, I have already established that crossing the threshold does not affect labour market outcomes (from section 5.1), even though it does seem to have a large positive effect on their education choices. Table 7 provides some suggestions as to why this might be the case. Model 2 shows that achieving a level 3 or above has a positive impact on earnings, whilst model 3 suggests that this is driven primarily by achieving level 3 vocational or above, with level 3 academic or above still being statistically significant but having smaller coefficients. Model 4 shows that, for men, higher education (i.e., level 4 or above) does not have positive and significant returns to earnings on average. This model simply includes a dummy variable for if someone completed a level 4 or above qualification. Model 5 adds variables that control for the institution that an individual went to, their academic performance and what they studied. It shows that when adding these controls (for university 'status', degree classification and subject studied) higher education has strong and significant positive returns. This means that on average, for men, completing a degree does not impact upon labour market outcomes, but completing a degree at a higher 'status' university, performing better academically and/or studying a subject with higher returns, leads to higher earnings. Putting steps 2 and 3 of the mediation analysis together, I find little evidence of an earnings return to crossing the five GCSE A*-C threshold, despite the fact that it significantly increases the proportion of men achieving level 3 academic qualifications.

This suggests that the reason why more men who just cross the threshold go on to take academic qualifications, but don't go on to earn more, is because that men in my sample achieve little benefit in the form of earnings from achieving academic qualifications at level 3 and above. The academic findings contrast with the sizeable returns they achieve for the acquisition of level 3 vocational qualifications. Strong, positive returns to vocational qualifications are seen in the literature (Battiston et al 2019, Buscha and Urwin 2013, Greenwood et al 2007, McIntosh and Morris 2021). A similar study focusing on returns to education for marginal learners (Dearden et al 2004) found that women had greater returns than men for continuing in education beyond 16.

In terms of higher education, the results may seem surprising initially. This does however change, in model 5, when I control for degree classification and institution quality. This could be interpreted that achieving a degree does not have an effect on earnings for men on average, but when accounting for the institution and academic performance at university, it does. Put another way, those in my sample tend to go to 'lower status' institutions and have lower attainment, but the ones that do go to higher status universities and get a good classification see it reflected in their earnings. Studies focusing on higher education (i.e., level 4 and above) have generally shown men have lower returns to academic education than women (Belfield et al 2018b, Walker and Zhu 2003, Dearden et al 2004). It should also be noted that my sample is a marginal group of students, notably without higher achievers, and this could explain what might seem to be a counterintuitive result. This is further investigated in section 6, the discussion.

The story for women is much more straightforward. Model 1 shows the results from step 1 (section 5.1) that crossing the threshold has a positive impact on earnings. Model 2 in Table 7 shows that achieving a level 3 or above is associated with higher earnings and crucially (in model 3) achieving an academic level 3 or above is associated with higher earnings (than vocational). Achieving a level 4 or above is associated with higher earnings also, whilst for achieving a level 3 vocational or above, the size of coefficient is smaller than the academic route (though still

statistically significant). As crossing the threshold leads to more women achieving an academic level 3 or above and level 4 or above, and both mediators are linked to higher earnings, it might be assumed this is driving the effect on labour market outcomes. In both models 3 and 4 the (crossing the threshold) coefficient decreases by a small amount, but remains statistically significant. This suggests that crossing the threshold is affecting the post 16 education outcomes, which in turn are leading to higher earnings. However, this is only part of the picture and there appears to be a residual impact of crossing the threshold. It should be noted though that the coefficient does not reduce to zero (or close to it) and in fact remains statistically significant. Thus, perhaps suggesting that crossing the threshold remains important in itself, and not just as a facilitator to higher levels of education. Reasons for why there may be such differences between men and women are explored in the discussion (section 6).

I carry out a robustness check where level 2 vocational achievement (or above) at age 25 is included in the analysis, which is shown in table A5 in the appendix. This changes the baseline from level 2, both academic and vocational, and below (as in table 7) to level 2 academic and below (as in table 5). The results show the overall findings hold which is important. However, the differences in returns for vocational level 2 are interesting and add context. The negative returns to vocational level 2 achievement for women help to explain why crossing the threshold leads to higher earnings, i.e., because crossing the threshold means women are less likely to do a vocational level 2 qualification. For men they also aid in explaining why crossing the threshold does not lead to higher earnings. This is because crossing the threshold leads to being less likely to do a level 2 vocational qualification, which has strong and significant earnings. Therefore, this effect is creating a counter effect to crossing the threshold leading to higher levels of education which are associated with higher earnings. A notable difference is that in table A5 (for men) the returns to level 3 academic and level 4 and above become larger (and statistically significant for the latter) but are not as sizeable as level 2 vocational. Thus, in table 7 the returns to academic level 3 and level four and above are negatively affected by being compared to level 2 vocational achievement (which is included in the baseline).

Table 7: Effect on earnings age 27 to 29 of crossing the threshold and post 16 education (as a mediator) – regression output

	Model 1 **(Original model)**	**Model 2** **(Level 3 achievement as mediator)**	**Model 3** **(Level 3 academic and vocational achievement as mediators)**	**Model 4** **(Level 4 or above achievement as mediator)**	**Model 5** **(Level 4 or above achievement as mediator – with controls)**
<u>Men</u>					
Five GCSEs A* to C (explanatory variable)	0.001 (0.006)	-0.002 (0.006)	0.000 (0.006)	-0.002 (0.006)	-0.002 (0.006)
Level 3 or above achievement		0.049*** (0.007)		0.047*** (0.007)	0.047*** (0.007)
Level 3 or above academic achievement			0.022** (0.009)		

Level 3 or above vocational (only) achievement			0.054*** (0.007)		
Level 4 or above				0.008 (0.009)	0.062*** (0.014)
Sample size	23,798				
<u>Women</u>					
Five GCSEs A* to C (explanatory variable)	0.031*** (0.008)	0.024*** (0.008)	0.019** (0.008)	0.019** (0.008)	0.019** (0.008)
Level 3 or above achievement		0.125*** (0.008)		0.044*** (0.009)	0.042*** (0.009)
Level 3 or above academic achievement			0.172*** (0.011)		
Level 3 or above vocational (only) achievement			0.061*** (0.009)		
Level 4 or above				0.218*** (0.010)	0.245*** (0.015)
Sample size	22,367				

** significant at 5% level
This model includes those that achieved five full GCSEs A* to C, including at least one C and those who just missed out, i.e., achieving 4 full GCSEs A* to C and at least one D.
The dependent variable is log of earnings for all models.
Models 1 is step 1 from the mediation analysis, i.e., as shown in Table 5. Model 2 adds achieved a dummy variable for if someone achieved level 3 or above as a mediator. Model 3 replaces achieved level 3 with two dummy variables for if someone achieved a level 3 academic or above or level 3 vocational or above. Model 4 adds a dummy variable for achieved level 4 or above to model 2. All models include those employed for 365 of tax year. Model 5 adds control variables for university 'status', degree classification and subject studied.
The coefficients shown are for the dummy variable indicating if someone just achieved five GCSEs A* to C, i.e., the main explanatory variable of interest (T). All models use OLS as estimator.
All regressions include a set of controls described in section 4.1, research design.
The baseline for these returns is achieved level 2 highest level, which can be both academic or vocational, and below level 2 (i.e., those who just missed out on achieving five GCSEs A* to C and have not subsequently completed level 2 or above).

The analysis above is supplemented by an IV approach which uses crossing the threshold as an instrument for achieving level 3 or above and how this relates to earnings for my cohorts in 2017-18. Table A6 in the appendix produces findings for models evaluating equations (4) and (5). It shows that for men returns are not statistically significant, yet they are sizeable and statistically significant for women. It should be noted that the output from the instrumental variables approach produces a LATE which is different to the ATE produced from OLS and Probit analysis in previous sections. The LATE produces an estimate for the instrumented variable on the outcome variable, based on the change in the instrument. Thus, Table 8 shows the effect of achieving a level 3 or above on earnings in 2017-18, for those who were induced to achieve more education as a result of crossing the threshold. In light of this, the findings complement those from the mediation analysis (step 3) in section 4.3. Table 6 showed that for both men and women crossing the threshold leads to an

increase in achieving level 3 academic education. Thus, the change in the instrumented variable based on the instrument is the change in level 3 or above coming from crossing the threshold, i.e., the increase in level 3 academic or above achieved. The insignificant result for men shows that this increase in level 3 academic or above (from crossing the threshold) does not lead to higher earnings. Conversely, women see large and positive returns for this uptake in level 3 academic or above achievement. All of this said, the results from Table 7 cast doubt over the exclusion restriction for women. The fact that the coefficient for crossing the threshold remains positive and statistically significant after level 3 achievement (the mediator) is introduced undermines the validity of the instrument. This is because it implies that crossing the threshold is linked with greater earnings (and productivity) for men.

5.4 Heterogeneity

Sub models are run for different socioeconomic and demographic characteristics. Results are shown in Table A7 in the appendix. This shows that the same effect of crossing the threshold holds for males across socioeconomic and demographic characteristics. For women it illustrates that the effect of crossing the threshold on earnings was driven by those who are white, non-FSM eligible and not identified as SEN.

5.5 Robustness checks

I run several robustness checks on my findings. The first check is to see if using employment outcomes at years 11 to 13, i.e., ages 27 to 29 makes a difference to results. I run three separate models on labour market variables at age 27, 28 and 29 and see similar patterns (Table 8), though coefficients differ slightly. Secondly, the definition of employment and earnings is checked to see if this makes a difference to results. As explained in the data section, I restrict my analysis of earnings returns to those who are in employment for the whole of the tax year 2017-18, i.e., 365 days. For the robustness check I use daily earnings of those in employment for at least one day in the tax year (i.e., relaxing the assumption someone needs to be employed for full tax year). This is therefore likely to bring in casual and temporary work. Though the coefficients seen in Table 8 are slightly different to my main model, the patterns are the same as my primary model: for men there is no effect on crossing the threshold, but for women there is an impact upon earnings.

Table 8: Robustness checks - regression output

Model	**Men**				**Women**		
	Model 1 - Employed	**Model 2 - Claiming out of work benefits**	**Model 3 - Earnings**		**Model 1 - Employed**	**Model 2 - Claiming out of work benefits**	**Model 3 - Earnings**
Different years							
Age 27	0.002 (0.004)	0.000 (0.002)	0.002 (0.006)		0.001 (0.004)	-0.005** (0.003)	0.022*** (0.008)
Sample size	43,034		23,721		39,104		22,293
Age 28	-0.002	-0.001	-0.003		0.006	-0.006**	0.021**

	(0.005)	(0.002)	(0.008)		(0.005)	(0.003)	(0.010)
Sample size	27,970		16,842		25,781		15,871
Age 29	-0.011 (0.007)	-0.001 (0.003)	0.008 (0.010)		0.012 (0.007)	-0.002 (0.004)	0.038*** (0.013)
Sample size	14,933		9,052		14,012		8,583
Different definitions							
Daily earnings from employment (at least one day)	0.001 (0.004)		0.004 (0.006)		0.004 (0.004)		0.016** (0.007)
Sample size	43,043		31,619		39,104		29,113

** significant at 5% level, *** significant at 15 level
This is the primary model including those that achieved five full GCSEs A* to C, including at least one C and those who just missed out, i.e., achieving 4 full GCSEs A* to C and at least one D.
Models 1 and 2 are for all the sample. Model 3 is those employed for 365 of the tax year except for daily earnings from employment (at least one day) which is those employed for at least one day in the tax year. This model is also a Probit, thus calculating marginal effects. Whilst al other models are OLS.
The coefficients shown are for the dummy variable indicating if someone just achieved five GCSEs A* to C, i.e., the main explanatory variable of interest (T).
All regressions include a set of controls described in section 4.1, research design.

6. Discussion

This paper finds that crossing the threshold in these high stakes exams at age 16, despite leading to higher post 16 education levels for both men and women, has little impact on labour market outcomes for men, but some for women. I go on to show that this is because men who just cross the threshold do not appear to benefit from taking advanced and higher-level academic qualifications relative to their counterfactual (vocational) education choices, whilst women do. This section focuses on the contributions to the literature and policy implications introduced in the introduction in slightly more depth. I start by examining contributions to the literature: 1) some methodological reflections, 2) a holistic examination of the link between high stake exams, different educational pathways and labour market outcomes and 3) the human capital versus signalling debate. Subsequently, I discuss key policy contributions around ensuring those who do not cross the threshold continue in education (if they wish) and high stakes exams and allocative efficiency.

Starting with a few methodological reflections, it is worth returning to the discussion around causality and whether there are any unobservables between the treatment and the control group driving outcomes. It was observed that the observable differences between those crossing the threshold, and those not, were similar for both men and women. These differences raised concerns around potential selection bias and crossing the threshold being non-random. Yet, the differences in results for men and women offer an interesting insight into this. Should factors such as ability, motivation or other unobservables be driving any difference in labour market outcomes it would be expected to do so for both men and women. However, the fact just achieving five GCSEs A* to C affects women's labour market outcomes and not men's suggests that the unobservables may not be that important. Of course, there is potential that any unobservable factors may work differently for men and women, but

it is difficult to see a rationale for this. All in all, this backs up my key assumptions and the robustness of results presented.

Another point on my empirical strategy is the age when I capture labour market outcomes could be important. Though early career outcomes are important, evidence suggests that the earnings of those with higher education levels are likely to increase more than those with lower education levels in their 30s and 40s (Hodge et al 2021, Britton et al 2020, Hayward et al 2014). Lifetime earnings for marginal students are less explored and particularly those crossing important thresholds. The question becomes whether the higher education levels for men in my sample, and their academic qualifications, have an effect later in their career (or not).

This paper is the first to examine crossing an important high stakes threshold and the effect on education and labour market outcomes in a holistic way, i.e., by examining both academic *and* vocational education routes and their returns. I find that crossing a grade threshold in high stakes qualifications has a significant impact on education outcomes, for both men and women. This almost mirrors Machin et al (2020) who find that those just crossing the C/D English language exam threshold are more likely to have achieved a level 3 qualification, achieved a level 3 from an academic route and started higher education by 19 years old. Though focusing on a different threshold (their analysis is a subset of the one I cover) I find very similar short-term (education) results. I extend this to age 25 and observe that those just crossing the threshold are more likely to complete higher levels of education and do it via an academic route. This is important as it allows higher education to take place and the control group to potentially 'catch up', i.e., close the gap in education outcomes shown at 20 years old. Crossing this threshold does however not impact men's early career outcomes, whilst for women it leads to higher earnings. For women I show that both academic and higher levels of education (associated with crossing the threshold) appear to be driving this. For men, my results are quite different. Vocational qualifications have higher returns than academic qualifications for the marginal student, whilst academic education and higher education levels do not have a (statistically significantly) positive impact on earnings.

Before understanding what is driving these differences between males and females, it is important to understand my sample and how this relates to the wider population and the literature. My sample are more likely to attend a university with lower 'status' and more likely have poorer attainment (to get 2:2 or below[17]) when compared with other graduates. Other studies (Belfield et al 2018a) show returns to different institutions vary hugely and degree classification is important (Naylor et al 2016, de Vries 2014, Boero et al 2020). The average A-level results of my sample are DDD, again at the lower end of the distribution, especially compared with BBB for the rest of the population that did academic level 3 (derived from level 3 participation age 20 points in Table 4). This helps to show why returns to (academic and) higher education might not be what we might normally expect and show why my results are different to other studies focusing on academic level 3 (Hayward et al 2014, McIntosh 2006). Their analysis focused on average returns to A levels rather than for the lower achieving sample I focus on. However, whilst this might explain why my sample might not do as well in the labour market as others with similar levels of academic and higher education, it does not explain the differences between men and women.

Why is it that women at the five GCSE A* to C threshold have positive returns to academic education, but men do not? And why do men have higher returns from

[17] Degree classifications in England are first class honours above 70%, upper class second (2:1) 60-69.9%, lower class second 50-59.9% and third class 40-49.9%.

vocational qualifications? The two are linked. My results correspond with another study (Dearden et al 2004) that focuses on the marginal learner and show women have greater returns to continuing in education and higher education than men. Other literature (Belfield et al 2018b, Walker and Zhu 2003) show that women have greater returns to academic and higher education. When I add in controls for degree classification and university status the level 4 and above coefficient turns positive, which is not hugely different from Belfield et al 2018, who also use the LEO data, and estimate returns to degrees at 8 percent for men and 28 percent for women. The baseline and comparator groups are likely to be important too, especially as men tend to have greater returns from vocational qualifications than women (Buscha and Urwin 2013, Greenwood et al 2007, McIntosh and Morris 2021). Comparisons in the model(s) shown in table 7 are made against level 2 or below, which includes those that did not achieve five GCSEs A* to C, those that did achieve five GCSEs A* to C and did not move beyond it and (crucially) those that completed a vocational level 2. When I split vocational level 2 out of the baseline and include in the analysis (as shown in table A5 in appendix) it helps to explain these results. Women have negative returns to achieving level 2 vocational, whilst men have large positive returns. For women the effect of crossing the threshold is all pulling the same way, whilst for men there is a counter effect. By this I mean women who cross the threshold benefit from greater returns to level 3 academic and level 4 or above, with those not crossing the threshold being more likely to do level 2 vocational which pays less. Whereas for men the level 2 vocational has strong positive returns, which benefits the control group, and the effect of crossing the threshold and earning more due to higher academic education is counter acted. Differences in returns for vocational qualifications across the sexes can be explained by differences in the vocational area studied (Battiston et al 2019, Buscha and Urwin 2013, Greenwood et al 2007, McIntosh 2006), with men more likely to take vocational qualifications in higher earning sector areas (such as engineering, construction, and ICT), whilst women are more likely to take qualifications in lower returns sector areas (such as health and education). All the factors mentioned above contribute to explaining my main findings. That is, why for men in my sample taking a vocational route seems financially the best option, whilst for women it is academic.

The findings in this paper also contribute to the signalling versus human capital debate. I am able to observe signalling at two points and to two different receivers: 1) education institutions and 2) employers. The latter is signalling as usually referred to in economics, i.e., a signal of unobserved ability and productivity to employers to overcome information asymmetry. Starting with the education outcomes, for both men and women, it appears that crossing the threshold is sending a signal to education institutions that individuals are more capable of continuing down an academic route. Though alternatively it could be a self-signal, i.e., that individuals believe they are not capable and changes their choices. For employers, i.e., the more ‘classic sense’ of signalling things, things are more complex because education paths diverge. Other studies, for example Clark and Martorell’s (2014) have focused simply on the link of crossing the threshold and labour market outcomes. They argue that just achieving a high school diploma is simply a piece of paper and would not increase productivity. Conversely, it would be expected to send a signal to employers, but they find it does not. As achieving five GCSEs A* to C is similar to a achieving a high school diploma and was seen a certificate for ‘passing’ secondary school, I could apply the same logic here. The complication comes from the differences in education choices. However, I am able to examine potential signals to employers after taking account of these different education pathways. For women, there is tentative evidence that supports signalling theory. We’ve seen that the treatment and control group are very similar in their socioeconomic and demographic characteristics and education profiles, the only difference being one crossed the

threshold and the other did not. Crossing the threshold leads to a higher uptake in academic and higher education, which has positive and significant returns. However, whilst the coefficient for crossing the threshold decreases, it remains significant and sizeable. This suggests there may be a residual, and direct effect, of crossing the threshold on labour market outcomes and therefore offers some support to signalling theory. For men I find no support for signalling theory. This is because after accounting for post-compulsory education outcomes the coefficient for crossing the threshold remains small and insignificant.

This paper also makes an important policy contribution for marginal learners, in this case some (policy) considerations for supporting those who just miss out on crossing important thresholds. Whilst the impact on early career labour market outcomes is fairly minimal, for men at least, we have seen that the impact upon education outcomes is substantial. As mentioned in the above paragraph, crossing the five GCSE A* to C threshold may send a signal (which is received) to education institutions that an individual is likely to be more successful in completing academic education. Schools and colleges (and to a lesser extent universities) appear to be using the number of GCSEs A* to C achieved (or at least around the four/five threshold to focus on) to 'screen' candidates (or conversely candidates just passing send a signal). The alternative is that individuals receive a signal themselves about their own ability to continue with their academic studies and making decisions to take other routes. Or both aspects could contribute to explaining the large difference in academic education. Policymakers and leaders in the education sector, especially in England, may want to reflect on this. Should a candidate wish to continue on the academic route, education providers are encouraged to ensure they are allowed to do so, if they just missed out. And alternatively, individuals should be supported to make the choice they want and encouraged to continue down an academic route, even if they just missed out on crossing an important threshold. They may want to reflect on the financial aspect of this, particularly men, but should have the choice. For women, this could be important in terms of allocative efficiency. Should a policy solution be found to allow women just missing out to access higher levels of academic education, it should (at least partly) reduce the gap between the treatment and the control group. Thus, it would allow those just missing out (for factors potentially outside of their control) to increase their human capital and improve the overall productivity of the economy.

All this said, the findings in this paper offer some food for thought for pupils, parents, careers advisors, school leavers and policymakers around whether it might be best to take an academic or vocational route. This paper can only offer an insight into the financial and value for money side of things, and on early career outcomes, which is one element. But it does challenge the assumption that an academic (i.e. A-levels and a degree in England) is the best route and perhaps the 'default' option for those crossing a threshold (particularly men) should not be to continue with an academic education. In reality choices are complex, and individuals should be encouraged to do what they are good at and passionate about (as much as possible). This is where I advocate that the marginal learner is given a choice as to which route they can go down, or be supported to take, even if they just miss out. I say this from a system perspective and thus making sure education institutions offer flexibility, it is built into career guidance, teachers and parents are aware etc. However, the results in this paper on academic and vocational qualification returns should be one thing that mid attaining pupils use as one aspect of their decision making.

Another key policy contribution is around high stakes exams and allocative efficiency, especially for men. This evidence is timely, as many countries across the world cancelled exams in 2021, including the United Kingdom, France, Belgium, Ireland,

the Netherlands, Norway, Mexico, Pakistani, Saudi Arabia and India. Some of these countries may be considering exams and qualifications reform and potentially, as part of this, the future of exam systems and high stakes exams. The findings in this paper suggest that just missing out on a particular exam grade may not necessarily have serious consequences on the life chances of all individuals, or at least for men. This is reassuring, as at the margins, those not crossing the threshold may have done so because of random factors. Many individuals achieve marks or grades they deserve and factors such as ability and motivation play a key role in this. However, for the marginal student (i.e., those just either side of a threshold) there is a risk this is not the case. I have already established in the introduction that external factors and measurement error have been shown to affect exam performance. It is feasible that other factors outside of the individual's control may affect their grade or marks achieved (for example, cognitive acuteness has been shown to be affected by factors such as sleep deprivation and noise (Angus et al 1985 and Goriounova and Mansvelder 2012), thus feasibly affecting exam performance). If this in turn affects an individual's labour market outcomes, it would lead to allocative inefficiency. This is because these individuals may not reach their full potential from a human capital point of view, i.e., they are earning less money and doing lower skilled jobs than they have the potential for. Thus, they would be less productive than they could be. It should, however, be noted that I only focus on one aspect of the high-stakes exams debate. [18]

[18] There are various factors that could be consider and the key one is what the alternative is. I do not explore the case for high stake exams versus alternatives in this paper. Those interested in the wider considerations of high stakes exams compared to alternatives can see a well-structured and balanced summary of the evidence by Wyness et al 2021

References

Aakvik, A., Salvanes, K. G., & Vaage, K. (2010). Measuring heterogeneity in the returns to education using an education reform. *European Economic Review*, *54*(4), 483-500.

Anelli, M. (2020). The returns to elite university education: A quasi-experimental analysis. *Journal of the European Economic Association*, *18*(6), 2824-2868.

Angrist, J. D., & Keueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings?. *The Quarterly Journal of Economics*, *106*(4), 979-1014.

Angus, R. G., Heslegrave, R. J., & Myles, W. S. (1985). Effects of prolonged sleep deprivation, with and without chronic physical exercise, on mood and performance. *Psychophysiology*, *22*(3), 276-282.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology, 5, 1173-1182

Battiston, A., Patrignani, P., Hedges, S., & Conlon, G. (2019). *Labour market outcomes disaggregated by subject area using the Longitudinal Education Outcomes (LEO) data*. CVER Discussion Paper 021.

Belfield, C., Britton, J., Buscha, F., Dearden, L., Dickson, M., Van Der Erve, L., ... & Zhu, Y. (2018a). The relative labour market returns to different degrees.

Belfield, C., Britton, J., Buscha, F., Dearden, L., Dickson, M., Van der Erve, L., ... & Zhu, Y. (2018b). The impact of undergraduate degrees on early-career earnings.

BOERO, G., COOK, D., NATHWANI, T., NAYLOR, R., & SMITH, J. (2020). HOW DOES THE RETURN TO A DEGREE VARY BY CLASS OF AWARD?

Britton, J., Dearden, L., van der Erve, L., & Waltmann, B. (2020). The impact of undergraduate degrees on lifetime earnings: Research report, February 2020.

Boliver, V. (2015). Are there distinctive clusters of higher and lower status universities in the UK?. *Oxford review of education*, *41*(5), 608-627.

Buscha, F., & Urwin, P. J. (2013). Estimating the labour market returns to qualifications gained in English further education using the Individualised Learner Record (ILR).

Canaan, S., & Mouganie, P. (2018). Returns to education quality for low-skilled students: Evidence from a discontinuity. *Journal of Labor Economics*, *36*(2), 395-436.

Caplan, B. (2018). The case against education: Why the education system is a waste of time and money. Princeton University Press.

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), 176: 431-457

Clark, D., & Martorell, P. (2014). The signaling value of a high school diploma. *Journal of Political Economy*, *122*(2), 282-318.

Crawford, C. (2014). *The link between secondary school characteristics and university participation and outcomes: CAYT research report*. Department for Education, Runcorn, England.

Dearden, L., McGranahan, L., & Sianesi, B. (2004). *Returns to Education for the 'Marginal Learner': Evidence from the BCS70* (No. 45). Centre for the Economics of Education, London School of Economics and Political Science.

Dee, T., W. Dobbie, B. Jacob, and J. Rockoff (2016). The Causes and Consequences of Test Score Manipulation: Evidence from the New York Regents Examinations (No. w22165). National Bureau of Economic Research.

De Philippis, M. (2016). *Studying science: the impact of school curriculum on degree choice* (No. 493). Centre for Economic Performance, LSE.

De Vries, R. (2014). Earning by degrees: differences in the career outcomes of UK graduates.

Diamond, R., and P. Persson (2016). The long-term consequences of teacher discretion in grading of high-stakes tests (No. w22207). National Bureau of Economic Research.

Ebenstein, A., V. Lavy, and S. Roth (2016). The Long Run Economic Consequences of High Stakes Examinations: Evidence from Transitory Variation in Pollution. American Economic Journal: Applied Economics, 8(4), 36-65

Galindo-Rueda, F., Marcenaro-Gutierrez, O., & Vignoles, A. (2004). The widening socio-economic gap in UK higher education. *National Institute Economic Review*, *190*, 75-88.

Gayle, V., D. Berridge and R. Davies (2002), Young people's entry into higher education: quantifying influential factors, Oxford Review of Education, vol. 28, pp. 5–20.

Goriounova, N. A., & Mansvelder, H. D. (2012). Short- and long-term consequences of nicotine exposure during adolescence for prefrontal cortex neuronal network function. Cold Spring Harbor perspectives in medicine, 2(12), a012120.

Hastings, J. S., Neilson, C. A., & Zimmerman, S. D. (2013). *Are some degrees worth more than others? Evidence from college admission cutoffs in Chile* (No. w19241). National Bureau of Economic Research.

Hayward, H., Hunt, E., & Lord, A. (2014). *The economic value of key intermediate qualifications: estimating the returns and lifetime productivity gains to GCSEs, A levels and apprenticeships*. London: Department for Education.

Hodge, Louis, Allan Little, and Matthew Weldon. "GCSE attainment and lifetime earnings: June 2021." (2021).

Hoekstra, M. (2009). The effect of attending the flagship state university on earnings: A discontinuity-based approach. *The review of economics and statistics*, *91*(4), 717-724.

Jenkins, A., Greenwood, C., & Vignoles, A. (2007). *The returns to qualifications in England: updating the evidence base on level 2 and level 3 vocational qualifications* (No. 89). Centre for the Economics of Education, London School of Economics and Political Science.

Jia, R., & Li, H. (2021). Just above the exam cutoff score: Elite college admission and wages in China. *Journal of Public Economics*, *196*, 104371.

Lemieux, T., & Card, D. (2001). Education, earnings, and the 'Canadian GI Bill'. *Canadian Journal of Economics/Revue canadienne d'économique*, *34*(2), 313-344.

Machin, S., McNally, S., & Ruiz-Valenzuela, J. (2020). Entry through the narrow door: the costs of just failing high stakes exams. *Journal of Public Economics*, *190*, 104224.

Marshall, A. (2009). *Principles of economics: unabridged eighth edition*. Cosimo, Inc.

McIntosh, S. (2006). Further analysis of the returns to academic and vocational qualifications. *Oxford bulletin of economics and statistics*, *68*(2), 225-251.

McIntosh, S., & Morris, D. (2021). Variation in the labour market rewards to vocational qualifications in the UK. *Scottish Journal of Political Economy*, *68*(5), 535-552.

Mincer, J. (1974). Schooling, Experience, and Earnings. Human Behavior & Social Institutions No. 2.

Naylor, R., Smith, J., & Telhaj, S. (2016). Graduate returns, degree class premia and higher education expansion in the UK. *Oxford Economic Papers*, *68*(2), 525-545.

Rhead, Stephen, Beth Black, and Anne Pinot de Moira. "Marking consistency metrics: 14 November 2016." (2016).

Rhead, S., Black, B., & Pinot de Moira, A. (2018). Marking consistency metrics: an update. *Coventry: Ofqual*.

Park, J. (2017). Hot temperature and high stakes exams: Evidence from new york city public schools. *URL: https://scholar.harvard.edu/files/jisungpark/files/papernycaejep. pdf*.

Rees, A. (1965). Human capital: A Theoretical and empirical analysis with special reference to education.

Riley, J. G. (2001). Silver signals: Twenty-five years of screening and signaling. *Journal of Economic literature*, *39*(2), 432-478.

Saavedra, J. E. (2008). The returns to college quality: A regression discontinuity analysis.

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: new procedures and recommendations. Psychological Methods, 7, 422-445.

Smith, A. (1937). *The wealth of nations [1776]* (Vol. 11937).

Spence, M. (1973). Job Market Signalling": The Quarterly Journal of Economics".

Walker, I., & Zhu, Y. (2003). Education, earnings and productivity: recent UK evidence. *Labour Market Trends*, *111*(3), 145-152.

Weiss, A. (1995). Human capital vs. signalling explanations of wages. *Journal of Economic perspectives*, *9*(4), 133-154.

Wyness, Gill, (2021), Should we abolish GCSEs?, No 14, CEPEO Briefing Note Series, UCL Centre for Education Policy and Equalising Opportunities, https://EconPapers.repec.org/RePEc:ucl:cepeob:14.

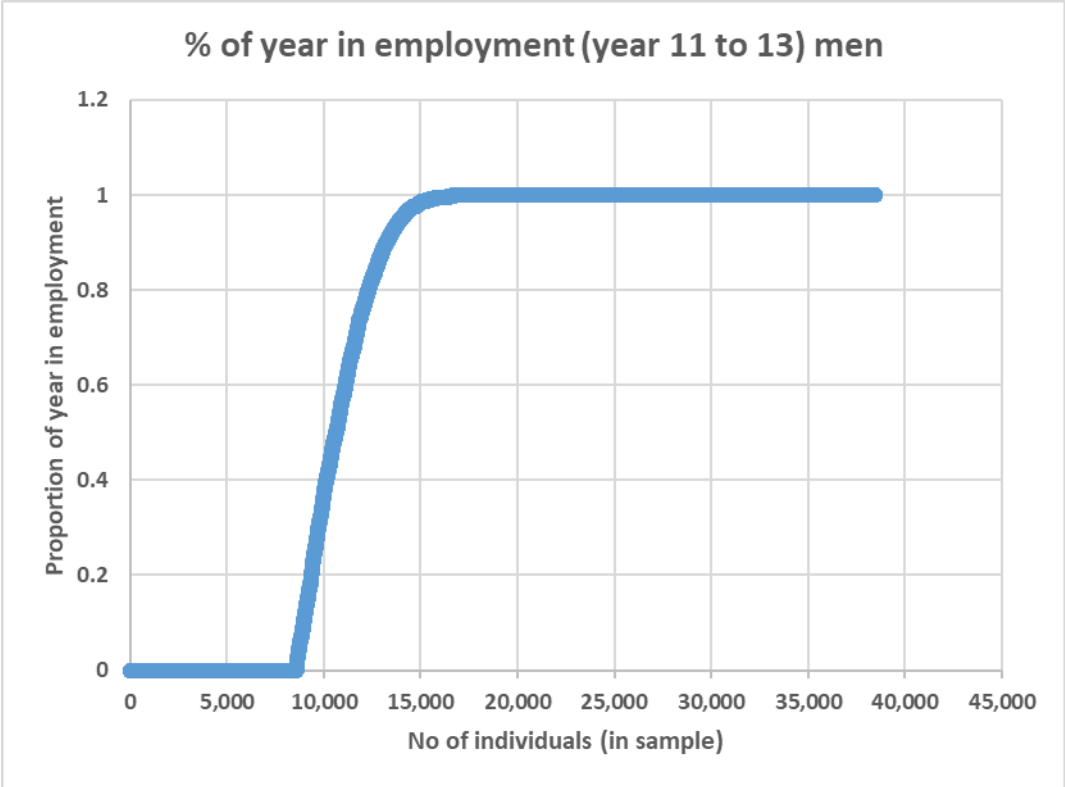
Zimmerman, S. D. (2019). Elite colleges and upward mobility to top jobs and top incomes. *American Economic Review*, *109*(1), 1-47.

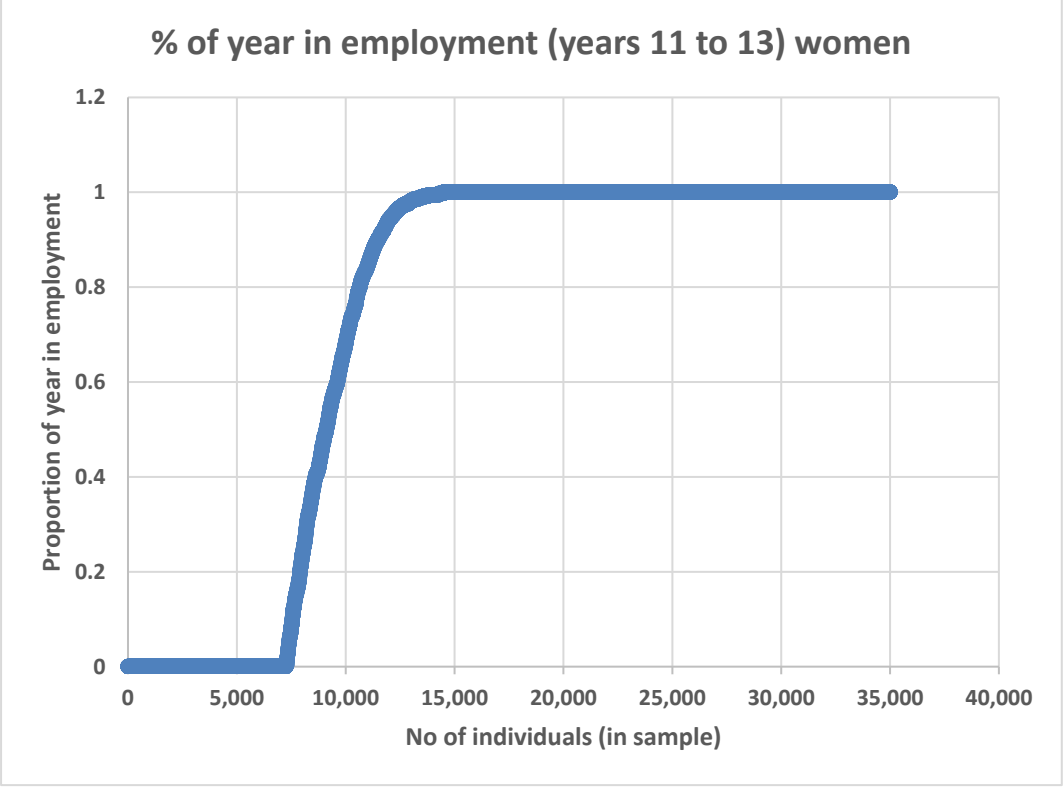
Appendix – additional figures and tables

Table A1: Descriptive statistics on number and proportion of sample in employment age 27 to 29

	Number in employment		Percentage in employment	
	Men	**Women**	**Men**	**Women**
0 days employed	9,624	8,210	22.4%	21.0%
1 to 364 days employed	9,128	8,071	21.2%	20.6%
365 days employed	24,282	22,823	56.4%	58.4%
Total	43,034	39,104		
Analysis includes those that achieved five full GCSEs A* to C, including at least one C and those who just missed out, i.e., achieving 4 full GCSEs A* to C and at least one D.				

Figure A1: Mapping proportion of sample in employment age 27 to 29 for men and women





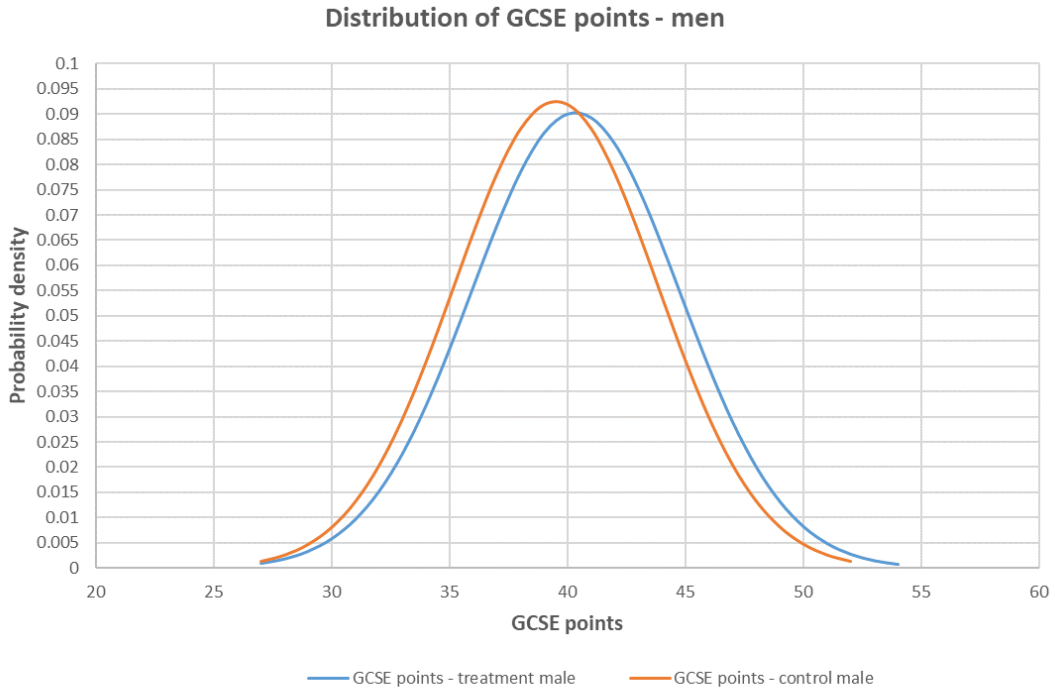
Note – includes all those that achieved five full GCSEs A* to C, including at least one C and those who just missed out, i.e., achieving 4 full GCSEs A* to C and at least one D.

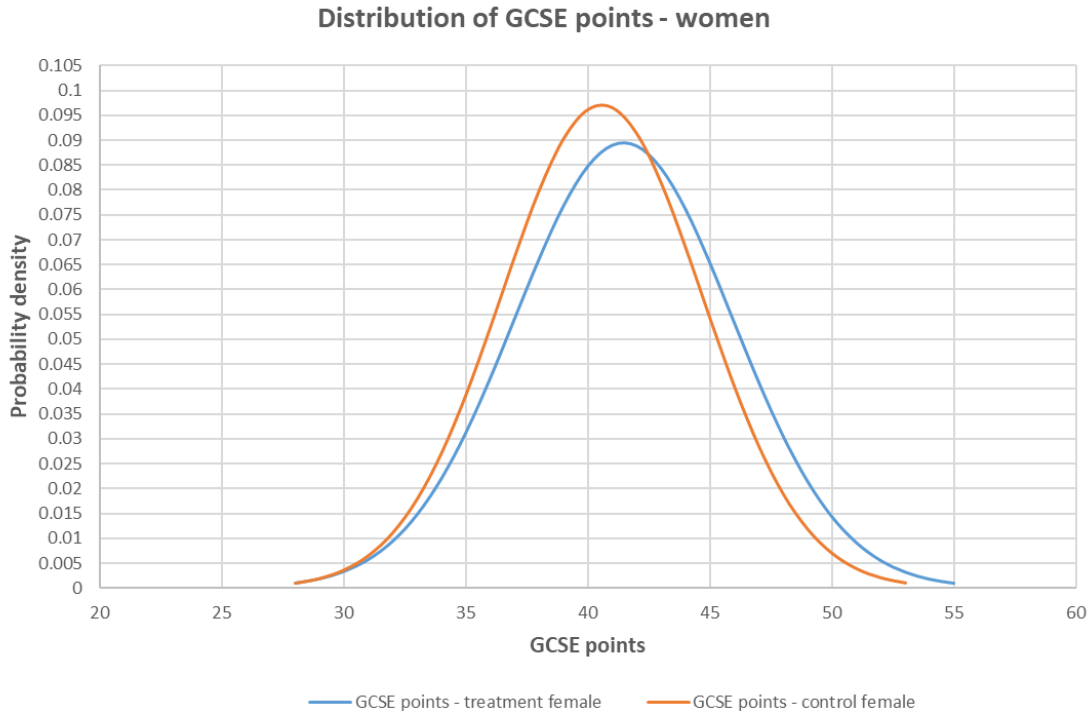
Table A2: Descriptive statistics (mean) - GCSE subject entries

Subject	Men			women		
	Treatment group	Control group	T-C	treatment group	control group	T-C
English	1.000	1.000	0.000	1.000	1.000	0.000
Maths	1.000	1.000	0.000	1.000	1.000	0.000
English literature	0.902	0.898	0.005	0.945	0.940	0.005
Double science	0.900	0.894	0.007**	0.880	0.875	0.007**
Single science	0.094	0.091	0.002	0.119	0.114	0.004
Chemistry	0.037	0.030	0.008***	0.019	0.016	0.003***
Biology	0.036	0.033	0.009***	0.027	0.021	0.007***
Physics	0.041	0.032	0.009***	0.018	0.014	0.004***
History	0.368	0.362	0.008	0.334	0.327	0.013***
Geography	0.389	0.381	0.008	0.276	0.275	0.000
French	0.419	0.418	0.002	0.473	0.459	0.014***
German	0.171	0.158	0.013***	0.169	0.164	0.005
Spanish	0.069	0.066	0.002	0.096	0.097	0.001
Art	0.304	0.297	0.008	0.429	0.429	0.001
Physical education	0.372	0.365	0.008	0.232	0.226	0.007
Drama	0.151	0.144	-0.007	0.257	0.250	0.005
Music	0.096	0.088	0.007***	0.081	0.078	0.003
Technology electronics	0.062	0.059	0.002	0.003	0.004	0.000
Technology food	0.106	0.111	0.004	0.253	0.264	-0.012***
Technology resistant materials	0.311	0.329	-0.018***	0.071	0.067	0.004
Technology textiles	0.007	0.006	0.000	0.191	0.186	0.003
Information Technology	0.272	0.263	0.009**	0.183	0.177	0.006
Religious Education	0.242	0.236	0.006	0.317	0.303	0.015**
Business studies	0.197	0.203	-0.008**	0.151	0.150	-0.001

Analysis includes those that achieved five full GCSEs A* to C, including at least one C and those who just missed out, i.e., achieving 4 full GCSEs A* to C and at least one D. Sample means shown in columns 2, 3, 4 and 5.

Figure A2: Descriptive statistics – GCSE points distribution charts





Note – includes those that achieved five full GCSEs A* to C, including at least one C and those who just missed out, i.e., achieving 4 full GCSEs A* to C and at least one D. GCSE points are adjusted to account for the crossing the threshold, i.e., the C/D difference (one GCSE point).

Table A3: descriptive statistics (means): socioeconomic, demographic and education factors

Variable	Men		Women	
	Treatment (T)	**Control (C)**	**Treatment (T)**	**Control (C)**
Socioeconomic and demographic characteristics	%	%	%	%
IDACI	0.182	0.189	0.199	0.205
FSM eligible	0.084	0.095	0.105	0.115
SEN	0.101	0.112	0.060	0.071
Minority ethnic	0.118	0.115	0.139	0.137
State school (non-grammar)	0.926	0.949	0.944	0.960
First language English	0.078	0.076	0.092	0.088
Region at GCSEs: London	0.123	0.124	0.137	0.128
Region at GCSEs: North East	0.049	0.048	0.046	0.047
Region at GCSEs: North West	0.168	0.168	0.163	0.171
Region at GCSEs: Yorkshire and the Humber	0.099	0.099	0.104	0.099
Region at GCSEs: East Midlands	0.075	0.077	0.074	0.085
Region at GCSEs: West Midlands	0.085	0.088	0.087	0.082
Region at GCSEs: South East	0.173	0.166	0.160	0.166
Region at GCSEs: South West	0.100	0.104	0.103	0.099
Region at GCSEs: East of England	0.110	0.108	0.112	0.108
Education factors				
National test age 11 total marks (English and maths)	105.6	103.4	102.9	100.4
State school (non-grammar)	0.919	0.942	0.938	0.956
School performance measure	0.062	0.026	0.050	0.083
Peer effect: % of FSM pupils in school	0.126	0.131	0.135	0.0137
Peer effect: % of SEN pupils in school	0.139	0.142	0.138	0.140
Peer effect: % of pupils achieving five GCSEs A* to C in cohort	0.571	0.548	0.565	0.546
Analysis includes those that achieved five full GCSEs A* to C, including at least one C and those who just missed out, i.e., achieving 4 full GCSEs A* to C and at least one D.				

Table A4: Medium term (post 16) education outcomes - regression outputs

	Model 1) Achieved level 2 vocational by age 20	Model 2) Achieved level 2 vocational or above at age 25	Model 3) Achieved level 3 or above at age 25	Model 4) Achieved level 3 academic or above at age 25	Model 5) Achieved level 3 vocational or above at age 25	Model 6) Achieved level 4 or above at age 25
Men	-0.060*** (0.004)	-0.065*** (0.005)	0.060*** (0.005)	0.066*** (0.004)	-0.005 (0.005)	0.038*** (0.004)
Sample size	43,034					
Women	-0.075*** (0.005)	0.077*** (0.005)	0.046*** (0.005)	0.070*** (0.004)	-0.024*** (0.005)	0.040*** (0.004)
Sample size	39,104					

** significant at 5% level, *** significant at 1% level

This is the primary model including those that achieved five full GCSEs A* to C, including at least one C and those who just missed out, i.e., achieving 4 full GCSEs A* to C and at least one D. Models 1 to 4 include the whole sample.

The coefficients shown are for the dummy variable indicating if someone just achieved five GCSEs A* to C, i.e., the main explanatory variable of interest (T). Estimated using Probit model and marginal effects calculated.

All regressions include a set of controls described in section 4.1, research design.

Table A5: Robustness check - Effect on earnings age 27 to 29 of crossing the threshold and post 16 education (as a mediator) including level 2 vocational – regression output

	Model 1 (Original model)	Model 2a (Level 2 vocational achievement as mediator)	Model 2b (Level 3 achievement as mediator)	Model 3 (Level 3 academic and vocational achievement as mediators)	Model 4 (Level 4 or above achievement as mediator)	Model 5 (Level 4 or above achievement as mediator – with controls)
Men						
Five GCSEs A* to C (explanatory variable)	0.001 (0.006)	0.005 (0.006)	0.002 (0.006)	0.003 (0.006)	0.002 (0.006)	-0.002 (0.006)
Level 2 or above vocational achievement		0.068*** (0.007)	0.083*** (0.007)	0.080*** (0.007)	0.086*** (0.007)	0.073*** (0.007)
Level 3 or above achievement			0.067*** (0.007)		0.060*** (0.007)	0.075*** (0.007)
Level 3 or above academic achievement				0.055** (0.009)		
Level 3 or above vocational (only) achievement				0.063*** (0.007)		
Level 4 or above					0.023*** (0.009)	0.115*** (0.014)
Sample size	23,798					
Women						
Five GCSEs A* to C (explanatory variable)	0.031*** (0.008)	0.024*** (0.008)	0.021*** (0.008)	0.017** (0.008)	0.017** (0.008)	0.018** (0.008)
Level 2 or above vocational achievement		-0.078*** (0.009)	-0.048*** (0.009)	-0.035*** (0.009)	-0.025*** (0.008)	-0.032*** (0.009)
Level 3 or above achievement			0.112*** (0.008)		0.039*** (0.009)	0.042*** (0.009)

Level 3 or above academic achievement				0.157*** (0.011)		
Level 3 or above vocational (only) achievement				0.056*** (0.009)		
Level 4 or above					0.215*** (0.010)	0.255*** (0.015)
Sample size	22,367					

** significant at 5% level

This model includes those that achieved five full GCSEs A* to C, including at least one C and those who just missed out, i.e., achieving 4 full GCSEs A* to C and at least one D.

The dependent variable is log of earnings for all models.

Models 1 is step 1 from the mediation analysis, i.e., as shown in Table 5. Model 2a adds a dummy variable for if someone achieved level 2 vocational as a mediator and subsequently model 2b adds achieved a dummy variable for if someone achieved level 3 or above as a mediator. Model 3 replaces achieved level 3 with two dummy variables for if someone achieved a level 3 academic or above or level 3 vocational or above. Model 4 adds a dummy variable for achieved level 4 or above to model 2. All models include those employed for 365 of tax year. Model 5 adds control variables for university 'status', degree classification and subject studied.

The baseline is those achieving level 2 academic, or those just missing out on achieving level 2 academic (i.e. five GCSEs A* to C). The difference between this and table 7 is that table 7 also includes level 2 vocational achievement in the baseline, whereas in this table it is split out and included in the analysis.

The coefficients shown are for the dummy variable indicating if someone just achieved five GCSEs A* to C, i.e., the main explanatory variable of interest (T). All models use OLS as estimator.

All regressions include a set of controls described in section 4.1, research design.

Table A6: Effect of achieving level 3 or above on labour market outcomes - Instrumental variables regression output

	Log (earnings) age 27-29 – IV
Men	0.019 (0.109)
Sample size	23,798
Women	0.613*** (0.168)
Sample size	22,367

** significant at 5% level, *** significant at 1% level.
This is the IV model focusing on equations (4) and (5) in section 4.1. It includes those that achieved five full GCSEs A* to C, including at least one C and those who just missed out, i.e., achieving 4 full GCSEs A* to C and at least one D.
The model includes those that were employed for 365 days of tax year.
The coefficients shown are for the dummy variable indicating if someone just achieved level 3 or above, conditional on the instrument, five GCSEs A* to C.
All regressions include a set of controls described in section 4.1, research design.

Table A7 – Heterogeneity regression results: socioeconomic and demographic characteristics

Model	Men				Women		
	Employed	Claiming out of work benefits	Earnings		Employed	Claiming out of work benefits	Earnings
Lower socioeconomic status: FSM eligible	0.004 (0.015)	0.011 (0.009)	-0.008 (0.023)		0.024 (0.014)	-0.020 (0.011)	0.008 (0.026)
Sample size	3,434		1,765		3,833		1,983
Not FSM eligible	0.002 (0.005)	-0.001 (0.002)	0.009 (0.007)		0.000 (0.005)	-0.005 (0.003)	0.022** (0.008)
Sample size	35,030		26,736		31,185		24,385
Lower socioeconomic status: upper quintile IDACI	-0.002 (0.011)	0.004 (0.006)	-0.010 (0.017)		0.015 (0.011)	-0.013 (0.08)	0.039 (0.020)
Sample size	5,552		3,015		5,948		3,272
IDACI lower quintile	0.003 (0.010)	-0.004 (0.003)	0.002 (0.014)		-0.019 (0.010)	-0.009 (0.005)	0.025 (0.018)
Sample size	7,895		6,017		6,583		5,232
IDACI interquintile range (2nd to 4th combined)	0.002 (0.006)	0.001 (0.002)	0.010 (0.008)		0.008 (0.006)	-0.005 (0.004)	0.019 (0.010)
Sample size	22,864		15,570		20,767		16,221
BAME	0.003 (0.013)	-0.003 (0.006)	-0.001 (0.023)		0.000 (0.012)	-0.009 (0.008)	0.026 (0.024)
Sample size	4,479		2,169		4,845		2,478
White	0.002 (0.005)	0.000 (0.002)	0.005 (0.007)		0.002 (0.005)	-0.006** (0.003)	0.020** (0.009)
Sample size	33,985		25,293		30,173		23,518
SEN	-0.004 (0.014)	0.005 (0.007)	-0.012 (0.022)		0.032 (0.018)	0.005 (0.013)	0.019 (0.034)
Sample size	4,085		2,098		2,284	1,161	4,085
Not identified as SEN	0.003 (0.005)	-0.001 (0.002)	0.010 (0.007)		0.000 (0.005)	-0.007** (0.003)	0.021*** (0.008)
Sample size	34,379		26,359		32,734		25,532

 @cepeo_ucl

ucl.ac.uk/ioe/cepeo