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 at the impact of crossing an important high stakes threshold on both academic and vocational education choices and ultimately early career labour market outcomes. It does so by comparing those either side of an important threshold in the English education system at the end of compulsory schooling which was commonly regarded as the minimum benchmark for continuing into post-compulsory education. I find that crossing this threshold led to a 5.8-6.0 percentage point increase in the proportion of men and women continuing to the next education level. Crossing the threshold had a positive effect on earnings for women (3.4 percent) but not for men. The results for men can be explained by the fact that more marginal learners, who just crossed the threshold, were more likely to go on to low return academic qualifications, while those not crossing the threshold instead opted for relatively high return vocational education. I extend the analysis to other thresholds to test the uniqueness of the salience of this threshold on labour market outcomes (for women) and find that crossing other thresholds leads to better labour market outcomes. Equally, this can be explained by women crossing different thresholds leading to a greater likelihood of continuing in academic education, which has positive returns.

REVISED: July 2024

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 at the impact of crossing an important high stakes threshold on both academic *and* vocational education choices and ultimately early career labour market outcomes. It does so by comparing those either side of an important threshold in the English education system at the end of compulsory schooling which was commonly regarded as the minimum benchmark for continuing into post-compulsory education. I find that crossing this threshold led to a 5.8-6.0 percentage point increase in the proportion of men and women continuing to the next education level. Crossing the threshold had a positive effect on earnings for women (3.4 percent) but not for men. The results for men can be explained by the fact that more marginal learners, who just crossed the threshold, were more likely to go on to low return academic qualifications, while those not crossing the threshold instead opted for relatively high return vocational education. I extend the analysis to other thresholds to test the uniqueness of the salience of this threshold on labour market outcomes (for women) and find that crossing other thresholds leads to better labour market outcomes. Equally, this can be explained by women crossing different thresholds leading to a greater likelihood of continuing in academic education, which has positive returns.

Acknowledgements

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, Landuad et al 2022, Bjork and Karhunen 2019). Equally just passing a high stakes exam, or just attaining a certain grade, has been shown to have a long-term impact on labour market outcomes in several settings (Canaan and Mouganie (2018) in France, Landuaud et al (2022) in Norway, Ebenstein in Israel and Bjork and Karhunen (2019) in Finland), though not all (the exception is Clark and Martorell (2014) in the United States). 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. Yet, understanding how just passing or failing a high stakes exam affects individuals' education pathways and labour market outcomes is important because of the implications for allocative efficiency, i.e. individuals not achieving their full potential (potentially due to factors outside their control) and this impacting upon productivity.

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. Crucially, I am able to explore whether post 16 education choices and outcomes mediate any observed differences in earnings. I leverage the uniqueness of the English education system, where compulsory high stakes exams are taken at 16 (not 18). This is important because individuals make choices of whether to continue in academic or vocational education at this age. I am able to do this using detailed administrative records on compulsory education in England linked to subsequent education and tax records. 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). The paper examines a (formerly) extremely important threshold in the English exam system: achieving five General Certificate of Secondary Education (GCSE) passes A* to C. This threshold (concentrated around the middle of the overall results distribution) 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).

I have rich administrative data allowing me to track individuals through the compulsory education system, into post compulsory education and the transition into the labour market. Though I do not have access to exam marks (which are not available) I 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 in the same way as running variable (though I do not undertake a regression discontinuity design approach) 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 (conditionally) and thus crossing the threshold is due to arguably random factors which are not fully under the control of the individual. Should falling either side of the threshold impact upon labour market outcomes, it would lead to allocative efficiency and fairness issues, as individuals would not be achieving their full human capital due to factors outside of their control.

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. For example, crossing the threshold leads to a 3.4 percent increase in earnings at age 27 to 29, i.e. 11 to 13 years later. I go on to show that this is because women who just cross the threshold benefit from taking academic and higher-level academic qualifications. However, there is a tension for men, as whilst men do have positive returns to academic and higher level of education, those not crossing the threshold are more likely to take lower-level vocational qualifications which have strong returns. These findings are 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).

I extend the analysis to focus on other (GCSE A* to C) thresholds to see if crossing the five GCSE A* to C threshold is exclusively salient in terms of early career earnings (for women). I find that crossing a range of other different thresholds (one, two, three, four, six, seven, eight, nine and ten) has an effect on earnings for women, between 2 to 3 percent at age 27 to 29. Thus, it appears that 'just' achieving an additional C grade (seen as a 'good' pass) is salient in terms of labour market outcomes, rather than it being exclusively the five GCSE A* to C threshold. This is an important finding as it suggests the (five GCSEs A* to C) threshold is not actually very special in terms of salience to employers, when compared with other thresholds (e.g. one, two etc GCSE grade C).

This paper makes an important contribution to the literature, by offering 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. The English context is important here, as taking high stakes exams at 16 (not 18) gives the opportunity for individuals to continue in academic education, take a vocational education pathway or even join the labour market. 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

2

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. Landaud et al (2022) use the unique Norwegian context to show that luck in exam content has a substantial and impact upon wages, but similarly these are exams taken at 18 and they explore the link to higher education only[1]. By focusing on both academic and vocational pathways I am able to offer new insight.

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'.[2]

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 (post compulsory) 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 education 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

---

[1] In addition, there are several papers that find crossing high stakes has an impact on labour market (Hoekstra 2009, Saavedra 2014, Hastings et al 2013, Zimmerman 2019, Ebenstein et al 2016, Anelli 2020, Jia and Li 2021 in China). 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.

[2] 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.

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 and removed tiering, most subjects had a two-tier system, with a higher or foundation tier. There were some exceptions which did not have a tier (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[3] (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 going to 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[4]. 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, meaning schools were put into league tables

---

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

based on the proportion of students achieving five GCSEs A* to C. Though the accountability system changed, the 'five good passes' threshold continued to be reported on for many years afterwards. It is, or was, the closest thing to passing secondary school, in a way a high school diploma in the US might be seen. Even now, studies often refer to the number of pupils receiving five GCSEs 9-4 (the equivalent since GCSEs have been reformed). Additionally, as mentioned above, five GCSEs A* to C (or now 9-4) is seen as a full level 2 qualification and a level 2 qualification is expected to continue to level 3 education and beyond (such as A-levels). Thus, the achieving five GCSEs A* to C threshold could be argued to have left quite a legacy on the English education system and hence I wanted to understand if this was important outside of the education, i.e., in the labour market.

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 including almost everyone who went to school in England (including those attending private, selective, and special schools) and is not based on sampling.

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 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)[5], 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 (school) 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

---

[5] 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.

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. 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[6]. This includes more 'classic' unemployment benefits and disability out of work benefits[7]. 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 employment in the tax year. Earnings are only included for those in employment for the whole tax year[8] in order to filter out temporary work and ensure like for like comparison. Earnings are also adjusted to the latest tax year of data available (2017-18) using the Consumer Prices Index including owner occupiers' housing costs (CPIH) as I use multiple cohorts. 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

---

[6] Further education: outcome-based success measures, Methodology – Explore education statistics – GOV.UK (explore-education-statistics.service.gov.uk)
[7] 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)
[8] 365 days or 366 in a leap 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,[9] 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[10] 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 in the same way as running variable (though I do not use a regression discontinuity approach), 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 factors could have contributed to them 'just passing' or 'just missing out'. This is a particular problem for maths[11] and reduces the sample slightly further, to 82,138

---

[10] 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.

[11] As introduced in the background section, maths previously had a three-tier system. The

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.

I subsequently extend the analysis, beyond the five GCSE threshold, to focus on a range of C/D thresholds. I have established there is a relationship between crossing the five GCSE A* to C threshold and earnings for women. I subsequently test the salience of these results for crossing the (five GCSE A* to C) threshold on labour market outcomes, i.e. the one, two, three, four, six, seven, eight, nine and ten (GCSE A* to C) thresholds. The same principle is applied as in the previous paragraph, with those achieving X number of A* to C's and at least one C in the treatment group and those achieving X number of A* to C and at least one D in the control group. The numbers in treatment and control groups are shown in table A5 in the appendix.

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. The sample I have selected focuses on a marginal sample in the middle of the GCSE attainment distribution (explored in Table 2 below). As individuals with the aforementioned socioeconomic and demographic characteristics tend to be towards the bottom of the GCSE attainment distribution, it explains why they are slightly over represented in my sample.

Table 1: Comparison of my sample and wider population – socioeconomic and

---

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 Assessments and Qualification Alliance (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.

demographic characteristics and education factors

	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						
English and maths marks age 11	104.6	103.9	0.745***	101.6	106.4	-4.776***
GCSE points	40.3	40.5	-0.168	41.4	46.5	-5.087***
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.

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	0.360	0.199	0.169***	0.318	0.155	0.159***

but not completion ('drop out')[12]						
Degree classification – 2:1 or above	0.447	0.626	-0.179***	0.445	0.672	-0.225***
University 'status'[13] 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.

Similar checks are undertaken on the balance of the treatment and control groups for additional threshold analysis. These are not included in the paper due to space restraints. The different treatment and control group are relatively well balanced in terms of socioeconomic and demographic characteristics, with slightly larger differences in attainment (English and maths marks age 11 and GCSE points). Thus, they mirror the results Table 1.

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 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. 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).

---

[12] 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.

[13] Using clusters from Boliver (2015)

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

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. As we have dichotomous dependent variables, the Probit model calculates the probability of achieving an education level. Formally this can be expressed as:

(2) $P_i = \Phi (\beta_0 + \beta_1 T_i + \beta_2 X_i)$

Where P_i is the probability of achieving the mediator, i.e., the short- or medium-term education variable and Φ is the cumulative distribution function of a standard normal variable which ensures $0 \leq p_i \leq 1$. The mediators are: achieved a level 2 vocational qualification by age 2, 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 2 vocational by 25, 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. Thus P_i is the probability of achieving each of these

education levels. 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. I also choose to focus on level 2 vocational. It would not be appropriate to look at level 2 academic, as all of the treatment group already have it. However, level 2 vocational qualifications are taken after GCSEs and are thus a subsequent, post compulsory education outcome. Hence, it is appropriate to see the relationship between crossing the threshold and vocational level 2. 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 2 vocational or above by age 25 as my mediator. I then add level 3 or above to the model. Level 3 or above 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).

This approach is initially applied to the five GCSE A* to C threshold before being extended to other A* to C thresholds, i.e. one, two, three, four, six, seven, eight, nine and ten. This is to observe whether the effect of crossing the five GCSE A* to C threshold on labour market outcomes is salient or whether crossing other (A* to C) thresholds has similar results.

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 (adjusted for the threshold crossing) 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 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 5 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 focus on the results that make up just achieving or just missing out. 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. For GCSE points the difference of 1.1 and 1.2 points for men and women equates to just over one grade. Considering individuals took on average nine exams (with a range of between 5 and 17) this one grade difference could be argued to be small. Focusing on the best five results is perhaps a more salient measure as the other D/E-G grades could be considered as noise. For the adjusted KS4 points score for the 'best five' GCSE grades, there is a 0.4 and 0.5 difference between the treatment and control group for men and women respectively. This is around half a grade, suggesting the other 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 overall distribution is plotted 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.

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	9.0	9.0	0.042	9.1	9.1	0.029

GCSE points (adjusted)[14]	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.

I also show the distribution of GCSE points from the best five GCSEs (adjusted for the one grade difference, i.e. the C/D) in Table 4 (below). Here we see the modal number of points is 24 meaning the treatment group results were CCCCC vs CCCCD in the control group. The vast majority of both men and women from both the treatment and control group have fewer than 28 GCSE points (93 and 96 percent of men and 88 and 95 percent of women, respectively). 27 GCSE points is the equivalent of BBBCC for the treatment and BBBCD for the control group. Thus, whilst it is possible to have A*AAC in the treatment and A*AAAC there are very few cases like this.

Table 4: descriptive statistics - attainment (GCSE points) within treatment and control group

GCSE points (adjusted)	Men				Women			
	Number		**Proportion**		**Number**		**Proportion**	
	Treatment	**Control**	**Treatment**	**Control**	**Treatment**	**Control**	**Treatment**	**Control**
24	8,272	10,299	36.8%	50.1%	5,581	7,481	27.4%	39.9%
25	6,357	5,662	28.3%	27.5%	5,602	5,712	27.5%	30.4%
26	4,012	2,790	17.9%	13.6%	4,095	3,079	20.1%	16.4%
27	2,172	1,149	9.7%	5.6%	2,589	1,499	12.7%	8.0%
28	989	415	4.4%	2.0%	1,460	641	7.2%	3.4%
29	393	170	1.7%	0.8%	596	237	2.9%	1.3%
30+	265	89	1.2%	0.4%	416	116	2.0%	0.6%

As seen in table 4, whilst the treatment and control group are 'one grade away' from each other, there is a 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

---

[14] 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.

variable, T as the explanatory variable, and controls for GCSE points, number GCSEs entered, GCSE cohort, exam tier of subject crossing the threshold and subjects entered. The results are presented in Table 5. Table A3 in the appendix shows the descriptive statistics for the treatment and control group based on sample means for completeness and transparency. Table 5[15] shows that, after accounting for GCSE performance, the background characteristics of the two groups are very similar. For the most part there are no statistically significant differences between the treatment and control group for the socioeconomic and demographic factors I can account for. The exceptions FSM eligible for both men and women and IDACI and SEN for women. Thus, conditional on other GCSE results the background characteristics of the treatment and control are very similar. There was a difference in education performance at 11 between those that crossed the threshold and those that did not. There is around a 2.2 mark difference for men and 2.7 women, out of 200, in combined English and maths scores (national standard attainment tests). 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. I cannot rule it out.

Table 5: Comparison of treatment and control group: socioeconomic, demographic and education factors

Variable	Men	Women
Socioeconomic and demographic characteristics		
IDACI	-0.003 (0.002)	-0.003** (0.002)
FSM eligible	-0.006** (0.003)	-0.006** (0.003)
SEN	-0.001 (0.003)	-0.005** (0.003)
Minority ethnic	-0.001 (0.003)	-0.002 (0.004)
First language English	0.000 (0.003)	-0.001 (0.003)
Region at GCSEs: London	-0.005 (0.003)	-0.002 (0.004)
Region at GCSEs: North East	0.004** (0.002)	0.003 (0.002)
Region at GCSEs: North West	0.005 (0.004)	0.000 (0.004)
Region at GCSEs: Yorkshire and the	0.004	0.004

---

[15] I carry out similar validity checks for other C/D thresholds includes in the analysis, i.e. tables 3 and 5 for thresholds one, two, three, four, six, seven, eight, nine and ten. They are not included in the paper as it would take up too much space. However, the results are similar to the five GCSE A* to C threshold. Conditional on other GCSE results there only very minor differences in socioeconomic and demographic characteristics, however there remain small but significant differences in GCSE points and (English and maths) test marks at 11 years old.

Humber	(0.003)	(0.003)
Region at GCSEs: East Midlands	-0.003	-0.006**
	(0.003)	(0.003)
Region at GCSEs: West Midlands	-0.003	0.003
	(0.003)	(0.003)
Region at GCSEs: South East	0.001	-0.003
	(0.004)	(0.004)
Region at GCSEs: South West	-0.003	0.004
	(0.003)	(0.003)
Region at GCSEs: East of England	-0.001	-0.007**
	(0.003)	(0.003)
Education factors		
National test age 11 total marks (English and maths)	2.229***	2.688***
	(0.170)	(0.189)
National test age 11 English marks	0.726***	0.939***
	(0.097)	(0.119)
National test age 11 maths marks	1.502***	1.748***
	(0.143)	(0.153)
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, findings from Machin et al (2020) give confidence this is unlikely to be a major risk for my study. They 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. They also found around one percent of the overall population was upgraded (the average

probability of requesting a remark was 10 percent and 12 percent of these cases were upgraded).

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. Subsequently, I extend the analysis to other thresholds (one, two, three, four, six, seven, eight, nine and ten full GCSEs A* to C). Results for earnings are presented in this section, with other results shown in the appendix. Whether there are any heterogenous effects to crossing the threshold, by background characteristics, are also explored. Finally, I present robustness checks.

5.1 Labour market outcomes

Table 6 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 -0.008 implies that those in the treatment group are likely to be claiming out of work benefits for 0.8 percentage points less of the year than the control group. This implies that just achieving five GCSEs A* to C leads almost three days (2.92) fewer claiming out of work benefit on average. Crossing the threshold is also associated with a statistically significant increase in the log of daily earnings for women, around 3.4 percent (i.e., 1.034, derived from the exponential of 0.033). On average, this equates to around £1.82 per day or around £666 gross annual earnings (i.e., before tax).

17

Table 6 – Effect of achieving five GCSEs A* to C on early career labour market outcomes

	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	-0.001 (0.004)	0.000 (0.002)	0.003 (0.006)
Sample size	43,034		23,798
Women	0.003 (0.004)	-0.008*** (0.003)	0.033*** (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

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 7 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, yet a lower likelihood of doing a lower-level vocational qualification. Models 2, 3 and 5 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). Whilst model 1 shows that by 20 those in the control group are more likely to have done a level 2 vocational qualification, though crossing the threshold does not impact (negatively) on achieving level 3 vocational (model 4). The results from model 3 and 4 imply that crossing the threshold 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 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 (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, suggest 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 in academic education (i.e. they do not believe they have the competence as they struggled to achieve the prerequisite grades).

Table 7: Effect of achieving five GCSEs A* to C on short term education outcomes

	Model 1) Achieved level 2 vocational by 20 – Probit	Model 2) Achieved level 3 by 20 – Probit	Model 3) Achieved level 3 academic by 20 – Probit	Model 4) Achieved level 3 vocational by 20 – Probit	Model 5) Started level 4 by 20 – Probit
Men	-0.055*** (0.004)	0.060*** (0.005)	0.060*** (0.004)	0.001 (0.004)	0.033*** (0.004)
Sample size	43,034				
Women	-0.074*** (0.004)	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

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 8 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 8 provides some suggestions as to why this might be the case. Model 1 shows that a level 2 vocational qualification is associated with higher earnings. Model 3 shows that achieving a level 3 or above has a positive impact on earnings, whilst within that model 4 suggests that returns are similar for level 3 academic and vocational, and notably lower than level 2 vocational. Model 5 shows that, for men, higher education (i.e., level 4 or above) does has positive and significant returns to earnings on average, but lower than other (lower) education levels. This model simply includes a dummy variable for if someone completed a level 4 or above qualification. Model 6 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 has a limited impact upon labour market outcomes (at age 27 to 29), 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.

Table 8 suggests there may be a 'push' and 'pull' effect of crossing the threshold on earnings, which is associated with switching from level 2 vocational to level 3 academic. Those crossing the threshold are more likely to achieve an academic level 3 and a level 4 or above, which are both associated with higher earnings (push effect). Yet, crossing the threshold leads to a lower chance of achieving a vocational level 2, which has very strong earnings (pull effect). 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.

20

Notably model 6 shows that returns to level 3 and level 4 or above become higher than level 2 vocational when adding controls. These controls include the sector of the vocational qualification, which is likely to be driving men's level 2 vocational returns.

In terms of higher education, the results may seem surprising initially. However, in model 6, when I control for degree classification and institution quality, the results are in line with expectations. The difference between models 5 and 6 could be interpreted that achieving a degree does not have a large effect on earnings for men on average (at age 27 to 29) 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 in Table 8 shows the results from step 1 (section 5.1) that crossing the threshold has a positive impact on earnings. Model 2 shows that women have strong, negative returns to vocational level 2 qualifications, in stark contrast to men. Model 3 shows that achieving a level 3 or above is associated with higher earnings and crucially (in model 4) 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 (models 5 and 6), 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. Equally, crossing the thresholds leads to lower chances of achieving a level 2 vocational qualification which is associated with lower earnings. Think of the 'push' and 'pull' effect for men, everything is pushing in the same (positive direction for women). It should be noted that in model 6 the negative returns to vocational level 2 for women decrease significantly and thus the sector of vocational qualifications may explain some of the results.

In models 2 to 5 the (crossing the threshold) coefficient decreases by a small amount, as post 16 education mediators are added, but remains statistically significant. This suggests that for women 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. In model 6 the coefficient decreases further and becomes statistically insignificant, suggesting education institution status and subject/sector

21

choices are important. Reasons for why there may be such differences between men and women are explored in the discussion (section 6).

Table 8: Effect on earnings age 27 to 29 of crossing the threshold and post 16 education (as a mediator)

	Model 1 (Original model)	Model 2 (Level 2 vocational achievement as mediator)	Model 3 (Level 3 achievement as mediator)	Model 4 (Level 3 academic and vocational achievement as mediators)	Model 5 (Level 4 or above achievement as mediator)	Model 6 (Level 4 or above achievement as mediator – with controls)
Men						
Five GCSEs A* to C (explanatory variable)	0.003 (0.006)	0.007 (0.006)	0.004 (0.006)	0.004 (0.006)	0.003 (0.006)	0.000 (0.006)
Level 2 or above vocational achievement		0.070*** (0.007)	0.084*** (0.007)	0.082*** (0.007)	0.087*** (0.007)	0.045*** (0.009)
Level 3 or above achievement			0.069*** (0.007)		0.060*** (0.007)	0.058*** (0.011)
Level 3 or above academic achievement				0.055** (0.009)		
Level 3 or above vocational (only) achievement				0.056*** (0.007)		
Level 4 or above					0.029*** (0.009)	0.117*** (0.014)
Sample size	23,798					
Women						
Five GCSEs A* to C (explanatory variable)	0.033*** (0.008)	0.027*** (0.008)	0.023*** (0.008)	0.020** (0.008)	0.020** (0.008)	0.015 (0.008)
Level 2 or above vocational		-0.075*** (0.008)	-0.051*** (0.008)	-0.036*** (0.009)	-0.024*** (0.008)	-0.004** (0.010)

achievement						
Level 3 or above achievement			0.116*** (0.009)		0.043*** (0.009)	0.058*** (0.013)
Level 3 or above academic achievement				0.143*** (0.010)		
Level 3 or above vocational (only) achievement				0.033** (0.008)		
Level 4 or above					0.214*** (0.010)	0.241*** (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 2 vocational or above as a mediator. Model 3 adds achieved a dummy variable for if someone achieved level 3 or above as a mediator (to model 2). Model 4 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 5 adds a dummy variable for achieved level 4 or above to model 2. All models include those employed for 365 of tax year. Model 6 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).

5.4 Different thresholds

I extended the analysis to focus on a range of thresholds to test to see if the five GCSE A* to C threshold is exclusively salient, in terms of labour market outcomes (for women). Whilst the importance of the five GCSE A* to C threshold in the education system is clear (as outlined in background section) I wanted to see whether it had a salient effect on labour market outcomes. Does the five GCSE A* to C threshold really matter or is it just crossing any C/D threshold that is important in terms of labour market outcomes? A table of full regression results is shown in Table A5 in the appendix. In this section I focus on the link between crossing different

24

thresholds and earnings. Figure 1 shows the effect of crossing different thresholds on earnings age 27 to 29 for men and women. We see that crossing thresholds at the lower end of the distribution, i.e. fewer than five GCSEs has little effect on earnings, with the notable exception of crossing the one A* to C threshold. However, at the upper end of the attainment distribution (i.e. thresholds six to ten) crossing thresholds has a strong and positive effect on earnings.

Figure 1: The effect of crossing different A* to C thresholds on earnings age 27 to 29
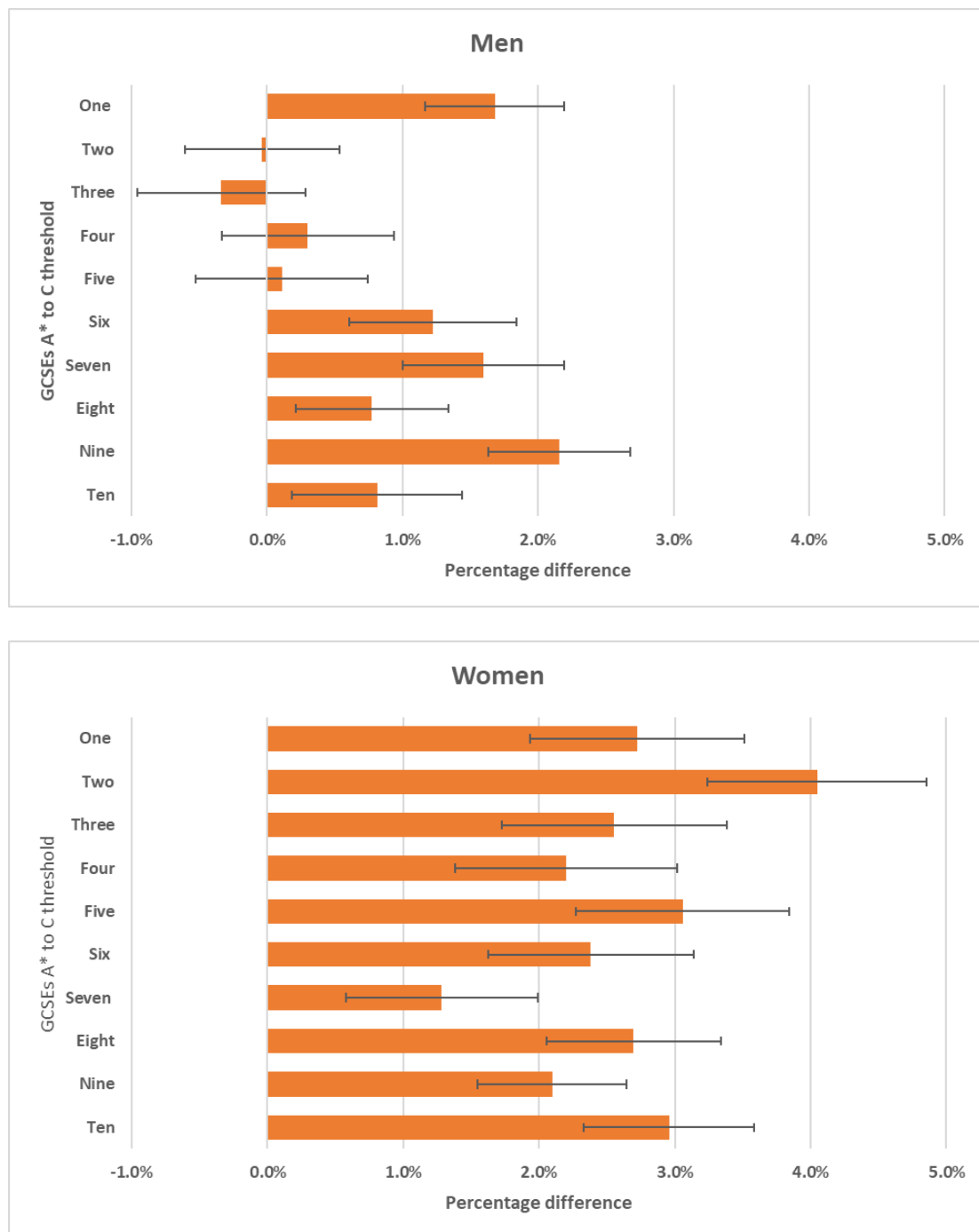
The difference in effect between the lower and upper end of the attainment spectrum (for men) can be explained by differences in post compulsory education pathways and their subsequent returns. Those not crossing the thresholds (with the exception of one A* to C) are more likely to undertake vocational level 2 qualifications which have strong, positive returns to earnings age 27 to 29. Whist those crossing the threshold are more likely to undertake level 3 academic and level 4 or above. Those at the lower end of the attainment spectrum seem benefits from this 'vocational effect', but less so from the uptake in academic education. Whilst at the upper end of

the attainment spectrum, those crossing the threshold have higher returns to level 4 or above qualifications. The effect of crossing the threshold on education outcomes age 20 can be seen in Figure A3 in the appendix and returns to these education levels can be seen in Figure A4 in the appendix.

The effect of crossing different thresholds on earnings age 27 to 29 for women is shown in Figure 1. It can be observed that crossing each threshold has a strong and significant effect on early career earnings. Thus, the salience of crossing the five GCSE A* to C threshold on women's earnings is not unique.

As for men, the effect of crossing different thresholds can be explained by differences in post compulsory education outcomes and their associated returns. The narrative for women is more simple than men. As seen in Figure A3 in the appendix, women who cross the (five GCSE A* to C) threshold are more likely to undertake level 3 academic and level 4 or above qualifications, which are linked to higher earnings age 27 to 29. There is no 'vocational effect' as we see for men, as women benefits more financially from academic and higher education than vocational education (as shown in Figure A4). With men see a tension in that not crossing thresholds leads to vocational level 2 education, which pays well, whilst crossing the threshold leads to academic and higher education which also has strong returns. However, for women we do not see this tension, i.e. everything is pulling in the same direction (crossing the thresholds leads to education which is linked to stronger early career earnings).

5.5 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.

I considered heterogeneity analysis focusing on the subject crossing the threshold, however this is not possible. As seen in Table 4 and discussed in the narrative most of the treatment group have more than one C grade (the modal results are CCCCC and the mean and median results are BCCCC) which makes isolating the subject very tricky. Thus, I am assuming it could be one (or more) of the (C) subjects the threshold was crossed in.

5.6 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 9), 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 9 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. Thirdly, and finally, I add a robustness check for different attainment levels to show this is not affecting results. Table 4 showed the distribution of GCSE points for the best five GCSE grades, and we saw the majority had below 28 points. Thus, I remove those with 28 points or above (essentially as a type of outlier) and run the model on those with 27 points or fewer. Thus, the model is comparing only those with CCCCC vs CCCCD, BCCCC vs BCCCD, BBCCC vs BBCCD and BBCCD (as an exemplar) as the model controls for GCSE points. The results in Table 9 show that the main findings hold for this robustness check, thus it is not some high attainers driving the results.

Table 9: Robustness checks

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
Main findings							
Age 27 to 29	-0.001 (0.004)	0.000 (0.002)	0.003 (0.006)		0.003 (0.004)	-0.008*** (0.003)	0.033*** (0.008)
Sample size	43,034		23,798		39,104		22,367
Different years							
Age 27	0.000 (0.004)	0.000 (0.002)	0.007 (0.006)		0.001 (0.004)	-0.005 (0.003)	0.024*** (0.008)
Sample size	43,034		24,245		39,104		22,878
Age 28	-0.004 (0.005)	-0.001 (0.002)	-0.002 (0.010)		0.005 (0.005)	-0.006 (0.003)	0.025** (0.008)
Sample size	29,773		16,842		25,185		15,871
Age 29	-0.014** (0.007)	-0.001 (0.003)	0.009 (0.010)		0.011 (0.007)	-0.003 (0.004)	0.043*** (0.013)
Sample size	15,780		9,052		14,686		8,583
Different definitions							
Daily earnings from employment (at least one day)	0.001 (0.002)		-0.005 (0.001)		0.006 (0.004)		0.026** (0.008)

Sample size	43,043		37,742	39,104		30,278
Different levels of attainment: GCSE points from best five GCSEs adjusted						
Less than 28 GCSE points from best five GCSEs	-0.004 (0.004)	0.000 (0.002)	0.002 (0.007)	0.007 (0.004)	-0.007*** (0.003)	0.033*** (0.008)
Sample size	40,713		22,801	35,638		20,667

** significant at 5% level, *** 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 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 five GCSE A* to C 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 women who just cross the threshold benefit from taking advanced and higher-level academic qualifications. However, for men there is a tension associated with crossing the threshold which leads to switching from vocational level 2 to academic level education (and subsequently higher education). Both have strong positive returns and thus cancel each other out and lead to no overall effect. Extending the analysis, I show that these crossing other A* to C thresholds (one, two, three, four, six, seven, eight, nine and ten) have a similar effect on labour market outcomes.

This section focuses on policy implications introduced in the introduction in slightly more depth.

I start with a limitation of my current study, 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). However, these studies do not split lower-level qualifications into academic qualifications, vocational qualifications or other qualifications. And as I am not aware of any studies focusing on lifetime earnings for vocational qualifications, it is difficult to know whether the trends would hold. 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). Equally, another question may be whether the effect seen on women in my sample is an effect at the time (i.e., some form of child penalty or down to part time working) or would it last over a working lifetime. I considered estimating lifetime income in the way other studies (mentioned above) have done but this requires using the labour force survey (LFS), which has a much smaller sample size and does not have detailed information on education level and type. Considering the key findings on differences in education level and type (i.e. academic and vocational) between treatment and control, I do not believe projecting lifetime income using the LFS would produce robust and reliable results This is explored below.

There are several points that I think merit discussion from a policy point of view.

The first policy implication is that crossing high stakes exams thresholds has an effect on women's early career earnings, yet not men's. Is this a genuine effect or driven by difference in working patterns? Women with lower levels of education work fewer hours than those with higher levels of education (Trostel and Walker 2020). Many women reduce their hours after having their first child and whilst this holds across education levels, it is particularly pronounced for women with lower levels of education (Buchanan et al 2023). But women with higher levels of education tend to have children later[16]. I cannot disentangle this from the data as I do not have hours worked and though I use daily earnings, this is derived from the days in employment during the tax year (e.g. 365 days in the year) and not the hours worked (e.g. 37 hours per week). As I am focusing on labour market outcomes at age 27 to 29, the observed effect for women could be down to them working fewer hours, as they have had children earlier than their better education peers. Nevertheless, there is an observed effect whether this is a child penalty, or a permanent effect.

The difference in results for men and women can be explained by the role of academic and vocational training. The patterns are clear for women, that they benefit from greater returns to academic and higher education. Whilst for men this effect is not as strong and they benefit from vocational education, in a way that women do not. This is supported by the literature (Dearden et al 2004, Belfield et al 2018b, Walker and Zhu 2003, Belfield et al 2018, Buscha and Urwin 2013, Greenwood et al 2007, McIntosh and Morris 2021). 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). Notably, in model 6 (of Table 8) we see that the positive and negative returns to vocational level 2 (for men and women respectively) decrease dramatically when sector (of vocational training) controls are added.

16

https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/conceptionandfertilityrates/adhocs/008981meanageofmotheratbirthoffirstchildbyhighestachievededucationalqualification1996to2016englandandwales

Secondly, a key policy contribution is around high stakes exams and allocative efficiency. Many countries reconsidered high stakes exams as a form of assessment in the aftermath of the Covid-19 pandemic. This debate has not completely gone away and is likely to resurface in the future. 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. For example, a number of studies have shown that external factors can impact upon exam performance (Park 2020, Ebenstein et al 2016, Poutvaara and Ropponen 2018, Landaud et al (2022), Rhead et al 2016, Rhead et al 2018). 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.

My findings may initially seem like they support the case for abolition of high stakes exams as a form of assessment, in that they negatively impact on allocative efficiency. This is primarily driven by the results for women. However, are a few important factors to note. One is the first policy implication around whether the results show a permanent or temporary (e.g. child penalty) affect. It should be noted that I only focus on one aspect of the high-stakes exams debate[17]. Also, the analysis shows that almost all of the differences in labour market outcomes can be explained by differences in education trajectories after crossing the threshold. There is a no, or a very limited direct effect, of crossing the threshold on labour market outcomes. Differences are driven by those who cross the threshold being able to access higher levels of education and higher quality education. (Though not included in the paper for lack of space) This is also the case for other thresholds explored. Thus, policy interventions aimed at addressing this could be (potentially) more efficient, in negating risks associated with allocative inefficiency, than scrapping high stakes exams.

Thirdly, 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. It is clear that these individuals, particularly women, need support or interventions targeted at them and policymakers and leaders in the education sector, may want to reflect on this. Having more flexibility in admissions policies, having access to timely and easily accessible GCSE exams resits and support (careers advice, teacher and leader support, parental guidance etc) could address this and ultimately decrease allocative inefficiency.

Finally, I return to the importance of the five GCSE A* to C threshold. Achieving this was seen like passing high school. It is the equivalent of a full level 2 qualification and needed to continue to the next education level (for example A-levels). The threshold also had a legacy because it was used in accountability procedures. However, despite this, crossing this threshold does not appear to be more salient in

---

[17] 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)

influencing labour market outcomes than crossing other thresholds. Achieving an additional 'good pass' has a similar effect on (education and) labour market outcomes. Thus, the 'all important' five GCSE A* to C threshold is nothing special after all (at least in terms of employment prospects).

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.

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.

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.

Buchanan, I., Pratt, A., Francis-Devine, B. (2023), Women and the UK economy, House of Commons Research briefing.

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.

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.

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.

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.

Hannu Karhunen & Artturi Björk, 2019. "The long shadow of high stakes exams: Evidence from discontinuities," Working Papers 329, TyÃ¶n ja talouden tutkimus LABORE, The Labour Institute for Economic Research LABORE.

Landaud, Fanny and Maurin, Eric and Willage, Barton and Willén, Alexander, Getting Lucky: The Long-Term Consequences of Exam Luck (2022). CESifo Working Paper No. 9570, Available at SSRN: https://ssrn.com/abstract=4036986 or http://dx.doi.org/10.2139/ssrn.4036986

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.

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.

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*.

Poutvaara, P. & Ropponen, O. (2018). "Shocking news and cognitive performance," European Journal of Political Economy, Elsevier, vol. 51(C), pages 93-106.
Rees, A. (1965). Human capital: A Theoretical and empirical analysis with special reference to education.

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.

Trostel, P. & Walker I (2006). "Education and Work." Education Economics 14 (2006): 377 - 399.

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

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

% of year in employment (year 11 to 13) men

% of year in employment (years 11 to 13) women
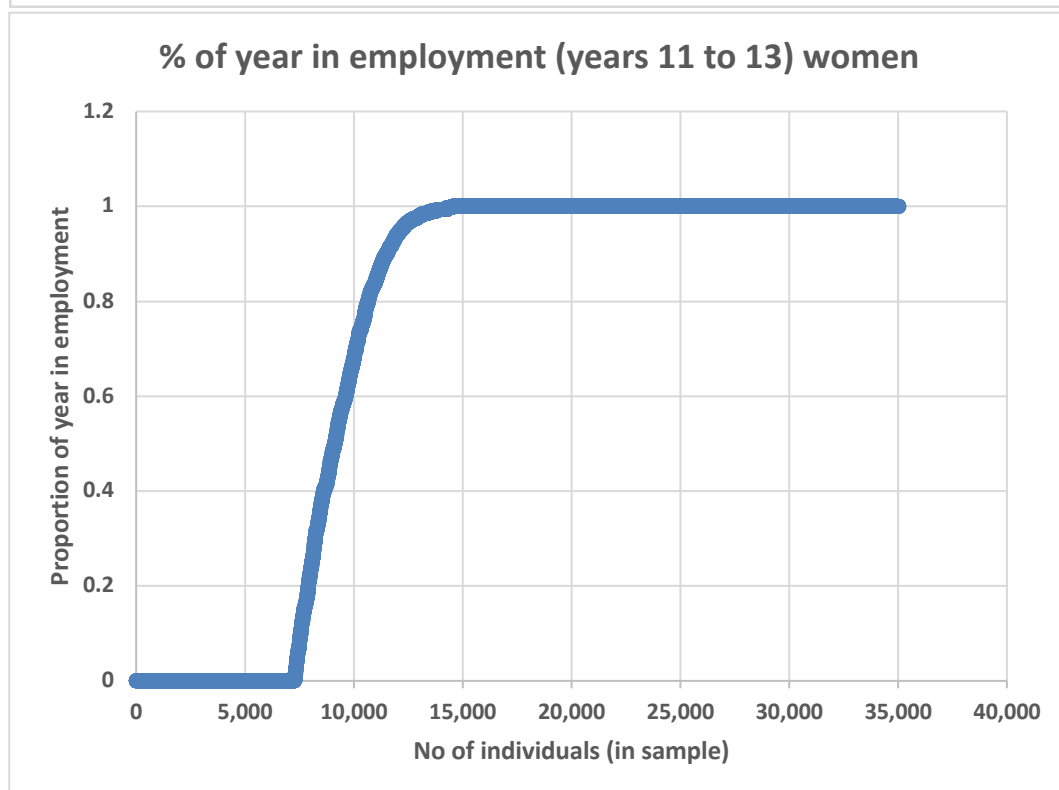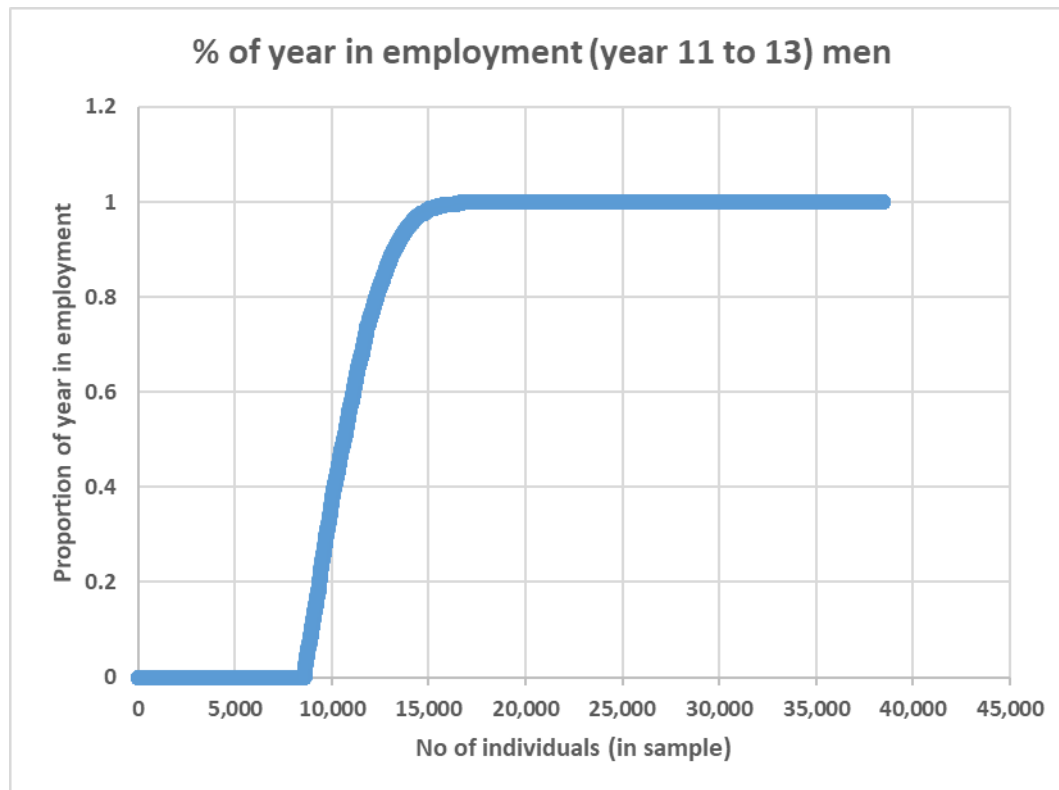
Note – includes all those that achieved five full GCSEs A* to C, including at least one C and those who just missed out, i.e., achieving 4 full GCSEs A* to C and at least one D.

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

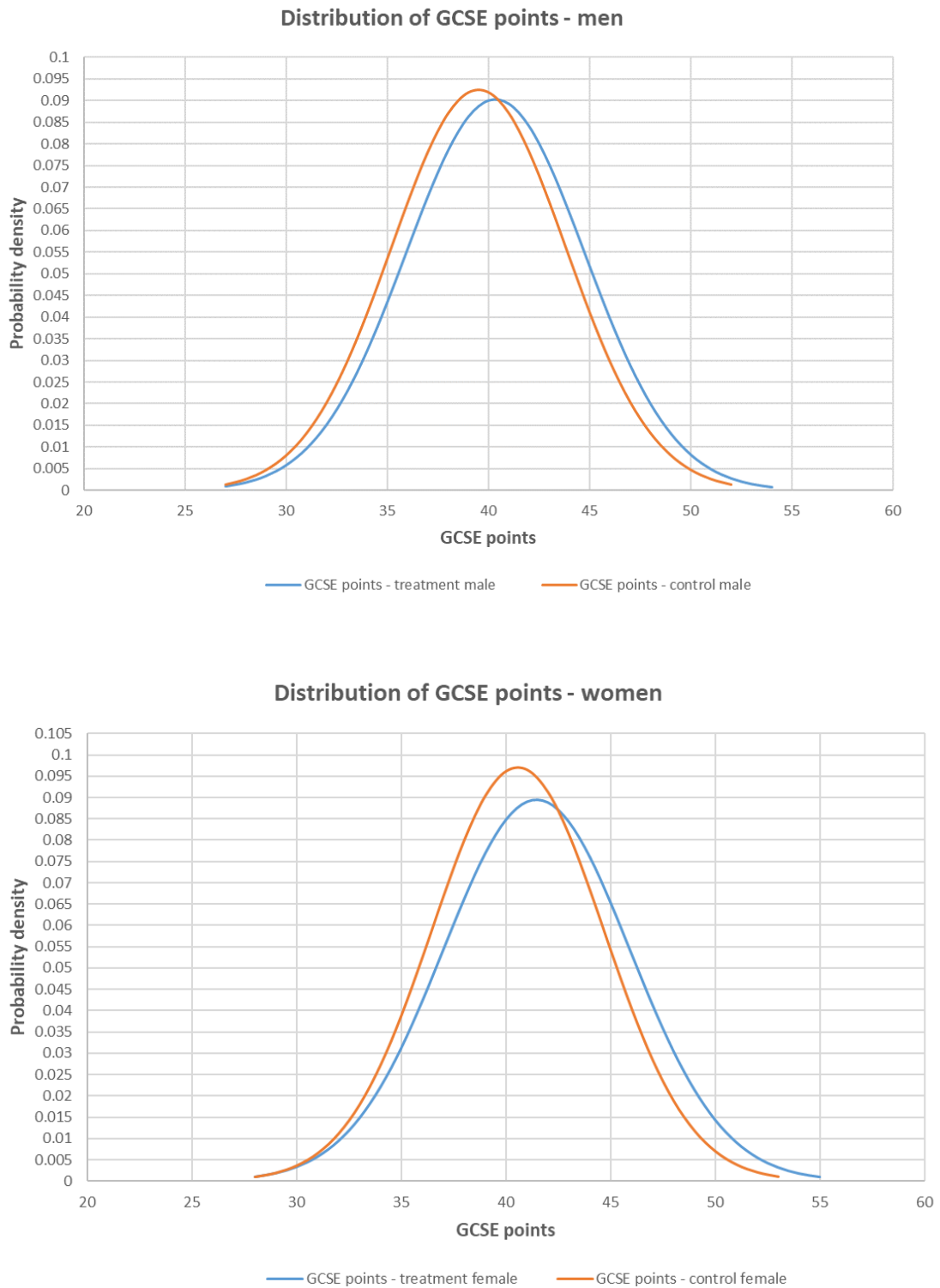| Subject	Men			women		
	Treatment group	Control group	T-C	treatment group	control group	T-C
English	1.000	1.000	0.000	1.000	1.000	0.000
Maths	1.000	1.000	0.000	1.000	1.000	0.000
English literature	0.902	0.898	0.005	0.945	0.940	0.005
Double science	0.900	0.894	0.007**	0.880	0.875	0.007**
Single science	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

Distribution of GCSE points - men

Legend: GCSE points - treatment male; GCSE points - control male

Distribution of GCSE points - women

Legend: GCSE points - treatment female; GCSE points - control female

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 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	-0.059*** (0.005)	0.058*** (0.005)	0.062*** (0.004)	-0.004 (0.005)	0.036*** (0.004)
Sample size	43,034				
Women	0.074*** (0.005)	0.046*** (0.005)	0.069*** (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.					

Table A5: Crossing different thresholds and labour market outcomes

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 thresholds							
One GCSE A* to C	-0.009*** (0.003)	-0.007*** (0.002)	0.017*** (0.005)		0.010** (0.004)	-0.017*** (0.003)	0.027*** (0.008)
Sample size	87,896		42,506		60,863		27,690
Two GCSEs A* to C	-0.003 (0.004)	0.004** (0.002)	0.000 (0.006)		0.010** (0.004)	-0.014*** (0.003)	0.040*** (0.008)
Sample size	43,692		30,496		35,459		25,014
Three GCSEs A* to C	0.008** (0.004)	0.002 (0.002)	-0.003 (0.006)		0.009** (0.004)	-0.005 (0.003)	0.026*** (0.008)
Sample size	38,197		20,358		34,109		18,589
Four GCSEs A* to C	0.005 (0.004)	0.002 (0.002)	0.010 (0.006)		0.004 (0.002)	-0.005** (0.002)	0.022*** (0.008)
Sample size	33,124		23,412		30,537		21,807
Five GCSEs A* to C (main model)	-0.001 (0.004)	0.000 (0.002)	0.003 (0.006)		0.003 (0.004)	-0.008*** (0.003)	0.033*** (0.008)
Sample size	43,034	23,798		39,104	22,367		30,278
Six GCSEs A* to C	-0.004 (0.004)	0.002 (0.002)	0.010 (0.006)		0.011*** (0.004)	-0.004 (0.002)	0.024*** (0.008)
Sample size	45,413		25,540		40,406		23,754
Seven GCSEs A* to C	0.006 (0.004)	-0.001 (0.001)	0.016*** (0.006)		0.010** (0.004)	-0.012*** (0.002)	0.013 (0.007)
Sample size	50,436		29,521		43,636		26,834
Eight GCSEs A* to C	0.007** (0.004)	-0.004*** (0.001)	0.008 (0.006)		0.004 (0.004)	-0.001 (0.002)	0.027*** (0.006)
Sample size	60,836		36,444		53,943		33,761
Nine GCSEs A*	0.008***	-0.002	0.022***		0.017***	-0.007***	0.021***

to C	(0.003)	(0.001)	(0.005)		(0.003)	(0.001)	(0.006)
Sample size	76,130		46,626		80,037		51,827
Ten GCSEs A* to C	0.010*** (0.004)	-0.002** (0.001)	0.008 (0.006)		0.012*** (0.004)	-0.002 (0.001)	0.030*** (0.006)
Sample size	54,184		33,838		61,813		40,764

** significant at 5% level, *** 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 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.

Figure A3: Effect of crossing different thresholds on different short term education outcomes (age 20)
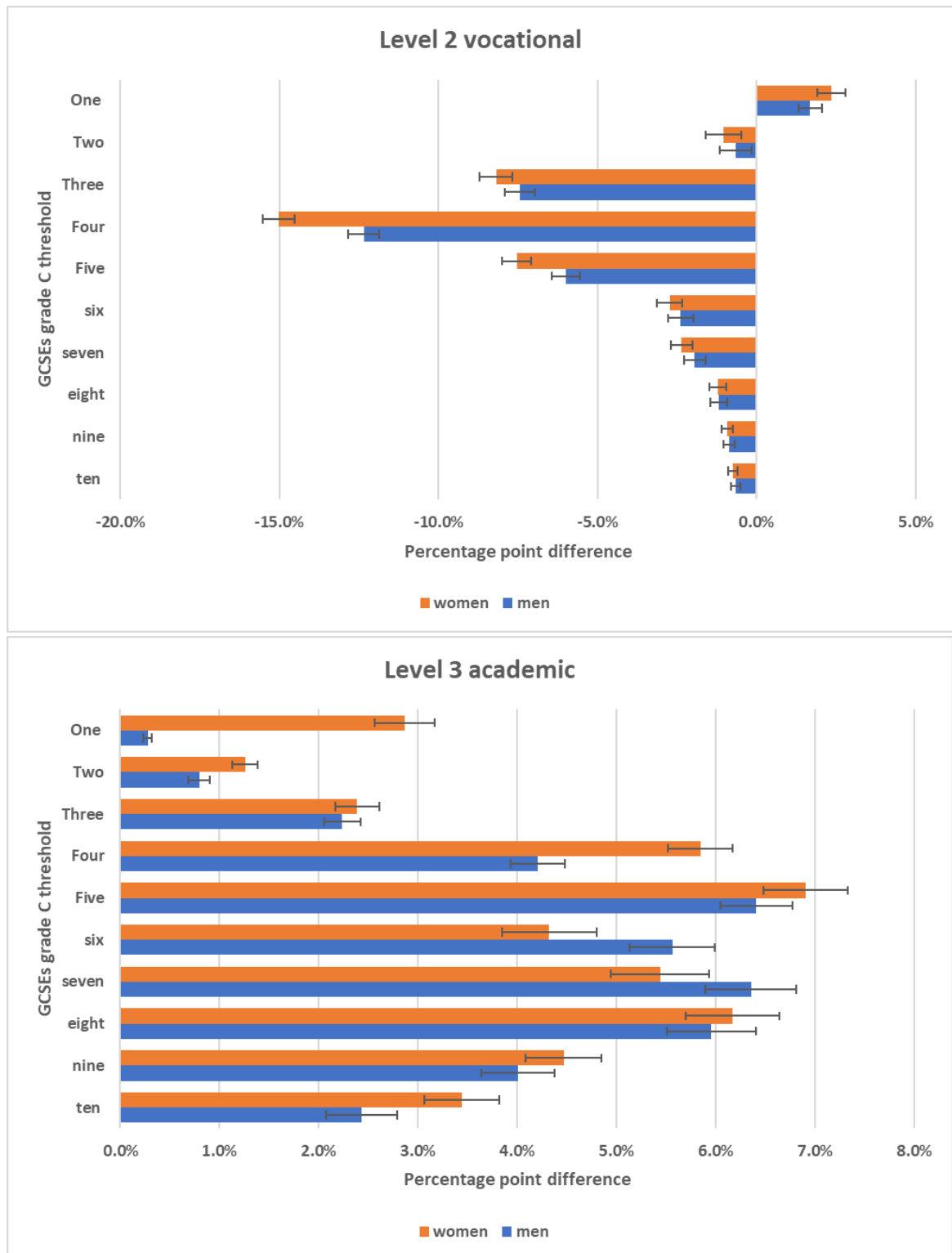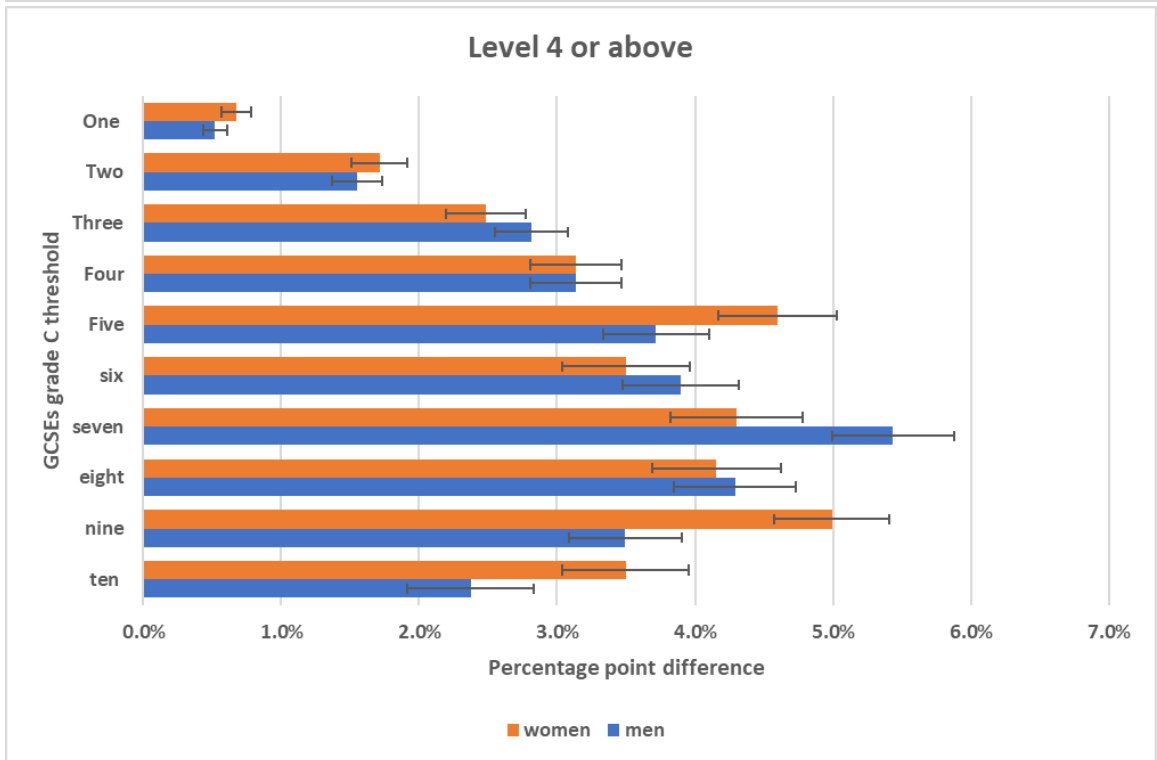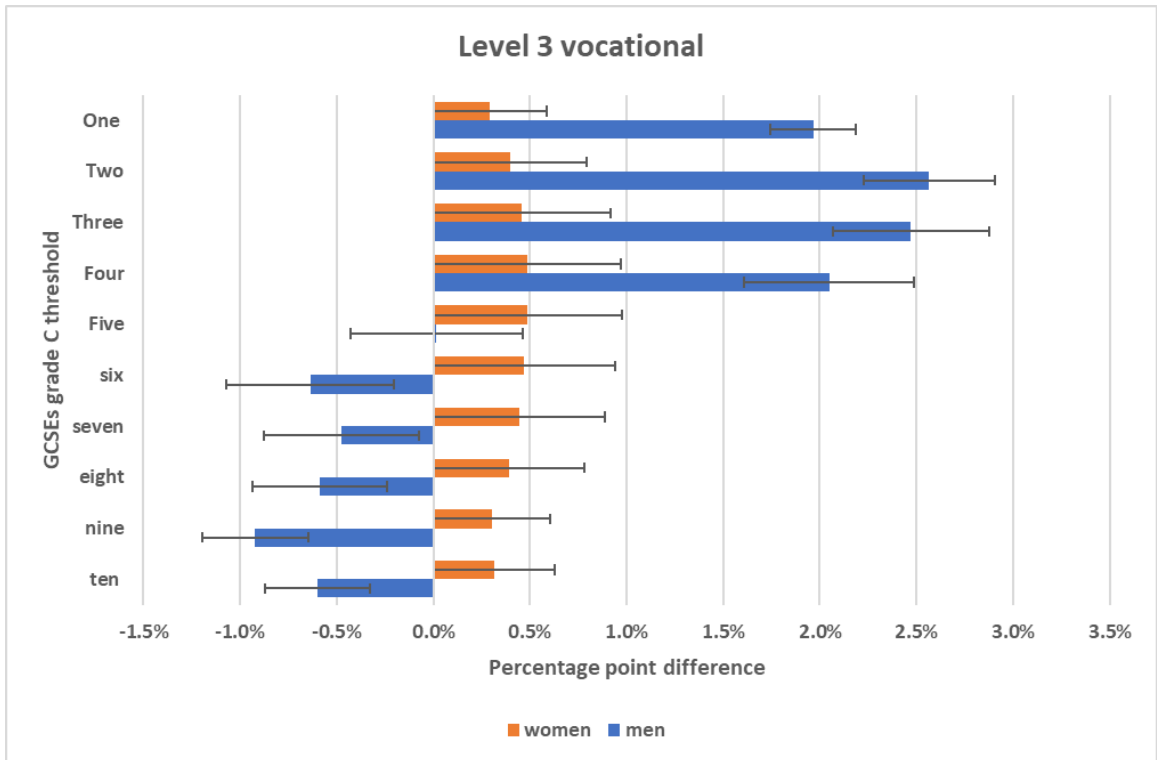
Level 3 vocational

Level 4 or above

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

46

Figure A4: Returns to (log) earnings age 27 to 29 for different education outcomes at different GCSE thresholds
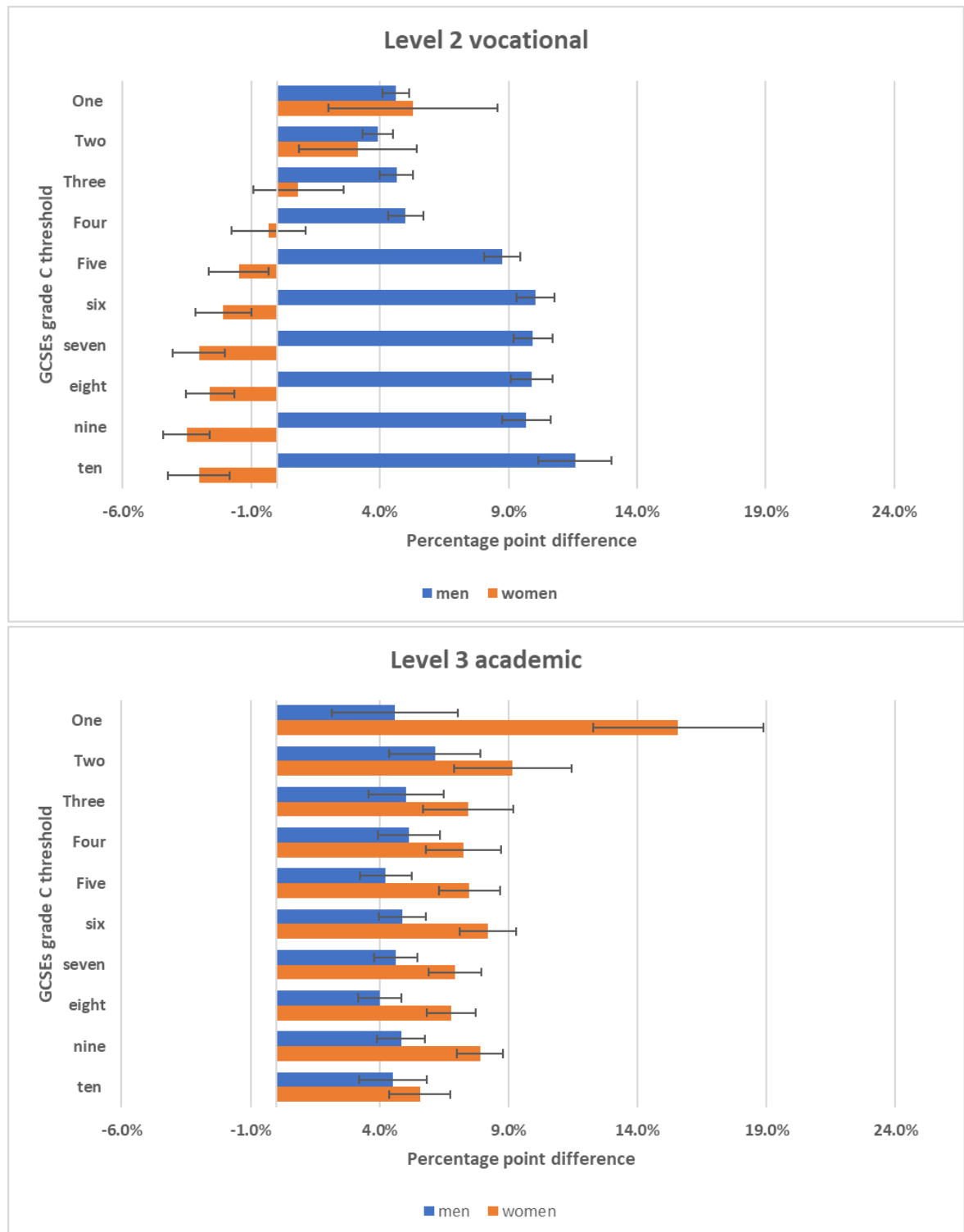
Level 3 vocational

GCSEs grade C threshold (y-axis): One, Two, Three, Four, Five, six, seven, eight, nine, ten

x-axis: Percentage point difference (-6.0%, -1.0%, 4.0%, 9.0%, 14.0%, 19.0%, 24.0%)

Legend: men, women

Level 4 or above

GCSEs grade C threshold (y-axis): One, Two, Three, Four, Five, six, seven, eight, nine, ten

x-axis: Percentage point difference (-6.0%, -1.0%, 4.0%, 9.0%, 14.0%, 19.0%, 24.0%)
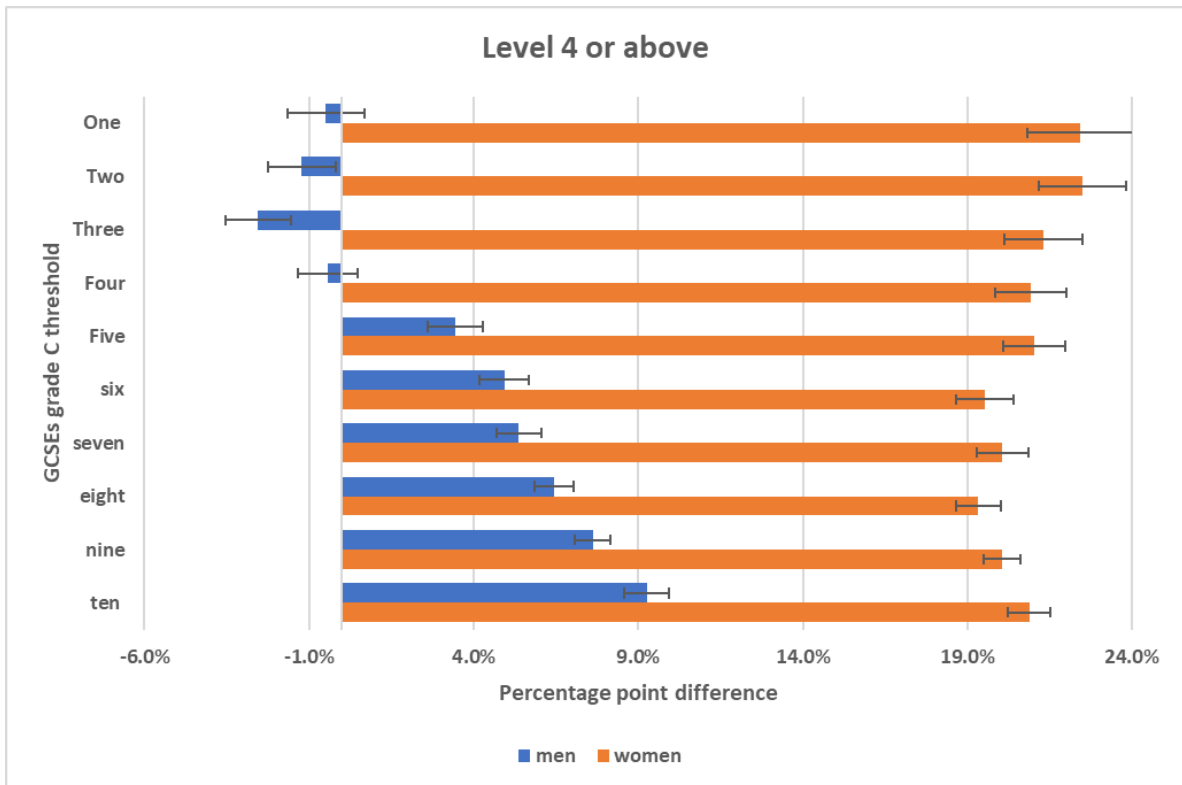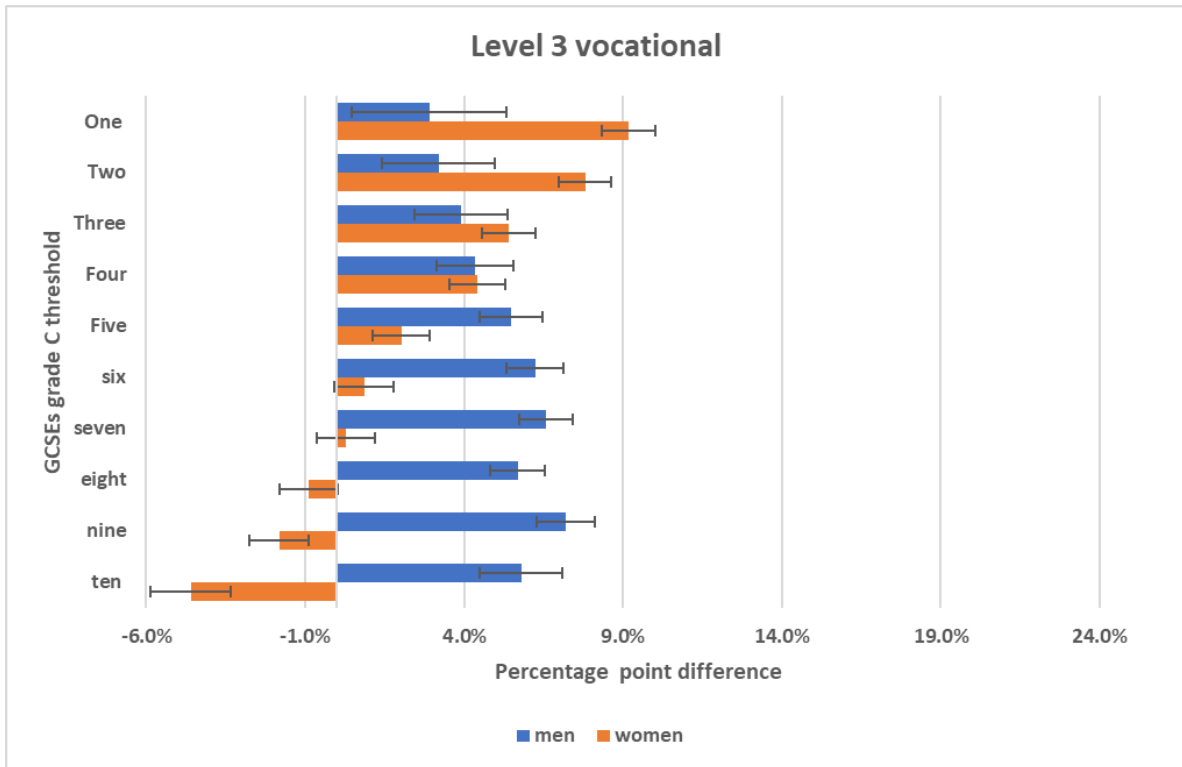
Legend: men, women

Note – includes those that achieved X full GCSEs A* to C, including at least one C and those who just missed out, i.e., achieving Y full GCSEs A* to C and at least one D.
Models are from mediation analysis model 6 equivalent (i.e. with full controls).

Table A6 – Heterogeneity regression results (labour market outcomes): 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