



THE BENEFITS OF MEETING KEY GRADE THRESHOLDS IN HIGH-STAKES EXAMINATIONS. NEW EVIDENCE FROM ENGLAND

John Jerrim

To cite this article: John Jerrim (2022): THE BENEFITS OF MEETING KEY GRADE THRESHOLDS IN HIGH-STAKES EXAMINATIONS. NEW EVIDENCE FROM ENGLAND, British Journal of Educational Studies, DOI: [10.1080/00071005.2022.2033692](https://doi.org/10.1080/00071005.2022.2033692)

To link to this article: <https://doi.org/10.1080/00071005.2022.2033692>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



[View supplementary material](#)



Published online: 10 Mar 2022.



[Submit your article to this journal](#)



Article views: 148




[View related articles](#)



[View Crossmark data](#)



THE BENEFITS OF MEETING KEY GRADE THRESHOLDS IN HIGH-STAKES EXAMINATIONS. NEW EVIDENCE FROM ENGLAND

by JOHN JERRIM , UCL Social Research Institute, London, UK

ABSTRACT: In England, failing to achieve a ‘good pass’ (C/4 grade) in key subjects is thought to have serious negative implications. Yet evidence on this issue remains relatively sparse. This paper therefore presents new evidence on the link between meeting a key threshold on high-stakes examination and a wide array of future outcomes. Using Next Steps survey data collected from around 4,000 young people in England, we explore the short-to-medium term benefits of achieving a ‘good pass’ (grade C/4) in English Language, double science and mathematics. Results from our regression analyses point towards a sizable association with future educational attainment; those who achieve a good pass in GCSE mathematics are around 5 percentage points more likely to hold a university degree by age 26 than observationally similar individuals who fail to meet this threshold. No link is found with future wellbeing and mental health, while results for labour market outcomes are somewhat mixed. The findings potentially motivate the need for GCSEs to move away from awarding a set of discrete grades and towards a continuous measurement scale. Alternatively, if discrete grades are to be retained, computer adaptive testing should be introduced for GCSEs to increase measurement precision around high-stakes grade boundaries.

Keywords: high-stakes testing, examinations, mental health, wellbeing

1. INTRODUCTION

High-stakes examinations are now a prominent part of education systems across the globe (Suto and Oates, 2021). Performance on such examinations opens doors for young people, providing a gateway to post-secondary education, university, the labour market and beyond. Such certification of teenagers’ skills, documenting what they have learned throughout their time at school, has important benefits. This includes providing information to potential employers about young people’s academic strengths and acting as a clear target for pupils to work towards (Saminsky, 2011). Yet the negative consequences of high-stakes testing are also well documented, including the stress they cause young people and the negative impact this may have upon their mental health (Cho and Chan, 2020; UK Education Select Committee, 2017).

England – the empirical setting for this paper – is an example of a country where performance on high-stakes examinations carries particular importance.

At age 16 teenagers sit a series of General Certificate of Secondary Education (GCSE) examinations that help to determine educational progression and access to labour market opportunities (Machin, McNally and Ruiz-Valenzuela 2020). Much has been written about the pros and cons of these examinations, with recent calls for them to be scrapped (Carr, 2021; Coughlan, 2019). Yet GCSEs have – and continue to be – a central pillar of the education landscape in England, being widely used to certify young people’s academic competencies and for the purposes of school accountability (Prior *et al.*, 2021).

Given the importance of high-stakes examinations to individuals – both in England and internationally – it is surprising that relatively little is known about the benefits or ‘value’ of achieving a particular high-stakes grade, and how this varies across subjects. For instance, does it really matter if someone achieves a C grade in GCSE mathematics rather than a D grade? How does this compare to achieving a C versus a D grade in another subject, such as English or science? And are such benefits only short-lived, or do they continue latter into young people’s lives? Having high-quality information on these issues has wide-ranging implications for both policy and practise. For instance, if obtaining a certain grade in a specific subject carries particular importance for young people’s futures, then examination boards should devote additional resources into ensuring that these grades are correct. Similarly, it is crucial that young people sitting examinations understand the implications of achieving (or failing to achieve) a key GCSE threshold.

Yet, despite the importance of this issue, the existing evidence on the consequences of meeting key grade thresholds remains relatively sparse. This includes in England, despite the academic, media and policy attention given to GCSEs. From an international perspective, Ebenstein *et al.* (2016) find there to be long-term consequences of random fluctuations in high-stakes examination performance in Israel, impacting future educational attainment and labour market earnings. Based upon a sample of higher education students, Arnold (2017) finds that narrowly missing a high-stakes threshold – and thus being forced to resit an examination – has little impact upon future learning. Analysing data from Massachusetts, Papay *et al.* (2010) find no overall association between meeting a high-stakes grade threshold and future educational achievement. They do, however, find a negative impact upon selected sub-groups (most notably low-income students from urban neighbourhoods). Similarly, based upon data from Californian exit examinations, Reardon *et al.* (2010) find ‘*no evidence of any significant or sizeable effect of failing the exam on high school course-taking, achievement, persistence, or graduation.*’ In terms of psychological consequences, Cornell *et al.* (2006) analysed data from 911 students (who were between 13/14 and 17/18 years old) who were wrongly informed they had failed the Minnesota Basic Standards Test. They found more than 80% felt depressed, worried or embarrassed by the results, with four percent dropping out of school as a result. Focusing specifically on the

evidence for England, Hayward *et al.* (2014) find that ‘*individuals who just cross the five good GCSE threshold have considerable lifetime productivity returns compared to those who don’t.*’ Specifically, they estimate that men who achieve between 5 and 7 good passes in their GCSEs earn around £73,000 more over their lifetime than observationally comparable individuals who achieve 3–4 good GCSE passes, with the returns for women lower (£55,000). Similarly, Hodge *et al.* (2021), find a one-grade increase in GCSE mathematics is associated with a £14,500 discounted labour market return, compared to £7,300 for English language. They also note that the greatest marginal labour market returns are for those young people who achieve grade C rather than grade D in a subject.

One particularly relevant paper to this study is Machin, McNally and Ruiz-Valenzuela (2020). Drawing upon data supplied by one of England’s examination boards (AQA) linked to administrative records, the authors studied the effect of narrowly obtaining a ‘good pass’ (C grade) in GCSE English Language. They find that pupils who just fail to achieve a C grade are at greater risk of dropping out of education, and to not be in employment, up to three years later. Critically, the magnitude of the effects they report are summarised as ‘*moderately high*’ noting how this has ‘*high potential long-term consequences for those affected*’. The administrative data used by Machin McNally and Ruiz-Valenzuela (2020) has not historically been easy for researchers to access, and hence few other studies have used such resources to investigate the impacts of high-stakes testing. This, however, may begin to change over the next few years, with recent development of the Grading and Admissions Data for England (GRADE) database.

The work of Machin, McNally and Ruiz-Valenzuela (2020) has some key strengths in terms of a very large sample size and a research design that attempts to control for potential confounding by unobserved characteristics. Yet clear gaps in our knowledge and the existing evidence remain. Firstly, most previous work has a short-term focus, typically concentrating upon the link between achieving a specific high-stakes threshold and short-term educational progression. In contrast, this paper provides evidence on both short and medium-term outcomes, up to ten years after the high-stakes examinations have taken place. Secondly, most existing research considers a relatively narrow set of outcomes. We advance this aspect of the literature by considering a broader array of measures, including educational achievement, labour market attainment, well-being and mental health. Indeed, the ‘psychological impacts’ of missing a C grade was highlighted by Machin, McNally and Ruiz-Valenzuela (2020) as an area where existing evidence in England is lacking. Thirdly, in England, existing work has focused upon young people achieving a GCSE C grade in English language. Again, the evidence presented here is broader – drawing comparisons across mathematics, English language and science. This paper thus attempts to build upon the foundations laid by Machin, McNally and Ruiz-Valenzuela

(2020) by exploring the link between meeting the high-stakes GCSE C threshold (a) across a range of different subjects; (b) upon a wider array of outcome measures and (c) tracking outcomes over a longer time horizon.

In doing so, this paper also feeds into broader debates about formal assessment and examination outcomes, including the purpose of education. In particular, throughout the education literature there has been growing concern surrounding curriculum design and teaching approaches being unduly affected by the pressure teachers and schools face to get pupils to achieve certain examination results (Berliner, 2011; Marshall, 2017; Natriello, 2009). This, in turn, is more broadly related to issues such as school and teacher accountability and the ‘datafication’ of learning. The notion that young people ‘need’ to achieve certain high-stakes examination results treats education as a means to an end (ultimately, credentials that are used by employers) rather than as an important outcome in itself. This has only been exacerbated in recent years by increasing competition amongst countries based upon performance in international assessments such as PISA and PIRLS. In other words, the topic of focus within this paper is to some extent driven by the normative assumption that the purpose of examinations are primarily a measure of successful schooling, and that meeting a key threshold (such as a C grade in GCSE in English and mathematics) is how this is often judged.

To preview findings, strong evidence emerges of a substantial association between failing to obtain a C grade and educational progression and attainment, including the probability of holding a degree by age 26. On the other hand, there is no evidence of a link between failing to achieve a ‘good pass’ in GCSEs and young people’s future wellbeing and mental health. Those who do not achieve a C grade in mathematics are however found to reflect more negatively about their experiences at school, including whether they felt school had adequately prepared them for life. The same was not true, however, for those who failed to achieve a C in English Language or science. Similarly, evidence of a link between long-lasting labour market outcomes are greatest for mathematics, where a failure to achieve a ‘good pass’ is linked to employment and earnings outcomes at age 26.

The paper now proceeds as follows. Section 2 describes the Next Steps (NS) dataset, with our empirical methodology set out in section 3. Results follow in section 4. Conclusions and policy implications are then discussed in section 5.

2. DATA

Next Steps (formerly the Longitudinal Study of Young People in England) began in 2004 collecting data from a school year group born in 1989/1990. Note that these data refer to a sample of pupils living in England only, with it not possible to conduct an equivalent analysis for other parts of the UK. In the first wave, schools were selected with probability proportional to size. Then,

within each school, around 35 Year 9 (age 13/14) pupils were randomly selected. This resulted in a baseline sample of 15,770/13/14 year-olds – a 74% response rate. Annual follow-ups were conducted for the next six years (until age 19/20) and then again at age 26. In the latest (age 26) survey sweep, 7,707 young people took part. See Silverwood *et al.* (2020) for a discussion of these data, including an in-depth analysis of non-response. The only previous work we are aware of to use these data to explore the impacts of high stakes testing upon pupils in Benton (2013). He uses the data to investigate how the division of pupils into different assessment tiers is linked to their aspirations for the future.

Next Steps has been linked to the National Pupil Database (NPD); administrative records that include detailed information on national test and examination performance. These data provide two key pieces of information. Firstly, Key Stage 2 (age 11) and Key Stage 3 (age 14) national test scores, providing high-quality measures of young people's achievement in English, science and mathematics prior to sitting GCSEs.¹ Secondly, the precise grade that they obtained in each GCSE subject – high-stakes, externally marked national examinations taken at age 16. The Next Steps cohort sat pre-reformed GCSEs, and so have received grades using the old eight-point letter scale (A*, A, B, C, D, E, F and U).² Although any grade above a U on this eight-point scale was technically a pass, the C grade was in practise the minimum standard young people were expected to achieve. In this paper we are particularly interested in whether young people achieved a C or a D grade across three subjects: (a) mathematics; (b) English language and (c) double science.³

2.1 Outcome Measures

The outcomes we focus upon fall into the three groups. Firstly, future educational trajectories:

- Whether the student took three A-Levels. A-Levels are the main post-16 academic qualifications in England. We code this variable as one if the cohort member went on to take A-Levels in at least three subjects, and zero otherwise.
- University entry. In the sixth (age 18/19) and seventh (age 19/20) sweeps, young people were asked about their current activities. From this information, we create a binary variable coded as one if the cohort member reported being a full-time student in higher education studying for a degree-level qualification, and zero otherwise.
- High-tariff university entry. During the sixth and seventh survey sweep, young people were asked the name of the university they attend. The survey organisers have then derived a categorical variable, indicating whether the young person attends a Russell Group university. These

universities are the UK's most research-intensive higher education institutions and have high academic entry requirements. A binary variable is derived, coded as one if the young person attends a high-status university at any point during the sixth and seventh survey sweep, and zero otherwise.

- University graduation. As part of the age 26 survey sweep, young people were asked to report the educational qualifications they hold. We derive a binary variable, coded as one if they hold a bachelor's degree or higher, and zero otherwise.

Secondly, young people's labour market outcomes.

- Not in Education, Employment or Training – NEET (ages 17–20, 26). During the fourth-eighth survey sweeps, respondents were asked about their education, employment and training activities. Yet around eight percent of 16–18 year olds (and 14% of 16–24 year-olds) are not involved in any of these activities. Using the information on young people's activities at these ages we derive a binary variable, coded as one if the young person was NEET and zero otherwise.
- Routine/semi-routine job. From the age 17 sweep onwards, young people in employment were asked for the title of their job and their duties. The survey organisers have used this information to derive National Statistics Socio-Economic Classification (NS-SEC) groupings. We derive a binary variable, coded as one if the young person was in routine/semi-routine employment, and zero otherwise. Note that analysis using this variable is restricted to only those individuals who left education.
- Earnings at age 26. During the age 26 sweep, cohort members reported their gross weekly pay in their current main job.

Thirdly, we also investigate a set of wellbeing and socio-emotional outcomes:

- Mental health at ages 17 and 26. Cohort members twice completed the General Health Questionnaire (GHQ-12). This encompasses 12 statements such as *'have you recently felt constantly under strain'*, *'have you recently felt you couldn't overcome your difficulties'* and *'have you recently been feeling unhappy or depressed'*, with four possible responses (*'not at all'* to *'much more than usual'*). It has been widely used to detect minor psychiatric conditions (e.g. anxiety, depression) within the general population (Gnambs and Staufenbiel, 2018). Total scores for each cohort member have been standardised to mean zero and standard deviation one, meaning results can be interpreted in terms of effect sizes. We also investigate the probability of young people having an 'elevated' GHQ

score (defined as a score as three and above on the GHQ scale).

- Life-satisfaction at age 20 and 26. Cohort members were asked ‘*how dissatisfied or satisfied are you about the way your life has turned out so far?*’ with five options (very satisfied to very dissatisfied). We convert this into a binary format, indicating whether the respondent selected fairly/very satisfied (1) or not (0).
- Attitudes towards school. The year after young people left school, they were asked a series of questions about their attitudes towards their time in Year 11. This included five questions such as ‘*School has helped give me confidence to make decisions*’, ‘*School has done little to prepare me for when I leave school*’ and ‘*school has taught me things which would be useful in a job*’, each answered using a four-point scale (strongly disagree to strongly agree). From young people’s responses, the survey organisers derived an ‘attitudes towards school’ scale, which we have standardised to mean zero and standard deviation one. We consider differences between C and D grade pupils on this scale, as well as for the three exemplar questions above.
- Locus of control. At age 20 and 26, respondents were asked a series of questions such as ‘*how well you get on in this world is mostly a matter of luck*’ and ‘*if you work hard at something you’ll usually succeed*’. From responses, locus of control scales have been derived and standardised to mean zero and standard deviation one.

2.2 Background Controls Measured Prior to Young People Receiving Their GCSE Results

The following variables are used in a selection of models as background controls (see the methodology section for further details):

- Attitudes towards school (age 14, 15 and 16). Respondents were asked a series of questions such as ‘*School is a waste of time for me*’ and ‘*School work is worth doing*’. This has been converted into a single continuous scale score by the survey organisers.
- Attitudes towards individual subjects (age 14 and 15). At age 14, young people were asked ‘*How much do you like or dislike maths/English/Science*’. At both ages 14 and 15, they were also asked to name their favourite and least favourite subject. These data are used to control for subject-specific differences in young people’s attitudes.
- Mental health (age 15). Measured via the GHQ scale, as described above.

- Locus of control (age 15). Measured via the locus of control scale as described above.
- Future educational plans (age 14, 15 and 16). Young people were asked about their future educational plans, such as whether they plan to leave school after year 11, whether they plan to do A-Levels and whether they plan to apply to university.
- Truancy from school (age 15 and 16). Respondents were asked whether they had missed school or lessons without permission over the last academic year and how frequently this occurred.

3. METHODOLOGY

Our focus is pupils who achieved a C grade versus a D grade. The same analytic process will be followed for each outcome.

Using GCSE mathematics as an example, the analysis will begin by restricting the sample to young people who achieved a C or D grade in the subject. The raw, unconditional values of each outcome will first be compared across these two groups. We will then use OLS regression to establish the benefits of achieving a C grade (relative to a D grade) via the following regression model:

$$O_{ij} = \alpha + \beta.C_i + \gamma.D_i + \delta.S_i + \theta.P_i + \emptyset.KS4_i + \vartheta.T_i + \tau.5AC_i + \varepsilon_{ij} \quad (1)$$

Where:

O_i = One of our outcomes of interest.

C_i = A binary variable coded as one if the young person achieved a C grade in the subject being investigated (e.g. mathematics) and zero if they achieved a D grade.

D_i = A vector of demographic background characteristics, including gender, ethnicity and socio-economic status.

S_i = A vector of measures of the young person's prior academic achievement. This includes Key Stage 2 and (age 11 and 14) examination scores in English, science and mathematics.

$KS4_i$ = A vector of variables capturing performance in other GCSE subjects (other than in the subject in question – e.g. mathematics). This includes (a) the total GCSE points score across all other GCSE subjects (other than in mathematics); (b) GCSE English language grade; (c) GCSE double science grade.

P_i = Characteristics and attitudes of the child at age 14, 15 and 16, before GCSE results have been received.

T_i = Whether the young person sat the higher or lower GCSE tier paper.

$5AC_i$ = Whether the young person achieved five A*-C grades in their GCSEs.

ε_{ij} = Random error term.

i = Pupil i

j = School j

Six specifications of this model are estimated, with multiple imputation using chained equations used to account for missing covariate data. All standard errors will be clustered at the school level and weights applied.

Model M1 only includes controls for demographic background characteristics. Key Stage 2 and 3 (age 11 and 14) scores are added in model M2. Model M3 adds further controls for characteristics and attitudes of young people *before* they have received their GCSE results (this includes all the ‘background control’ measures described in the data section above⁴). Performance in other GCSE subjects (i.e. other than in the subject of interest) are then added in model M4 – these provide our headline results. In particular, M4 will compare the association between getting a grade C versus a D in a given subject, amongst those young people with similar levels of prior achievement, who hold similar attitudes to school during Years 9 to 11 and who achieved very similarly across other GCSE subjects.

Although we are able to include subject-specific test score controls at age 11 and 14 (Key Stage 2 and 3 scores) that are strongly correlated with GCSE grades (e.g. Pearson $r = 0.83$ for mathematics for the full sample and $r = 0.50$ amongst those who achieve a C or D grade), these only imperfectly capture subject-specific skills at age 16 when young people sit their GCSEs. The main implication is that the variables included in vector S in equation (1) are unlikely to fully capture the confounding effect of differences in academic skills at age 16 between GCSE C/D grade pupils. Consequently, our estimates are likely to provide an upper bound of the link between obtaining a higher grade in a key subject and future outcomes (independent of their skills in the subject).

Two further specifications are then estimated to provide further context to the results (. In some subjects, schools may enter pupils into ‘higher’, ‘intermediate’ and ‘lower’ GCSE papers. These papers contain questions of different difficulty, meaning they require pupils to answer a different proportion of questions correctly to obtain a C grade. Although these are statistically equated to ensure comparability, there has been much previous work into the link between paper tier and examination outcomes (Barrance, 2020; Elwood, 2005; Elwood and Murphy, 2002; Vitello and Crawford, 2018). In Next Steps, the tier of the GCSE paper is available for those taking papers from the AQA and (occasionally) the OCR examination boards (see Benton, 2013).⁵ Specifically, we know the tier of the GCSE paper for 77% of cohort members for English Language, 80% for Double science, but just 16% in mathematics. In M5, we include an additional control for paper tier to investigate whether this changes our results.

In model M6, we also control for whether the young person achieved five A*-C grades. As noted by Machin, McNally and Ruiz-Valenzuela (2020), failing to achieve a C grade in a single, particularly important subject such as English Language or mathematics means a young person is also less likely to achieve another high-stakes threshold – gaining five or more ‘good’ GCSEs. In particular, they highlight how – by failing to gain a grade C in English language – pupils ‘*may face the double whammy of failing to obtain a “good” grade in a core subject and failing to achieve a sufficient number of “good” GCSEs*’. By controlling for whether young people achieve the five good GCSE threshold in model M6, we attempt to establish the extent that this is the mechanism via which any apparent association emerge (rather than failing to meet the C threshold in a particular subject – such as mathematics – per se).

3.1 Robustness Tests

Two sets of robustness tests will be conducted. Firstly, our use of OLS regression provides a set of easy-to-interpret and communicate estimates. However, for binary outcomes, linear probability models (i.e. OLS regression with binary outcomes) have some well-known limitations (Mood, 2010). Consequently, for binary outcomes, we also present estimates from logistic regression models in the form of odds-ratios. Secondly, one of the limitations with using regression modelling is that it does not enforce ‘common support’ on the analytic sample. In other words, there may be some young people (e.g. those at the upper end of the C grade in mathematics) for whom there is no observationally similar young person that achieved a grade D. Our second robustness test hence uses propensity score matching as an alternative analytic approach. Specifically, we estimate a logistic regression model including controls for demographic characteristics, Key Stage 2 scores, Key Stage 3 scores, attitudes to school in Years 9–11 and GCSE performance in other subjects (i.e. roughly equivalent to OLS regression model specification M4 described above). Propensity scores are then derived and used to match pupils who achieved a C grade to an equivalent pupil who achieved a D grade. This is done using one-to-one nearest neighbour matching with replacement and caliper length set to 0.02. These results will provide an estimate of the average treatment effect (ATE) of obtaining a C grade. These alternative estimates are presented in full in Appendix B and Appendix C.

3.2 A Note about Standard Errors

An anonymous reviewer of the manuscript has suggested that all reference to significance testing and standard errors be removed. However, given that the data were originally randomly sampled from the population, we follow

convention in the quantitative social science literature and continue to report standard errors in the results tables. These provide a guide to the amount of uncertainty in the estimates due to sampling variation. For a discussion of the arguments against reporting standard errors, see Gorard (2015).

4. RESULTS

4.1 *Comparison of Unadjusted Outcomes*

A descriptive comparison of outcomes between teenagers who received C and D grades in mathematics, English Language and Double Science is presented in Table 1. There are clearly sizable differences in both short-term and medium-term educational outcomes, with obtaining a good pass linked to a higher probability of completing A-levels, going to university, entering a high-tariff institution and graduating with a degree. Likewise, those who obtain a C grade tend to achieve better labour market outcomes than their peers with a D grade, with sizable differences in NEET status at all ages, while gaps in occupational outcomes and earnings are most apparent at age 26. Conversely, there is little sign of a substantive difference between C and D grade pupils in terms of their mental health and life satisfaction. The final panel reveals that achieving a good pass in a key GCSE subject is associated with attitudes towards school, particularly in mathematics (e.g. 81% of those who achieved a grade C said they felt school had given them confidence, compared to 72% who achieved a grade D). Yet there is little sign of a major difference – even in these unconditional estimates – for the other social-emotional outcomes considered.

4.2 *Educational Outcomes*

Table 2 presents results from our OLS models, based upon specification M4. Analogous results from all six model specifications can be found in Appendix A. These refer to the difference in probability of experiencing each outcome if a young person achieves a C grade in the subject compared to a D grade.

There is clear and consistent evidence that achieving a C grade confers substantial advantages for educational progression and long-term educational attainment. This holds true across English Language, mathematics and double science. For instance, obtaining a C grade in any of these subjects is associated with somewhere between a 5 and 10 percentage point increase in the probability of studying A-Levels at age 17/18. This is broadly consistent with the findings of Machin, McNally and Ruiz-Valenzuela (2020: Table 6), who reported around a 10-percentage point increase in the probability of studying for A-Levels at age 17 from obtaining a C grade in English Language. Yet we also find evidence of longer-term educational associations as well. For each of subject, achieving a C grade is strongly associated with the probability of applying to university and

TABLE 1. Descriptive statistics. Raw outcomes for young people who achieve C versus D grades

	Maths			English			Science		
	Obs	D	C	Obs	D	C	Obs	D	C
Educational outcomes									
Doing A-Levels (age 17)	4362	27%	57%	4790	19%	52%	3683	60%	34%
Doing A-Levels (age 18)	4003	22%	46%	4391	15%	41%	3409	49%	25%
Retaking some GCSEs (age 17)	4362	9%	3%	4790	9%	4%	3683	8%	3%
Retaking some GCSEs (age 18)	4019	12%	4%	4402	11%	6%	3422	10%	5%
In education (age 17)	4044	54%	63%	4440	51%	60%	3447	55%	65%
Received university offers (age 18)	4050	11%	32%	4441	8%	28%	3445	15%	36%
Applied to university (age 18)	4054	13%	36%	4446	10%	32%	3451	19%	39%
Studying for a degree (age 19)	3791	9%	24%	4151	6%	23%	3257	13%	27%
Studying for a degree (age 20)	3374	17%	37%	3648	12%	32%	2903	23%	39%
At first-choice university (age 19)	3787	11%	23%	4156	7%	21%	3255	13%	25%
Studying first-choice of subject at university (age 19)	3790	11%	25%	4159	8%	23%	3261	15%	28%
Holds a degree (age 26)	3049	15%	27%	3312	11%	26%	2617	19%	30%
Attends high-tariff university (age 19)	3793	0.20%	2%	4158	0.30%	1.90%	3261	0.6%	2%
Attends high-tariff university (age 20)	3376	0.20%	3%	3649	0.50%	2.20%	2899	1%	3%
Holds degree from high-tariff university (age 26)	3033	0.20%	1.40%	3296	0.50%	0.80%	2605	0.7%	1.5%
Labour market outcomes									
Routine/semi-routine employment (age 18)	1451	57%	54%	1625	58%	56%	1219	56%	55%
Routine/semi-routine employment (age 20)	2801	33%	24%	2978	39%	23%	2466	32%	20%
Routine/semi-routine employment (age 26)	2553	29%	19%	2715	33%	19%	2255	27%	17%
NEET (age 17)	4427	9%	5%	4880	11%	5%	3737	9%	3%
NEET (age 19)	3758	11%	6%	4117	14%	7%	3240	9%	6%
NEET (age 20)	3361	13%	7%	3643	16%	8%	2891	11%	6%
NEET (age 26)	3037	16%	8%	3295	16%	11%	2608	11%	8%
Claiming benefits (age 20)	3359	16%	10%	3629	19%	11%	2888	15%	8%
Log income (age 20)	1351	0.09	0.09	1511	0.09	0.09	1105	0.09	0.09

	2930	297	315	3173	290	311	2518	308	318
Income (age 26)									
Socio-emotional outcomes									
GHQ score (age 17)	4154	-0.05	0.00	4557	-0.15	-0.01	3542	-0.02	0.03
GHQ score (age 26)	2915	-0.03	-0.01	3161	-0.01	-0.02	2521	-0.04	-0.03
Elevated GHQ score (age 17)	4154	27%	31%	4557	23%	30%	3542	29%	31%
Elevated GHQ score (age 26)	2915	32%	32%	3161	32%	32%	2521	30%	31%
Life satisfaction (age 20)	3337	77%	78%	3614	76%	77%	2871	75%	80%
Life satisfaction (age 26)	2941	72%	72%	3194	69%	72%	2536	71%	74%
Attitude towards school (age 17)	4362	-0.26	0.06	4790	-0.28	-0.01	3683	-0.17	0.08
Felt school gave them confidence (age 17)	4323	72%	81%	4739	73%	77%	3656	75%	80%
Felt school did little to prepare them (age 17)	4334	32%	20%	4753	29%	22%	3658	27%	19%
Felt school was useful (age 17)	4336	71%	77%	4761	72%	76%	3659	75%	78%
Locus of control (age 20)	3360	0.02	0.06	3636	-0.01	0.04	2890	0.00	0.02
Locus of control (age 26)	2923	-0.04	-0.03	3172	-0.07	-0.04	2521	-0.05	-0.07

Obs refers to the number of observations, D refers to the results for those young people who obtain a grade D, while C refers to outcomes for those who obtain grade C.

TABLE 2. *The association between getting a GCSE C grade and educational outcomes*

	Maths		English		Double science	
	Diff	SE	Diff	SE	Diff	SE
Doing A-Levels (age 17)	7.3%	1.9%	10.1%	1.7%	7.8%	1.9%
Doing A-Levels (age 18)	5.5%	2.1%	7.3%	1.7%	7.1%	1.9%
Retaking some GCSEs (age 17)	-5.1%	1.1%	-5.5%	1.0%	-1.2%	1.2%
Retaking some GCSEs (age 18)	-7.5%	1.4%	-5.3%	1.2%	1.0%	1.2%
In education (age 17)	1.0%	2.2%	1.1%	1.9%	5.3%	2.1%
Received university offers (age 18)	9.9%	1.7%	7.7%	1.3%	7.9%	1.9%
Applied to university (age 18)	9.8%	1.9%	8.6%	1.5%	6.2%	2.1%
Studying for a degree (age 19)	5.6%	1.9%	6.7%	1.3%	5.3%	1.8%
Studying for a degree (age 20)	6.1%	2.0%	7.3%	1.6%	3.7%	2.4%
At first-choice university (age 19)	2.9%	1.9%	4.8%	1.4%	3.5%	1.8%
Studying first-choice of subject at university (age 19)	4.8%	1.8%	5.7%	1.5%	3.3%	1.8%
Holds a degree (age 26)	6.2%	2.0%	5.4%	1.7%	5.2%	2.2%
Attends high-tariff university (age 19)	0.9%	0.4%	0.4%	0.3%	0.5%	0.5%
Attends high-tariff university (age 20)	1.4%	0.5%	0.3%	0.4%	0.6%	0.7%
Holds degree from high-tariff university (age 26)	0.8%	0.4%	-0.1%	0.4%	-0.2%	0.8%
Controls						
Demographics				Yes		
Age 11/14 test scores				Yes		
Age 14–16 controls				Yes		
GCSE outcomes in other subjects				Yes		

See Appendix Table A1 for details on the number of observations. Estimates refer to the change in the probability of the outcome if a young person achieves a C grade relative to a D grade. Estimates based upon model M4. See Appendix A (OLS), B (PSM) and C (logistic regression) for full results across all models and robustness tests. Green shading indicates statistical significance at the five percent level.

receiving an offer of a place at age 18, studying for a degree at age 19/20 and, critically, obtaining a university degree by age 26. Specifically, achieving a C in any of the three subjects is associated with a 5 to 6 percentage point increase in the probability of gaining a degree, amongst young people with similar Key Stage 2 and Key Stage 3 scores, who performed similarly in other GCSE subjects, who held similar attitudes towards schools and with similar socio-emotional outcomes at ages 14, 15 and 16, and who have similar demographic characteristics. These are sizeable, substantively important associations.

On the whole, the magnitude of the associations observed are broadly similar across the three subjects. There are, however, some noteworthy exceptions. Firstly, attending – and graduating from – a high-tariff university. Here we observe an association (if relatively small) for gaining a C grade in mathematics (around one percentage point), but not for English Language or Double science. Secondly, failing to achieve a C grade in mathematics or English Language makes young

people much more likely to retake some GCSEs at age 17/18 (between a 5 and 7 percentage point difference), while the same is not true for double science.⁶

The parameter estimates from all model specifications (see Appendix A) provide two further insights. From model M5 one can see that these results are not being driven by young people being allocated to different GCSE tiers; controlling for this factor does not meaningfully alter our results. Similarly, results from model M6 suggest that the above findings are not being primarily driven by the fact that obtaining a grade C in a key subject means that young people will also be more likely to achieve another important examination threshold (achieving 5 A*-C grades across all GCSE subjects). Indeed, for most outcomes, estimates from M6 are only marginally smaller than in M4.⁷

Appendix B and C presents key findings from our sensitivity analyses, using propensity score matching (PSM) and logistic regression (Logit) rather than OLS. Reassuringly, the direction and magnitude of the estimated associations are similar across the three approaches. On the whole, these results support the finding that failing to achieve a C grade in a key GCSE subject has sizeable, long-lasting links with educational progression and attainment.

4.3 Labour Market Outcomes

Table 3 presents analogous results for labour market outcomes. The evidence here is mixed.

Starting with mathematics, the estimated association between obtaining a C are generally small when looking at short-term labour market outcomes (i.e. between ages 17 and 20). There is little clear and consistent difference between those who achieved a C versus a D in terms of their income, occupation, claiming of benefits, attitudes about their job or whether they are not in education, employment or training (NEET). These shorter-term labour market outcomes are of course somewhat complicated by labour market selection, with many young people continuing in education at this point in their lives. Yet, by age 26, some substantively important differences have emerged. In particular, those who gain a C grade in GCSE mathematics are around 5 percentage points less likely to be working in a routine/semi-routine job, 4 percentage points less likely to be NEET and with an average income around £7 higher per week (mean = £300, standard deviation = £73). The results from our robustness tests (see Appendix B and C) suggest that the direction and magnitude of these estimated associations are similar across different analytic approaches. Moreover, the results from models M5 and M6 indicate that the findings are not changed by the addition of extra controls (see Appendix A). There is hence some evidence of moderate medium-term labour market benefits from gaining a GCSE C grade in mathematics.

For English Language, most of the estimated coefficients have the expected sign, though the associations seems relatively weak. The one potential

TABLE 3. *The association between getting a GCSE C grades and labour market outcomes*

	Maths		English		Double science	
	Diff	SE	Diff	SE	Diff	SE
Routine/semi-routine employment (age 18)	-1.4%	3.3%	-3.7%	3.3%	1.5%	3.6%
Routine/semi-routine employment (age 20)	-2.0%	2.5%	-8.0%	2.5%	-6.8%	2.4%
Routine/semi-routine employment (age 26)	-5.4%	2.5%	-6.3%	2.3%	-0.6%	2.2%
NEET (age 17)	-0.4%	1.0%	-1.8%	1.0%	-2.9%	0.9%
NEET (age 19)	-3.8%	3.6%	-2.3%	3.2%	0.8%	3.4%
NEET (age 20)	-1.8%	1.5%	-2.8%	1.5%	-1.2%	1.3%
NEET (age 26)	-4.0%	1.6%	-1.4%	1.5%	1.2%	1.5%
Claiming benefits (age 20)	0.6%	1.6%	-2.5%	1.6%	-3.7%	1.4%
Log income (age 20)	-0.04	0.04	0.03	0.03	0.02	0.04
Income (age 26)	6.8	2.6	0.0	2.5	-3.5	2.6
Controls						
Demographics				Yes		
Age 11/14 test scores				Yes		
Age 14–16 controls				Yes		
GCSE outcomes in other subjects				Yes		

See Appendix Table A2 for details on the number of observations. Estimates refer to the change in the probability of the outcome if a young person achieves a C grade relative to a D grade. Estimates based upon model M4. See Appendix A (OLS), B (PSM) and C (logistic regression) for full results across all models and robustness tests. Green shading indicates statistical significance at the five percent level.

exception, however, is with respect to working in a routine/semi-routine job; gaining a C grade in English Language reduces the probability of working in such a job (relative to working in intermediate/professional employment) by around 8 percentage points at age 20 and 6 percentage points at age 26 (the latter being similar to the findings for mathematics). This finding again appears robust to alternative analytic approaches and to model specification (see Appendix A).

Results for science are less clear. The direction and magnitude of the estimated association varies across the different outcomes. Moreover, where sizable associations are observed, there is again inconsistency in terms of either the outcome measure or the age. Overall, the results for science are somewhat inconclusive.

Thus, in summary, we find robust evidence of moderate medium-term associations between failing to achieve a C grade in GCSE mathematics and labour market outcomes. For English Language, the results suggest a C grade is linked to reducing the probability of working in routine employment. For science, evidence of labour market benefits from achieving a C grade is less clear.

TABLE 4. The association between getting a C grade and future socio-emotional outcomes

	Maths		English		Double science	
	Diff	SE	Diff	SE	Diff	SE
GHQ score (age 17).	-0.01	0.04	0.08	0.03	0.00	0.04
GHQ score (age 26).	-0.02	0.05	0.01	0.04	0.04	0.05
Elevated GHQ score (age 17)	1.5%	2.0%	4.2%	1.6%	0.5%	2.0%
Elevated GHQ score (age 26)	-0.7%	2.6%	-0.7%	2.2%	1.9%	2.6%
Life satisfaction (age 20)	1.7%	2.1%	-1.0%	2.0%	2.7%	2.1%
Life satisfaction (age 26)	2.3%	2.4%	1.1%	2.0%	1.2%	2.3%
Attitude towards school (age 17)	0.09	0.04	0.07	0.03	0.03	0.04
Felt school gave them confidence (age 17)	5.0%	1.8%	-0.3%	1.7%	0.4%	1.8%
Felt school did little to prepare them (age 17)	-4.9%	1.8%	-1.4%	1.7%	-1.8%	1.9%
Felt school was useful (age 17)	1.6%	1.8%	0.3%	1.6%	-0.7%	1.9%
Locus of control (age 20)	-0.01	0.05	0.05	0.04	0.00	0.05
Locus of control (age 26)	0.04	0.05	0.04	0.05	-0.04	0.05
Controls						
Demographics			Yes			
Age 11/14 test scores			Yes			
Age 14–16 controls			Yes			
GCSE outcomes in other subjects			Yes			

Estimates refer to the change in the probability of the outcome if a young person achieves a C grade relative to a D grade. Estimates based upon model M4. See Appendix A (OLS), B (PSM) and C (logistic regression) for full results across all models and robustness tests. Notes: See Appendix Table A3 for details on the number of observations. Green shading indicates statistical significance at the five percent level.

4.4 Socio-emotional Outcomes

Estimates for socio-emotional outcomes can be found in Table 4. In terms of our measures of wellbeing, a clear and consistent pattern emerges; there is no evidence that failing to meet a key, high-stakes threshold in GCSE examinations is linked to a decline in young people's wellbeing and mental health. The vast majority of estimates are small in terms of magnitude. This holds true across our headline estimates, different model specifications (Appendix A) and alternative analytic approaches (Appendix B and C). The only potential exception is with respect to English Language and GHQ outcomes at age 17. This is, however, opposite to the expected direction; teenagers who achieve a C grade in their English Language GCSE appear to have slightly *worse* mental health outcomes (i.e. are 4 percentage points more likely to have an elevated GHQ score) than those who achieved a grade D. Hence, overall, failing to achieve a 'good pass' in key GCSE subjects such as English Language and mathematics does not seem to be related to young people's wellbeing and mental health (either in the

short or the longer term). Additional analysis of data from the Millennium Cohort Study (MCS) cohort support this finding (further details are provided in Appendix E and in the sub-section below).

Turning to other socio-emotional outcomes, the only area where there is reasonably consistent evidence is for the relationship between achieving a ‘good pass’ (grade C) in mathematics and young people’s attitudes towards school. Those teenagers with a grade C were more positive about school than their observationally equivalent peers who achieved grade D (effect size = 0.09), being around five percentage points *more* likely to agree that school gave them confidence and five percentage points *less* likely to believe that school did little to prepare them for life. The same findings do not hold for science, where all parameter estimates are small, while the analogous results for English Language are somewhat mixed. For all other outcomes (e.g. locus of control) there is no evidence of any difference.

Thus, evidence of broader links between failing to achieve key high-stakes grade thresholds and future outcomes is limited, restricted to short-term associations with attitudes towards school from failing to achieve a C grade in mathematics.

4.5 Comparison to the B/C Grade Boundary

The ‘good pass’ (C) threshold is thought to have particular importance in the English education system. If young people do not obtain this grade, then their future educational and labour market opportunities are thought to diminish. But is there any evidence of similar results at other grade boundaries? We explore this issue in Appendix D where we present a comparison of (a) obtaining a C grade compared to a D grade to (b) obtaining a B grade versus a C grade. These estimates are all based upon OLS regression model M4.⁸

To summarise the findings, the overall direction, magnitude of the estimated associations are similar across the D/C and C/B grade comparisons. In other words, we do not find any evidence that the good pass (C/D) threshold is unusually important; similarly sized associations with our outcomes are observed for the B/C threshold. There are, however, a handful of noteworthy exceptions. For instance, obtaining a C rather than a D is particularly important in terms of the probability of young people retaking any GCSEs the following academic year (much more so than the C versus B distinction). Similarly, for English Language, the difference between obtaining a C instead of a D has a much stronger link with the probability of being in routine employment at age 20 and 26 than the difference between a C and a B. Finally, achieving a C rather than a D in mathematics seems to be linked to young people’s reflections of their time at school (e.g. whether they felt school had prepared them for life), with no such difference observed for those at the B/C boundary.

However, outside of these exceptions, estimates for the C/D comparison are similar to those for B/C. There are two potential interpretations of this finding. One is that falling either side of the good pass (C/D) boundary holds little extra importance than falling either side of another (e.g. B/C) grade boundary. Another is that unobservable characteristics may be confounding both our D/C and C/B comparisons. It is not possible to tease these two potential explanations apart with the data available.

5. CONCLUSIONS

High-stakes examinations at the end of secondary school are now a common feature of education systems across the world. Performance on these examinations – particularly in key subjects such as English, science and mathematics – may have long-lasting consequences for the rest of young people’s lives. It is therefore little wonder that such high-stakes examinations have been linked to stress and anxiety amongst teenagers (Banks and Smyth, 2015), with some young people feeling they are the be-all and end-all when taking them (Young, 2020). England is a prime example, where performance on the high-stakes GCSE examinations often considered pivotal in shaping young people’s futures.

But is this really true? Previous work from England (Machin, McNally and Ruiz-Valenzuela 2020) has shown how young people who just manage to achieve a ‘good pass’ in their GCSE English Language examinations have better educational outcomes up to three years later. This paper has built upon this evidence in three ways. Firstly, we produce evidence for – and compare results across – three key subjects in the English education system (English, mathematics and double science). Secondly, we consider differences in outcomes up to ten years after the examinations took place, tracking outcomes from the short to the medium term. Finally, our analysis considers a broader set of outcomes than previous work, including occupation held, income, measures of wellbeing and reflections upon experiences at school. We have thus added further depth and breadth to our understanding of the consequences of obtaining a ‘good pass’ in key GCSE subjects.

Consistent with findings from the existing literature, our results point towards large and long-lasting associations between obtaining a good pass on future educational outcomes. This continues through to the chances of graduating from university, with associations being of similar magnitude across English Language, double science and mathematics. There is some suggestion that this then feeds through into labour market outcomes, with the evidence stronger for mathematics than the other two subjects, particularly at age 26. On the other hand, failing to achieve a good pass in a key GCSE subject has no link with future levels of wellbeing or other socio-emotional outcomes. Similarly, although there is some suggestion that failing to achieve a good pass may

mean teenagers reflect somewhat more negatively upon their time at school, this seems confined to mathematics only.

These findings should be interpreted in light of the limitations of the work. All our analytic approaches invoke an untestable selection-upon-observables assumption. Although we have been able to condition upon a wide-array of characteristics, including high-quality subject-specific measures of prior achievement and outcomes in other GCSE subjects – the possibility of there being some residual confounding cannot be ruled out. A prudent interpretation of our estimates might therefore be that they provide an upper-bound to any potential effects. Another limitation is the moderate Next Step sample size, meaning that there has been limited statistical power to detect potentially small associations. Our analysis of labour market outcomes is also arguably still at a relatively early age, as young people’s careers are developing. Future research should seek to follow participants through into their thirties, as careers (and age-education-earnings profiles) stabilise. Finally, there has recently been a major change to GCSE qualifications in England. In supplementary analysis of another dataset (the Millennium cohort study) we have replicated our finding that failing to achieve a grade 4 (the new ‘good pass’) is linked to wellbeing outcomes at age 17 (details available from the authors upon request). Yet further work is needed to establish whether longer-term outcomes may have been impacted by recent changes to the qualification system in England.

Despite these limitations, our findings may hold some important implications for education policy and practice. Given that just managing to achieve a good pass matters (at least in some dimensions), one may question why a relatively small set of discrete grades are used in the first place. Indeed, this might suggest that an alternative grading metric (e.g. percentile rank within the cohort) should be used instead. However, if discrete grades are to remain in place, it is vital that decisions around any particularly high-stakes thresholds (e.g. the C/4 ‘good pass’) are as robust as can be. This might motivate the introducing of high-stakes digital GCSE examinations, with adaptive testing used to increase measurement accuracy for those who fall around high-stakes grade thresholds. It is also notable how the Department for Education in England now require those 16-year-olds who fail to achieve a good pass in English Language and mathematics to retake examinations in these subjects the following academic year. Yet, given that we observe similarly sized associations in some areas for science, the Department for Education might consider extending this policy to other key subjects as well. Also, we should not forget what this implies for young people, the vast majority of whom are acutely aware of the importance of GCSEs, and who put themselves under pressure to do well. It is vital that they understand what the implications are of narrowly missing out on a ‘good pass’. Yes, it may mean that certain educational opportunities will not be as accessible to them, and that this may be linked to the job they hold in the future. But this needs to be put into a broader perspective, with no evidence to suggest that their mental health, wellbeing or overall satisfaction with life will be affected by the results.

ACKNOWLEDGEMENTS

This research was funded by UK Economic and Social Research Council (ESRC) grant ES/T003677/1.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the author(s).

FUNDING

This work was supported by the Economic and Social Research Council [ES/T003677/1]. The paper has used data from the UK Data Service Secure Lab. Please see data citations below for further details

SUPPLEMENTARY MATERIAL

Supplemental data for this article can be accessed [here](#).

NOTES

- ¹ For the Next Steps cohort, children in state schools in England all took externally marked tests at the end of Year 9 (known as Key Stage 3 tests). These ceased in 2009.
- ² For this cohort, some subjects also involved coursework, which would also typically make a minor contribution towards the final grade.
- ³ Pupils in this cohort could either take ‘double science’ (equivalent to two GCSEs) or separate GCSEs in Biology, Chemistry and Physics (each equivalent to a single GCSE). Of the 15,770 cohort members in the baseline sample, 10,519 have a GCSE double science grade available.
- ⁴ 99% of wave 3 Next Steps interviews took place between April and August – and were thus prior to when young people received their GCSE results. Responses to the survey questions in wave 3 would therefore not have been influenced by the GCSE grades young people achieved.
- ⁵ For the AQA and OCR exam boards, qualification codes are prefixed by either ‘F’, ‘I’ or ‘H’ – referring to the Foundation, Intermediate and Higher papers. Within the Next Steps data, this can be identified via the variable ‘KS4_QAN’.
- ⁶ Since 2015, those who fail to achieve a C/4 grade in English Language and mathematics are obliged to retake these courses at age 17/18 if they remain in publicly funded education. This was not true, however, for the Next Steps cohort.
- ⁷ One notable exception is for achieving a C grade in double science and the probability of taking A-Levels at age 17/18. In this instance, controlling for whether the young person achieved 5 A*-C grades reduces the estimates by half.
- ⁸ Comparisons across further grade boundaries were not made due to more limited sample sizes.

ORCID

John Jerrim  <http://orcid.org/0000-0001-5705-7954>

REFERENCES

- Arnold, I. (2017) Resitting or compensating a failed examination: does it affect subsequent results?, *Assessment & Evaluation in Higher Education*, 42 (7), 1103–1117. [10.1080/02602938.2016.1233520](https://doi.org/10.1080/02602938.2016.1233520).
- Banks, J. and Smyth, E. (2015) Your whole life depends on it': academic stress and high-stakes testing in Ireland, *Journal of Youth Studies*, 18 (5), 598–616. [10.1080/13676261.2014.992317](https://doi.org/10.1080/13676261.2014.992317).
- Barrance, R. (2020) Tiering in the GCSE: a children's rights perspective, *British Educational Research Journal*, 46 (6), 1210–1231. [10.1002/berj.3629](https://doi.org/10.1002/berj.3629).
- Benton, T. (2013) *Examining the impact of tiered examinations on the aspirations of young people*. British Educational Research Association conference paper, <https://www.cambridgeassessment.org.uk/Images/145926-examining-the-impact-of-tiered-examinations-on-the-aspirations-of-young-people.pdf> (Accessed 15 June 2021).
- Berliner, D. (2011) Rational responses to high stakes testing: the case of curriculum narrowing and the harm that follows, *Cambridge Journal of Education*, 41 (3), 287–302. [10.1080/0305764X.2011.607151](https://doi.org/10.1080/0305764X.2011.607151).
- Carr, J. (2021) *Speed read: think tank calls for GCSEs to be scrapped by 2025*. *Schools Week*, <https://schoolsweek.co.uk/speed-read-think-tank-calls-for-gcse-to-be-scrapped-by-2025> (Accessed 08 June 2021).
- Cho, E. and Chan, T. (2020) Children's wellbeing in a high-stakes testing environment: the case of Hong Kong, *Children and Youth Services Review*, 109 (1), 104694. [10.1016/j.childyouth.2019.104694](https://doi.org/10.1016/j.childyouth.2019.104694).
- Cornell, D., Krosnick, J. and Chang, L. (2006) Student reactions to being wrongly informed of failing a high-stakes test: the case of the Minnesota Basic Standards Test, *Educational Policy*, 20 (5), 718–751. [10.1177/0895904805284123](https://doi.org/10.1177/0895904805284123).
- Coughlan, S. (2019) 'Pointless' GCSEs should be scrapped, says senior MP. *BBC News*, <https://www.bbc.co.uk/news/education-47149808> (Accessed 08 June 2021)
- Data citation**
- Ebenstein, A., Lavy, V. and Roth, S. (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.
- Education Select Committee. (2017) *High-stakes testing harming teaching and learning in primary schools*. <https://www.parliament.uk/business/committees/committees-a-z/commons-select/education-committee/news-parliament-2015/primary-assessment-report-published-16-17> (Accessed 13 Apr 2020).
- Elwood, J. (2005) Gender and achievement: what have exams got to do with it?, *Oxford Review of Education*, 31 (3), 373–393. [10.1080/03054980500222031](https://doi.org/10.1080/03054980500222031).
- Elwood, J. and Murphy, P. (2002) Tests, Tiers and Achievement: gender and performance at 16 and 14 in England, *European Journal of Education*, 37 (4), 395–416. [10.1111/1467-3435.00118](https://doi.org/10.1111/1467-3435.00118).
- Gnamb, T. and Staufenbiel, T. (2018) The structure of the General Health Questionnaire (GHQ-12): two meta-analytic factor analyses, *Health Psychology Review*, 12 (2), 179–194. [10.1080/17437199.2018.1426484](https://doi.org/10.1080/17437199.2018.1426484).
- Gorard, S. (2015) Rethinking 'quantitative' methods and the development of new researchers, *Review of Education*, 3 (1), 72–96. [10.1002/rev3.3041](https://doi.org/10.1002/rev3.3041).
- Hayward, H., Hunt, E. and Lord, A. (2014) *The economic value of key intermediate qualifications: estimating the returns and lifetime productivity gains to GCSEs, A levels and apprenticeships*. Department for Education Research Report. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/387160/RR398A_-_Economic_Value_of_Key_Qualifications.pdf (Accessed 23 June 2021).
- Hodge, L., Little, A. and Weldon, M. (2021) *GCSE attainment and lifetime earnings*. Department for Education Research Report. <https://assets.publishing.service.gov.uk/>

- government/uploads/system/uploads/attachment_data/file/993202/GCSE_Attainment_and_Lifetime_Earnings_PDF3A.pdf (Accessed 29 July 2021).
- Machin, S., McNally, S., & Ruiz-Valenzuela, J. (2020). Entry through the narrow door: the costs of just failing high stakes exams. *Journal of Public Economics*, 190, 104224.
- Marshall, B. (2017) The politics of testing, *English in Education*, 51 (1), 27–43. [10.1111/eie.12110](https://doi.org/10.1111/eie.12110).
- Mood, C. (2010) Logistic Regression: why We Cannot Do What We Think We Can Do, and What We Can Do About It, *European Sociological Review*, 26 (1), 67–82. [10.1093/esr/jcp006](https://doi.org/10.1093/esr/jcp006).
- Natriello, G. (2009) High Stakes Testing and Teaching to the Test. L. J. Saha and A. G. Dworkin. (eds) *International Handbook of Research on Teachers and Teaching*. Springer International Handbooks of Education 21 (Boston, MA, Springer 1101–1111), [10.1007/978-0-387-73317-3_72](https://doi.org/10.1007/978-0-387-73317-3_72)
- Papay, J., Murnane, R. and Willett, J. (2010) The consequences of high school exit examinations for low-performing urban students. Evidence from Massachusetts, *Educational Evaluation and Policy Analysis*, 32 (1), 5–23. [10.3102/0162373709352530](https://doi.org/10.3102/0162373709352530).
- Prior, L., Jerrim, J., Thomson, D. and Leckie, G. (2021) A review and evaluation of secondary school accountability in England: statistical strengths, weaknesses, and challenges for ‘Progress 8’ raised by COVID-19, *Review of Education*, 9 (3), [10.1002/rev3.3299](https://doi.org/10.1002/rev3.3299).
- Reardon, S., Arshan, N., Atteberry, A. and Kurlaender, M. (2010) Effects of failing a high school exit exam on course taking, achievement, persistence, and graduation, *Educational Evaluation and Policy Analysis*, 32 (4), 498–520. [10.3102/0162373710382655](https://doi.org/10.3102/0162373710382655).
- Saminsky, A. (2011) High-stakes standardized testing: a panacea or a pest?, *Inquiries Journal/Student Pulse*, 3, 01. Retrieved from <http://www.inquiriesjournal.com/a?id=373> Accessed 01 March 2022
- Silverwood, R. J., Calderwood, L., Sakshaug, J. W. and Ploubidis, G. B. (2020) A data driven approach to understanding and handling nonresponse in the Next Steps cohort. CLS Working Paper, London UCL Centre for Longitudinal Studies, 5, 2020
- Suto, I. and Oates, T. (2021) *High-stakes testing after basic secondary education: how and why is it done in high-performing education systems? Cambridge Assessment Research Report*. (Cambridge, UK, Cambridge Assessment).
- University College London, UCL Institute of Education, Centre for Longitudinal Studies. (2021). *Next Steps: Linked Education Administrative Datasets (National Pupil Database), England, 2005-2009: Secure Access*. [data collection]. 6th Edition. UK Data Service. SN: 7104, <http://doi.org/10.5255/UKDA-SN-7104-6>
- University College London, UCL Institute of Education, Centre for Longitudinal Studies. (2021). *Next Steps: Sweeps 1-7, 2004-2010: Secure Access*. [data collection]. UK Data Service. SN: 8656, <http://doi.org/10.5255/UKDA-SN-8656-1>
- University College London, UCL Institute of Education, Centre for Longitudinal Studies. (2021). *Next Steps: Sweeps 1-8, 2004-2016*. [data collection]. 15th Edition. UK Data Service. SN: 5545, <http://doi.org/10.5255/UKDA-SN-5545-7>
- University College London, UCL Institute of Education, Centre for Longitudinal Studies, Department for Education. (2021). *Millennium Cohort Study: Linked Education Administrative Datasets (National Pupil Database), England: Secure Access*. [data collection]. UK Data Service. SN: 8481, <http://doi.org/10.5255/UKDA-SN-8481-1>
- University of London, Institute of Education, Centre for Longitudinal Studies. (2020). *Millennium Cohort Study: Longitudinal Family File, 2001-2018*. [data collection]. 4th Edition. UK Data Service. SN: 8172, <http://doi.org/10.5255/UKDA-SN-8172-4>

- University of London, Institute of Education, Centre for Longitudinal Studies. (2020). *Millennium Cohort Study: Sixth Survey, 2015*. [data collection]. 7th Edition. UK Data Service. SN: 8156, <http://doi.org/10.5255/UKDA-SN-8156-7>
- University of London, Institute of Education, Centre for Longitudinal Studies. (2021). *Millennium Cohort Study: Fifth Survey, 2012*. [data collection]. 5th Edition. UK Data Service. SN: 7464, <http://doi.org/10.5255/UKDA-SN-7464-5>
- University of London, Institute of Education, Centre for Longitudinal Studies. (2021). *Millennium Cohort Study: First Survey, 2001-2003*. [data collection]. 13th Edition. UK Data Service. SN: 4683, <http://doi.org/10.5255/UKDA-SN-4683-5>
- University of London, Institute of Education, Centre for Longitudinal Studies. (2021). *Millennium Cohort Study: Fourth Survey, 2008*. [data collection]. 8th Edition. UK Data Service. SN: 6411, <http://doi.org/10.5255/UKDA-SN-6411-8>
- University of London, Institute of Education, Centre for Longitudinal Studies. (2021). *Millennium Cohort Study: Second Survey, 2003-2005*. [data collection]. 10th Edition. UK Data Service. SN: 5350, <http://doi.org/10.5255/UKDA-SN-5350-5>
- University of London, Institute of Education, Centre for Longitudinal Studies. (2021). *Millennium Cohort Study: Seventh Survey, 2018*. [data collection]. 2nd Edition. UK Data Service. SN: 8682, <http://doi.org/10.5255/UKDA-SN-8682-2>
- University of London, Institute of Education, Centre for Longitudinal Studies. (2021). *Millennium Cohort Study: Third Survey, 2006*. [data collection]. 8th Edition. UK Data Service. SN: 5795, <http://doi.org/10.5255/UKDA-SN-5795-5>
- Vitello, S. and Crawford, C. (2018) Which tier? Effects of linear assessment and student characteristics on GCSE entry decisions, *British Educational Research Journal*, 44 (1), 94–188. [10.1002/berj.3320](https://doi.org/10.1002/berj.3320).
- Young, M. (2020) *How I'm coping with exams being cancelled*. <https://youngminds.org.uk/blog/how-im-coping-with-exams-being-cancelled> (Accessed 21 June 2021).

*Correspondence John Jerrim UCL Social Research Institute, University College London 20 Bedford Way, London, WC1H 0AL, UK
Email: J.Jerrim@ucl.ac.uk*