Prediction or Prejudice? Standardized Testing and University Access

Nagisa Tadjfar Kartik Vira

(Job Market Paper)

November 2025

Click here for latest version

Abstract

Do high-stakes standardized tests expand or inhibit opportunity for low-SES students? We answer this question in the context of the UK's staggered elimination of pre-university exams in favor of teachers' predicted exam grades. Eliminating testing increases the university enrollment of low-income students by 3 percentage points (7%), while leaving wealthy students' enrollment unchanged. Marginal students induced to enroll in university attain employment at better firms and, in expectation, earn £50,000—£100,000 more over their careers, in net present value. Paradoxically, standardized exams exhibit no calibration bias against marginal low-income students—accurately predicting their university success—whereas teacher-supplied grades are systematically biased in their favor. Despite proper calibration, standardized tests inhibit low-SES students by deterring human capital investment. When tests are eliminated, 5% of low-income students shift into academic tracks. These findings highlight how disparate impacts can arise even when screening algorithms are unbiased. When the measurement of information itself poses a direct disutility, standardized tests generate disparities that commence earlier in the pipeline.

Tadjfar: Massachusetts Institute of Technology, ntadjfar@mit.edu; Vira: Massachusetts Institute of Technology, kvira@mit.edu. We are deeply grateful to David Autor, Jonathan Gruber, Nathan Hendren, and Frank Schilbach for their guidance throughout this project. We thank Daron Acemoglu, Abigail Adams-Prassl, Isaiah Andrews, Josh Angrist, Peter Arcidiacono, Esteban Aucejo, Zach Bleemer, Amy Finkelstein, Lindsey Macmillan, Parag Pathak, Nina Roussille, Gill Wyness, and Basit Zafar for very helpful comments. We received constructive comments from numerous participants at the MIT Behavioral, Labor, and Public Finance Lunches and the CEP Education Conference at LSE. We are also grateful to Nick May, Mike Tucker, and Jody Pink from Office for National Statistics for helping us navigate the Saferoom. We also thank David Burnett and Philippa Norgrove at the Department for Education (DfE) and Ben Jordan at the Universities and Colleges Admissions Service (UCAS) for their insightful comments. We gratefully acknowledge the generous financial support from the George and Obie Shultz Fund at MIT, the Horowitz Foundation for Social Policy, the American Institute for Boys and Men, the Jerry A. Hausman Graduate Dissertation Fellowship, the NBER PhD Fellowship for Identifying and Developing Mathematical Talent among Youth, and the National Science Foundation Graduate Research Fellowship.

I. Introduction

Standardized tests face widespread criticism for systematically disadvantaging low-SES students. In the last decade, institutions across numerous countries and contexts have scaled back testing requirements. This broad movement away from standardized exams accelerated during the 2020 pandemic, when many US universities cemented test-optional admissions policies for the foreseeable future. More recently in 2024, Massachusetts voted to end the requirement that students pass the Massachusetts Comprehensive Assessment System (MCAS) to graduate from public high schools, with teachers arguing that the MCAS "unfairly punishes students with disabilities or fewer financial resources" (Creamer, 2024). Yet proponents of standardized testing maintain that such assessments are important predictors of academic success, and concerns about grade inflation in teacher-assigned grades have prompted several US universities to reinstate SAT requirements. Despite this growing debate, there is little empirical evidence on whether standardized testing inhibits opportunity for low-income students.

Assessing the impact of standardized testing on opportunity is challenging for three reasons. First, testing is often optional, and lower-ability students may select out of standardized tests, confounding efforts to isolate the role of testing. Second, tracing the impact of testing requirements for higher education requires data linking high school, university, and downstream labor market outcomes, a linkage which rarely exists. Finally, a full assessment of standardized exams as a potential impediment to opportunity requires an empirical test for whether any such effects run through calibration bias from the exam itself. A test for calibration bias in screening algorithms is difficult to implement in practice because university performance is not observed for individuals who are *screened out* of university. Moreover, comparing average university performance across demographic groups yields biased estimates when underlying ability distributions differ across groups (Arnold et al., 2018).

This paper overcomes these challenges to estimate the labor market consequences of eliminating pre-university testing requirements and empirically test whether standardized exams are calibrated for low-income students. First, we use quasi-random variation in pre-university testing requirements resulting from the staggered rollout of a UK educational reform. In the final two years of high school, students in England, Wales, and Northern Ireland choose between two tracks: A-level classes, which are academic classes akin to Advanced Placement (AP) classes in the US, and coursework-based vocational classes. Prior to the reform, all students taking A-level classes were required to take high-stakes standardized exams in each of their A-level subjects in addition to an endline exam after receiving university offers. Between 2015–2017, the UK government rolled back pre-university testing require-

ments for students in England. This elimination was staggered across subjects and did not apply to students in Wales or Northern Ireland. Many schools in England stopped offering the newly optional pre-application exams altogether, and virtually all schools in England discontinued the exams by 2018. Our empirical methodology therefore relies on three sources of variation in standardized testing requirements available in this context: geographic variation, variation in timing across subjects, and variation in timing across schools within England. Second, we draw on comprehensive administrative data, which links individual-level data on high school records from the Department for Education (DfE); university applications from the Universities and Colleges Application Service (UCAS); university attendance, graduation, and performance from the Higher Education Statistics Agency (HESA); and earnings and employment data from tax records (Office for National Statistics, 2023). This allows us to examine how reduced testing affects university decisions for low-income students and to estimate the returns to higher education for students on the margin of university enrollment. Third, we use the reform as an instrument for university attendance and measure calibration bias *at the margin* to directly test whether standardized exams are systematically biased for or against low-income students.

We begin by documenting how this switch from standardized tests to teacher grades affects academic sorting and university access. Bundling standardized tests with academic classes, as in the case of A-levels, can shape students' educational choices even *before* the decision to apply to university. We show that the elimination of tests shifted students, particularly low-income students, out of the vocational track and into the academic A-level track. Students in the bottom two income quintiles are 5-10pp more likely to take A-levels after the reform and 3-5pp more likely to take STEM A-levels. By contrast, students in the top two income quintiles are no more likely to take A-levels as a result of the reform but are more likely to take STEM A-levels. Consistent with negative selection, we find evidence that students with lower prior test scores opt into A-levels and STEM subjects in the absence of tests. A-level students at treated schools are also 5-6pp more likely to fail an A-level after the reform, suggesting that this selection led to worse average academic performance.

Next, we show that these changes in high school subject choices carry through to university outcomes. Low-income students at treated schools are 5-10pp more likely to apply to university and 3-5pp more likely to attend university compared to low-income students at control schools. High-income students become slightly more likely to apply to university, but do not see a statistically significant uptick in university enrollment. These increases in university enrollment are driven by students who were *previously unlikely* to take A-levels. We estimate a logit model using a sample of students before the reform and apply it to predict students' propensities to take A-levels after the reform. We find that increases in university

application and enrollment are driven by students with below-median predicted propensity of taking A-levels, suggesting that the path dependency between high school coursework and university attendance is a key mechanism for the rise in university enrollment.

Consistent with the negative selection into more ambitious high school coursework, we find that the marginal low-income university entrant is substantially academically weaker than the average low-income university student. These marginal entrants also primarily attend academically non-selective universities. Low-income students shifted into university are twice as likely to have a below-median prior math score compared to the average low-income university student. These marginal entrants also drop out of university at higher rates and are less likely to complete a degree on time. Students from treated schools that eliminated pre-university tests are 2pp less likely to graduate within 3 years and 1-3pp more likely to drop out conditional on attending university. Students from these schools also graduate at a lower position within their university cohort, as measured by 'degree classes'. These patterns are in line with previous findings that students induced into more ambitious coursework and university enrollment are academically worse off (Arcidiacono et al., 2011; Riehl, 2023). Eliminating pre-university testing brings more low-income students into universities, as critics of standardized testing predict. However, these marginal students have lower university performance compared to the typical low-income university student.

Despite these poor academic outcomes, we find substantial improvements in early-career labor market outcomes and lifetime earnings for marginal students. This shift into university also results in a net increase in university degrees among low-income students. The probability of obtaining a degree within 3 years, the standard undergraduate duration in the UK, increases by 1pp among low-income students, a 5% increase relative to the group mean. Three years after high school, low-income students are 1pp more likely to be employed and 2pp less likely to receive out-of-work benefits. Crucially, these students begin their careers at *better* firms—low-income students at treated schools are less likely to work at non-degree firms or low-paying firms.

We characterize the marginal potential outcome distributions for firm characteristics and demonstrate that these improvements in average early-career outcomes for low-income students reflect positive gains *across the distribution*. If a bimodal distribution of outcomes underlies these positive average effects, a handful of students may be substantially worse off as a result of the increase in attempts at more challenging academic pathways. Instead, we find that low-income students shifted into university have a near-zero probability of working at a firm in the bottom three deciles of firm characteristics, as measured by the average education level and earnings of previously hired employees. Counterfactually, these students have around a 40% chance of working at these low-paying firms. This change in career

trajectory is reflected in higher projected earnings. Net of tuition fees, marginal low-income university students gain between £50,000 to £100,000 in lifetime earnings.

These large private gains for marginal university entrants, however, are not the result of eliminating a biased or otherwise uninformative test. Comparing the informativeness and calibration of teacher grades and test scores in predicting university performance, we document that tests explain a much higher share of university performance compared to teacher grades, especially for low-income students. In pre-reform cohorts, adjusted R^2 values for regressions using test scores explain 21% more of the total variation in university graduation rates and 40% more for graduating in the top fifth of the university cohort. We find smaller increases in explanatory power for high-income students. Moreover, teacher grades add very little explanatory power in regressions that already include test scores. Standardized tests are also more predictive of out-of-sample university performance compared to teacher-assigned grades, and this gain in precision is larger for low-income students.

We then empirically test whether standardized exams in our context are differentially miscalibrated against low-income students. First, we evaluate calibration among university enrollees using cross-fitted models trained on test scores. Among enrollees, predicted university performance based on test scores is calibrated overall, with no economically meaningful differential miscalibration by income. However, this comparison does not capture population-level calibration due to selective labels (Lakkaraju et al., 2017). If the relationship between test scores and university performance is different for the students who are screened out, examining calibration among the set of students admitted to university is insufficient to draw conclusions about calibration for students on the margin.

One solution to overcome this challenge of selective labeling is to instrument for university attendance. Using the reform as an instrument for university attendance among low-income students, we are able to measure calibration bias *at the margin*. We estimate the average complier university performance and the complier mean exam-predicted performance. Concretely, we test for whether the marginal mean performance among low-income compliers is statistically distinguishable from the prediction generated by test scores and find that they are *not* distinguishable. Thus, standardized tests in our context are neither miscalibrated nor differentially calibrated for low-income students who are marginal to university enrollment. If anything, these students slightly *underperform* relative to the calibration for inframarginal students. Moreover, the complier mean university performance is significantly worse than the mean predicted performance using teacher grades. Our results suggest that the reduced testing environment did not eliminate miscalibration or bias, and if anything, switched to a more miscalibrated assessment system that *favors* marginal low-income students.

We conclude with a discussion of the efficiency of using standardized testing as a university screening tool. Although reduced testing lowered conditional graduation rates, the policy also resulted in improvements in early-career outcomes for marginal entrants. Due to the recency of the reform, we are unable to directly measure long-run earnings and employment outcomes, but in the short term, we observe gains for low-income students that come at no cost to high-income students. Conservative estimates for projected lifetime earnings gains suggest private returns of £50,000 for low-income students shifted into university. Importantly, these gains to disadvantaged students do not come at the expense of another group—high-income students have null effects in both average earnings at age 21 and projected earnings. Because aggregate undergraduate enrollment increased during our study period, observed increases in university attendance did not necessarily displace other students. These private returns to marginal university entrants also *exceed* the average per-student cost to government for a university degree, implying positive social returns.

Our results highlight one potentially overlooked downside of pre-university testing: the direct disutility of these screening algorithms for students on the margin of attending university. Students, especially low-income students, may inefficiently sort into high school classes and underapply to university in the presence of standardized testing. Under these conditions, eliminating pre-university exams can enhance efficiency from a social planner's standpoint. Moreover, these efficiency gains exist even when standardized exams are unbiased and therefore accurately capture the potential academic performance for disadvantaged students. In our context, reduced standardized testing promotes socioeconomic mobility among low-income students by *increasing attempts* into more challenging academic pathways with substantial private and social returns.

Contribution to the Literature. This paper makes three main contributions. First, while recent studies have examined the consequences of test-optional policies at US universities, these studies typically focus on the direct effects at a handful of institutions (Belasco et al., 2015; Saboe and Terrizzi, 2019; Bennett, 2022) or consist of more theoretical work predicting the consequences of eliminating the SAT (Borghesan, 2023; Dessein et al., 2025). More recently, Avery et al. (2025) examine test-optional policies and their effect on applications and enrollment across universities in the US. A related strand of literature documents strategic behavioral responses to grading policy changes and socioeconomic heterogeneity in this response (Astorne-Figari and Speer, 2019; McEwan et al., 2021; Exley et al., 2024; Ahn et al., 2024; Sacerdote et al., 2025; Aucejo et al., 2025). Our paper uses quasi-experimental variation and longitudinal administrative data to empirically estimate the aggregate impact of switching from standardized tests to teacher-assigned grades. We

extend this literature by directly estimating effects on university applications, enrollment, degree attainment, and labor market outcomes.

Second, our study contributes to the body of literature examining the disparate impacts of screening algorithms (Kleinberg et al., 2016, 2018; Arnold et al., 2018, 2022) and the role of standardized testing in predicting university performance (Rothstein, 2004; Geiser and Santelices, 2007; Allensworth and Clark, 2020; Chetty et al., 2023; Friedman et al., 2025). We also relate to the literature on teacher stereotyping and discrimination (Carlana, 2019; Carlana et al., 2022) and relative demographic bias between human judgments and algorithms (Autor and Scarborough, 2008; Angelova et al., 2025). We examine calibration bias in both pre-university tests and teacher assessments for low-income students using quasi-experimental variation in university attendance. Our paper highlights the importance of considering both bias and the direct disutility of screening algorithms when examining the potential for disparate impacts.

Finally, our paper contributes to the vast literature examining academic mismatch and student success, particularly among students on the margin of university enrollment. Prior studies on mismatch and returns to university have largely focused on the impacts of affirmative action policies, top-percent admissions plans, or financial aid expansions (Altonji, 1993; Zimmerman, 2014; Arcidiacono and Lovenheim, 2016; Arcidiacono et al., 2016; Dillon and Smith, 2020; Bertrand et al., 2021; Angrist et al., 2022; Bleemer, 2022; Black et al., 2023; Mountjoy, 2025). However, there is limited evidence on mismatch arising from reductions in standardized testing itself. Closest to our context, Riehl (2023) examines a Colombian reform that increased grading coarseness on national exams, finding that low-income students entered more selective universities but subsequently experienced reduced earnings due to elevated dropout rates. We bridge the literature on academic mismatch and returns to university with the more nascent literature examining test-optional policies and algorithmic bias. Our findings expand on this literature by directly documenting improved outcomes for disadvantaged students when standardized testing requirements are eliminated, despite reduced screening precision.

II. Background

The UK educational system is a particularly attractive setting to study the effects of standardized testing on university access and opportunity. In contrast to the US where students face large heterogeneity and uncertainty in tuition and financial aid, UK students face uniform tuition fees of around $12,000 per year and have access to standardized government loans. Beyond uniform prices, both secondary and higher education are highly standard-

ized and centralized. Students in England, Wales, and Northern Ireland follow a standardized national curriculum in school between ages 5-16, and students apply to universities through a centralized application platform called Universities and Colleges Admissions Service (UCAS).[1] These unique institutional features combined with rich administrative data allow us to isolate the educational and labor market effects of standardized testing.

II.A *Secondary education in the UK*

At the end of Year 11, typically at age 16, students take the General Certificate of Secondary Education (GCSE) exams.[2] After completing GCSEs, students in our sample period (academic years 2011-12 through 2018-19) predominantly chose between two types of classes to pursue in their last two years of high school during Year 12 and Year 13: A-Level classes and Business and Technology Education Council classes (BTECs). A-Levels are two-year academic classes akin to Advanced Placement (AP) courses in the United States and culminate in externally assessed, subject-specific standardized exams. BTECs are vocational and course-work based. Taking A-levels in three or more subjects is the standard entry requirement for most UK universities, although some BTEC students apply to and attend non-selective universities.[3] Throughout this paper, we will refer to taking three or more A-levels as "taking A-levels" and focus our analysis on this outcome as the university-relevant educational decision.

All A-level subjects consist of two rounds of standardized exams graded by an external agency. Halfway through the class at the end of Year 12, students take the first set of exams, called "AS-levels". Students then take a second round of exams, called "A2-levels", in June after Year 13. These second round of exams take place *after* prospective students apply to university and receive offers. Because of this pre-qualification admissions system, university applicants additionally receive teacher-assigned grades, called "predicted grades", in each A-level subject. Students applying to universities send teacher-assigned grades, GCSE grades, and AS-level grades as part of their application. The A-level reform, described in Section II.B, eliminated the requirement for the first set of pre-application exams.

Unlike many OECD countries, the UK secondary school system is not tracked. In 2012, 99% of secondary schools in England offered A-level courses, so students are typically

[1] Scotland uses a different exam system from the rest of the UK, and Scottish students are excluded from the analysis in this paper as a result.

[2] GCSEs in Mathematics, English and two science GCSEs are compulsory, and students additionally take GCSEs in other specific subjects such as history, foreign languages, art, or business studies.

[3] A small number of academically ambitious students take four or five A-levels. Around 16% of students entering university between 2014 through 2017 had only BTECs, while another 7% entered with a combination of both A-levels and BTECs (Dilnot et al., 2023).

able to choose between the academic and vocational courses regardless of the school they attend (Gill, 2013). Moreover, almost all secondary schools offer A-levels in major subjects such as Mathematics and Biology. As a result, the vast majority of students in principle have access to these advanced courses, unlike in the United States, where less than half of public high schools offer Calculus (US Department of Education, 2024).

Despite this wide availability of A-level classes, there is substantial socioeconomic heterogeneity in A-level take-up, which is highly correlated with university attendance. Table 1 shows that only 30% of students who completed GCSEs between 2009-2013 went on to take A-levels during their final two years of high school. Students in the top two income quintiles are twice as likely to take three or more A-levels compared to students in the bottom two income quintiles. Students who take A-levels are also significantly more likely to apply to and attend university compared to vocational students. While 77% of A-level students obtain an undergraduate degree within 3 years of graduating high school, only 17% of students who did not take A-levels do so. Although students are not explicitly tracked in the UK, these patterns suggest that the decision to take A-levels is strongly associated with downstream university attendance and outcomes.

II.B *Staggered reduction in testing requirements*

In an effort to reduce "teaching to the test", Education Secretary Gove eliminated the requirement that A-level students take AS-level exams before applying to university.[4] This reform first took effect for the A-level cohort who started A-level classes in academic year 2015-16 and therefore completed them in 2016-17. The elimination of mandatory AS-levels was highly unpopular among universities. Most universities used these exams to make admission decisions and argued that these exams had higher predictive validity compared to teacher-assigned grades (Partington, 2011). AS-level requirements were lifted in England, but not in Wales and Northern Ireland, where AS-levels remained a mandatory component of A-levels in all subjects.[5] The reform was also staggered across subjects, and Table C1 shows the timing of the reform for each A-level subject, split into three tranches of 2017, 2018, and 2019. This gradual roll-out across subjects was due to a government-announced delay and generated logistical hurdles for schools, with many schools expressing frustrations around "knowing what the standard will be to be in line with national trends" and "confusion among parents, staff, and students" (UCAS, 2015).

[4]This reform was introduced in 2013 under the Conservative Party. The Labour party pledged to reverse the "policy that undermines both rigour and equity" if they won the elections in May 2015.

[5]In Northern Ireland and Wales, the weight of AS-levels in the final A-level grade was reduced from 50% to 40%.

Amid this uncertainty, many schools in England rapidly stopped offering AS-level exams in reformed subjects, while other schools continued to require AS-level exams for all A-level students in the early years of the reform. This resulted in substantial heterogeneity in school policies around offering the newly-optional AS-level exams and rapidly shifting policies within schools (UCAS, 2015, 2016, 2017, 2018), documented by a series of surveys UCAS conducted with schools between 2015 to 2018. Initially, only 14% of schools surveyed indicated they were no longer offering AS-level exams in reformed subjects in academic year 2015-16, but this number rose to 21% over the course of the year. Of the remaining schools, 59% continued offering AS-level exams in *all* reformed subjects. The number of schools offering AS-levels continued to decline in subsequent academic years. Although the policy was intended to maintain AS-levels as standalone optional qualifications, by 2020, nearly all schools in England had stopped offering AS-level examinations. This inadvertent retirement was dubbed the "death of the AS-levels" by a 2018 news article in The London Standard.[6] Thus, even schools that initially maintained the status quo by requiring all students to take AS-levels stopped offering them entirely within a year or two, citing a variety of factors such as teaching time and policies of other schools in the area as drivers of their decision (UCAS, 2017):

> *"We are phasing out the AS as subjects become reformed, moving to internal end of year exams, to gain more teaching time."* – School 1, UCAS 2017 Survey

> *"We offered AS qualifications for most linear subjects for 2016 exams, but decided against doing so for 2017 due to a variety of factors and an overview of what other colleges were doing locally."* – School 2, UCAS 2017 Survey

Figure C1 shows that the number of schools requiring the pre-application exams in reformed subjects declined rapidly as the reform took effect. In the first year of the reform, around half of all schools in England still required all A-level students taking 2017-reformed subjects to take the pre-application exams. By 2020, nearly all schools in England no longer required the exams in reformed subjects. This change in school policies is reflected in the declining share of A-level students who take the pre-application AS-level exams. Panel A in Figure A1 shows that prior to 2017, around the same number of students take A-levels and AS-levels in 2017-reformed subjects. After 2017, the gap between the number of A-level students and AS-level students steadily grows, and by 2020, the vast majority of A-level students *do not* take AS-levels. Panels B and C in Figure A1 show similar patterns for subjects reformed

[6] https://www.standard.co.uk/news/education/aslevel-killed-off-by-alevel-reforms-as-take-up-falls-by-52-per-cent-a3913146.html; date accessed July 18, 2025.

in 2018 and 2019, respectively. The resulting variation of AS-level policies across schools in England equips us with another source of variation and is central to the empirical strategy used in this paper.

III. Data

To study the long-run impact of reduced pre-university testing in the UK on educational and labor market outcomes, we link five UK administrative datasets using the Longitudinal Education Outcomes database developed by the Department for Education (Office for National Statistics, 2023). These data are accessed via the Secure Research Service (SRS) of the Office for National Statistics. Due to the highly sensitive nature, these datasets are accessible only in person at SRS-designated Saferooms or Safepods located in the UK. We restrict our analysis to students who are 18 years of age when finishing high school between 2012 and 2019, which correspond to academic years 2011-12 through 2018-2019 respectively.[7] Figure B1 summarizes our data linkages and Appendix B provides further information on each data sources and data cleaning procedures and sample restrictions.

Our primary dataset is the National Pupil Database (NPD), with detailed educational attainment data throughout high school for students in England. We then merge our sample from the NPD to the Universities and Colleges Admissions Service (UCAS), which includes the universe of students from England who applied to UK universities. The Higher Education Statistics Agency (HESA) dataset is a student-year level dataset on the universe of students at higher education providers in the UK. Our HESA sample begins in the academic year 2006-2007 through 2019-2020, and includes students matriculating from high schools in England, Wales, Northern Ireland and Scotland. For each student, we observe their degree program (university, major, degree type), the level of program (e.g., undergraduate or graduate), program start date, completion date, and whether the student obtained a qualification associated with that program (i.e., degree). His Majesty's Revenue and Customs (HMRC) is an individual-year level dataset with data on earnings, and anonymized firm ID and Standard Industrial Classification (SIC2007) codes for primary firm of employment, spanning the tax years 2003-04 through 2020-21.[8] Finally, we link to Department for Work and Pensions (DWP) data that contain information on out-of-work benefits that individuals in England receive each tax year, spanning from 1999-2000 through 2020-21.

[7] We exclude 2020 because students did not sit the actual exams in the 2020-2021 academic year due to Covid-19, and instead received grades from a combination of algorithmic and teacher input. We begin our analysis in the academic year 2011-12 because in 2012, tuition fees for UK universities were raised from £3,000 per year to £9,000 per year, which led to an immediate decline in university enrollment.

[8] We winsorized all earnings at the 99% and 1% levels within a tax year and inflation adjusted to 2015 GBP using the CPIH index.

III.A *Construction of the school-year treatment variable*

Surveys conducted by UCAS and Ofqual document the overall shares of schools that stopped offering pre-application exams at various points in time, but we do not directly observe these school policies in our data. We therefore infer school policies from the share of A-level students in reformed subjects for whom we observe pre-application exams. Many schools adopted 'mixed' policies, such as requiring AS-levels in some reformed subjects or allowing students to choose whether they take AS-levels. We exclude these 'mixed-policy' schools from our main analysis and instead focus on policy 'switchers'—schools that went from *requiring all* students to take AS-levels in year $t-1$ to *no longer offering* AS-levels in year t.

We classify schools based on when they stopped offering the exams (i.e., 'switched'): schools that switched in academic year 2016-17, schools that switched in 2017-18, schools that switched in 2018-19, and control schools.[9] Mechanically, all schools that stopped offering AS-levels in A-level qualification year 2017 (and hence AS-level qualification year 2016), are switchers, since AS-levels were mandatory across the UK prior to 2017. We then identify schools that stopped offering AS-levels as follows. If less than 1% of the subject-student pairs in reformed A-levels are observed to have an associated AS-level exam beginning in qualification year t, the school is classified as no longer offering AS-levels starting in academic year t. These schools are then classified as switchers in year t, if more than 90% of the subject-student pairs in reformed A-levels were observed to have an associated AS-level in year $t-1$. Schools are classified as control schools if they required AS-levels through 2019, i.e., if over 90% of the subject-student pairs in reformed A-levels have an associated AS-level in all years between 2017 through 2019. Schools in Northern Ireland and Wales are also classified as control schools.

Figure C2 visualizes these two sources of variation, plotting in each panel the share of A-level students taking subjects reformed in qualification years 2017, 2018, and 2019, respectively, who additionally take AS-levels by year. Each panel in turn plots those shares separately by the year the school switched policies. In Panel A, we see that the share of students taking AS-levels in 2017-reformed subjects drops sharply at 2017. Similarly, Panel B shows a steep drop in the share of AS-level takers for both 2017-reformed subjects and 2018-reformed subjects. For schools that switched in 2019, Panel C shows a sharp drop in the share of AS-level takers across *all reformed* subjects. Panel D shows that the share of AS-level takers across all three tranches of subjects are flat prior to 2020.

[9] Of the 3,335 schools that existed throughout 2011-2019 and had at least five students per year taking A-levels, 560 schools are classified as switching in 2017, 390 in 2018, and 115 in 2019. Another 165 are classified as control schools that required all their A-level students to continue taking AS-levels throughout the period.

IV. Effects on Educational Outcomes

Reduced standardized testing requirements may impact educational choices and outcomes in several ways. If tests are miscalibrated for certain demographic groups and inaccurately predict the academic potential for low-SES students, then eliminating testing requirements changes the demographic composition of screened-out students. Thus, university enrollment among low-SES students may increase. Without standardized tests, students no longer receive a potentially miscalibrated signal *prior* to the decision of whether to apply to university. If tests, but not teacher grades, are biased against low-SES students, then university applications will also increase among disadvantaged students. When exams are bundled with academic classes, as in our context, the elimination of testing requirements can also lead to changes in educational choices earlier in the pipeline. Eliminating a high-stakes exam reduces the effort cost associated with taking academic-track classes. All else equal, this reduction in effort cost lowers the threshold of prior beliefs about academic ability at which a student would choose to do A-levels. Thus, removing pre-application testing requirements may also bring more students into the academic track.

To estimate the impact of reduced standardized testing on student decisions and outcomes at each of these stages, we exploit two sources of variation: school-level event studies and subject-level event studies. We estimate a staggered event-study specification at the school-level comparing schools in England that stopped offering AS-levels in reformed subjects to schools that still required *all* of their students to take AS-levels in reformed subjects. By focusing on these 'policy switchers', we estimate the dynamic treatment effects of a binary treatment, where a treated school is a school where *no student* takes the pre-application AS-level exams while a control school is one where *all students* take the exams. We estimate effects on outcomes such as A-level take-up, university application, and university enrollment for student i attending school $s(i)$:

$$Y_i = \alpha_{s(i)} + \gamma_{t(i)} + \sum_{\tau=-5}^{2} \theta_\tau I(t(i) - T_{s(i)} = \tau) + \varepsilon_i \tag{1}$$

where $\alpha_{s(i)}$ and $\gamma_{t(i)}$ are school and year fixed effects respectively, and $T_{s(i)}$ is the treatment year, which is the academic year school $s(i)$ stopped offering AS-levels in reformed subjects and is one of 2017, 2018, or 2019. Our never-treated control group includes English schools that continued requiring AS-level exams through 2019 and schools in Northern Ireland and Wales, where data is available. We estimate results from Equation 1 using the IW estimator developed by Sun and Abraham (2021) to allow for heterogeneous treatment effects across adoption years. To document heterogeneity by socioeconomic status, we estimate event

studies separately by low- and high-income students. Throughout the paper, we define low-income students as those in the bottom two quintiles of the Index of Multiple Deprivation (IMD), and high-income students as those in the top two quintiles.[10]

Our empirical strategy relies on the assumption that the exact year between 2017 through 2019 that each school stopped offering AS-levels was arbitrary and uncorrelated with student outcomes conditional on school and academic year fixed effects. We believe that this assumption is reasonable in our context for the following reasons. First, we observe no statistically significant pre-trends in academic outcomes or student composition. Second, schools that stopped offering AS-levels between 2017-2019 are very similar prior to the reform. Table C2 presents summary statistics for schools that stopped offering AS-levels by A-level cohort and schools that continued to require reformed AS-levels throughout this period. Third, surveys of schools conducted by Ofqual and UCAS during this period document that schools changed their exam policies due to a variety of reasons and within three years nearly all schools converged to the same policy of no longer offering the exams. Appendix C provides more details about the reform.

IV.A *Shifts into academic-track classes and changes in subject choice*

Figure 2 presents estimates from Equation 1 on the takeup of academic-track classes and STEM classes. Panels A and B report effects on takeup among all students, while Panels C and D report effects on takeup separately by low-income and high-income students. When schools stop offering pre-application AS-level exams, takeup of academic-track classes increases by around 3pp compared to students at schools still requiring AS-level exams. This increase is driven entirely by low-income students. Low-income students in treated schools become 5pp more likely to take A-levels compared to low-income students at control schools. Although there is a visually small uptick among high-income students, the effects are statistically indistinguishable from zero. Effect sizes appear to increase over time, almost doubling for A-level take-up two years after the reform, although estimates are noisier. This may reflect the fact that by academic year 2018-19, all A-level subjects were reformed, effectively removing AS-level exams entirely for students attending treated schools.

In addition to this shift into academic-track A-levels, the policy induced a relative shift across subjects within A-levels.[11] Takeup of STEM A-level classes increases by 3pp

[10] IMD is a UK-wide composite measure of neighborhood deprivation, defined at the Lower Layer Super Output Area (LSOA) level, which is roughly comparable to the US Census Block Group in geographic size and population.

[11] We restrict our definition of STEM A-levels to be Biology, Chemistry, and Physics because these were the subjects reformed in 2017. Mathematics and Further Mathematics were reformed in 2019 and we therefore only observe one post-reform period for those subjects.

after schools remove AS-levels, and this shift into STEM is visible for *both* low- and high-income students. Estimates at $t = 2$ are noisier because we only observe treatment effects at $t = 2$ for schools that stopped offering AS-levels in academic year 2016-17.

IV.B *Negative selection into academic-track classes*

If students' priors on ability are monotonic in GCSE scores, students shifted into A-levels and STEM would have lower GCSE scores. Figure 3 shows the change in academic composition as measured by the average GCSE percentile in mathematics and English among students taking A-levels (panels A and C) and STEM A-levels (panels B and D). We find that students taking A-levels in treated schools after the reform have average GCSE math scores and GCSE English scores that are 1-2 percentiles lower. We find a similar negative compositional shift among students taking STEM A-levels, albeit with larger effects on their GCSE math scores. Students taking STEM A-levels at treated schools have GCSE math scores that are 2-3 percentiles lower on average and have English GCSE scores that are 1-2 percentiles lower.

We also consider how this negative selection can translate into reduced academic performance in A-levels. In this second set of outcomes, we define treatment at the subject-school level, leveraging the staggered roll-out of the policy across subjects. For student i attending school $s(i)$ taking A-level subject $j(i)$, we estimate using the estimator developed by Sun and Abraham (2021):

$$Y_{i,j(i)} = \alpha_{s(i)j(i)} + \gamma_{t(i)} + \sum_{\tau=-5}^{2} \theta_\tau I(t(i) - T_{s(i)j(i)} = \tau) + \varepsilon_{i,j(i)} \tag{2}$$

where $\alpha_{s(i)j(i)}$ and $\gamma_{t(i)}$ are school-subject and year fixed effects respectively, and $T_{s(i)j(i)}$ is the treatment year. The treatment year now incorporates both the school-level adoption year and the subject-level variation in reform year. This specification is used to estimate student-subject level regressions on outcomes such as failing the A-level exam, getting an A or A* in the A-level, and having a teacher-assigned grade that was a different letter grade than the final grade achieved in the A-level examination.

Student performance on final A-level exams also declines. Figure D2 presents estimates on A-level performance from Equation 2, using school-subject level variation. Panel A in Figure D2 shows that students taking A-levels at treated schools are 2pp more likely to *fail* an A-level after the reform. In Panel B, we present these results separately by STEM subjects and non-STEM subjects, and find that the increase in failure is substantially larger among STEM subjects, with an increase of around 3pp while in non-STEM subjects the increase is round 1pp. We also find that students are less likely to obtain top scores in

these A-levels by considering the outcome that the student obtains an A* or an A shown in Panels C and D in Figure D2. Again, the decline in academic performance appears larger among STEM subjects. In Figure D3, we present estimates that control for GCSE scores and demographics and find that around half of the decline in receiving a top grade (A* or A) can be explained by selection on observables, while almost none of the increase in failure is explained. These results suggest that other factors, such as private information on ability, reduced motivation to study for the A-level exam, or reduced learning as a result of the removal of AS-levels may explain part of the decline in performance.

IV.C *Shifts into university*

Next, we examine how university applications and enrollments respond. Figure 4 presents estimates of Equation 1 on whether or not students apply to university at age 18 (Panel A) and whether they enroll in university at age 18 (Panel B). We find that low-income students are around 5pp more likely to apply to university and attend university. This increase is very similar in magnitude to the increase in A-level take-up among this group. On the other hand, high-income students are 3pp more likely to apply to university, but do not see a statistically significant uptick in university enrollment.

Conceptually, this increase in university applications and enrollment may be due to the increase in the university-eligible population that resulted from an increase in A-level take up, but may also partly reflect the change in signal structure that changes which students choose to apply and which applicants are screened out of university. To disentangle these mechanisms, we examine heterogeneity by students' ex-ante likelihood of taking academic-track classes. We develop a measure of students' predicted propensity to pursue A-levels based on their observed characteristics in the pre-reform period. Specifically, we estimate a logit model on the binary outcome of whether or not a student takes A-levels, using a set of predictors including prior test scores and demographics. Our approach is motivated by the prediction model in (Black et al., 2023). To mitigate concerns about overfitting bias resulting from endogenous stratification (Abadie et al., 2018), on a training sample and then use those estimates to predict the propensity to take academic-track classes in the remaining holdout sample. The training sample is a randomly selected sample of all GCSE students who completed KS4 between 2009–10 and 2013–14 and constitutes about one-third of the original sample (approximately 960,000 students out of 3 million). We exclude the training sample from all event study analyses using the predicted propensity measure. We estimate the following model on the training sample using the Broyden, Fletcher, Goldfarb, and Shanno (BFGS) algorithm:

$$\Pr\big(Y_i = 1\big) \;=\; \frac{1}{1 + e^{-(\alpha + \beta X_i + \gamma_f \mathbf{1}\{\text{Female}_i=1\} + \gamma_{eth(i)})}} \tag{3}$$

where Y_i is an indicator for whether the student takes three or more A-levels, X_i are student attributes including fully saturated interactions between GCSE Math score deciles $\times$ income deciles and GCSE English score deciles $\times$ income deciles, γ_f is the coefficient on an indicator that the student is female, and $\gamma_{eth(i)}$ are ethnicity fixed effects. The model yields accurate predictions of A-level take-up in the holdout sample during the training period with a mean squared error of 0.104. Table C3 reports summary statistics for students by predicted propensity, with notable differences in socioeconomic status, prior attainment, and demographic composition.

Panels C and D in Figure 4 present the estimates of university applications and enrollment effects separately by students who were *previously likely* to take academic-track classes and students who were *previously unlikely* to take academic-track classes.[12] The increase in university application and attendance among low-income students is concentrated among students who were previously unlikely to have taken A-levels. This suggests that the path dependency between high school coursework and university attendance is a key mechanism behind the rise in university enrollment among low-income students.

Why are some students, particularly high-income students, more likely to apply to university but no more likely to enroll? Figure A2 shows that it is specifically the low-propensity high-income students that become more likely to apply to university but not enroll. Figure D4 shows that conditional on applying to university, low-propensity high-income students at treated schools aim higher in their application portfolios, applying to universities where students have average A-level percentiles that are around 2 percentile points higher. Furthermore, they are 2-5pp more likely to apply to STEM majors. These changes in application patterns are particularly striking since those same students have lower GCSE grades on average and perform worse on A-levels compared to their counterparts in untreated schools. We do not find any changes in application patterns among low-income students, regardless of predicted propensity of taking A-levels, as shown in Panels A and C in Figure D4. These results suggest some "wasteful" university application attempts that may be the result of eliminating an informative signal of ability prior to university applications.

[12] We define low and high propensity as below or above median predicted propensity on the sample excluding the training sample.

IV.D *Marginal university entrants and academic mismatch*

Outcomes such as university graduation at age 21 are available only for the first treated cohort in our data. Because the average duration of an undergraduate degree in the UK is 3 years, we estimate effects of the reform on university performance for students who were treated in 2017. We estimate a two-way fixed effects model using OLS:

$$Y_i = \alpha_{s(i)} + \gamma_{t(i)} + \gamma_{um(i)} + \sum_{\tau=2012}^{2017} \theta_\tau I(t(i) = \tau) \times I(T_{s(i)} = 2017) + \varepsilon_i \tag{4}$$

where $\alpha_{s(i)}$ and $\gamma_{t(i)}$ are school and year fixed effects respectively, $\gamma_{um(i)}$ are university-major fixed effects, and $T_{s(i)}$ is the treatment year, which is the academic year school $s(i)$ stopped offering AS-levels. Our control group includes English schools that continued requiring AS-level exams in 2017 and schools in Northern Ireland and Wales, where data is available.

We find that average university performance among students from high schools that eliminated exams *declines* relative to students from control schools after the reform. Table 3 shows effects on university graduation rates, university performance, and dropout rates. Columns 1 through 6 present estimates from Equation 6 among students who began university between 2012 through 2017 and columns 7 and 8 present results from a staggered event study estimating Equation 1. Conditional on attending university, students from schools without exams are 2pp less likely to graduate after AS-level requirements were lifted. Turning to performance at university, students are 1pp less likely to graduate with first-class honors (which corresponds to graduating in the top 20% of the graduating cohort). These students are also more likely to be at the lower-end of their graduating cohort, as they are 2pp less likely to graduate with upper-second class honors or higher (which corresponds to graduating in the top 80% of the graduating cohort). Dropout rates, which we observe for students treated in not only 2016-17 but also in academic years 2017-18 and 2018-19, increase by 1-3pp. Moreover, the increase in dropout rates grows over time, which may be partly due to the fact that an increasing number of subjects are reformed between 2017-2019. Our estimates are largely stable across specifications with and without university-major fixed effects.

These results are robust across different specifications including ones that exclude schools in Northern Ireland and Wales (control group is restricted to England) as well as versions that only compare treated schools in England to schools in Northern Ireland and Wales (exclude English control schools). Results from these alternate specifications, as well as a placebo analysis comparing untreated schools in England over the same time period to schools in Northern Ireland and Wales, are provided in Appendix Table D2.

The decline in university performance also varies by the extent to which teacher

grades are correlated with the eliminated pre-application tests. To test for this heterogeneity, we separately examine changes in university performance at schools with high pre-reform teacher grade inflation relative to the tests and at schools with low pre-reform teacher grade inflation.[13] Table D4 presents estimates from this analysis. We find that when comparing schools with above-median teacher grade inflation to schools in Northern Ireland and Wales, the pure control schools, the likelihood of graduation declines by over 3pp. The effects at schools with below-median teacher grade inflation, on the other hand, are statistically indistinguishable from zero. Similarly, effects on graduating with first-class honors are twice as large at schools with high pre-reform teacher grade inflation. This is suggestive evidence that the academic mismatch effects on university performance are exacerbated at schools with historically higher grade inflation, i.e., larger differences between standardized test scores and teacher-provided grades.

These academic mismatch effects may reflect academically less-prepared marginal university entrants. Students who would have attended university regardless, such as high-income students with high predicted propensity to take A-levels prior to the reform, may also shift to higher-ranked institutions after their schools eliminate exams. This decline in average university performance may also arise from both mechanisms simultaneously. Estimates in Table D1 show that high-income and high-propensity students are no more likely to attend universities ranked in the top 10, top 20, or top 50 universities when their schools no longer offer exams.[14] Thus, we find no evidence that students who would have attended university prior to the reform are more vertically mismatched post-reform.

The marginal entrant induced into university by the reform, however, is significantly weaker academically and performs worse at university compared to the average low-income university student. We characterize the marginal low-income university entrant using an IV framework (Imbens and Angrist, 1994; Angrist et al., 1996). Under this framework, *compliers* into university in our context are students whose university enrollment is affected by the reform. Specifically, compliers are students who attend university ($D_i = 1$) if their school stopped offering pre-university tests after the reform ($Z_i = 1$), but would not attend university ($D_i = 0$) otherwise (i.e. if $Z_i = 0$). Consistent with our event study results in Figure 4, we do not observe a statistically significant first stage for high-income students. We therefore restrict our attention to low-income compliers. For each characteristic X, the

[13] We group schools in low and high teacher grade inflation groups as follows. First, we regress teacher grades on AS-level grades and A-level grades in years 2010-2012 with school and subject-by-year fixed effects among students taking A-levels reformed in 2017. Schools with above-median fixed effects are classified as "high teacher grade inflation", while schools with below-median fixed effects are classified as "low teacher grade inflation".

[14] We construct university rankings using the average standardized A-level grades of university attendees in the 2011 cohort.

distribution within low-income compliers is:

$$w_x = \frac{E[1\{X_i = x\}D_i|Z_i = 1] - E[1\{X_i = x\}D_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]} \tag{5}$$

Identification requires the assumptions outlined in Imbens and Angrist (1994) to hold, and in particular, that the instrument is independent of the error term in the decision to attend university. The IV framework in our setting imposes more stringent assumptions compared to our previous event study analysis. These analyses additionally require that the instrument Z_i assignment is good-as-random conditional on school and year fixed effects, as well as the standard monotonicity and first stage assumptions. We estimate Equation 5 using two-stage least squares (2SLS) to allow for school and academic year fixed effects, where university attendance is the endogenous variable. Table 2 presents marginal mean characteristics for low-income compliers (Column 1) as well as average characteristics for low-income university students between 2012-2016 (Column 2). Demographically, we see that students induced into university are slightly less likely to be white, more likely to be female, and more likely to be first-generation university goers. The marginal university entrant is also substantially academically weaker: 54% of compliers have below median GCSE math scores compared to 28% among low-income university students. These students primarily attend academically non-selective universities, with an average university ranking of around 100. The marginal university student is more likely to drop out and less likely to graduate on time. Overall, we find that only 41% of these marginal university students graduate university on time (within 3 years), while 21% drop out and another 38% remain in university without graduating.

V. Labor Market Returns for Marginal Entrants

A substantial share of marginal university entrants in our setting fail to complete their degrees. Would these students achieve better outcomes at less selective universities or by forgoing university altogether? To address this question, we analyze labor market outcomes three years after high school graduation, at ages 21-22. Due to limitations in our data, labor market outcomes at age 21 are available only for the first treated cohort, i.e. students who were treated in 2017. This is because our sample of the HESA dataset containing information about university graduation ends in 2020 and our sample of the HMRC dataset with information on employment and earnings outcomes ends in tax year 2020-21. For these outcomes, we estimate a two-way fixed effects model:

$$Y_i = \alpha_{s(i)} + \gamma_{t(i)} + \sum_{\tau=2012}^{2017} \theta_\tau I(t(i) = \tau) \times I(T_{s(i)} = 2017) + \varepsilon_i \tag{6}$$

where $\alpha_{s(i)}$ and $\gamma_{t(i)}$ are school and year fixed effects respectively and $T_{s(i)}$ is the treatment year, which is the academic year school $s(i)$ stopped offering AS-levels. Because we observe age 21 outcomes only for students in the first treated cohort, treated schools are schools that stopped offering AS-levels in the academic year 2016-17, while the control schools are schools that stopped offering AS-levels in subsequent years, or continued requiring them through 2018-19.

For earnings at age 21, we estimate Equation 7 using Poisson pseudo-MLE as recommended by Chen and Roth (2024), since our data include individuals with 0 earnings at age 21:

$$Y_i = \exp\left(\alpha_{s(i)} + \gamma_{t(i)} + \sum_{\tau=2012}^{2017} \theta_\tau I(t(i) = \tau) \times I(T_{s(i)} = 2017)\right) \varepsilon_i \tag{7}$$

where $\alpha_{s(i)}$ and $\gamma_{t(i)}$ are school and year fixed effects respectively and $T_{s(i)}$ is the treatment year, which is the academic year school $s(i)$ stopped offering AS-levels. Our control group includes English schools that continued requiring AS-level exams in 2017.

V.A *Early-career outcomes*

We estimate Equation 6 on the following outcomes at age 21: (i) an indicator for having a university degree; (ii) an indicator for receiving out-of-work benefits; (iii) an indicator for being employed; (iv) firm characteristics. To examine whether these students who were induced into university begin their careers at different firms, we consider characteristics of the firm at which they are employed at age 21. First, we look at whether or not the firm is a "non-degree firm", which we define as a firm where none of the employees in tax years 2007-08 through 2010-11 had graduated from university. Second, we look at whether or not the firm is a "bottom 10%" firm based on average earnings of employees hired before tax year 2010-11. The third firm characteristic we consider is whether a firm is an "elite" firm, which we define as a firm where 50% or more of the employees in tax years 2007-08 through 2010-11 had graduated from selective universities.[15] For completeness, we also estimate effects where the outcome is an indicator of whether or not the student took A-levels, applied to university, and attended university. We also look at effects on earnings in the tax year corresponding to three years after the first treated cohort graduated high school, which we estimate using Equation 7 to allow for zeroes.

[15] We define selective here as high-tariff universities following the UCAS classification in the corresponding year. This corresponds to roughly the top third of universities in the UK.

Figure 5 summarizes the difference-in-differences coefficients for outcomes and Table D5 reports event study coefficients to rule out pre-trends. Low-income students at schools that eliminated testing see statistically significant improvements in early-career outcomes. These students are 1pp more likely to have graduated with a university degree by age 21, which is consistent with an average graduation rate of 30-40% among marginal entrants. This small increase in university graduation is associated with substantial improvements in early labor market outcomes, as low-income students are 1pp more likely to be employed and 2pp less likely to receive out-of-work benefits at age 21. Low-income students also begin their careers at *better firms*, on average, as they are 1pp less likely to be employed at a non-degree firm or a low-paying firm post-reform. However, we find a null effect on earnings at age 21. We also find no statistically significant effects on working at an elite firm, which is unsurprising for students who were marginal to attending university.

We interpret the null effect on age-21 earnings with caution, however, for several reasons. First, in the UK context, workers who do not attend university outearn university attendees until age 22. Figure A3 shows historical earnings profiles by education level between age 18 through 30. Between ages 18 to 23, workers who did not attend university outearn both workers who attend university but never graduate (i.e. have partial university) and workers who attend university and eventually graduate. This pattern reverses, however, after age 23, where university graduates begin outearning workers with no university. Crucially, workers with partial university who never complete a degree also earn more than workers who never attended university.

This pattern is consistent with evidence in the literature that earnings do not stabilize until individuals are in their late 20s (Haider and Solon, 2006; Chetty et al., 2014, 2023) and that an increase in university participation rate typically is associated with lower earnings early on compared to individuals who began full-time work at age 18. Individuals who enter the workforce immediately after high school will have more experience and seniority within their position by age 21 compared to a university graduate who began full-time work at 21, even if the latter is on a different career trajectory and will outearn the former in several years. These foregone earnings during university may cancel out the positive returns to university early in their careers (Angrist et al., 2022).

Panel B in Figure 5 reports the coefficients for high-income individuals. Among high-income students, we generally see minimal effects, except for a slight reduction in the likelihood of receiving out-of-work benefits and a slight reduction in the probability of being employed at age 21. Collectively, our results suggest positive labor-market outcomes, particularly for the low-income students who were induced into attending university with minimal downstream effects for high-income students.

Of the students induced into taking A-levels and subsequently enrolling in university, many of them *succeed*. On average, low-income attending schools that eliminated pre-university testing are more likely to graduate university by age 21 and begin their careers at better firms. However, we do observe substantial leakage between the enrollment stage and the graduation stage for these marginal students, which is clearly visible in Panel A of Figure 5. This leakage is consistent with findings in the literature (Arcidiacono and Lovenheim, 2016; Black et al., 2023).

V.B *Distributional effects of early-career outcomes*

Low-income students are *on average* better off after their school eliminates pre-university standardized testing. At the same time, our findings also indicate that many of the marginally induced students do not succeed, reflected in both weaker endline A-level performance and lower on-time degree completion rates. If a bimodal distribution of outcomes underlies these positive average effects, a handful of students may be substantially worse off as a result of the increase in attempts. What do the distributional effects on early-career outcomes look like for the low-income students induced into university?

We estimate marginal potential outcome distributions for age 21 firm characteristic deciles for low-income compliers as established by Abadie (2002) in an extension of Imbens and Angrist (1994). Here, compliers are students who attend university at age 18 ($D_i = 1$), if they attended treated schools in the years after the reform ($Z_i = 1$), but do not attend university ($D_i = 0$) otherwise (i.e., if $Z_i = 0$). Because we are interested in the marginal distributions, we group potential outcomes into deciles, where Y_i^{21} is the age 21 outcome decile. For firm characteristics, we group the individual's primary firm of employment into deciles based on the following attributes of *previously hired* employees: (i) average earnings of employees[16] and (ii) the share of employees with university degrees. For each firm characteristic decile c, we estimate the distribution of age-21 firm decile Y_i^{21} for compliers who *attended university*:

$$E[1\{Y_{1i}^{21} < c\}|D_{1i} > D_{0i}] = \frac{E[1\{Y_i^{21} < c\}D_i|Z_i = 1] - E[1\{Y_i^{21} < c\}D_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]} \quad (8)$$

We also estimate the distribution of Y_i^{21} for compliers who *did not* attend university:

[16]In order to capture the long-term potential of the firm, we restrict to take the average of earnings of employees with 5-10 years of general work history, meaning they were first observed in the HMRC dataset between 5-10 years ago but employees need not necessarily have been at the firm for 5-10 years.

$$E[1\{Y_{0i}^{21} < c\}|D_{1i} > D_{0i}] = \frac{E[1\{Y_i^{21} < c\}(1 - D_i)|Z_i = 1] - E[1\{Y_i^{21} < c\}(1 - D_i)|Z_i = 0]}{E[1 - D_i|Z_i = 1] - E[1 - D_i|Z_i = 0]} \tag{9}$$

We estimate Equations 8 and 9 using 2SLS to allow for school and academic year fixed effects where university attendance is the endogenous variable. Consistent with the average treatment effects in our difference in differences estimates, we find that treated low-income students are less likely to end up in the far left tail of early-career and longer-term labor market outcomes. Figure 6 visualizes this shift by showing the cumulative distribution functions of outcomes for low-income compliers who attend university alongside the distribution for those who did not. Crucially, for both the age 21 firm education decile (Panel A) and firm earnings decile (Panel B), low-income students shifted into university appear to have a near zero likelihood of working at firms in the bottom three deciles. This suggests that the average gains in early-career outcomes do not mask an underlying bimodal set of outcomes wherein some students gain while others lose. Contrary to such a pattern, we find that university attendance pulls these students out of low-paying firms that typically employ workers without university degrees. Moreover, as shown by the potential outcome distributions of compliers who did not attend university, many of these students (around 40%) would have counterfactually worked at these low-wage firms.

V.C *Lifetime earnings gains for marginal entrants to university*

To quantify the long-run returns for students marginally shifted into university, we estimate the effect of university on the present discounted value of lifetime earnings. We estimate lifetime earnings effects for the first treated cohort of students who completed high school in the academic year 2016-2017 and therefore typically began university in August 2017, with expected completion in June 2020. As before, because we observe first stage effects of the reform on university enrollment only for low-income students, we focus our attention on low-income marginal entrants for statistical power.

First, we directly estimate the effect of university enrollment on earnings at ages 18 through 21 using 2SLS and controlling for school and year fixed effects. Next, we use early-career outcomes to project lifetime earnings until age 67, which is the standard retirement age in the UK. We do not observe earnings beyond age 21 because the HRMC data currently extend up to tax year 2020-2021. We project earnings at age 27, which is 9 years after the expected high school completion and is the oldest age we observe earnings for in our

pre-reform sample. To project age-27 earnings, we use two related approaches: (i) the surrogate index method developed by Athey et al. (2025b) and (ii) the experimental selection correction estimator of Athey et al. (2025a). For earnings between ages 22 through 27, we linearly interpolate between observed age 21 earnings and projected age 27 earnings. We take a conservative approach to extending earnings until age 67 by assuming no *real* earnings growth beyond age 27. Because earnings typically grow substantially between ages 30 to 40 in the UK, especially for men (Britton et al., 2022), our lifetime earnings gains estimates likely represent a lower bound.

For the surrogate index projection, we fit a linear model of age 27 earnings on demographics and observed employment variables between ages 18 through 21. Specifically, we estimate:

$$Y_i^{27} \;=\; \beta X_i \;+\; \gamma^{21}_{firm(i)} \;+\; \gamma^{21}_{SIC(i)} \;+\; \sum_{\mathrm{a}\,=\,18}^{21} \left(\delta_1^{\mathrm{a}} Y_i^{\mathrm{a}} \;+\; \delta_2^{\mathrm{a}} E_i^{\mathrm{a}} \;+\; \delta_3^{\mathrm{a}} B_i^{\mathrm{a}}\right) \;+\; \varepsilon_i \tag{10}$$

where the outcome Y_i^{27} is the observed earnings from a worker's primary firm of employment during the tax year 11 years after KS4 completion and X_i are demographic variables that include an indicator for female, ethnicity indicators, IMD deciles, and GCSE math score deciles. γ^{21}_{firm} and $\gamma^{21}_{SIC(i)}$ are fixed effects for the firm identifier and 2007 Standard Industrial Classification (SIC) codes for the firm of employment at age 21.[17] For ages $a \in \{18, 19, 20, 21\}$, Y_i^a is observed earnings, E_i^{a} is an indicator for missing or zero earnings, and B_i^{a} is an indicator for receiving out-of-work benefits. In a second linear model, we estimate:

$$Y_i^{27} \;=\; \beta X_i \;+\; \gamma X^{21}_{firm(i)} \;+\; \sum_{\mathrm{a}\,=\,18}^{21} \left(\delta_1^{\mathrm{a}} Y_i^{\mathrm{a}} \;+\; \delta_2^{\mathrm{a}} E_i^{\mathrm{a}} \;+\; \delta_3^{\mathrm{a}} B_i^{\mathrm{a}}\right) \;+\; \varepsilon_i \tag{11}$$

where instead of firm and industry fixed effects, we control for age-21 firm characteristics. Concretely, $X^{21}_{firm(i)}$ includes indicators for whether the firm is in the bottom or top quartile for previously hired workers' average earnings, the share of previously hired workers with university degrees, and the share of previously hired workers with university degrees from a selective university.[18] We estimate Equations 10 and 11 on a sample of students who completed KS4 in 2008 and 2009 and who were 18 years old at the start of tax years 2010-11 and 2011-12. The coefficients for both models are reported in Appendix Table A2.

[17] We assign a common identifier to small firms with fewer than 3 employees and draw on variation from the SIC 2007 codes instead.

[18] As in Section V.A, we define selective here as high-tariff universities following the UCAS classification in the corresponding year. This corresponds to roughly the top third of universities in the UK.

Using these estimates, we then follow Athey et al. (2025b) to project age-27 earnings by extrapolating the relationship between these surrogate "early-career" outcomes and age-27 earnings from our historical sample to our quasi-experimental sample. We then directly estimate the effect of attending university on age-27 earnings by using the reform as an instrument for attending university.

These earnings projections are based on demographics and observed earnings and employment history without explicitly factoring in human capital investment decisions made after GCSEs such as A-levels, university attendance, choice of major, and degree completion. By using only observed employment outcomes, we do not make any assumptions about similar returns to university between our pre-reform sample and our treated sample. Instead, we rely on the treated sample following similar earnings trajectories conditional on observed early-career outcomes such as earnings and firm characteristics. Any changes in projected earnings in our results will be directly driven by observed employment outcomes at ages 18 through 21 rather than A-level take-up, university enrollment or degree completion. For these earnings effects to be valid causal estimates of university on age-27 earnings, any effect of university attendance on future earnings must operate solely through the secondary early-career outcomes used in Equations 10 and 11 (i.e., firm characteristics, employment indicators, benefits, and earnings at ages 18 through 21). This is a strong assumption that is relaxed in our second approach to earnings using the experimental selection correction (ESC) method.

For the second approach, we directly estimate the effect of university attendance on "secondary outcomes": earnings, employment, and unemployment benefits at ages 18 through 21 as well as age-21 firm characteristics. We estimated these effects using 2SLS with school and academic year fixed effects. Second, for all low-income students in the historical sample (individuals who are 18 years old at the start of tax years 2010-11 and 2011-12), we calculate the difference between observed secondary outcomes and predicted secondary outcomes; this residual is computed using the estimates of the effect of university attendance on secondary outcomes in the quasi-experimental sample. Lastly, we regress the primary outcome, age-27 earnings, in the historical data controlling for the residual computed in the second step. This approach yields projections that rely on less stringent assumptions compared to the surrogate approach. Specifically, the ESC permits the treatment to affect age 27 earnings through channels other than observed labor market outcomes at ages 18 - 21 (intermediary variables). This approach is better suited for our setting as A-level take-up, subject composition, and university attendance may impact downstream earnings in ways not fully captured by intermediary employment and earnings outcomes.

Figure 7 presents cumulative lifetime earnings gains (in 2015£) for low-income stu-

dents discounted at 3% and 5% for projections estimated using the ESC method as well as the Surrogate Index estimates from Equations 10 and 11. The trajectory is similar to the differences in earnings profiles by age between university attendees and non-attendees in Appendix Figure A3. In the early years, university attendance reduces earnings, but by age 26, university attendance has positive labor market returns. Even the most conservative of our estimates suggest positive private returns to university for low-income marginal entrants. At a 3% discount rate, our estimates yield more than a £100,000 increase in lifetime earnings for students induced into attending university as a result of reduced testing requirements. Thus, although marginal low-income entrants are only 30-40% likely to graduate on time, the improved early-career outcomes captured by firm quality and employment represent substantial financial gains in the long run.

VI. Calibration Bias in Standardized Tests

VI.A *Tests exhibit no calibration bias against low-income students*

Marginal entrants to university, in our context, see large increases in lifetime earnings from attending university. Why, then, were these low-income students not previously attending university? One hypothesis is that standardized tests are biased against low-income students, who are then screened out of university in admissions on the basis of these test scores. We now extend our analysis to examine *calibration* of tests and teachers in predicting university performance.

We begin by formalizing our notion of calibration and bias in an algorithmic prediction framework (Kleinberg et al., 2016, 2018). Let ζ denote a standardized test score (e.g., AS points) observed before admission. Let Y be a university outcome measured ex post, such as graduation or position within graduating cohort. Let $\hat{Y}(\zeta) \in [0,1]$ be a probabilistic prediction of Y built from ζ (and allowed controls) trained on pre-admission information. Test score ζ (and thus predictor $\hat{Y}$) is calibrated if, for any $p \in [0,1]$ in the support of $\hat{Y}$, $E[Y|\hat{Y}(\zeta) = p] = p$. Test scores are calibrated differentially for low-income and high-income students when $E[Y|\hat{Y}(\zeta), \text{Low-income} = 1] \neq E[Y|\hat{Y}(\zeta), \text{Low-income} = 0]$.

We proceed in two steps: (i) Are standardized test scores well-calibrated predictors of subsequent outcomes among university students? (ii) Are standardized tests in our setting *differentially* calibrated for low- vs. high-income students (i.e., biased against or in favor of one demographic group relative to another)? To make the grades comparable across individuals and cohorts, we standardize each grade within qualification year and subject. We then take the average standardized grades for both signals respectively and restrict our

analysis to students who took large A-level subjects.[19]

First, we examine overall calibration and differential calibration among UK university students who started university sometime between 2012-2017. We train a model on a random 20% subsample to predict university performance and outcome measure Y. In particular, we train a logit model on fully interacted standardized test scores for a student's best three A-level endline test scores. We then predict $\hat{Y}$ for the 80% holdout sample and bin the predictions $\hat{Y}$ and corresponding true outcomes Y into ventiles separately for low-income and high-income students. We can graphically depict this relationship by plotting Y on the y-axis and $\hat{Y}$ on the x-axis. If A-levels are calibrated for a given group, the points corresponding to that group, say low-income students, would lie along the 45-degree line. Similarly, if A-levels are calibrated for both groups, the points for all ventiles for both groups would lie on the 45-degree line. Figure 8 visualizes this relationship for two outcomes: (i) on-time graduation (Panel A) and (ii) on-time graduation with first-class degree (Panel B). A-levels appear well-calibrated for both low-income and high-income groups in predicting university graduation and degree class.

However, this comparison does not capture population-level calibration due to the selective labels problem (Lakkaraju et al., 2017). This challenge occurs in many contexts such as loan default rates or misconduct risk after release (Arnold et al., 2018, 2022). In our setting, the empirical challenge reflects the fact that true university performance Y is observed only among *students who attend university*, i.e. we observe $E[Y|\hat{Y}(\zeta), \text{University} = 1]$. However, students who enroll into university comprise a highly selected subsample of all students (or even university applicants). As such, in general, $E[Y|\hat{Y}(\zeta), \text{University} = 1] \neq E[Y|\hat{Y}(\zeta)]$. This selection into labelling can generate uninformative or misleading conclusion when extrapolating the relationship between test scores ζ and university performance Y among university goers to that of students who were screened out of university. Calibration, defined this way, captures how universities may extrapolate from the relationship between the signal and university performance metrics they care about to the marginal student they do not admit. If that relationship is different for the students who are screened out, examining calibration among the set of students admitted is insufficient to draw conclusions about calibration for students on the margin of university enrollment.

One solution to overcome this challenge to measure calibration bias *at the margin* is to use an instrument for university attendance. Specifically, we use schools' decisions to eliminate AS-level exams in response to the policy reform as an instrument for university

[19]We restrict this analysis to students taking the following A-level subjects: Art and Design, Biology, Business, Chemistry, Design and Technology, Drama, Economics, Film, French, Further Mathematics, Geography History, Information Technology, Language and Literature, Language, Law, Literature, Mathematics, Media Studies, Physical Education, Physics, Politics, Psychology, Religious Studies, Sociology, and Spanish.

enrollment. Let Y_{1i} be the potential outcome (e.g., probability of completing a university degree within three years) for individual i if i attends university, and Y_{0i} be the potential outcome for the same individual if i does not attend university. Let $Z_i = 1$ if student i has a school-cohort where AS-level testing was discontinued and $Z_i = 0$ if i has a school-cohort where AS-level testing is mandatory. Let $D_i \in 0, 1$ indicate university attendance ($D_i = 1$ when individual i attends university and $D_i = 0$ otherwise).

We assume that Z_i is conditionally exogenous, i.e., $Z(Y_{0i}, Y_{1i}, D_{0i}, D_{1i})$ conditional on school and year fixed effects. We also assume the standard IV conditions of monotonicity and first stage, which are consistent with our event study estimates in Figure 4. Under these assumptions, we can estimate the marginal mean outcomes Y_i and $\hat{Y}$ for a subpopulation of compliers:

$$E[Y_{1i}|D_{1i} > D_{0i}] = \frac{E[Y_{1i}D_i|Z_i = 1] - E[Y_{1i}D_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]} \tag{12}$$

We estimate Equation 12 for low-income students using 2SLS, controlling for school and year fixed effects. This specification satisfies the conditions for rich covariates outlined in Blandhol et al. (2025).[20] Thus, we interpret our estimates as marginal mean Y_i and $\hat{Y}$ among a set of individuals for whom university attendance is conditionally random, averting selective observation of university outcomes.

If the marginal mean Y for low-income compliers who are shifted into university lies along the line of the relationship between Y and $\hat{Y}$ among university goers, then standardized tests are not differentially calibrated for the screened-out population. On the other hand, if the marginal mean Y lies *above* the 45-degree line, low-income compliers are *outperforming* what their test scores would predict. This pattern of outperformance would be consistent with the notion that test scores are miscalibrated for low-income students and, in particular, that they are biased against low-income students. Symmetrically, if the marginal mean Y lies *below* the 45-degree line, test scores would be miscalibrated but in favor of low-income students.

Panels C and D in Figure 8 overlay the marginal mean Y for low-income compliers on the 45-degree line and the relationship between $\hat{Y}$ and Y among inframarginal students. For on-time graduation, we can nearly rule out that the average marginal graduation rate exceeds the rate predicted for the marginal students by their test scores $\hat{Y}^{test}$, shown in red. We also generate the predicted $\hat{Y}^{teacher}$ implied by the teacher grade with which these students applied to university. $\hat{Y}^{teacher}$ lies to the right of $\hat{Y}^{test}$ for the marginal university

[20] In particular, $E[\tilde{Z}|X] = 0$ where $\tilde{Z}$ are the residuals from a regression of Z on X and X are school and year fixed effects.

entrants, though the difference is not statistically distinguishable from zero. Results for first-class degree shown in Panel D are similar, though marginal students are less likely to underperform relative to their test score predictions $\hat{Y}^{test}$. Taken together, these findings suggest that standardized tests in our context are *not biased* against the low-income students who are marginal to university enrollment. If anything, these students slightly *underperform* relative to the calibration for inframarginal students.

From a social planner's perspective, we may be most interested in whether a test is calibrated to capture true ability, rather than university performance. Because we cannot observe true ability, we focus on an empirical proxy based on long-term academic outcomes. We acknowledge that university performance may also be impacted by bias, particularly when considering broader systemic discrimination a la Bohren et al. (2025).

As discussed in Section II., students in the UK generally apply to universities and receive offers based on grades assigned by teachers in each A-level subject. Nominally, these grades are meant to predict what grade the student will ultimately receive on the actual A-level examination. However, in practice, the process by which teachers assign these grades varies across schools and incorporate more subjective measures of performance. For example, one high school website describes their predicted grade policy as "a professional judgment formulated by the teacher using a holistic assessment of potential and performance across the first year of study" which "may consider some or all of the following factors: results from internal mock exams, general attitude to learning and commitment, performance in homework, GCSE results, the student's drive and passion for the subject."[21] Moreover, teacher-assigned grades have been shown to be systematically inflated relative to standardized test scores in A-level and AS-level exams (Murphy and Wyness, 2020; Leckie and Maragkou, 2024). Between 2012-2021, 55% of these grades were over-predicted, meaning that the predicted grade exceeded the actual A-level grade that the student eventually achieved, while only 37% were correctly predicted.

In this sense, it is unsurprising that our results indicate that teacher-assigned grades are less accurate measures of a student's potential to succeed at university compared to standardized tests. Our findings are also consistent with the recent literature showing that high-school GPAs are less predictive of university performance at Ivy plus universities in the US compared to SAT scores (Chetty et al., 2023; Friedman et al., 2025). Moreover, these results likely underestimate the overall decrease in signal precision as previously, teachers were able to observe the AS-level grades when forming their predicted grades. After the reform, teachers no longer had access to this anchor and instead relied on internal examinations and interactions with the students over the course of the academic year.

[21]Source: https://www.bhasvic.ac.uk/parents-carers-dashboard/higher-education as of February 12, 2025.

VI.B *Standardized tests are more informative of future outcomes*

Teachers may observe student traits such as creativity, persistence, or knowledge that may not be captured for students who are bad test takers. These attributes may in turn be predictive of future academic success. Do teacher grades reflect additional information *not captured* by standardized tests? To answer this question, we compare the predictive power and informativeness of teacher grades and endline (post-application) A-level exam scores on university performance measures. We focus on a pre-reform sample of university students pursuing undergraduate degrees who began university between 2012-2016 at large universities.[22] We begin by estimating OLS regressions of each university outcome separately by income group to compare the explanatory power of teacher grades and test scores:

$$Y_i = \delta_g \sigma_i + \gamma_{umg} + \varepsilon_i \tag{13}$$

where Y_i is the university outcome and σ_i is either the average standardized teacher-assigned grade, the average standardized A-level exam score, or the interaction of both grades, and γ_{umg} is a major-university fixed effect. Table 4 presents the adjusted R^2 values each set of regressions estimated in Equation 13. Across both low-income and high-income students, test scores explain a substantially higher share of the variance in university performance. Comparing columns 1 and 2 in Table 4, the adjusted R^2 increases by 21% (from 0.092 to 0.111) for low-income students when using test scores instead of teacher grades to predict on-time graduation. Test scores also explain a higher share of variance in on-time graduation compared to teacher grades for high-income students, but the increase is smaller in magnitude (+7%). Across both income groups, teacher scores add little explanatory power in regressions that already include test scores.

Next, we compare the out-of-sample predictive power of each set of predictors using a random forest classifier, split into a 70% training sample and a 30% testing sample.[23] Figure 9

[22] In order to avoid overfitting to very small major-university pairs, we restrict our analysis to just over 80 universities with 1,000 or more students beginning each year. We restrict this analysis to students taking the following A-level subjects: Art and Design, Biology, Business, Chemistry, Design and Technology, Drama, Economics, Film, French, Further Mathematics, Geography History, Information Technology, Language and Literature, Language, Law, Literature, Mathematics, Media Studies, Physical Education, Physics, Politics, Psychology, Religious Studies, Sociology, and Spanish. We use endline, post-application A-level exam scores since these are observed in both pre-reform and post-reform samples.

[23] The random forest classifier is a tree-based ensemble learning technique that constructs multiple decision trees on different boot-strap sub-samples. Our model was implemented using the sklearn package in python using 500 trees allowed to grow fully with no depth restrictions and a minimum sample size of 100 per leaf with balanced class weights. Because random forest classifiers generally perform best on continuous variables rather than fixed effects, we include a ranking of the average A-level score percentiles of matriculating students in 2010 at each university instead of university fixed effects. Both models include either the average standardized A-level exam scores or the average standardized teacher grades.

presents the precision-recall (PR) curves and the average precision (AP) measures, which are standard diagnostics used in the literature (Cengiz et al., 2022), for students matriculating between 2012-2016.[24] Panel A shows that the model trained using only exam grades slightly outperforms the model trained solely on teacher grades when predicting whether or not a student graduates from university, with an AP of 0.728 compared to 0.721. For low-income students, however, the increase in out-of-sample predictive power when using tests is larger, with an AP of 0.725 compared to 0.710. Consistent with our results on explained variation in Table 4, standardized tests have substantially higher out-of-sample predictive power for a student's position within the graduating cohort. Moreover, Panels C and D in Figure 9 show a visibly striking gap between the model trained using test scores and the model trained using teacher grades. The blue curve, showing the PR curve for the model trained on test scores, lies above the orange curve at all values of recall.[25]

Lastly, we examine how large discrepancies between teachers and tests are correlated with downstream university and postgraduate outcomes. For each A-level, we estimate a regression of the teacher-assigned grade on the AS-level grade, average GCSE percentile for core subjects, and the final A-level grade controlling for year fixed effects.[26] We then take the residuals of these regressions and standardize them. Conceptually, these residuals capture the components of the teacher-assigned grade that is *unexplained* by standardized test scores, such as subject-specific measures of ability recognized by teachers but not observed in standardized test scores. We estimate:

$$Y_{ig} = \alpha_{0g} + \delta_g \sigma_{ig} + \gamma_{sg} + \gamma_{tg} + \varepsilon_{ig} \tag{14}$$

where Y_{ig} denotes university outcomes, g denotes the subject group (i.e., STEM or non-STEM), σ_i is the standardized teacher-assigned grade residual for the student, γ_{sg} is school fixed effects, and γ_t is year fixed effects.[27] Table 5 presents estimates from Equation 14.

Although these residuals are strongly correlated with a student's university destination, they are not predictive of success. A one standard deviation increase in teacher-assigned grade residuals is associated with an 8-9pp increase in probability that the student applies to a selective university and a 5pp increase in the probability that the student attends a

[24]We additionally report receiver operating characteristic (ROC) curves in Figure D5 for completeness, although PR curves are preferable in our context because of the unbalanced nature of our outcomes (i.e. around 85% of students graduate with a degree and 30% of graduates do so with a first-class degree).

[25]Intuitively, this means that for any given level of recall (the ratio of students predicted by the model to graduate with first-class degrees over the students who actually graduated with first-class degrees), the model using tests has a higher true positive rate among the set of students it predicted to graduate with first-class degrees.

[26]Core GCSE subjects are Mathematics, English Language, English Literature, and Science.

[27]STEM subjects are Mathematics, Further Mathematics, Physics, Biology, and Chemistry.

selective university. In STEM subjects, higher residuals are associated with *lower* probability of degree completion and reduced university performance, as measured by degree class, conditional on graduating. In non-STEM subjects, these residuals have small but positive correlations with university performance. These findings suggest that teacher-assigned grades have limited predictive power for university performance and post-graduate outcomes beyond what is already captured by standardized tests, mirroring statements from many UK universities in response to the reform. Moreover, in STEM subjects, over-prediction is associated with *worse* outcomes, potentially due to academic mismatch.

VII. Standardized Testing and Efficiency

The reform effectively increased student *attempts*, which in turn also increased *failures*. More students apply to universities and enroll, but graduation rates conditional on attending university decline by 2pp. In the absence of labor market outcomes, a decline in graduation rates conditional on attending may be interpreted as a decrease in efficiency resulting from academic mismatch. However, graduation rates alone are insufficient measures of welfare, which depend instead on the relative returns of a university degree (or attending university without completing a degree) across different groups. In our setting, we observe employment, earnings, and benefits outcomes for affected individuals, which permits direct engagement with welfare notions beyond those captured by graduation rates.

Although reduced testing lowered conditional graduation rates, the policy also resulted in improvements in early-career outcomes. Even our conservative estimates for projected lifetimes earnings gains suggest £50,000 in private returns for low-income students who were shifted into university. Do these private returns also imply positive *social* returns? We calculate lifetime earnings gains net of tuition fees and the average per-student cost to the UK government for a three-year university degree. We adjust for three years of tuition (£9,250 per year during our sample years) and one third of £45,000 paid annually at ages 18, 19, and 20 (Drayton et al., 2025).[28] Figure 10 presents net social returns after these adjustments for the lifetime earnings projections using both the Surrogate Index and the ESC methods at 3% and 5% discount rates. Across all specifications, the net social benefit (in net present value) is large and positive for these marginal university entrants. We note that these estimates for social returns represent a lower bound, as the per-student average cost is typically higher than the marginal cost. Computing the marginal value of public

[28] Drayton et al. (2025) report that in 2025, the UK government spent around £20 billion for 480,000 university students to attend university over the duration of their degrees. Back of the envelope calculations lead to this figure of £45,000 in 2025 GBP. We divide this value by three and deflate it to 2015 GBP for each corresponding year.

funds (MVPF) using a marginal tax rate of 20%, we obtain between 2.5 to 7.1. This implies that £1 of public spending on these marginal students' university education generates between £3 to £7 of private benefits, putting it well above the median of estimated MVPFs of cost-effective university aid programs (Hendren and Sprung-Keyser, 2020).

These gains to disadvantaged students do not come at the cost of another group with *decreased* earnings. This may be due to the lack of rigid capacity constraints at UK universities during our sample window. Enrollment caps were fully lifted in 2014 and aggregate undergraduate enrollment increased from around 160,000 in 2012 to 200,000 in 2019 as shown in Figure A4.[29] This shift into university may eventually dilute labor market premia for university graduates, as has been documented in previous educational expansions (Bianchi, 2020). While this is possible, back of the envelope calculations suggest that the reform shifted approximately 1,400 low-income students into university in the first year of the reform, which is substantially smaller in magnitude compared to prior expansions.[30] Nevertheless, labor market returns for marginal entrants may decrease as adoption by schools increases over the course of the reform. We might also expect university capacity constraints to start binding in the event of larger shifts into university. We therefore interpret our findings as a short-run effect that is net of any university-side adjustments admission policies.

In our context, relaxing testing requirements brought us closer to social optimum as we document positive social returns to university for marginal students. Our findings suggest that standardized testing may discourage disadvantaged students from attempting paths with large returns, consistent with findings in the undermatching literature (Hoxby and Avery, 2012; Campbell et al., 2022).

VIII. Conclusion

Critics of standardized testing argue that tests are inherently biased against low-income students and serve as a barrier to higher education access. This paper provides empirical evidence on the educational and labor-market consequences of reduced standardized

[29]Beginning in 2012, enrollment caps for UK universities no longer included students who obtained grades AAB or higher in A-levels. In 2014, enrollment caps were lifted at all universities (except for alternative providers, which are private "for-profit" entities in the UK or other independent institutions not receiving funding from the UK government and represent less than 2% of undergraduate students in the UK and are not considered in our results). In 2015, enrollment caps were lifted at all universities including alternative providers (Hillman, 2014). Universities with remaining vacancies fill spots by indicating availability on the UCAS website for the Clearing process.

[30]This is calculated based on a 3pp increase in university applications in the first year of the reform among low-income students. Around 560 schools with an average size of 267 students adopted the reform in the first year (i.e., for the high school cohort graduating in 2017) as shown in Appendix Table C2. On average, 32% of students at these schools are low-income, giving $560 \times 0.32 \times 267 \times 0.03 \approx 1{,}400$ students.

testing using the staggered elimination of pre-university standardized tests in the UK and administrative data on student outcomes. Our results demonstrate that students are highly responsive to relaxed testing requirements—students update their educational choices during high school as well as their decisions to apply to and attend university.

We find that the eliminating testing requirements expands participation in academic-track classes among disadvantaged, lower-ability students. High-income students and students who were previously likely to take academic-track classes shift toward STEM subjects. At the same time, we find that students become more likely to fail academic-track classes on average, particularly in STEM subjects. In the absence of pre-application exams, students are more likely to apply to university, but increases in enrollment and degree completion are concentrated among low-income applicants.

By age 21, low-income students who attended treated schools at the time of the reform are 1pp more likely to have completed an undergraduate degree. Treated low-income cohorts are also more likely to be employed, less likely to receive out-of-work benefits, and begin their careers at better firms. These early-career improvements translate to large private returns, as low-income students shifted into university see £50,000—£100,000 in lifetime earnings gains. These marginal university entrants reap large returns despite a low on-time degree completion rate of around 30—40%. Conditional on university matriculation, on-time degree completion declines by approximately 2pp among students who attended high schools that eliminated pre-application exams. Although the reform reduced conditional graduation rates, it expanded university access and degree completion on the extensive margin.

These benefits to low-income students did not arise from the elimination of a demographically *biased* exam that inefficiently *screened out* low-income students. Using the reform as an instrument for university attendance, we empirically test for calibration bias at the margin. Contrary to recent policy discourse, we find no evidence that standardized tests are biased against marginal low-income students or underpredict their performance at university. Our findings indicate that high-stakes standardized testing discourages participation in academic-track classes among disadvantaged students. When teachers assign inflated grades, reduced testing requirements expand university access for students on the margin of university enrollment. Because socioeconomically disadvantaged students are more likely to be on this margin, pre-university testing policies have distributional implications. In our setting, eliminating pre-application standardized testing improved outcomes at the lower end of the socioeconomic distribution with minimal losses on the upper end. Taken together, our findings suggest that if low-income students are underenrolling in university relative to the social optimum, reduced testing requirements can improve efficiency even in the absence of bias in standardized testing.

References

Abadie, A. (2002): "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models," *Journal of the American Statistical Association*, 97, 284–292.

Abadie, A., M. M. Chingos, and M. R. West (2018): "Endogenous Stratification in Randomized Experiments," *The Review of Economics and Statistics*, 100, 567–580.

Ahn, T., P. Arcidiacono, A. Hopson, and J. Thomas (2024): "Equilibrium Grading Policies With Implications for Female Interest in STEM Courses," *Econometrica*, 92, 849–880.

Allensworth, E. M. and K. Clark (2020): "High School GPAs and ACT Scores as Predictors of College Completion: Examining Assumptions About Consistency Across High Schools," *Educational Researcher*, 49, 198–211.

Altonji, J. G. (1993): "The Demand for and Return to Education When Education Outcomes are Uncertain," *Journal of Labor Economics*, 11, 48–83.

Angelova, V., W. Dobbie, and C. S. Yang (2025): "Algorithmic Recommendations and Human Discretion," *Review of Economic Studies*.

Angrist, J., D. Autor, and A. Pallais (2022): "Marginal Effects of Merit Aid for Low-Income Students*," *The Quarterly Journal of Economics*, 137, 1039–1090.

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996): "Identification of Causal Effects Using Instrumental Variables: Rejoinder," *Journal of the American Statistical Association*, 91, 468.

Arcidiacono, P., E. M. Aucejo, H. Fang, and K. I. Spenner (2011): "Does affirmative action lead to mismatch? A new test and evidence," *Quantitative Economics*, 2, 303–333.

Arcidiacono, P., E. M. Aucejo, and V. J. Hotz (2016): "University Differences in the Graduation of Minorities in STEM Fields: Evidence from California," *American Economic Review*, 106, 525–562.

Arcidiacono, P. and M. Lovenheim (2016): "Affirmative Action and the Quality–Fit Trade-off," *Journal of Economic Literature*, 54, 3–51.

Arnold, D., W. Dobbie, and P. Hull (2022): "Measuring Racial Discrimination in Bail Decisions," *American Economic Review*, 112, 2992–3038.

Arnold, D., W. Dobbie, and C. S. Yang (2018): "Racial Bias in Bail Decisions*," *The Quarterly Journal of Economics*, 133, 1885–1932.

Astorne-Figari, C. and J. D. Speer (2019): "Are changes of major major changes? The roles of grades, gender, and preferences in college major switching," *Economics of Education Review*, 70, 75–93.

Athey, S., R. Chetty, and G. Imbens (2025a): "The Experimental Selection Correction Estimator: Using Experiments to Remove Biases in Observational Estimates," *NBER Working Paper Series*.

Athey, S., R. Chetty, G. W. Imbens, and H. Kang (2025b): "The Surrogate Index: Combining Short-Term Proxies to Estimate Long-Term Treatment Effects More Rapidly and Precisely," *Review of Economic Studies*.

Aucejo, E., J. French, P. U. Araya, and B. Zafar (2025): "Understanding Gaps in College Outcomes by First-Generation Status," *NBER Working Paper Series*.

Autor, D. H. and D. Scarborough (2008): "Does Job Testing Harm Minority Workers? Evidence from Retail Establishments*," *Quarterly Journal of Economics*, 123, 219–277.

Avery, C., L. Shi, and P. Magouirk (2025): "Test-Optional College Admissions: ACT and SAT scores, applications, and enrollment changes," *NBER Working Paper Series*.

Belasco, A. S., K. O. Rosinger, and J. C. Hearn (2015): "The Test-Optional Movement at America's Selective Liberal Arts Colleges: A Boon for Equity or Something Else?" *Educational Evaluation and Policy Analysis*, 37, 206–223.

Bennett, C. T. (2022): "Untested Admissions: Examining Changes in Application Behaviors and Student Demographics Under Test-Optional Policies," *American Educational Research Journal*, 59, 180–216.

Bertrand, M., M. Mogstad, and J. Mountjoy (2021): "Improving Educational Pathways to Social Mobility: Evidence from Norway's Reform 94," *Journal of Labor Economics*, 39.

Bianchi, N. (2020): "The Indirect Effects of Educational Expansions: Evidence from a Large Enrollment Increase in University Majors," *Journal of Labor Economics*, 38, 767–804.

Black, S. E., J. T. Denning, and J. Rothstein (2023): "Winners and Losers? The Effect of Gaining and Losing Access to Selective Colleges on Education and Labor Market Outcomes," *American Economic Journal: Applied Economics*, 15, 26–67.

Blandhol, C., J. Bonney, M. Mogstad, and A. Torgovitsky (2025): "When is TSLS Actually LATE?" *Review of Economic Studies, forthcoming*.

Bleemer, Z. (2022): "Affirmative Action, Mismatch, and Economic Mobility after California's Proposition 209*," *The Quarterly Journal of Economics*, 137, 115–160.

Bohren, J. A., P. Hull, and A. Imas (2025): "Systemic Discrimination: Theory and Measurement," *The Quarterly Journal of Economics*.

Borghesan, E. (2023): "The Heterogeneous Effects of Changing SAT Requirements in Admissions: An Equilibrium Evaluation," .

Britton, J., L. van der Erve, C. Belfield, A. Vignoles, M. Dickson, Y. Zhu, I. Walker, L. Dearden, L. Sibieta, and F. Buscha (2022): "How much does degree choice matter?" *Labour Economics*, 79.

Campbell, S., L. Macmillan, R. Murphy, and G. Wyness (2022): "Matching in the Dark? Inequalities in Student to Degree Match," *Journal of Labor Economics*, 40, 807–850.

Carlana, M. (2019): "Implicit Stereotypes: Evidence from Teachers' Gender Bias*," *The Quarterly Journal of Economics*, 134, 1163–1224.

Carlana, M., E. La Ferrara, and P. Pinotti (2022): "Goals and Gaps: Educational Careers of Immigrant Children," *Econometrica*, 90, 1–29.

Cengiz, D., A. Dube, A. Lindner, and D. Zentler-Munro (2022): "Seeing beyond the Trees: Using Machine Learning to Estimate the Impact of Minimum Wages on Labor Market Outcomes," *Journal of Labor Economics*, 40, S203–S247.

Chen, J. and J. Roth (2024): "Logs with Zeros? Some Problems and Solutions," *The Quarterly Journal of Economics*, 139, 891–936, publisher: Oxford University Press (OUP).

Chetty, R., D. Deming, and J. Friedman (2023): "Diversifying Society's Leaders? The Causal Effects of Admission to Highly Selective Private Colleges," .

Chetty, R., N. Hendren, P. Kline, and E. Saez (2014): "Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States *," *The Quarterly Journal of Economics*, 129, 1553–1623.

Creamer, L. (2024): "Question 2 wins: Mass. voters end MCAS high school graduation requirement," *WBUR*.

Dessein, W., A. Frankel, and N. Kartik (2025): "Test-Optional Admissions," *American Economic Review*, 115, 3130–3170.

Dillon, E. W. and J. A. Smith (2020): "The Consequences of Academic Match between Students and Colleges," *Journal of Human Resources*, 55, 767–808.

Dilnot, C., L. Macmillan, and G. Wyness (2023): "The Path Increasingly Travelled: Vocational Entry Qualifications, Socioeconomic Status and University Outcomes," *British Educational Research Journal*, 49, 1142–1160.

Drayton, E., C. Farquharson, K. Ogden, L. Sibieta, D. Snape, and I. Tahir (2025): "Annual report on education spending in England: 2024-25," Tech. Rep. ISBN 978-1-80103-212-4, Institute for Fiscal Studies, London.

Exley, C., R. Fisman, J. Kessler, L.-P. Lepage, X. Li, C. Low, X. Shan, M. Toma, and B. Zafar (2024): "Information-optional Policies and the Gender Concealment Gap," *NBER Working Paper Series*.

Friedman, J. N., B. Sacerdote, D. O. Staiger, and M. Tine (2025): "Standardized Test Scores and Academic Performance at Ivy-Plus Colleges," *NBER Working Paper Series*.

Geiser, S. and M. V. Santelices (2007): "Validity Of High-School Grades In Predicting Student Success Beyond The Freshman Year:High-School Record vs. Standardized Tests as Indicators of Four-Year College Outcomes," *Center for Studies in Higher Education, Research and Occasional Papers Series (ROPS)*.

Gill, T. (2013): "Provision of level 3 qualifications in English schools 2008-2012," Statistics Report 58, Research and Development Cambridge Assessment.

Haider, S. and G. Solon (2006): "Life-Cycle Variation in the Association between Current and Lifetime Earnings," *American Economic Review*, 96, 1308–1320.

Hendren, N. and B. Sprung-Keyser (2020): "A Unified Welfare Analysis of Government Policies*," *The Quarterly Journal of Economics*, 135, 1209–1318.

Hillman, N. (2014): "A guide to the removal of student number controls," *Higher Education Policy Institute Report 69*.

Hoxby, C. M. and C. Avery (2012): "The Missing "One-Offs": The Hidden Supply of High-Achieving, Low Income Students," *NBER Working Paper Series*.

Imbens, G. W. and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467.

Kleinberg, J., J. Ludwig, S. Mullainathan, and A. Rambachan (2018): "Algorithmic Fairness," *AEA Papers and Proceedings*, 108, 22–27.

Kleinberg, J., S. Mullainathan, and M. Raghavan (2016): "Inherent Trade-Offs in the Fair Determination of Risk Scores," ArXiv:1609.05807 [cs].

Lakkaraju, H., J. Kleinberg, J. Leskovec, J. Ludwig, and S. Mullainathan (2017): "The Selective Labels Problem: Evaluating Algorithmic Predictions in the Presence of Unobservables," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Halifax NS Canada: ACM, 275–284.

Leckie, G. and K. Maragkou (2024): "Student sociodemographic and school type differences in teacher-predicted vs. achieved grades for university admission," *Higher Education*, 88, 2291–2317.

McEwan, P. J., S. Rogers, and A. Weerapana (2021): "Grade Sensitivity and the Economics Major at a Women's College," *AEA Papers and Proceedings*, 111, 102–106.

Melrose, K. and R. Mead (2018): "AS and A level decoupling: Implications for the maintenance of AS standards," Tech. Rep. Ofqual/18/6378/7.

Mountjoy, J. (2025): "Marginal Returns to Public Universities," *Quarterly Journal of Economics, forthcoming*.

Murphy, R. and G. Wyness (2020): "Minority report: the impact of predicted grades on university admissions of disadvantaged groups," *Education Economics*, 28, 333–350.

Office for National Statistics (2023): "Longitudinal Education Outcomes SRS Iteration 2 Standard Extract - England," .

Partington, R. (2011): "Predictive Effectiveness of Metrics in Admission to the University of Cambridge," Tech. rep.

Riehl, E. (2023): "Do less informative college admission exams reduce earnings inequality? Evidence from Colombia," *Journal of Labor Economics*.

Rothstein, J. M. (2004): "College performance predictions and the SAT," *Journal of Econometrics*, 121, 297–317.

Saboe, M. and S. Terrizzi (2019): "SAT optional policies: Do they influence graduate quality, selectivity or diversity?" *Economics Letters*, 174, 13–17.

Sacerdote, B., D. O. Staiger, and M. Tine (2025): "How Test Optional Policies in College Admissions Disproportionately Harm High Achieving Applicants from Disadvantaged Backgrounds," *NBER Working Paper Series*.

Sun, L. and S. Abraham (2021): "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," *Journal of Econometrics*, 225, 175–199.

UCAS (2015): "Unpacking Qualification Reform: Results from the UCAS survey on A level reform," Tech. rep.

——— (2016): "UCAS A level Survey 2016 Update," Tech. rep.

——— (2017): "UCAS Qualification Provision Survey 2017," Tech. rep.

——— (2018): "UCAS Qualification Provision Survey 2018," Tech. rep.

US Department of Education (2024): "Student Access to and Enrollment in Mathematics, Science, and Computer Science Courses and Academic Programs in U.S. Public Schools," New Data Release.

Zimmerman, S. D. (2014): "The Returns to College Admission for Academically Marginal Students," *Journal of Labor Economics*, 32, 711–754.

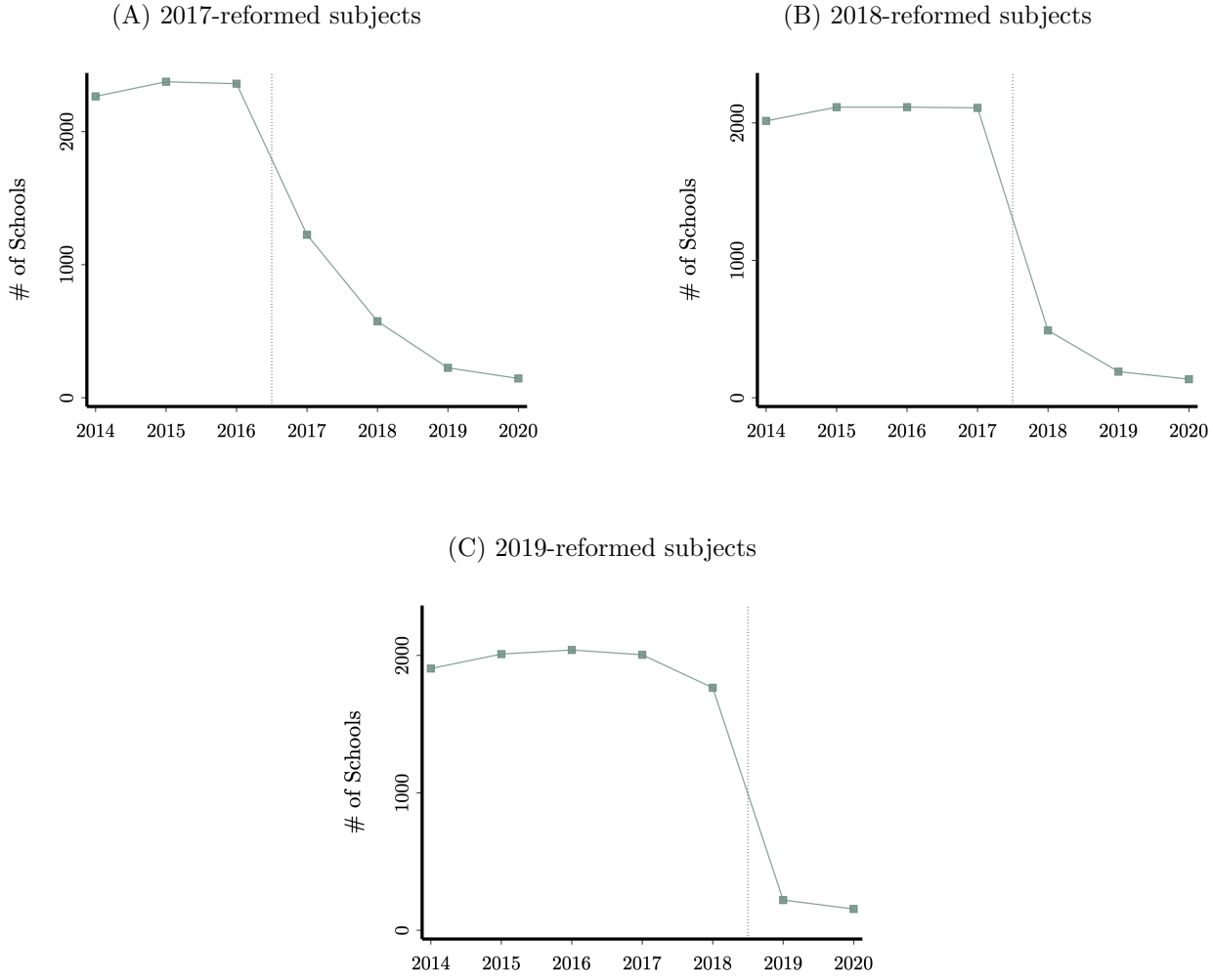
Exhibits

Table 1: High School Students in England (2010-2016 graduating cohorts)

	All	Academic Track	Nonacademic Track	Low-income	High-income
	(1)	(2)	(3)	(4)	(5)
Panel A: Education Outcomes					
Academic Track (A-levels)	0.28	–	–	0.16	0.38
Applies to University	0.48	0.93	0.31	0.38	0.57
Attends University	0.45	0.89	0.28	0.35	0.53
On-time Degree (within 3 yrs)	0.34	0.77	0.17	0.23	0.42
First-class Degree (top 20%)	0.09	0.23	0.03	0.05	0.12
First or 2:1 Degree (top 80%)	0.26	0.64	0.11	0.16	0.34
Panel B: GCSE Percentiles					
Math	50	77	40	41	57
English Language	50	77	40	41	57
English Lit	50	72	38	41	55
N	4,466,505	1,247,160	3,219,340	1,824,285	1,587,375

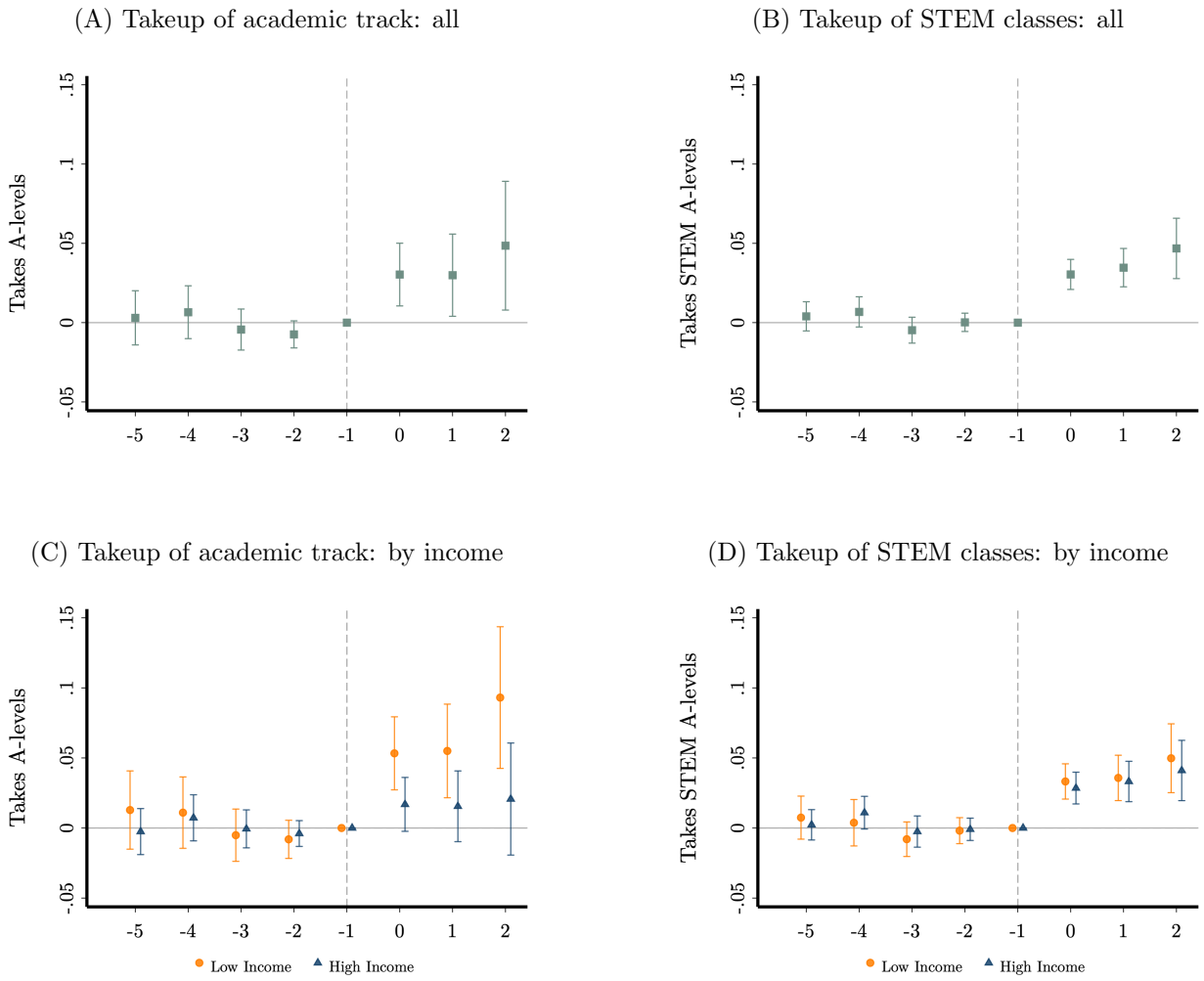
Notes. Summary statistics for students who complete GCSEs (mandatory National Curriculum until age 16) between 2008-2014 in England and graduate between 2010-2016 at age 18. Low-income students are defined as students in neighborhoods in the bottom two quintiles of the Index of Multiple Deprivation (IMD) and high-income students are defined as students in neighborhoods in the top two quintiles of the Index of Multiple Deprivation (IMD). Academic-track students are defined as students taking a complete course of A-levels (3 or more subjects). Nonacademic-track students are students taking vocational courses or a combination of vocational courses and fewer than 3 A-level subjects. GCSE percentiles are calculated within subject and qualification year.

Figure 1: School Requirements of Pre-application Standardized Tests (AS-levels)



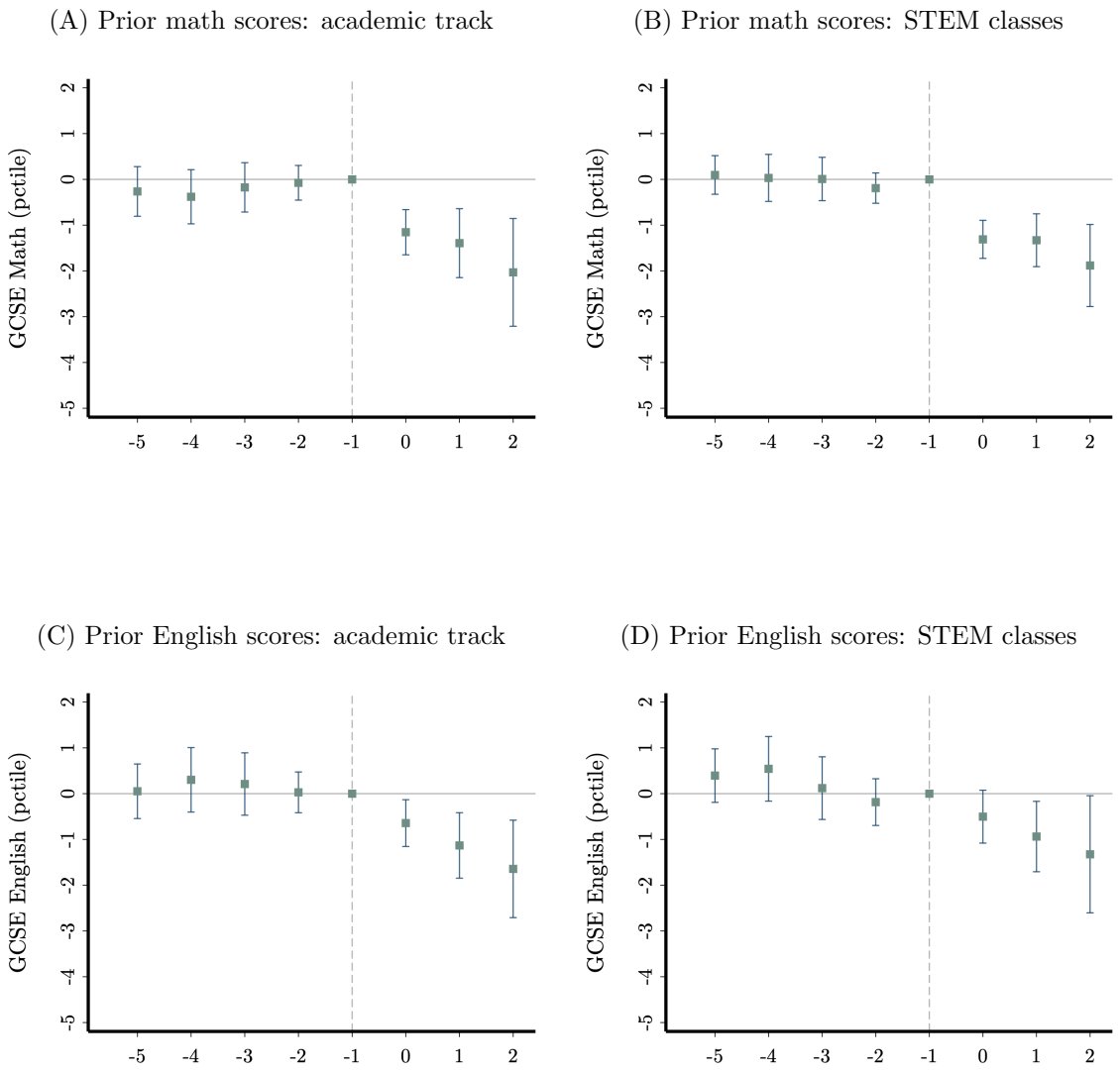
Notes. This figure shows the number of high schools in England that required pre-application exams (AS-levels) in reformed academic-track (A-level) subjects between 2014–2020. Panel A shows schools that required exams in 2017-reformed subjects, Panel B for 2018-reformed subjects, and Panel C for 2019-reformed subjects. We infer school-level requirements when more than 95% of academic-track (A-level) students in reformed subjects are observed to take pre-application exams in a given year at a school. This sample is restricted to schools with more than 5 students per year taking academic-track classes (A-levels) in the relevant subjects and students who are age 18 when they graduate high school. Table C1 presents the full list of subjects by year of reform.

Figure 2: Shifts into Academic-track and STEM Classes



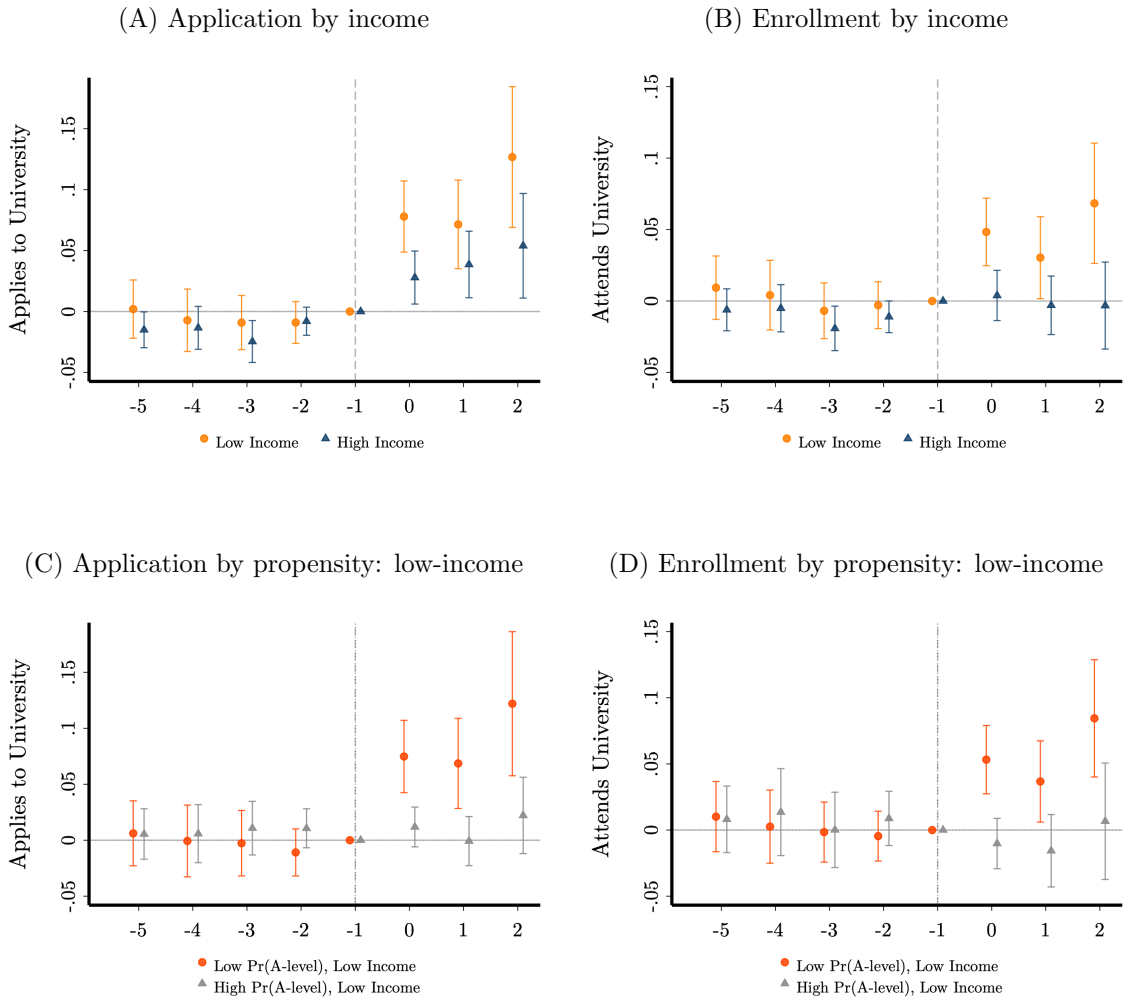
Notes. This figure presents estimates from Equation 1 using the estimator from Sun and Abraham (2021) on a sample of students 18 years of age at the end of high school in England between 2010-2019. Panels A and C present estimates when the outcome is an indicator for whether a student takes academic-track classes in the last two years of high school, defined as taking at least three or more A-level classes. Panels B and D present estimates when the outcome is an indicator for whether a student took a STEM class (A-levels in Biology, Chemistry, or Physics). In each plot, the teal square marker indicates estimates for all students, the orange circular marker restricts to students in neighborhoods in the bottom two quintiles of the Index of Multiple Deprivation (IMD), and the blue triangular marker restricts to students in neighborhoods in the top two quintiles of IMD. Treatment is at the school level and staggered by the year a school stops offering pre-application exams (AS-levels) in academic-track classes. Standard errors are clustered at the school level.

Figure 3: Negative Selection into Academic-track and STEM Classes



Notes. This figure presents estimates from Equation 1 using the estimator from Sun and Abraham (2021) on a sample of students 18 years of age at the end of high school in England between 2010-2019. The outcome is students' average prior math scores (GCSE Mathematics percentiles) in Panels A and B and students' average prior English scores (GCSE English percentiles) in Panels C and D. Treatment is at the school level. The sample is restricted to students taking academic-track classes (A-levels) in Panels A and C and taking STEM classes (A-levels in Biology, Chemistry, or Physics) in Panels B and D. Treatment is at the school level and staggered by the year a school stops offering pre-application exams (AS-levels) in academic-track classes. Standard errors are clustered at the school level.

Figure 4: Shifts into University by Income and Predicted Academic-track Propensity



Notes. This figure presents estimates from Equation 1 using the estimator from Sun and Abraham (2021) on a sample of students 18 years of age at the end of high school in England between 2010-2019. Panels A and C present estimates where the outcome is an indicator for whether a student applies to university at age 18. Panels B and D present estimates where the outcome is an indicator for whether a student enrolls in university at age 18. In Panels A and B, the orange circular marker restricts to students in neighborhoods in the bottom two quintiles of the Index of Multiple Deprivation (IMD), and the blue triangular marker restricts to students in neighborhoods in the top two quintiles of IMD. Panels C and D present estimates for low-income students separately by students with above-median predicted propensity of taking academic-track classes (A-levels) and below-median predicted propensity. Predicted propensities are based on predictions from Equation 3. Treatment is at the school level and staggered by the year a school stops offering pre-application exams (AS-levels) in academic-track classes. Standard errors are clustered at the school level.

Table 2: Marginal Low-income University Entrants

	Low-income Compliers (1)	Low-income University Students (2)
Panel A: Student Characteristics		
White	64%	76%
Female	65%	56%
First Generation	64%	48%
Below-median Math GCSE	54%	28%
Panel B: University Characteristics and Performance		
Avg. Ranking of University	98	74
Drops Out	21%	13%
On-time Degree (within 3 yrs)	41%	66%

Notes. This table presents marginal mean characteristics for the complier population, i.e., low-income students induced into attending university by the reform who finished high school between 2012-2017. Column 1 presents average demographic characteristics and marginal mean university outcomes for low-income compliers. Column 2 presents the same statistics for low-income university students who entered university between 2012-2017. Marginal mean characteristics are estimated based on Equation 5 using 2SLS to control for school and year fixed effects.

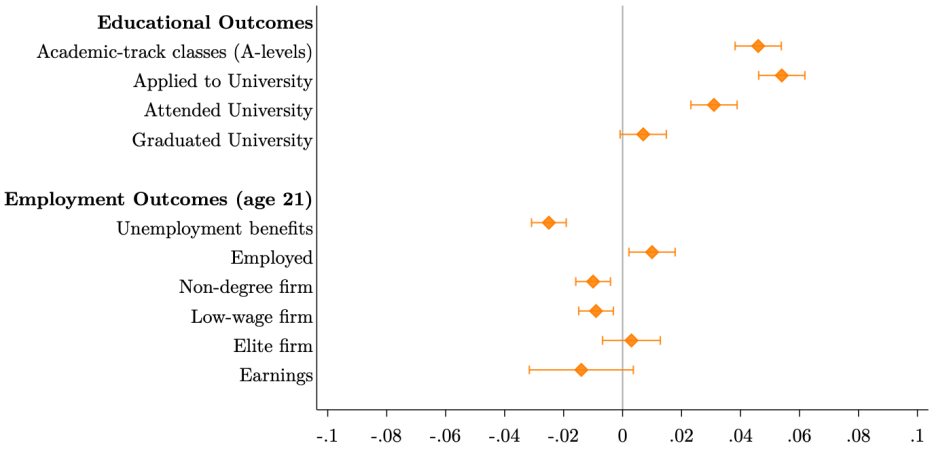
Table 3: Conditional University Performance Effects

	Any Degree *within 3 years*		First-class *Top 20%*		First-class or 2:1 *Top 80%*		Dropped Out *within 1-3 years*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$1\{t{=}{-}5\} \times$ Treated	-0.003	-0.002	-0.003	-0.001	0.005	0.007	0.000	-0.001
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.003)	(0.003)
$1\{t{=}{-}4\} \times$ Treated	-0.006	-0.005	-0.002	-0.001	-0.009*	-0.007	-0.005	-0.005
	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.004)	(0.004)
$1\{t{=}{-}3\} \times$ Treated	0.003	0.004	0.002	0.004	0.001	0.002	-0.008**	-0.008**
	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.004)	(0.004)
$1\{t{=}{-}2\} \times$ Treated	0.001	0.002	0.002	0.002	-0.002	-0.001	-0.005*	-0.005**
	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.003)	(0.003)
$1\{t{=}{-}1\} \times$ Treated	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$1\{t{=}\ 0\} \times$ Treated	-0.016***	-0.015***	-0.008**	-0.008**	-0.019***	-0.018***	0.010***	0.009***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)
$1\{t{=}\ 1\} \times$ Treated	—	—	—	—	—	—	0.017***	0.017***
							(0.004)	(0.004)
$1\{t{=}\ 2\} \times$ Treated	—	—	—	—	—	—	0.027***	0.026***
							(0.007)	(0.007)
University FE	N	Y	N	Y	N	Y	N	Y
N	485,875	485,875	485,875	485,875	485,875	485,875	652,275	652,275

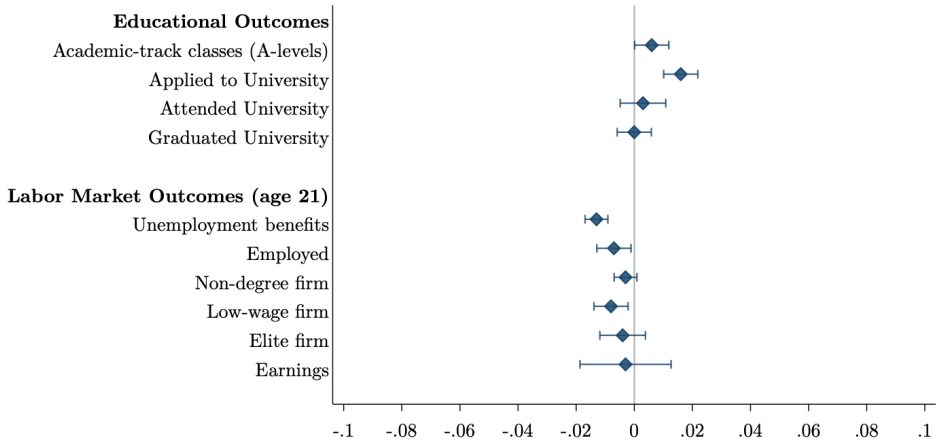
Notes. This table presents event study estimates on outcomes related to university performance. Columns 1 through 6 report estimates from Equation 6 on observed degree outcomes at age 21. Columns 7 and 8 report estimates from the staggered estimates from Equation 1 estimated using the estimator from Sun and Abraham (2021) on dropouts. Columns 1 through 6 compare university students from schools that stopped offering pre-application exams (AS-levels) beginning in academic year 2016-17 to schools that still required all students to take pre-application exams (AS-levels) in reformed subjects in 2016-17 and schools in Northern Ireland and Wales. Columns 7 and 8 compare university students from schools that stopped offering AS-levels at some year between 2017 through 2019 to schools that still required all students to take AS-levels in reformed subjects in 2019 and schools in Northern Ireland and Wales. Sample: university students who entered university between 2012-2017 (columns 1-6) and university students who started between 2012-2019 (columns 7-8). Treatment is at the school level. Standard errors are clustered at the school level.

Figure 5: Summary of Effects on Educational and Post-graduate Outcomes

(A) Low-income students



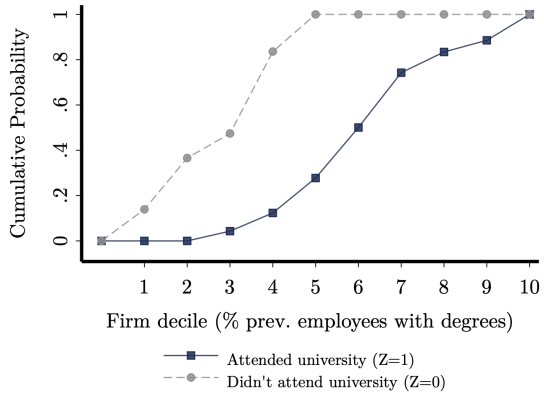
(B) High-income students



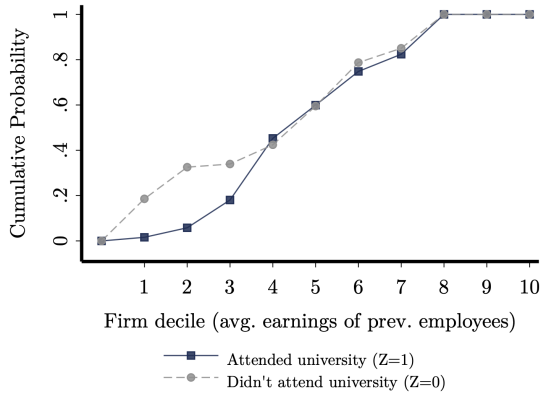
Notes. This figure reports difference-in-differences coefficients estimated in Equations 6 and 7 for educational and early-career labor market outcomes. Panel A reports estimates for low-income students, defined as students in neighborhoods in the bottom two quintiles of the Index of Multiple Deprivation (IMD). Panel B reports estimates for high-income students, defined as students in neighborhoods in the top two quintiles of the Index of Multiple Deprivation (IMD). Estimates for age-21 firm characteristics are conditional on employment. Non-degree firms are firms at which 0% of the employees hired prior to tax year 2010-11 had university degrees. Low-wage firms are firms in the bottom decile based on average earnings of employees hired prior to tax year 2010-11. Elite firms are firms where $>$ 50% of employees hired prior to tax year 2010-11 had degrees from higher-tariff universities, which are the top tercile of universities based on average grades of admits. Sample: Students age 18 at the end of high school in England between 2010-2017 in the NPD. Event study estimates are presented in Appendix Table D5.

Figure 6: Distributional Effects on Age-21 Firm Characteristics

(A) Age-21 firm decile: education



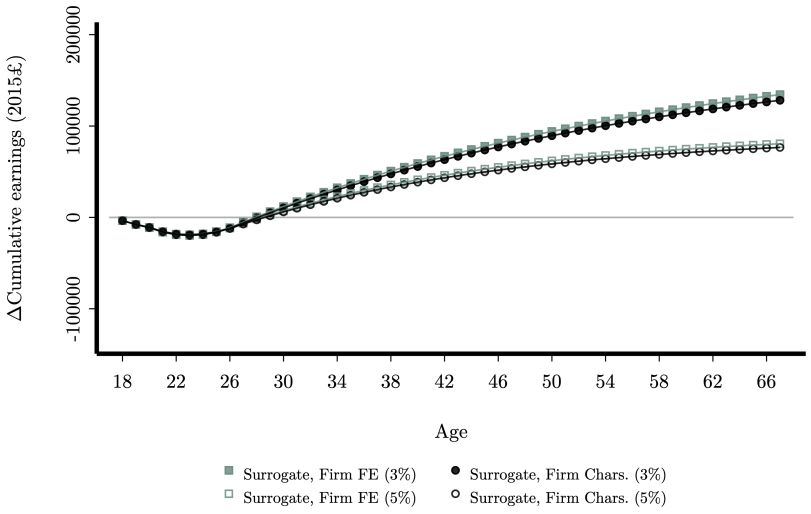
(B) Age-21 firm decile: pay



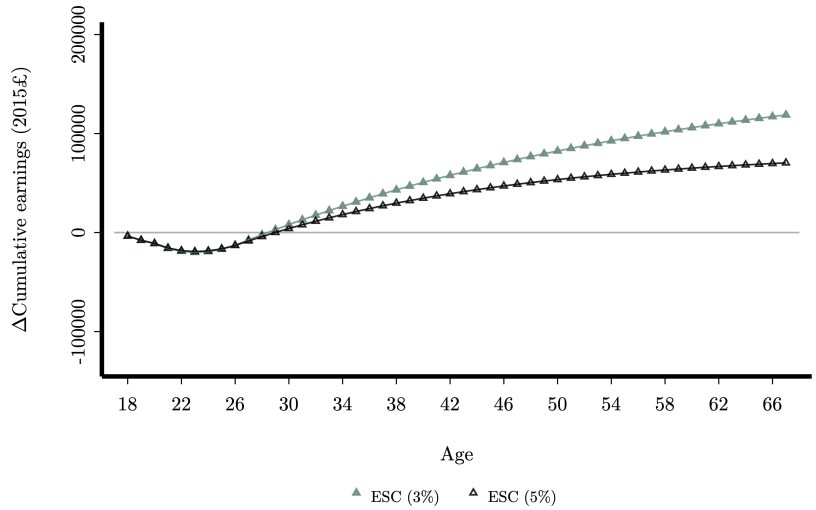
Notes. This figure presents the marginal potential outcome distributions of characteristics of the primary firm of employment four years after completing high school (age 21). In each figure, the solid blue markers report estimates from Equation 8 and the dashed gray lines report estimates from Equation 9. All estimates control for school and academic year fixed effects. Each figure compares the cumulative probability distribution for compliers induced into attending university to compliers not induced into university. Compliers are low-income students who attend university at age 18 if they attended treated schools in the years after the reform but do not attend university otherwise. Panel A presents estimates for the age-21 firm education decile, defined by the share of employees hired in tax years 2007-08 through 2010-11 with university degrees. Panel B presents estimates for the age-21 firm earnings decile, defined by the average earnings of employees hired in tax years 2007-08 through 2010-11.

Figure 7: Change in Cumulative Lifetime Earnings

(A) Surrogate projections at 3% and 5% discount rates



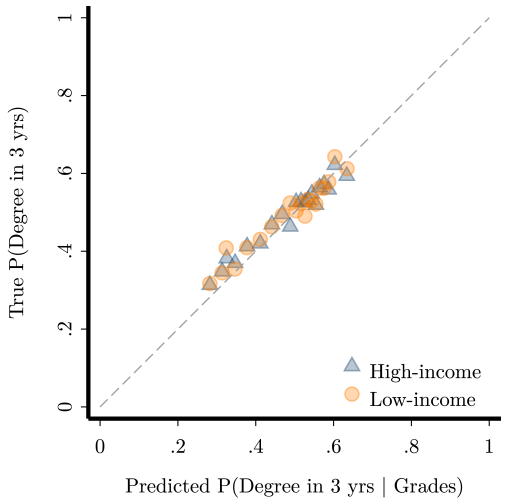
(B) ESC projections at 3% and 5% discount rates



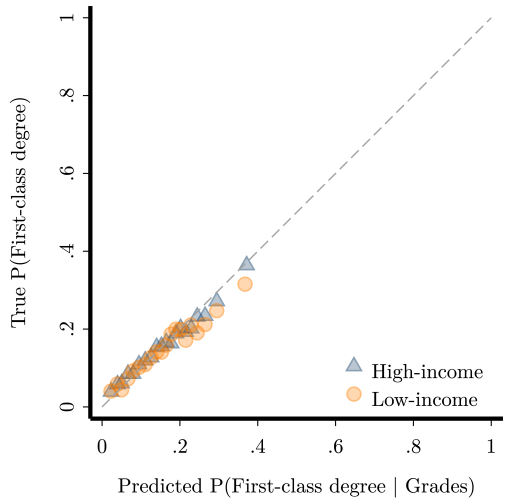
Notes. This figure presents cumulative lifetime earnings gains for low-income students discounted at 3% and 5%. Panel A presents projections using the surrogate index (Athey et al., 2025b) estimates from Equation 10 which includes age-21 firm identifiers and Equation 11 which includes age-21 firm characteristics instead. Panel B presents projections using the experimental selection correction (ESC) estimator (Athey et al., 2025a). We project age-27 earnings out until age 67. Earnings effects between ages 22 and 27 are linearly interpolated. Earnings effects for ages 18 through 21 are estimated using 2SLS on the low-income quasi-experimental sample. Earnings are deflated using the 2015 consumer price index to 2015 GBP.

Figure 8: Estimates of Calibration Bias for Marginal Entrants

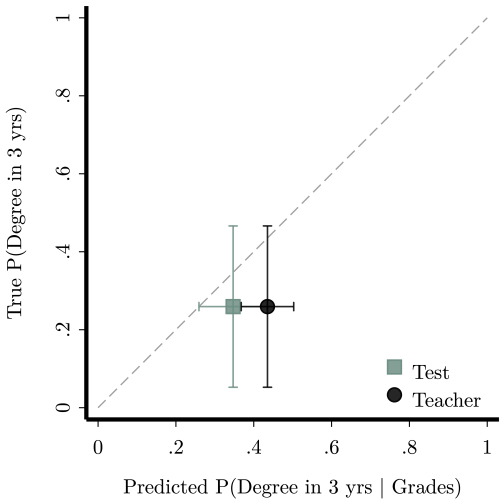
(A) On-time graduation: all entrants



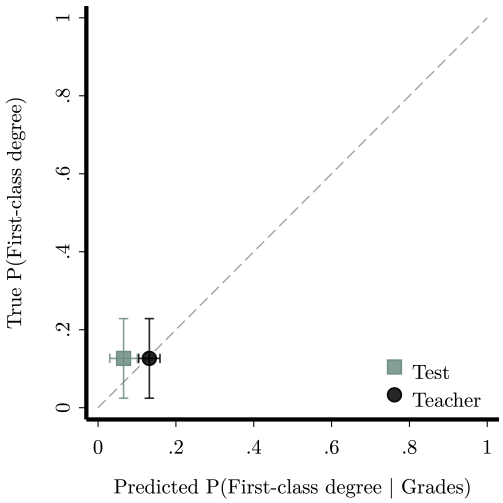
(B) First-class (top 20%): all entrants



(C) On-time graduation: marginal entrants



(D) First-class (top 20%): marginal entrants



Notes. Panels A and B report average university outcomes and predicted university outcomes based on test scores for low-income and high-income university students binned into ventiles. Panels C and D report the marginal mean university outcome, the predicted university performance based on test scores, and the predicted university performance based on teacher grades. Marginal means are estimated in Equation 12, controlling for high school and academic year fixed effects. The sample of students is a random 80% holdout sample from the set of university students who begin university between 2012-2017 and graduated high school in England at age 18. Whiskers denote 90% confidence intervals constructed from standard errors clustered at the school level.

Table 4: Variation in University Performance Explained by Tests and Teachers

	On-time Graduation (3 years)			First-class Honors (Top 20%)		
	Teacher	Test	Both	Teacher	Test	Both
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Low-income						
Adjusted R^2	0.092	0.111	0.112	0.066	0.092	0.094
N	75,535	75,535	75,535	75,535	75,535	75,535
Panel B: High-income						
Adjusted R^2	0.117	0.126	0.127	0.069	0.092	0.094
N	89,230	89,230	89,230	89,230	89,230	89,230

Notes. This table presents adjusted R^2 values for OLS regressions of university outcomes in Equation 13. Regressions in columns 1 and 4 under "Teacher" include only the average standardized teacher-assigned grades in academic-track (A-level) high school courses. Regressions in columns 2 and 5 under "Test" include only average standardized endline exam grades in academic-track (A-level) high school courses. Regressions in columns 3 and 6 under "Both" include average standardized teacher-assigned grades fully interacted with standardized exam grades. All regression specifications include university fixed effects interacted with major fixed effects. Sample: students in England who took academic-track high school classes (A-levels) and began university at age 18 between 2012–2016.

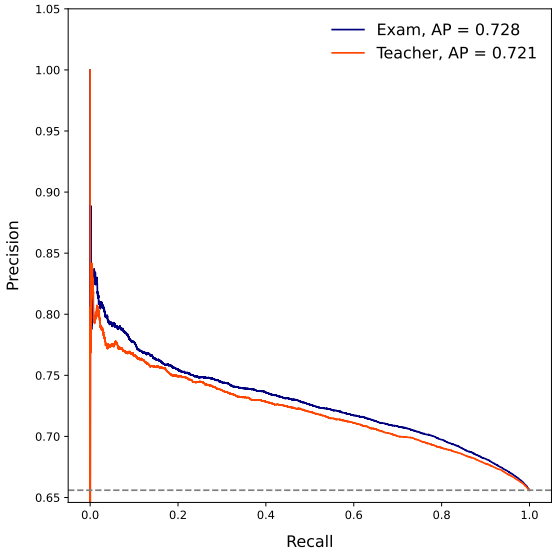
Table 5: Correlation of University Outcomes with Discrepancies between Teachers and Tests

	Applied to Selective Uni (1)	Attended Selective Uni (2)	On-time Degree *within 3 years* (3)	First-class Honors *Top 20%* (4)
Panel A: Students Taking STEM Subjects				
Teacher Grade Residual	0.081***	0.052***	-0.004***	-0.004***
	(0.001)	(0.001)	(0.001)	(0.001)
R^2	0.12	0.15	0.02	0.03
Mean Outcome	0.79	0.49	0.87	0.33
N	474,755	408,550	441,745	380,165
Panel B: Students Taking Non-STEM Subjects				
Teacher Grade Residual	0.090***	0.046***	0.005***	0.007***
	(0.001)	(0.001)	(0.001)	(0.001)
R^2	0.15	0.18	0.02	0.03
Mean Outcome	0.57	0.31	0.86	0.25
N	886,870	757,440	794,195	672,695

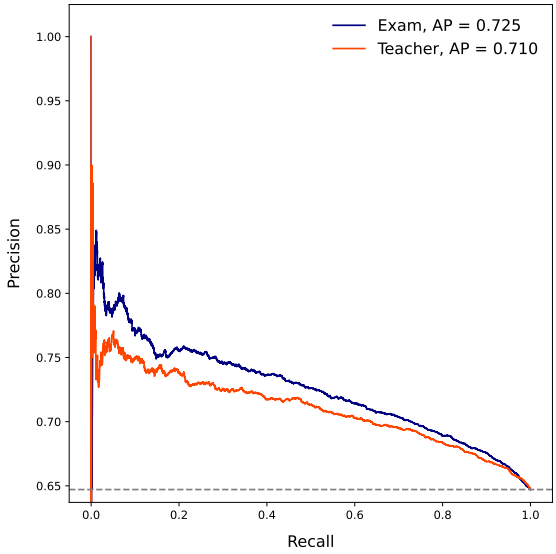
Notes. This table shows estimates from Equation 14 of university outcome and performance measures regressed on standardized residuals from regressions of teacher-assigned grades on standardized test scores. Panel A presents estimates for students who took STEM academic-track high school classes (A-levels) between 2010-2015. Panel B presents estimates for students who took non-STEM academic-track high school classes (A-levels) between 2010-2015. Sample restricted to students in England taking academic-track classes between 2010-2015 and are age 18 at the end of high school. Selective universities are defined as high-tariff universities following the UCAS classification in the corresponding year; this corresponds to approximately the top third of UK universities. All regressions include school and year fixed effects. Robust standard errors are reported in parentheses.

Figure 9: Out-of-sample Predictive Power of Teacher Grades and Test Scores

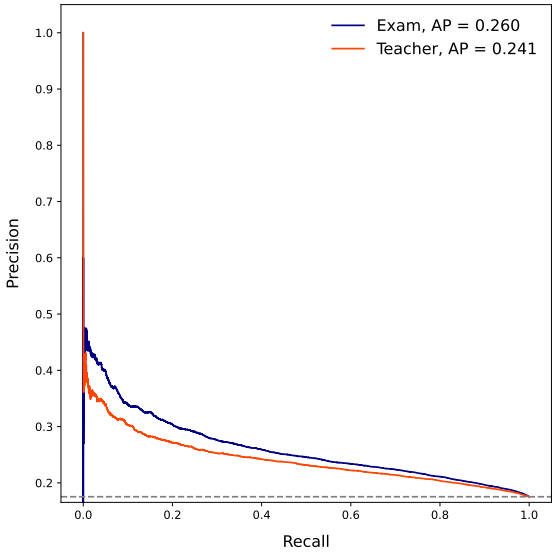
(A) On-time Degree: all students



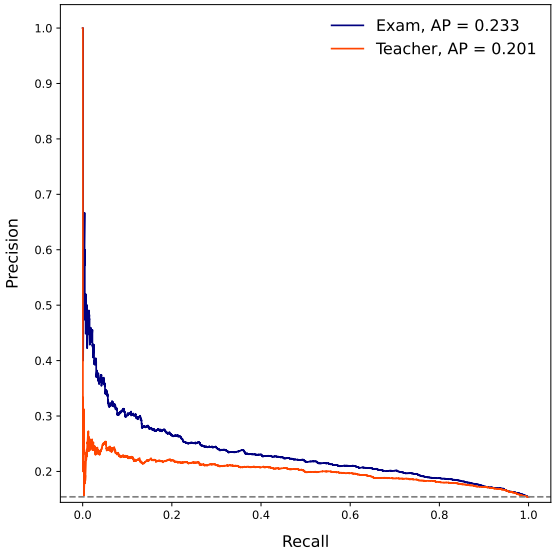
(B) On-time Degree: low-income



(C) First-class Honors: all students

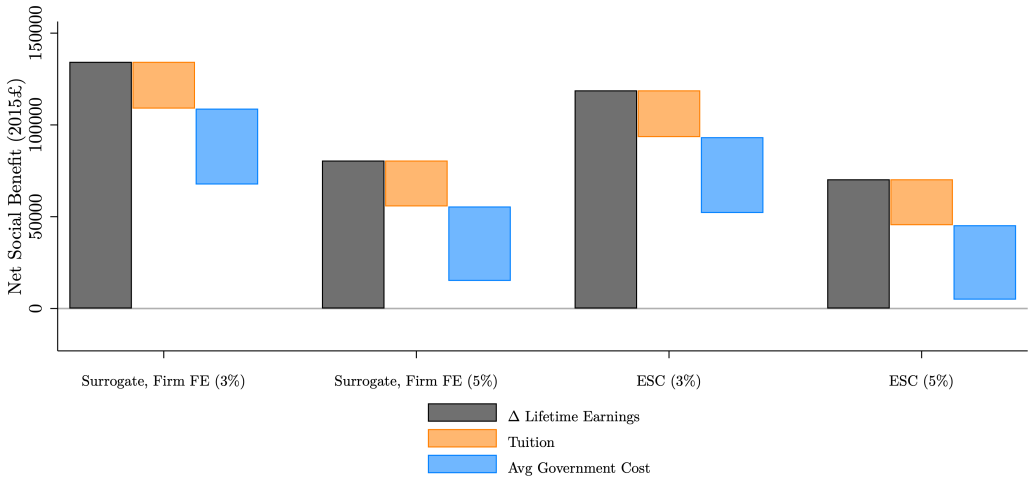


(D) First-class Honors: low-income



Notes. This figure presents precision-recall (PR) curves for out-of-sample predictions of random forest models trained on teacher-assigned grades only vs. endline A-level test scores only separately, by low-income and all students. Panels A and B show predictive power for indicators for whether students who started university between 2012-2016 graduated university within 3 years (the standard undergraduate duration in the UK). Panels C and D a show predictive power for indicators for whether students who graduated university and started between 2012-2016 graduated with a first-class degree (typically top 20% of graduating class). Average precision (AP) metrics for each model are reported. Dashed gray lines represent benchmarks for performance, which is the testing sample average positive class rate.

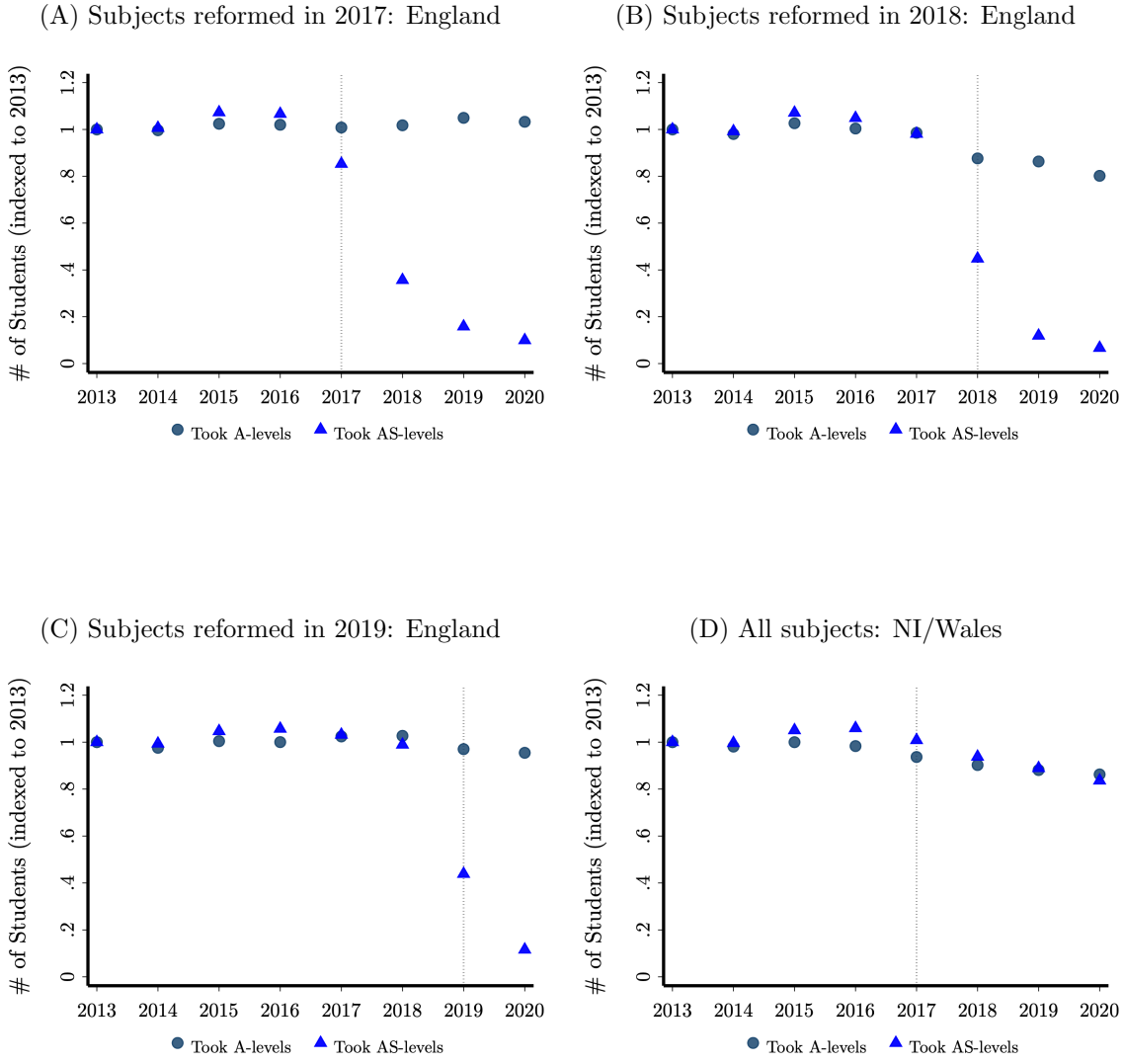
Figure 10: Net Social Benefit after Tuition (2015£)



Notes. This figure presents the net social benefit after adjusting for the average per-student government cost and tuition fees for low-income students. We report net present values discounted at 3% and 5%. Surrogate estimates present projections using the surrogate index estimated a la Athey et al. (2025b) in Equation 10, which includes age-21 firm identifiers. Total tuition amount is calculated as 3 × £9,250 which is deflated to 2015 GBP and discounted to its net present value. The average per-student cost to government is approximated as £45,000 based on estimates by (Drayton et al., 2025) and deflated to 2015 GBP.

Appendix A: Additional Exhibits

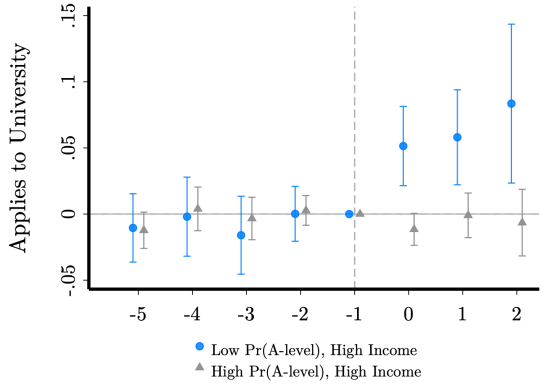
Figure A1: Academic-track (A-level) Students Taking Pre-application Exams



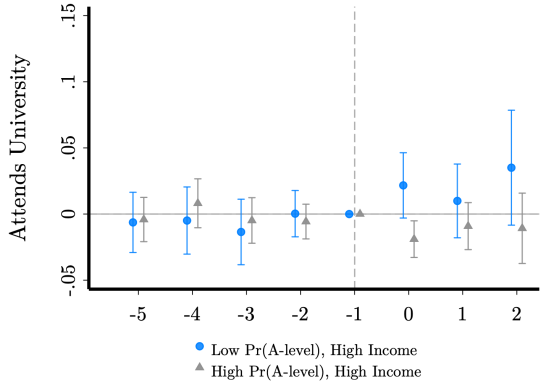
Notes. Time series of the total number of students A-level and AS-level exams taken indexed to 2013 numbers. Results presented separately by geographic region (England vs. Northern Ireland and Wales) and subject reform year. The complete list of subject titles and corresponding reform years can be found in Table C1. Panels A-C present the time series for A-level and AS-level take-up all schools in England and the A-level subjects reformed in 2017, 2018, and 2019 respectively. Panel D presents the time series of A-level and AS-level take-up for all schools in Northern Ireland and Wales across all A-level subjects. Data collected by the Joint Council for Qualifications (JCQ) for member awarding organisations (AQA, CCEA, City and Guilds, Eduqas, OCR, NCFE, Pearson and WJEC).

Figure A2: Shifts into University (High-income Students)

(A) University application: high-income

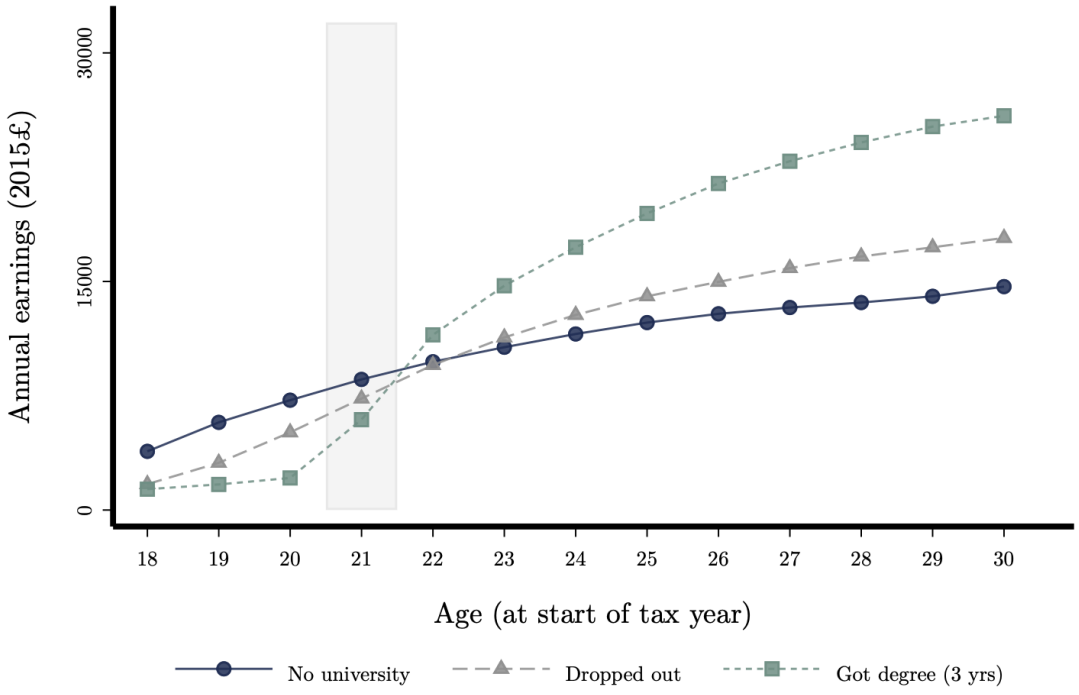


(B) University enrollment: high-income



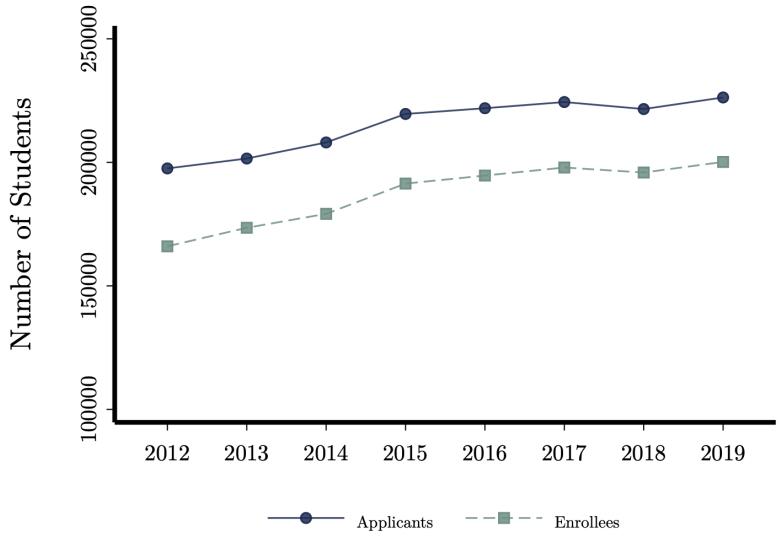
Notes. Event study estimates from Equation 1, with the outcome being an indicator for whether a student applied to university at age 18 in Panel A and an indicator for whether a student enrolled in university at age 18 in Panel B. In each plot, the blue markers restrict to high-income students, with the triangular marker additionally restricting to students with above-median predicted propensity of taking A-levels based on predictions from Equation 3 while the circular marker restricts to students with below-median predicted propensities. Treatment is at the school level. Sample: Students 18 years of age at the end of high school in England between 2010-2019 in the NPD. Standard errors are clustered at the school level.

Figure A3: Earnings Profile by Education



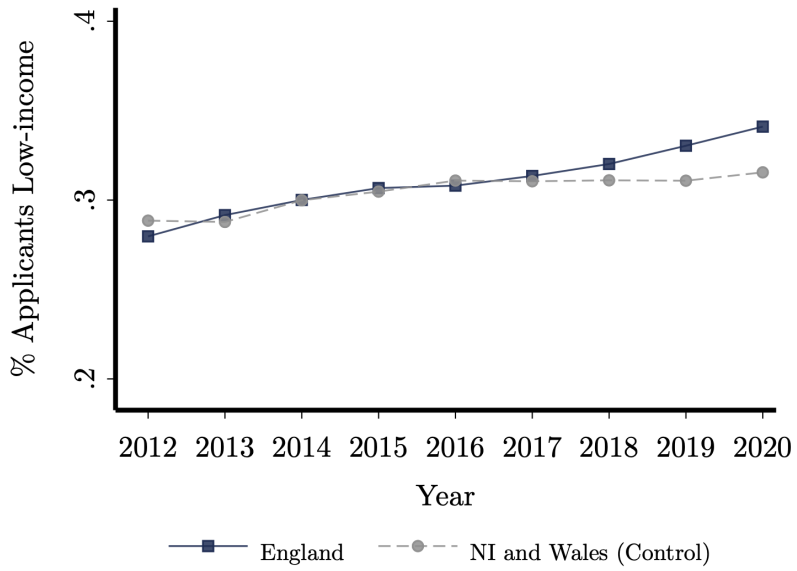
Notes. This figure shows earnings profiles for adults in England between ages 18 through 30, separately by (i) workers who never attended university; (ii) workers who attended university at 18 but never graduated; (iii) workers who attended university and completed a degree in the standard duration (i.e., within 3 years).

Figure A4: University Applications and Enrollment by Year



Notes. This figure shows a time series illustrating the number of UK university applicants and enrollees over time by year of enrollment and application. Sample restricted to students 18 years of age at time of enrollment who applied from England observed in UCAS.

Figure A5: Low-income University Applicants



Notes. Time series illustrating the share of university applicants in England, Northern Ireland, and Wales who were low-income between 2012-2020.

Table A1: Summary Statistics of University Applicants

	2012–2016 (1)	2017–2019 (2)
Panel A: Student Demographics		
Female	0.56	0.56
Black	0.04	0.04
Chinese	0.01	0.01
Indian	0.04	0.04
Other Asian	0.05	0.06
Panel B: Predicted Grade Outcomes		
Share of A-levels Correctly Predicted	0.37	0.33
Share of A-levels Over-predicted	0.55	0.59
Share of A-levels Under-predicted	0.08	0.08
Panel C: University Outcomes		
Number of Offers	3.80	3.95
Attends Firm Choice	0.62	0.61
Unplaced	0.15	0.14
Placed via Main Scheme	0.73	0.72
Placed via Clearing	0.11	0.13
N	1,112,608	674,662

Notes. This table presents summary statistics for students in the UK (excl. Scotland) applying to university between 2012 through 2019.

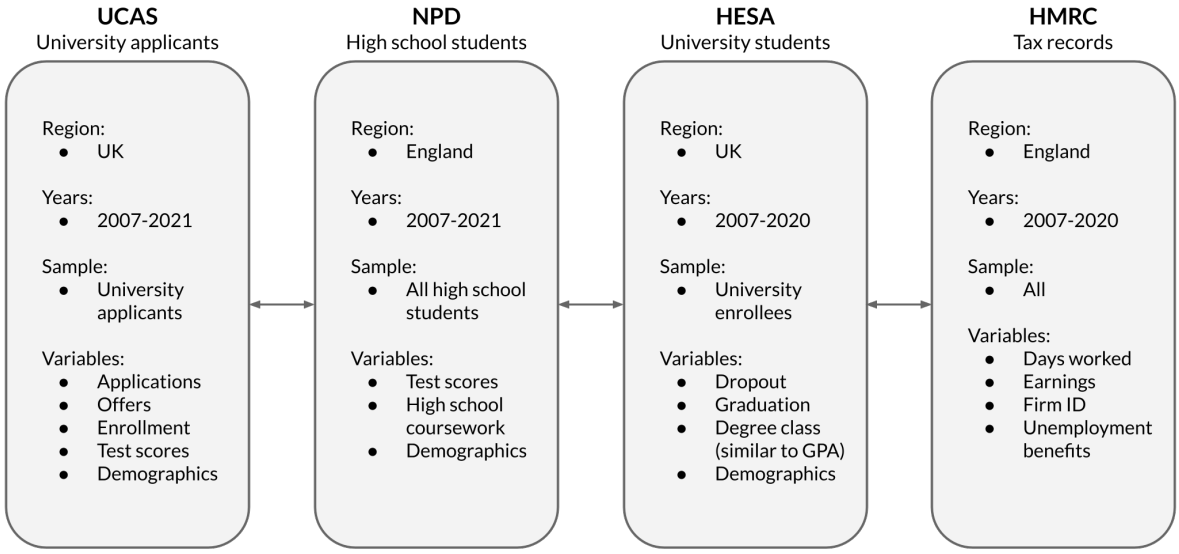
Table A2: Coefficients from Equations 10 and 11 used to project earnings

	Age 27 Earnings (GPB)	
	(1)	(2)
Earnings (Age 21)	0.418***	0.280***
	(0.003)	(0.003)
Earnings (Age 20)	0.022***	-0.023***
	(0.004)	(0.004)
Earnings (Age 19)	-0.100***	-0.066***
	(0.004)	(0.004)
Earnings (Age 18)	0.027***	0.032***
	(0.005)	(0.005)
Unemployed (Age 21)	834.926***	-185.715
	(43.904)	(694.018)
Unemployed (Age 20)	110.404***	-254.339***
	(40.181)	(42.164)
Unemployed (Age 19)	-296.462***	-402.536***
	(38.135)	(39.945)
Unemployed (Age 18)	-581.806***	-566.778***
	(33.955)	(35.644)
Out-of-work Benefits (Age 21)	-741.278***	-1,429.815***
	(39.864)	(42.151)
Out-of-work Benefits (Age 20)	-1,658.210***	-1,945.205***
	(44.813)	(47.077)
Out-of-work Benefits (Age 19)	-1,689.799***	-1,646.122***
	(42.763)	(44.843)
Out-of-work Benefits (Age 18)	-1,282.342***	-1,371.812***
	(38.132)	(39.983)
Q1 Firm (selective uni share)	-476.004***	
	(57.563)	
Q1 Firm (uni graduates share)	501.158***	
	(59.017)	
Q4 Firm (uni graduates share)	964.910***	
	(44.852)	
Q4 Firm (selective uni share)	2,056.851***	
	(46.250)	
Bottom 10% Firm (avg. earnings)	-498.050***	
	(56.464)	
R^2	0.235	0.306
N	899,185	897,580
Outcome Mean	20,029.12	20,027.42
Firm, SIC2007 FE	N	Y

Notes. This table presents regression coefficients for Equations 10 and 11 on observed age 27 earnings in the historical sample of individuals who were age 18 at the start of tax years 2010-11 and 2011-12. These estimates were used to project age 27 earnings for the quasi-experimental sample by treating age 18-21 earnings, employment, and benefit status as surrogates (Athey et al., 2025b). Both specifications include demographic and ability controls: GCSE Math score decile fixed effects, race fixed effects, indicator for female, and parent neighborhood income quintile fixed effects.

Appendix B: Data

Figure B1: Data linkages for main analysis sample



Notes. This figure presents the key variables, time period, and sample populations in the four datasets we merged in this paper.

National Pupil Database (NPD). The NPD is an administrative dataset provided by the Department for Education with detailed educational attainment data from KS1 through KS5 for students in England. The dataset spans academic years 2001 through 2021 and includes student demographics, anonymized school identifiers, and standardized test scores. Our analysis focuses on KS4 (GCSE) coursework and test scores as well as KS5 (AS-levels, A-levels, or more vocational courses). Our main measure of socioeconomic background is students' decile of the Index of Multiple Deprivation, which is a composite measure of deprivation constructed by the UK government based on incomes, unemployment, education, health, housing, and environment. This measure is defined based on the Lower-Level Super Output Area (LSOA) of a student's home residence; LSOAs are neighborhoods with an average population of 1500, roughly equivalent to a US Census Block Group. As shorthand, we refer to students in the bottom 4 (most deprived) deciles of IMD as 'low-income' or 'from low-income neighborhoods' throughout. We map the discrete GCSE, AS-level, predicted A-level, and achieved A-level to numerical grades and generate percentiles within academic year and subject.

During our sample period, GCSE grades were assigned in either discrete letter grades (U, F, E, D, C, B, A, A*) or numerical grades between 1-9. We mapped letter GCSE grades to numerical grades as follows: a grade of U is mapped to 1, F is mapped to 2, F to 3, E to 4, D to 5, C to 6, B to 7, A to 8 and A* to 9. A-levels and AS-levels are mapped as follows: U to 0, E to 1, D to 2, C to 3, B to 4, A to 5, A* to 6. To create a comparable measure across time, we convert grades in each core subject into a percentile within each year based on the distribution of grades in that subject across all students who complete GCSEs. We then take the mean of these percentiles for each student across their core subjects. In some cases, we use the subject-specific percentiles.

We restrict our classification to the sample of 3,335 schools that existed throughout 2011-2019 and had at least five students per year taking A-levels. Another sample restriction is that we focus on KS5 students who take no more than 5 A-levels ($< 0.05\%$ of the total sample takes more than 5 A-levels). Our main analysis sample based on 90% and 1% thresholds estimated in Equations 1 and 2 covers about a third of all schools, while our analyses estimating Equation 6 has a larger set of control schools to include around 67% of the total schools.

Higher Education Statistics Agency (HESA). The HESA dataset is a student-year level dataset on the universe of students at higher education providers in the UK. Our HESA sample begins in the academic year 2006-2007 through 2019-2020, and includes students matriculating from high schools in England, Wales, Northern Ireland and Scotland. For each student, we observe their degree program (university, major, degree type), the

level of program (e.g. undergraduate or graduate), and details of when the program began and whether the student has obtained a qualification associated with that program (i.e. degree) and when the student obtained that qualification. We focus our analysis on full-time undergraduate students. For each student, we identify the first program they began and link that to the first qualification they obtained. Our analysis sample is then restricted to students whose earliest program is an undergraduate program (i.e. excluding further education colleges) that entered university sometime between 2012 through 2019.

His Majesty's Revenue and Customs (HMRC). The HMRC is an individual-year level dataset with data on earnings, and anonymized firm ID and Standard Industrial Classification (SIC2007) codes for primary firm of employment, spanning the tax years 2003-04 through 2020-21. We winsorize all earnings at the 99% and 1% levels within a tax year. We additionally inflation-adjust all earnings to 2015 GBP using the CPIH index (Consumer Prices Index including owner occupiers' housing costs) from the Office for National Statistics (https://www.ons.gov.uk/economy/inflationandpriceindices/timeseries/l55o/mm23, date of access: November 1, 2025).

Our primary measure of earnings is the total earnings in GBP received in a tax year from an individual's primary employer, conditioned on receiving positive earnings. We exclude earnings from self-employment because these are only included for tax years after 2013, meaning that it would be impossible to consistently include self-employment earnings for the entire sample. Throughout, we assume that any individual observed in the educational data in an appropriate cohort but not observed in the tax data in the year corresponding to ages above 18 years had earnings of 0 and was not employed. This procedure excludes individuals who have positive earnings but did not work in the UK in the given tax year or do not appear in the tax data for some other reason. Because we do not observe hours of employment, annual earnings may reflect part-time work or employment spells lasting less than a full year and mean earnings can be below the annual equivalent of full-time minimum wage.

We restrict our analysis to exclude individuals who work at more than 10 firms in a tax year, which correspond to less than 0.01% of observations. We generate pre-2011 firm characteristics as follows: for tax years 2006-07 through 2011-12, we define the share of total employees at each firm that attended university, graduated from university, and the share of employees who graduated from higher tariff universities. We do not consider earlier years since the HESA dataset begins in 2005-06 so we are unable to observe university attendance.

Department for Work and Pensions (DWP). The DWP data contain information about government benefits, including out-of-work benefits, that individuals in England receive each tax year, spanning from 1999-2000 through 2020-21.

Universities and Colleges Admissions Service (UCAS). This dataset includes the universe of students from England who applied to UK universities between 2007 through 2020 at age 18, including the applications, offer status, and final student decisions for all courses (university-major pairs) students applied to. The dataset also includes the academic qualifications and grades students sent to universities including achieved GCSE grades, predicted and achieved BTEC grades and A-level grades, AS-level scores, as well as other smaller more specialized qualifications that we do not focus on in this paper (such as International Baccalaureate, SQA, etc.) In the UCAS dataset, we restrict our analysis to applicants who are 18 years of age at the time of application and are applying to university for the first time, collapsing it to the student level. We generate university-level attributes in 2010, including the average A-level percentiles of matriculating students, based on the applicant and enrollee composition in the period prior to our main sample.

Appendix C: Additional Background on the UK Education System and 2015 Reform

Prior to the A-level reforms described in more detail in Section II.B, all A-level subjects consisted of two rounds of subject-specific standardized exams graded by an external agency. The first set of exams, "AS-levels", were taken at the end of Year 12, awarding students with a grade that is one of U, E, D, C, B, A, in ascending order. During the months of May and June of Year 13, A-level students took the second round of externally-assessed exams, called "A-levels", and received a grade for each subject in August. A-level grades are one of U, E, D, C, B, A, A*, in ascending order. AS-levels also comprised 50% of the final grades for the full A-levels if the student continued to study the subject for a second year, but the two exams were distinct and independently administered. BTEC grades were assigned based on coursework and were one of Unclassified, Near Pass, Pass, Merit, Distinction, and Distinction* in ascending order. Around 5% of schools also offered the International Baccalaureate (IB) courses and 25% also offered other smaller vocational classes such as the VRQ and the OCR National in 2012. Most of our analysis will mostly examine the take-up of A-levels as the academic path rather than IB classes, as less than 1% of students in the UK take IB classes.

Students in the UK apply to university through the centralized application platform on the Universities and Colleges Admissions Service (UCAS). Applicants may apply to up to five "courses", where each course is a university-major pair. Students applying to study medicine can apply to a maximum of four different universities. While some students will only apply to 4 universities in this case, most students apply to four medical courses and one related but non-medical course such as Biology. Similar to admissions in the United States, university admissions decisions in the UK are made by the universities based on the student's application profile without knowledge of the other courses a student has applied to. In contrast to the US system, however, there is no marginal financial cost of application after the second course, and the total cost of applying to 5 courses is relatively small – in 2024, the fee was £27.50 ($\approx$ US$35) in total for 5 courses. Students do not rank their applications, and students may receive offers from any number of courses within their application portfolio.

Prospective students apply between October and January during Year 13 and receive their final BTEC and A-level grades in mid-August, at the end of Year 13. However, university courses make admissions offers *before* these grades are available, in October through May of Year 13. Because of this pre-qualification admissions system, university applicants received teacher-assigned grades, called "predicted grades", in each A-level subject and/or BTEC that were sent to universities as part of their application, along with GCSE grades

and AS-level grades. Offers from universities are one of two broad types: conditional offers (which require that the final grades in mid-August exceed a threshold) or unconditional offers (which guarantee admission regardless of final results). Offer conditions may stipulate that the top three A-level grades a student achieves must meet a certain threshold (e.g. A*AA) regardless of subject, or may require certain grades in a particular A-level subject. If students fail to meet their offer conditions, the university has discretion over whether to admit the student; universities are often lenient, within reason, for students whose final grades are slightly below the condition and are required to admit any student offered admission who meets the conditions.

Applicants respond to offers between May through June, by designating at most one offer as a "firm choice" and at most one offer as an "insurance choice". Firm choices are binding for both unconditional and conditional offers – as long as an applicant meets any conditions associated with the offer in August, they must attend that course. Insurance choices are similarly binding, where applicants who do not meet the conditions for their firm choice but do meet those of their insurance are required to attend their insurance choice. Students who receive no offers or miss both their firm and insurance offers, have the option to directly apply to universities with open slots in the secondary "Clearing" process.

University tuition is uniform across universities in England and Wales. Scottish universities have no tuition associated for Scottish students, but students from England who attend these universities pay the same tuition fees that they would pay at English universities. Tuition for domestic students was capped at £3,000 from 2006—2011 and increased to £9,000 in 2012, with irregular increases thereafter (generally below the rate of inflation). Essentially all courses charge tuition fees exactly at the cap, meaning that there is no variation in tuition between universities; financial considerations thus only enter into the choice *between* universities to the extent that a student's cost of living differs between different universities. The government also provides universal income-contingent loans covering all tuition costs to all students: under the current policy regime, students pay no tuition upfront and repay 9% of their annual income in excess of £25,000 after graduating. Interest rates are linked to inflation. Any debt remaining unpaid after 40 years is canceled. The government also provides maintenance loans to cover living expenses: the amount offered depends on parental income and whether students live with their parents, and there is a supplement for students living in London. These are paid back in the same way as tuition loans.

The typical length of an undergraduate degree in the UK is three years, although a substantial minority of courses last 4 years – particularly those that embed a requirement for a year abroad or a year in industry as part of the course – and medical courses last 6 years. When students graduate, they receive a degree with an honors class based on some weighted

average of the marks they receive over the course of their degree, which can be thought of as a coarse GPA: the available classifications are first-class honors, upper second-class honors (2:1), lower second-class honors (2:2) and third-class honors. At most universities around 20–30% of students are awarded first-class honors and the next 40–50% awarded a 2:1.

Table A1 presents summary statistics on university applicants between 2012-2016 and 2017-2019. Applicants received 4 offers on average, and over 70% of applicants are matched with a university through the main phase of applicants, with just over 60% attending their firm choice. Although around 1 in 4 students miss both of their offer conditions, only 15% of applicants are unmatched with universities at the end of each application cycle as 11-13% of applicants match through Clearing. During our sample period, the share of university applicants that are low-income was gradually increasing similarly across England, Northern Ireland, and Wales prior to 2017. Figure A5 shows that after 2017, the rise in low-income student share in England outpaces the share of low-income applicants from Northern Ireland and Wales.

Between 2015 through 2018, UCAS conducted surveys of schools in England documenting substantial heterogeneity in school policies around offering the newly-optional AS-level exams and rapidly shifting policies within schools (UCAS, 2015, 2016, 2017, 2018). Similarly, Melrose and Mead (2018) at the Office of Qualifications and Examinations Regulation (Ofqual) conducted two waves of interviews with teachers, once during academic year 2016-17 and again in academic year 2017-18. Their results document that schools generally adopted one of three policies following the first set of A-level reforms in academic year 2015-16: (i) continue to *require all* A-level students to take AS-level exams in reformed subjects; (ii)*no longer offer AS-levels* altogether in reformed subjects; and (iii) offer AS-levels in some reformed subjects and/or leave the decision at student discretion. Initially, only 14% of schools surveyed by UCAS indicated they were *no longer offering AS-level exams* in reformed subjects in academic year 2015-16, but this number rose to 21% over the course of the year. Of the remaining schools, 59% continued offering AS-level exams in *all* reformed subjects. The number of schools offering AS-levels continued to decline in subsequent academic years, with only 29% of schools offering AS-levels in all subjects in 2016-17 and only 14% in 2017-18. At the same time, an increasing share of schools had stopped offering AS-levels altogether, with 36% of schools in 2016-17. By academic year 2017-18, 55% of survey schools had retired AS-levels. Schools that initially continued requiring all students to take AS-levels cited reasons such as usefulness for university admissions, exam preparation, and student motivation (Melrose and Mead, 2018):

> *"We think it's a solid grade for universities to base offers on, as opposed to just an internal school exam."* – School 1, Ofqual Survey

> *"We actually feel that for the children it's quite beneficial with exam practise."* – School 2, Ofqual Survey

Schools that immediately stopped offering AS-levels cited reasons including the cost to the school, additional time required to prepare students for the AS-level, and the potential for a "bad AS-level grade" to hurt students in university admissions (Melrose and Mead, 2018):

> *"If we were going to have students doing ASs, and one or two of them having some disappointing grades in the mix, it was not fair on them to be filling in UCAS forms and being judged alongside students who might have been not as strong as them but who were coming from schools where the policy was not to take any."* – School 3, Ofqual Survey

The eventual phasing out of AS-levels meant that even schools that initially maintained the status quo by requiring all students to take AS-levels stopped offering them entirely within a year or two. These schools, in turn, cited teaching time, policies of other schools in the area, and student stress as drivers of their decision (UCAS, 2017):

> *"We are phasing out the AS as subjects become reformed, moving to internal end of year exams, to gain more teaching time."* – School 1, UCAS 2017 Survey

> *"We offered AS qualifications for most linear subjects for 2016 exams, but decided against doing so for 2017 due to a variety of factors and an overview of what other colleges were doing locally."* – School 2, UCAS 2017 Survey

> *"Removing them allowed us to take pressure off students in the midst of growing stress and anxiety levels."* – School 3, UCAS 2017 Survey

Table C1: Elimination of Pre-application Exam Requirement (AS-levels) by Subject

Qualification Year	A-level Subjects
2017	Art and Design, Biology, Business, Chemistry, Computer Science, Economics, English Language, English Language and Literature, English Literature, History, Physics, Psychology, Sociology
2018	Classical Greek, Latin, Dance, Drama and Theatre, Geography, French, German, Spanish, Music, Physical Education, Religious Studies
2019	Accounting, Ancient History, Classical Civilisation, Design and Technology, Electronics, Environmental Science, Film Studies, Further Mathematics, Geology, Government and Politics, History of Art, Law, Mathematics, Media Studies, Chinese, Italian, Russian, Music Technology, Philosophy, Statistics

Notes. This table presents the qualification year corresponding to the first A-level exam year for each subject at which AS-levels were no longer mandatory in England. Reproduced from https://www.gov.uk/government/publications/get-the-facts-gcse-and-a-level-reform/get-the-facts-as-and-a-level-reform, accessed on July 20, 2025.

Table C2: High Schools in England by Elimination Year

	Treated Schools (Switchers)			Control Schools	Excluded Schools
	2017	2018	2019	After 2019	Mixed Policies
	(1)	(2)	(3)	(4)	(5)
Panel A: Education Outcomes					
GCSE Math	67.68	68.10	68.57	62.84	61.45
GCSE English	67.24	68.02	68.69	62.97	61.73
Attended University	0.53	0.55	0.56	0.48	0.47
Graduated University	0.27	0.28	0.29	0.25	0.25
Panel B: Demographics					
Female	0.52	0.55	0.55	0.54	0.54
Low-income	0.32	0.30	0.28	0.37	0.37
White	0.77	0.78	0.76	0.75	0.79
School size	267.63	416.90	249.19	376.76	273.02
N of Schools	560	390	115	165	2,105

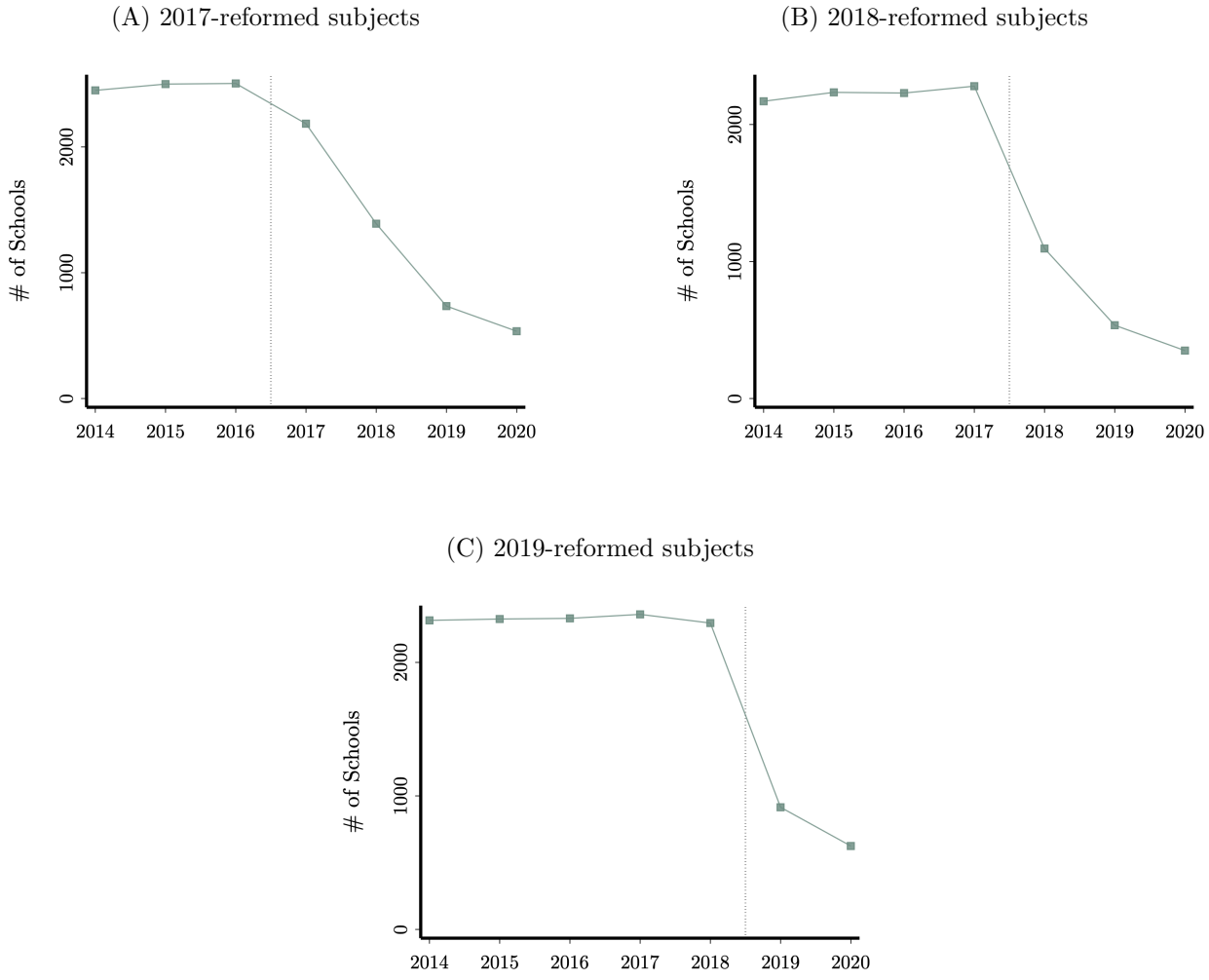
Notes. This table presents summary statistics in academic years 2011-12 through 2015-16 (pre-reform) for high schools in England by year adoption. The treated schools are schools that switched from having >90% of A-level students in reformed subjects taking AS-levels post-reform to <5% of A-level students in reformed subjects taking AS-levels post-reform in the switch year indicated. Control schools are schools that where >90% of A-level students in reformed subjects took AS-levels through 2019. Changes in A-level and AS-level take-up at each of these schools is presented in Figure C2. We exclude "mixed" schools from our analysis, where AS-levels were optional or offered for some reformed subjects but not others between 2017-2019.

Table C3: Summary Statistics by Predicted Academic-track Propensity

	Low Predicted P(A-levels)		High Predicted P(A-levels)	
	Low-income	High-income	Low-income	High-income
	(1)	(2)	(3)	(4)
Panel A: GCSE performance				
GCSE Maths percentile	27.87	31.36	68.13	70.68
GCSE English percentile	26.64	28.62	69.99	71.93
Panel B: Demographics				
Ever received free school meals	0.49	0.22	0.32	0.09
Female	0.46	0.40	0.57	0.55
White	0.78	0.91	0.63	0.87
Black	0.08	0.03	0.12	0.03
Indian	0.02	0.01	0.06	0.03
Other Asian	0.09	0.03	0.15	0.04
Mixed or Other	0.08	0.05	0.10	0.06
Panel C: Educational outcomes				
Take STEM A-level	0.02	0.03	0.26	0.29
Apply to university at 18	0.29	0.28	0.70	0.71
N	1,188,355	985,225	675,535	1,522,440

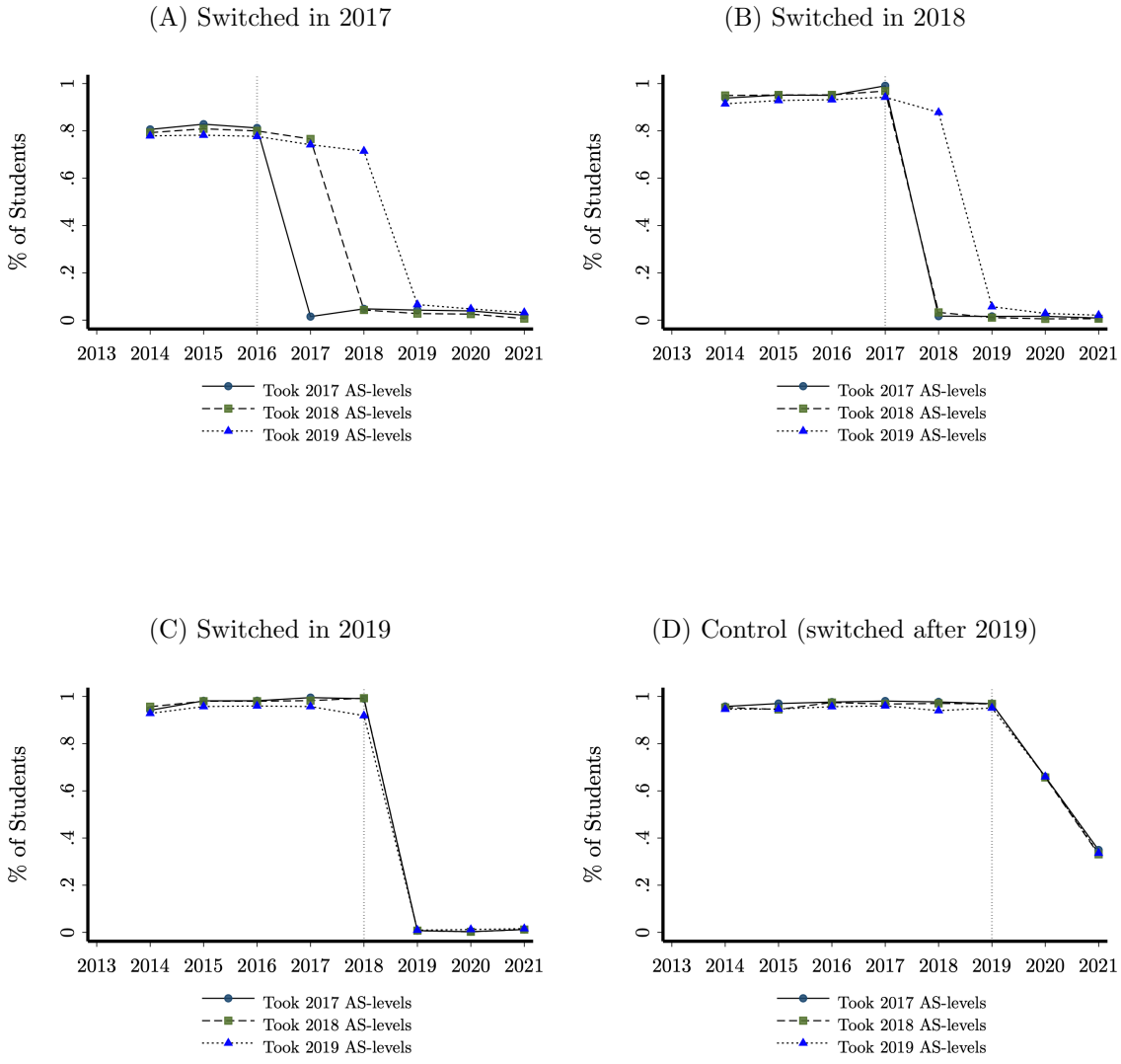
Notes. This table presents summary statistics for students across four cells based on their predicted propensity of taking A-levels as predicted by estimates from Equation 3 and income quintiles. Sample: GCSE students who completed KS4 between 2009-10 through 2016-2017.

Figure C1: Schools Offering Pre-application Exams (AS-levels) in Reformed Subjects



Notes. This figure shows the number of high schools in England that stopped offering pre-application exams (AS-levels) in reformed subjects between 2014–2020. Panel A shows schools that offered exams in 2017-reformed subjects, Panel B for 2018-reformed subjects, and Panel C for 2019-reformed subjects. We infer that a school offers reformed pre-application exams when we observe AS-levels for at least 5% of A-level student-subject pairs at the school in reformed subjects. This sample is restricted to schools with more than 5 A-level students per year in the relevant subject groups.

Figure C2: Academic-track Students with Pre-application Exams (AS-levels) by School

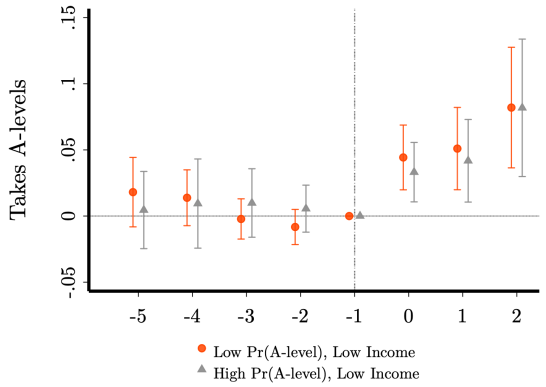


Notes. Time series of the average share of A-level students per school in England taking AS-levels in corresponding subjects, presented separately by subject reform year and school adoption year. Panel A presents the time series for schools that stopped offering AS-levels in reformed subjects beginning in academic year 2016-2017. Panel B presents the time series for schools that stopped offering AS-levels in reformed subjects beginning in academic year 2017-2018. Panel C presents the time series for schools that stopped offering AS-levels in reformed subjects beginning in academic year 2018-2019. Panel D presents the time series for schools that only stopped offering AS-levels in any subject *after* academic year 2019-2020 and are used as our within-England control sample. The complete list of subject titles and corresponding reform years can be found in Table C1.

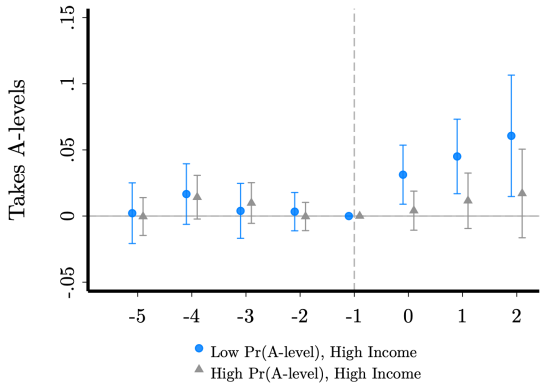
Appendix D: Additional Specifications and Robustness

Figure D1: Shifts into Academic-track by Predicted Propensity

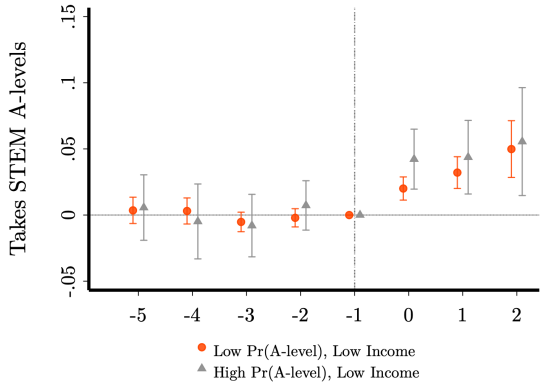
(A) Takes A-levels, Low-income



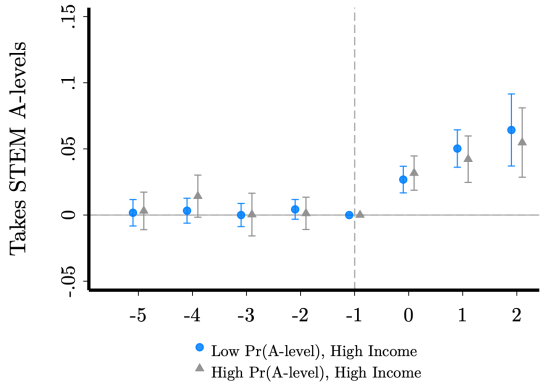
(B) Takes A-levels, High-income



(C) Takes STEM A-levels, Low-income



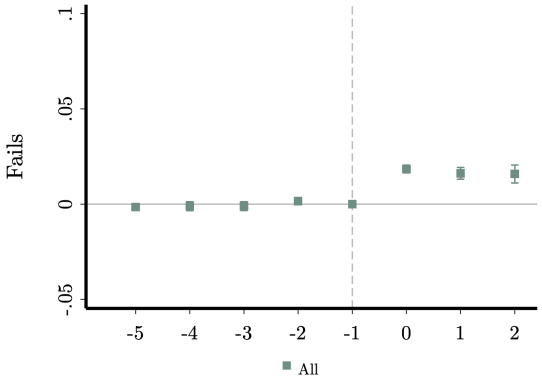
(D) Takes STEM A-levels, High-income



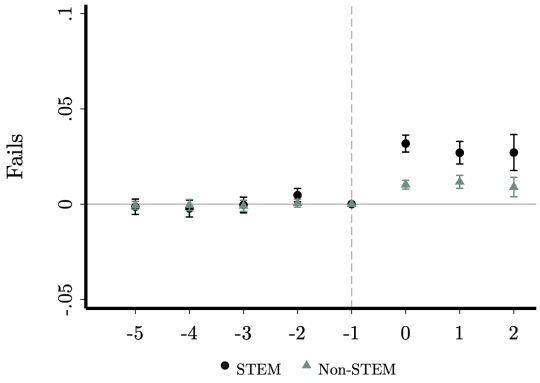
Notes. Event-study estimates from Equation 1, with the outcome being an indicator for whether a student took at least three or more A-levels in Panels A and B and an indicator for whether a student took a STEM A-level (Biology, Chemistry, Physics) in Panels C and D. In each plot, the orange markers restrict to low-income students, with the triangular marker additionally restricting to students with above-median predicted propensity of taking A-levels based on predictions from Equation 3 while the circular marker restricts to students with below-median predicted propensities. The blue markers restrict to high-income students, with the triangular marker additionally restricting to students with above-median predicted propensity of taking A-levels based on predictions from Equation 3 while the circular marker restricts to students with below-median predicted propensities. Treatment is at the school level. Sample: Students 18 years of age at the end of high school in England between 2010-2019 in the NPD. Standard errors are clustered at the school level.

Figure D2: Academic-track Performance (A-levels)

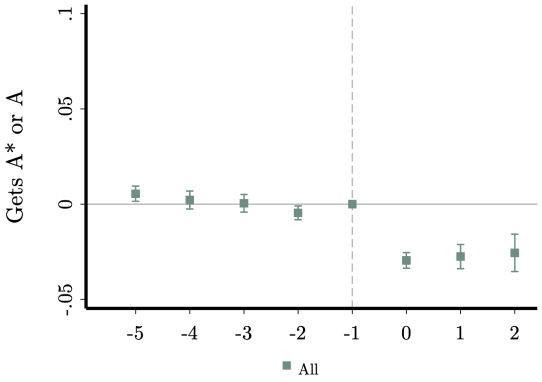
(A) Fails an A-level, pooled across all subjects



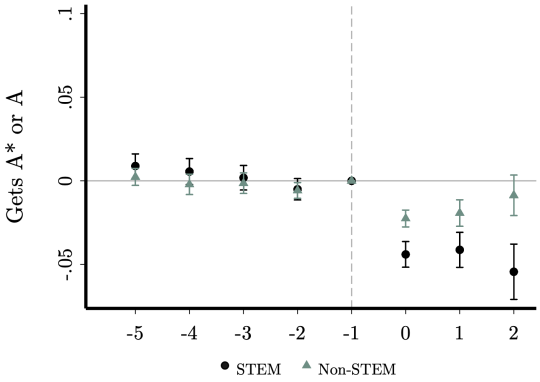
(B) Fails an A-level, by STEM vs. Non-STEM



(C) Gets an A*/A, pooled across all subjects

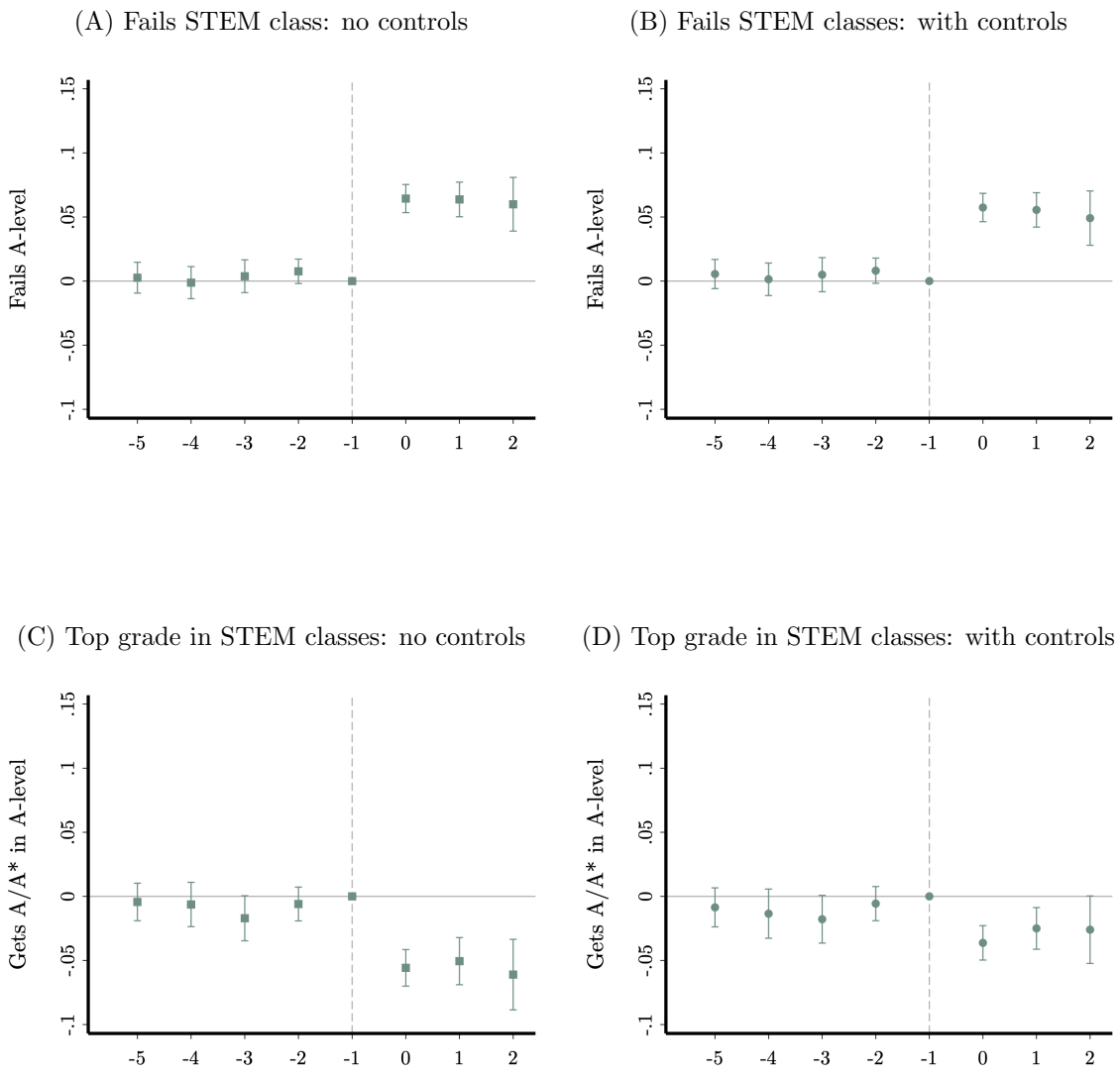


(D) Gets an A*/A, by STEM vs. Non-STEM



Notes. Event study using specification in (2), with the outcome being an indicator for whether a student failed an A-levels exam in Panels A and B and an indicator for whether a student got an A* or an A in Panels C and D. Treatment is at the school-subject level and fixed effects are included for student age. Sample: Students ages 18 or 19 at the end of high school in England between 2010-2019 applying to universities in UCAS. Standard errors are clustered at the school level.

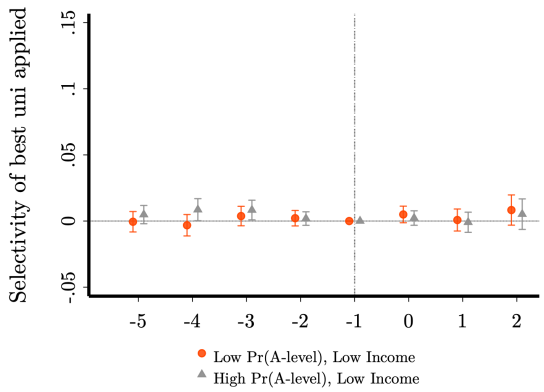
Figure D3: Academic-track Performance after Controlling for Selection



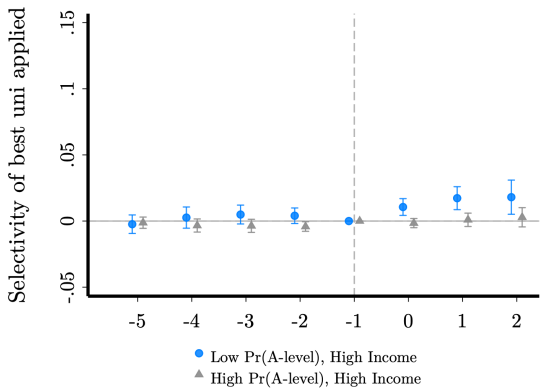
Notes. Event study using specification in (1), with the outcome being an indicator for whether a student taking STEM A-levels received a failing grade in Panels A and B and an indicator for whether a student taking STEM A-levels received a grade of A or A* in Panels C and D. Panels B and D include controls for IMD decile, gender, and GCSE Math percentiles. Sample: Students 18 years of age at the end of high school in England between 2010-2019 in the NPD. Standard errors are clustered at the school level.

Figure D4: University Application Patterns Conditional on Applying

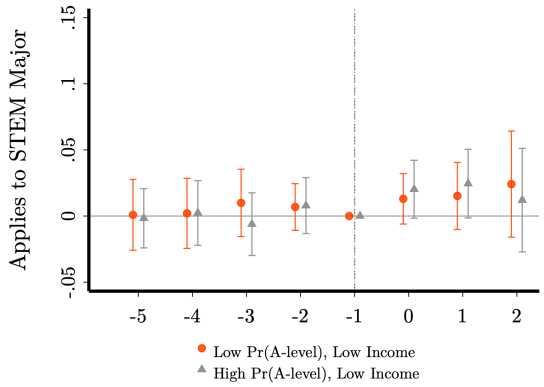
(A) Max selectivity in application portfolio: low-income



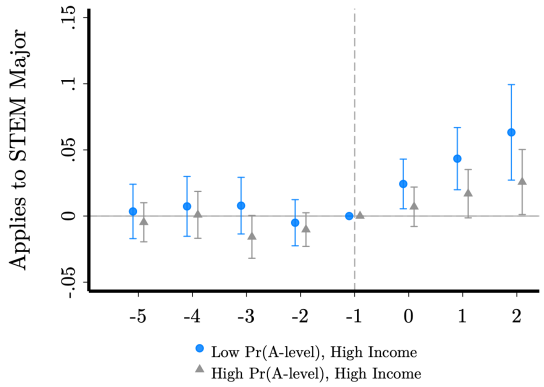
(B) Max selectivity in application portfolio: high-income



(C) Applies to STEM major: low-income



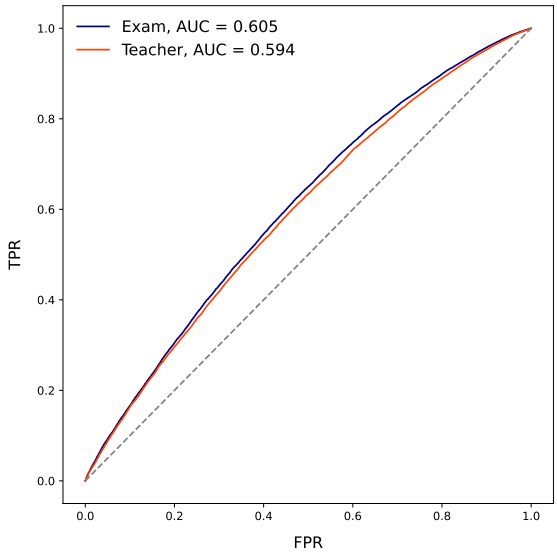
(D) Applies to STEM major: high-income



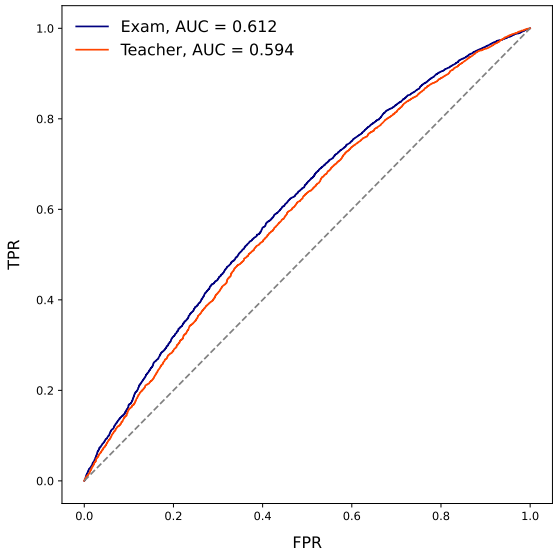
Notes. Event-study estimates from Equation 1, with the outcome being the average selectivity of the highest-ranked university in a student's application portfolio in Panels A and B and an indicator for whether a student applied to a STEM major conditional on applying to university at age 18 in Panels C and D. University selectivity is defined as the average A-level percentile of matriculating students in 2010. In each plot, the orange markers restricts to students in neighborhoods in the first two quintiles of the Index of Multiple Deprivation and bottom two terciles of predicted propensity to take A-levels. The blue triangular marker restricts to students in the top tercile of predicted propensity. Treatment is at the school level. Sample: Students 18 years of age at the end of high school in England between 2010-2019 in the NPD. Standard errors are clustered at the school level.

Figure D5: Out-of-sample Prediction of Teachers and Tests, ROC Curves

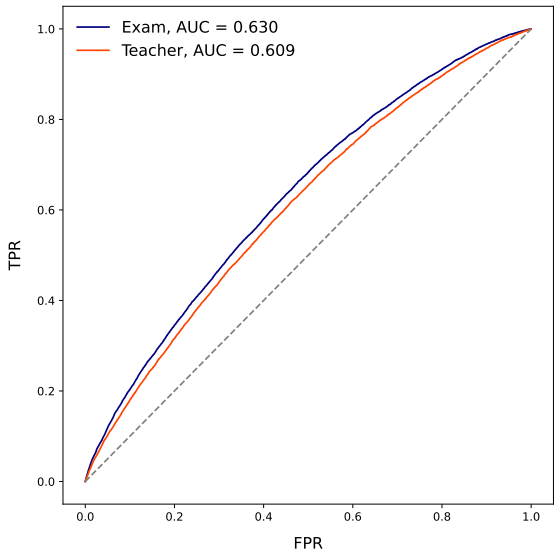
(A) Graduates (3yrs), All students



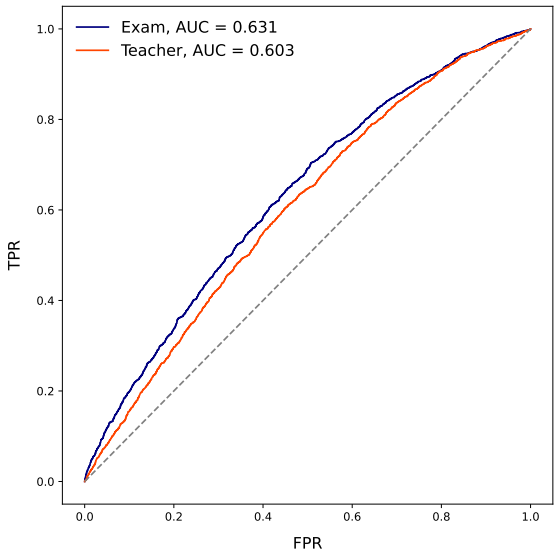
(B) Graduates (3yrs), Low-income



(C) Graduates with First, All students



(D) Graduates with First, Low-income



Notes. Receiver Operating Characteristic (ROC) curves for out-of-sample predictions of random forest models trained on teacher-assigned grades only vs. endline A-level test scores only separately by low-income and all students. Panels A and B show predictive power for indicators for whether students who started university between 2012-2016 graduated university within 3 years (the standard undergraduate duration in the UK). Panels C and D a show predictive power for indicators for whether students who graduated university and started between 2012-2016 graduated with a first-class degree (typically top 20-25% of graduating class). Area under the curve (AUC) measures are reported for each model fit. Dashed gray lines represent benchmark for performance, the 45-degree line in the ROC curves.

Table D1: Estimates of Vertical Mismatch

	High-income			High-income, High propensity		
	Top 10	Top 20	Top 50	Top 10	Top 20	Top 50
	(1)	(2)	(3)	(4)	(5)	(6)
$1\{t$=-5$\} \times$ Treated	-0.005	-0.003	-0.002	-0.006	-0.001	-0.005
	(0.003)	(0.006)	(0.007)	(0.004)	(0.007)	(0.009)
$1\{t$=-4$\} \times$ Treated	0.000	0.001	0.003	-0.000	0.003	0.001
	(0.004)	(0.006)	(0.008)	(0.005)	(0.008)	(0.010)
$1\{t$=-3$\} \times$ Treated	-0.002	-0.010*	0.002	-0.003	-0.009	0.002
	(0.003)	(0.006)	(0.007)	(0.005)	(0.008)	(0.009)
$1\{t$=-2$\} \times$ Treated	0.000	-0.009**	-0.006	-0.001	-0.009	-0.004
	(0.003)	(0.005)	(0.005)	(0.004)	(0.006)	(0.006)
$1\{t$=-1$\} \times$ Treated	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$1\{t$=0$\} \times$ Treated	-0.007**	-0.013***	-0.009	-0.009**	-0.013**	-0.010
	(0.003)	(0.005)	(0.006)	(0.004)	(0.006)	(0.008)
$1\{t$=1$\} \times$ Treated	-0.008*	-0.012*	-0.015	-0.008	-0.012	-0.017
	(0.004)	(0.007)	(0.009)	(0.005)	(0.009)	(0.011)
$1\{t$=2$\} \times$ Treated	-0.005	-0.015	-0.014	-0.002	-0.003	0.003
	(0.006)	(0.011)	(0.013)	(0.008)	(0.015)	(0.018)
N	241,385	241,385	241,385	170,650	170,650	170,650

Notes. This table presents OLS coefficients from event study regressions in Equation 6 on the rank of university attended. High-propensity students are students with above median predicted propensity to take academic-track classes prior to the reform as predicted by Equation 3. We construct university rankings using the average standardized A-level grades of university attendees in the 2011 cohort. Sample: Students age 18 at the end of high school in England between 2010-2017 in the NPD who attended university at age 18. High-income students are students who are in the top two quintiles of neighborhood income. Standard errors are clustered at the school level.

Table D2: University Graduation Effects, Robustness by Control Group

	Any Degree within 3 yrs		
Control Group:	NI/Wales	England	Placebo
	(1)	(2)	(3)
$1\{t=-5\}\times 1\{\text{Treated}\}$	-0.012	-0.001	-0.013
	(0.009)	(0.005)	(0.010)
$1\{t=-4\}\times 1\{\text{Treated}\}$	-0.007	-0.005	-0.002
	(0.009)	(0.004)	(0.010)
$1\{t=-3\}\times 1\{\text{Treated}\}$	-0.003	0.004	-0.007
	(0.009)	(0.004)	(0.010)
$1\{t=-2\}\times 1\{\text{Treated}\}$	-0.002	0.002	-0.005
	(0.009)	(0.004)	(0.009)
$1\{t=-1\}\times 1\{\text{Treated}\}$	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
$1\{t=0\}\times 1\{\text{Treated}\}$	-0.022***	-0.014***	-0.008
	(0.008)	(0.004)	(0.008)
Outcome Mean	0.71	0.72	0.71
N	215,680	455,542	289,901

Notes. This table presents robustness results for estimates from Equation 6. Column 1 compares university students from schools in England that eliminated AS-levels in 2017 to students from Wales and Northern Ireland. Column 2 compares university students from schools in England that eliminated AS-levels in 2017 to schools in England that eliminated AS-levels in 2018 or later. Column 3 presents placebo results comparing university students from schools in England that eliminated AS-levels in 2018 or later to students from Wales and Northern Ireland. All regressions include school, year, and university fixed effects. Sample: university students who started an undergraduate program in the UK between 2012 through 2017.

Table D3: University Performance Effects, Robustness by Control Group

	First-class Honors or 2:1 *Top 80%*		First-class Honors *Top 20%*	
Control Group:	NI/Wales	England	NI/Wales	England
	(4)	(5)	(6)	(7)
$1\{t=-5\}\times 1\{\text{Treated}\}$	0.003	0.007	-0.003	-0.001
	(0.011)	(0.005)	(0.008)	(0.003)
$1\{t=-4\}\times 1\{\text{Treated}\}$	-0.008	-0.007	0.001	-0.001
	(0.010)	(0.005)	(0.008)	(0.003)
$1\{t=-3\}\times 1\{\text{Treated}\}$	-0.002	0.003	0.005	0.004
	(0.010)	(0.005)	(0.008)	(0.003)
$1\{t=-2\}\times 1\{\text{Treated}\}$	0.002	-0.001	0.002	0.002
	(0.009)	(0.004)	(0.008)	(0.003)
$1\{t=-1\}\times 1\{\text{Treated}\}$	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
$1\{t=0\}\times 1\{\text{Treated}\}$	-0.037***	-0.016***	-0.018**	-0.006*
	(0.009)	(0.005)	(0.008)	(0.003)
Outcome Mean	0.57	0.58	0.21	0.21
N	215,680	456,790	215,680	456,790

Notes. This table presents robustness results for estimates from Equation 6. Columns 1 and 3 compare university students from schools in England that eliminated AS-levels in 2017 to students from Wales and Northern Ireland. Columns 2 and 4 compare university students from schools in England that eliminated AS-levels in 2017 to schools in England that eliminated AS-levels in 2018 or later. All regressions include school, year, and university fixed effects. Sample: university students who started an undergraduate program in the UK between 2012 through 2017.

Table D4: University Performance Effects, Heterogeneity by Grade Inflation

	Any Degree within 3 yrs	First-class or 2:1 *Top 80%*	First-class *Top 20%*
	(1)	(2)	(3)
Panel A: Low Prior Grade Inflation			
$1\{t=-5\} \times 1\{\text{Treated}\}$	-0.010	0.000	-0.005
	(0.010)	(0.012)	(0.010)
$1\{t=-4\} \times 1\{\text{Treated}\}$	-0.010	-0.016	0.001
	(0.010)	(0.011)	(0.009)
$1\{t=-3\} \times 1\{\text{Treated}\}$	-0.003	-0.003	0.012
	(0.009)	(0.011)	(0.009)
$1\{t=-2\} \times 1\{\text{Treated}\}$	0.001	0.000	0.010
	(0.009)	(0.011)	(0.009)
$1\{t=-1\} \times 1\{\text{Treated}\}$	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
$1\{t=0\} \times 1\{\text{Treated}\}$	-0.009	-0.024**	-0.008
	(0.009)	(0.010)	(0.009)
N	104,190	104,660	104,660
Panel B: High Prior Grade Inflation			
$1\{t=-5\} \times 1\{\text{Treated}\}$	-0.009	-0.016	-0.016*
	(0.014)	(0.012)	(0.010)
$1\{t=-4\} \times 1\{\text{Treated}\}$	-0.003	-0.013	-0.010
	(0.013)	(0.011)	(0.009)
$1\{t=-3\} \times 1\{\text{Treated}\}$	0.012	-0.015	-0.015
	(0.013)	(0.011)	(0.009)
$1\{t=-2\} \times 1\{\text{Treated}\}$	0.002	-0.008	-0.014
	(0.012)	(0.011)	(0.009)
$1\{t=-1\} \times 1\{\text{Treated}\}$	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
$1\{t=0\} \times 1\{\text{Treated}\}$	-0.030***	-0.051***	-0.025***
	(0.012)	(0.010)	(0.008)
N	103,673	108,725	108,725

Notes. This table presents estimates from Equation 6 comparing university students from schools in England that eliminated AS-levels in 2017 to students from Wales/NI. Panel A presents estimates for schools in England with below-median overprediction for 2017-reformed subjects between 2010-2011, while Panel B presents estimates for schools in England with above-median overprediction for 2017-reformed subjects between 2010-2011. All regressions include school, year, and university fixed effects. Sample: university students who started an undergraduate program in the UK between 2012 through 2017.

Table D5: Effects on Graduation and Early-career (Age 21) Outcomes

	Degree within 3 yrs	Employed	Earnings (GBP)	Out-of-work Benefits	Non-degree Firm	Low-wage Firm
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Low-income						
$1\{t{=}{-5}\} \times 1\{\text{Treated}\}$	-0.003	0.005	0.003	0.001	-0.006	-0.008
	(0.006)	(0.005)	(0.014)	(0.005)	(0.004)	(0.005)
$1\{t{=}{-4}\} \times 1\{\text{Treated}\}$	0.007	0.000	0.012	-0.003	-0.001	-0.004
	(0.005)	(0.005)	(0.013)	(0.004)	(0.004)	(0.005)
$1\{t{=}{-3}\} \times 1\{\text{Treated}\}$	0.007	0.007	0.025**	0.002	-0.001	-0.009*
	(0.005)	(0.005)	(0.013)	(0.004)	(0.004)	(0.005)
$1\{t{=}{-2}\} \times 1\{\text{Treated}\}$	0.005	-0.000	0.001	0.006	-0.004	-0.002
	(0.005)	(0.005)	(0.012)	(0.004)	(0.004)	(0.004)
$1\{t{=}{-1}\} \times 1\{\text{Treated}\}$	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$1\{t{=}0\} \times 1\{\text{Treated}\}$	0.010**	0.012***	-0.006	-0.023***	-0.012***	-0.013***
	(0.005)	(0.005)	(0.012)	(0.004)	(0.004)	(0.004)
N	525,615	525,615	525,680	525,615	353,220	342,562
Panel B: High-income						
$1\{t{=}{-5}\} \times 1\{\text{Treated}\}$	-0.003	-0.000	0.011	-0.003	0.002	-0.006
	(0.005)	(0.004)	(0.012)	(0.003)	(0.003)	(0.004)
$1\{t{=}{-4}\} \times 1\{\text{Treated}\}$	-0.005	0.002	-0.001	0.004*	0.003	-0.003
	(0.005)	(0.004)	(0.011)	(0.002)	(0.003)	(0.003)
$1\{t{=}{-3}\} \times 1\{\text{Treated}\}$	-0.007*	-0.000	0.006	0.002	0.002	-0.002
	(0.005)	(0.004)	(0.011)	(0.002)	(0.003)	(0.003)
$1\{t{=}{-2}\} \times 1\{\text{Treated}\}$	-0.006	-0.001	0.009	0.004*	0.002	-0.003
	(0.004)	(0.004)	(0.011)	(0.002)	(0.003)	(0.003)
$1\{t{=}{-1}\} \times 1\{\text{Treated}\}$	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$1\{t{=}0\} \times 1\{\text{Treated}\}$	-0.004	-0.006*	0.003	-0.012***	-0.002	-0.010***
	(0.004)	(0.004)	(0.011)	(0.002)	(0.003)	(0.003)
N	716,515	716,515	716,900	716,515	494,035	465,459

Notes. This table shows OLS coefficients from event study regressions in Equation 6 on observed graduation and post-graduate outcomes at age 21, separately by low-income and high-income students. Column 3 reports coefficients from the Poisson regression specified in Equation 7 for observed earnings at age 21. Results compare students at schools that stopped offering AS-levels beginning in academic year 2016-17 to schools that still required all students to take AS-levels in reformed subjects in 2016-17. Outcome of working at a non-degree firm is conditional on being employed. All regressions include school and year fixed effects. Robust standard errors are reported in parentheses and clustered at the school level. Sample: Students age 18 at the end of high school in England between 2010-2017 in the NPD. Earnings are CPI adjusted to 2015 GBP and winsorized at 1 percent above and below.