

Prediction or prejudice? Standardized testing and university access

Nagisa Tadjfar Kartik Vira
(Job Market Paper)

October 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 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 teacher 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 pre-university 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

requirements 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 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 from the Department for Education (DfE) on high school records; 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 effective 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 on a sample in the years prior to the reform and use the model 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 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 in 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 resulted in a net increase of university degrees among low-income students. The probability of obtaining a degree within 3 years, the standard undergraduate duration in the UK, increased by 1pp among low-income students, a 5% increase relative to the group mean. Three years after high school, these 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 early-career firm characteristics and demonstrate that these improvements in average 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. 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 in terms of average education level or 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 share 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 are 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 is instrumenting 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* using 2SLS. 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-optimal 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). More recently, Avery et al. (2025) examine test-optimal 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, particularly among socioeconomically advantaged students (Astorne-Figari and Speer, 2019; McEwan et al., 2021; Ahn et al., 2024; Sacerdote 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 algorithmic fairness (Kleinberg et al., 2016, 2018) and the disparate impacts of screening algorithms (Arnold et al., 2018, 2022). Previous literature has found that high school GPAs more effectively predict university performance at lower-ranked institutions, while SAT scores are more predictive at elite institutions (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 bias between interviews and hiring examinations (Autor and Scarborough, 2008). We use an instrument for university attendance to examine calibration bias in pre-university tests and teacher assessments for low-income students. Our paper highlights the importance of considering both bias and direct disutility of screening algorithms when examining the potential for disparate impacts of algorithms.

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 primarily focused on the impacts of affirmative action policies, top-percent admissions plans, or financial aid expansions (Zimmerman, 2014; Arcidiacono and Lovenheim, 2016; Dillon and Smith, 2020; Angrist et al., 2022; Bleemer, 2022; Black et al., 2023). 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 increased dropout rates. Our findings demonstrate the nuanced trade-offs in improved outcomes for disadvantaged students alongside reduced screening precision resulting from switching from standardized tests to teacher-assigned grades, bridging this literature with the more nascent literature examining test-optional policies and algorithmic bias literature.

II. Background

II.A *Secondary education in the UK*

Students in England, Wales, and Northern Ireland follow a standardized national curriculum in school between ages 5-16 organized into five “Key Stages” (KS1-KS5).¹ At the end of Year 11 (typically at age 16), students complete KS4 by taking General Certificate of Secondary Education (GCSE) exams, which are externally assessed, subject-specific standardized ex-

¹Scotland 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.

ams. 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. After completing GCSEs, students may elect to continue their education by choosing a small number of further qualifications, all externally graded, to pursue in their last two years of school during Year 12 and Year 13.

During our sample period of academic years 2011-12 through 2018-19, students predominantly chose between two types of qualifications: A-Levels and Business and Technology Education Council qualifications (BTECs). A-Levels are academic and exam-based, akin to Advanced Placement (AP) courses in the United States. BTECs are vocational and course-work based, although BTEC students may apply to and attend university.² Students typically take three A-Level subjects or an equivalent number of BTEC subjects,³ resulting in a high degree of specialization prior to attending university. Unlike many OECD countries, the UK secondary school system is not tracked. In 2012, 99% of secondary schools in England offered A-level courses and 55% of secondary schools offered BTECs, so students are typically 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.⁴ 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, there is substantial socioeconomic heterogeneity in A-level take-up and correlational path dependence on university access. Table 1 shows that only 30% of students who completed GCSEs between 2009-2013 went on to take 3 or more A-levels during their final two years of high school. Columns 4 and 5 of Table 1 highlight gaps in A-level take-up by income – students in the top two income quintiles are more than twice as likely to take three or more A-levels compared to students in the bottom two income quintiles. Columns 2 and 3 of Table 1 look at university trajectories separately for students who took 3 or more A-levels and students who did not. Students who take 3 or more A-levels generally have higher GCSE scores in Math and English and are significantly more likely to apply to and attend university compared to students who do not. While 77% of A-level students obtain an undergraduate degree within 3 years of graduating high school, only 17%

²Around 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).

³Most universities typically require three A-levels or equivalent units in BTECs, but a small number of academically ambitious students take four or five A-levels.

⁴Around 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 < 1% of students in the UK take IB classes.

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.

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 halfway through the class at the end of Year 12. The second round of exams, “A2-levels”, were taken in June after Year 13 and receive the final grade in August. Crucially, the second round of exams took place *after* prospective students applied and received offers to university. Because of this pre-qualification admissions system, university applicants received teacher-assigned grades, called “predicted grades”, in each A-level subject that were sent to universities as part of their application, along with GCSE grades and AS-level grades.

II.B *Staggered reduction in testing requirements*

In 2013, motivated by a desire to reduce “teaching to the test”, Education Secretary Michael Gove backed a reform to remove the requirement that students in England take AS-level exams halfway through A-level classes in Year 12. The reform was highly unpopular among many universities, who used AS-levels to make admission decisions, citing higher predictive validity compared to measures such as 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 In addition to this geographic variation, the reform was staggered across subjects between qualification years 2017 through 2020. Table C1 shows the timing of the reform for each A-level subject, split into three tranches of 2017, 2018, and 2019. The gradual roll-out across subjects was due to a government-announced delay, which generated substantial logistical hurdles for schools. Many schools expressed frustrations around “knowing what the standard will be to be in line with national trends” and “confusion among parents, staff, and students” (UCAS, 2015). Amid this uncertainty, schools in England rapidly stopped offering AS-level exams in reformed subjects. At the same time, some schools continued requiring AS-levels for all A-level students even in reformed subjects in the early years of the reform.

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).⁶

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

⁶Similarly, Melrose and Mead (2018) at the Office of Qualifications and Examinations Regulation (Ofqual)

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. 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.⁷ 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 1 shows the number of schools that required AS-levels in 2017-reformed subjects between 2014-2020. In the first year of the reform, for students taking A-levels in 2017, around half of the schools in England still required all students take AS-levels. However, this number rapidly falls and by 2020, nearly all schools in England no longer required the AS-level in reformed subjects. Figure 2 shows how the number of students taking A-levels and AS-levels evolved over time in the UK. In Panel A, we see 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 2 show similar patterns for subjects reformed in 2018 and 2019, respectively. In Panel D, we see that the the number of students taking A-levels and AS-levels across all subjects in Northern Ireland and Wales remain similar through 2020. 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.

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.

⁷<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.

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.⁸ Our primary dataset is the National Pupil Database (NPD), with detailed educational attainment data from KS1 through KS5 for students in England. We then merge these data at the student level to datasets containing prior test scores, university applications and enrollment, employment, earnings, and government benefits.

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

In the NPD, we restrict our analysis to students who are 18 years of age when completing KS5 and . Additionally, we focus on students completing their KS5 in the years 2012 through 2019, which correspond to academic years 2011-12 through 2018-2019 respectively, which leaves us with approximately 2.9 million KS5 students.⁹ As described in Section II.B, schools rapidly changed their AS-level policies as the AS-level exams became optional for various subjects between 2017-2019. Although surveys from UCAS and Ofqual provide a sense of the shares of schools in England that stopped offering AS-levels at various points in time, we do not directly observe these school policies in our data. Instead, we infer school policies from student A-level and AS-level records in the NPD data and classify schools into one of the following five categories based on when they stopped offering AS-levels: schools that switched in 2016-17, schools that switched in 2017-18, schools that switched in 2018-19, control schools, and schools with mixed policies.

Consistent with the patterns in the UCAS and Ofqual surveys, many schools in our data adopted ‘mixed’ policies, ranging from requiring AS-levels in some reformed subjects but not others to 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’. We define a switcher as a school that went from *requiring all* students to take AS-levels in $t - 1$ to no longer offering AS-levels in t . Mechanically, all schools that stopped offering AS-levels in A-level qualification year 2017 (and hence AS-level qualification year 2016), are switchers,

⁸Due to the highly sensitive nature, these datasets are accessible only in person at SRS-designated Safe-rooms or Safepods.

⁹We exclude 2020 because in the years 2020-2021 due to COVID-19, students did not sit the actual exams and received grades in 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.

since AS-levels were mandatory across the UK prior to 2017. We then identify schools that stopped offering AS-levels as follows. If $< 1\%$ 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 $> 90\%$ 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 $> 90\%$ 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.¹⁰

Figure C1 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. Table C2 presents summary statistics for schools that stopped offering AS-levels in each year between 2017-2019, schools that continued to offer reformed AS-levels throughout this period, as well as schools where a varying share of A-level students continued to take AS-levels. Schools that stopped offering AS-levels between 2017-2019 all appear quite similar, although the control schools as well as mixed schools are slightly academically weaker and lower income.

III.B University applications, enrollment, and performance

We link 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,

¹⁰Of 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 removing AS-levels 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. 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 5 has a larger set of control schools to include around 67% of the total schools.

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. The 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. All earnings were winsorized at the 99% and 1% levels within a tax year and inflation adjusted to 2015 GBP using the CPIH index. Finally, we link to the Department for Work and Pensions (DWP) data that contain information work benefits, that individuals in England receive each tax year, spanning from 1999-2000 through 2020-21. In a separate database, we also have access to another version of the UCAS dataset that additionally includes university applicants from Northern Ireland, Wales, Scotland, as well as outside the UK. We use this separate iteration of the UCAS dataset for additional robustness checks to some of our results to include students from Northern Ireland and Wales not available in the UCAS iteration within the Longitudinal Education Outcomes database. Section A provides more detail on data processing and sample restrictions.

IV. Effects on Educational Outcomes

To estimate the impact of reduced standardized testing on student decisions and outcomes, 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 treated 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. We use this specification to estimate effects on outcomes such as A-level take-up, university application and university enrollment. Specifically, for student i attending school $s(i)$:

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

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 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 estimator developed by Sun and Abraham (2021). 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.¹¹ 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.

IV.A *Changes in A-level take-up and subject choice*

Figure 3 presents estimates from Equation 1 separately by low-income and high-income students. Panel A in Figure 3 shows that when schools stop offering AS-level exams, low-income students in treated schools become 5pp more likely to take A-levels compared to low-income students at control schools that still require all students to take AS-level exams. Although there is a visually small uptick among high-income students as well, the effect sizes are not statistically distinguishable from zero.¹² In addition to the general increase in take-up of A-levels, the policy induced a relative shift into STEM subjects within A-levels.¹³ In Panel B of Figure 3, we find that take-up of STEM A-levels also increases by 3pp after schools remove AS-levels for *both* low- and high-income students. 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. 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. In Figure E1, we present estimates separately by low and high predicted propensity to take A-levels and find that the shift into A-levels is primarily driven by students who would have previously been unlikely to take A-levels, while the shift into STEM A-levels is driven by both students who were previously unlikely to take A-levels as well as the “always-takers” of A-levels.

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

These patterns are consistent with students being shifted into academic-track classes in high school once testing requirements are relaxed. This additional mass of students may reflect

¹¹ 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 U.S. Census Block Groups in geographic size and population.

¹² In our main findings, we define A-level take-up as taking three or more A-levels. This is considered a complete load of A-levels and is the minimum number required by many universities. However, 7% of university applicants take combinations of A-levels and BTECs, often taking fewer than three A-levels as a result.

¹³ In these results, 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.

reduced effort costs associated with taking these classes once students are no longer required to take a pre-university exam, but may also reflect students anticipating a change in the signal structure used in university admissions. This decrease in effort cost or increase in perceived university prospects shifts the threshold of priors on ability at which students choose to take A-levels (or STEM A-levels). If students' priors on ability are monotonic in GCSE scores, we would expect the students shifted into A-levels and STEM to have lower GCSE scores.

In the second set of outcomes on performance in A-levels, 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):

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

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.

In Figure 4, we examine 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.

Student performance on final A-level exams also decline. Figure E2 presents estimates on A-level performance from Equation 2, using school-subject level variation. Panel A in Figure E2 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 E2. Again, the decline in academic performance appears larger among STEM subjects. In Figure E3, we present estimates that control for GCSE scores and demographics and find that around half of the decline in getting an 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 5 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 present estimates of university applications and enrollment effects separately by students who were *previously likely* to take A-levels and students who were *previously unlikely* to take A-levels. This propensity is estimated on a training sample using Equation 3 and predicted for the remaining holdout sample. In order to examine heterogeneity by students' ex-ante likelihood of taking A-levels, 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 (Equation 3) on the binary outcome of whether or not a student takes 3 or more A-levels, using a set of predictors including prior test scores and demographics. Our approach is motivated by and similar to the prediction model in (Black et al., 2023). To mitigate concerns about overfitting bias resulting from endogenous stratification (Abadie et al., 2018), we estimate the model on a randomly selected training sample of all GCSE students who completed KS4 between 2009–10 and 2013–14. The training sample constitutes about one-third of the original sample (approximately 960,000 students out of 3 million) and is excluded 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:

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

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 decile 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. We then use the model to predict student A-level take-up in our entire sample. 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. We estimate our main event study specifications in Equations 1 through 6 separately across four subgroups, excluding the training sample: (i) low-income and below-median propensity; (ii) low-income and above-median propensity; (iii) high-income and below-median propensity; and (iv) high-income and above-median propensity. Table C3 reports summary statistics for students in each of the four cells, with notable differences in socioeconomic status, prior attainment, and demographic composition.

Figure 6 shows the estimates for these event studies run separately in four cells: low-income students with a low predicted propensity to take A-levels, low-income students with a high predicted propensity to take A-levels, high-income students with a low predicted propensity to take A-levels, and high-income students with a high predicted propensity to take A-levels.¹⁴ Panels A and C in Figure 6 show that the increase in university application and attendance among low-income students is concentrated among students who were previously unlikely to have taken A-levels. Panel B in Figure 6 shows that even among high-income students, there is a 5pp uptick in university applications among students who were previously unlikely to take A-levels. These students do not see a statistically significant increase in university enrollment, however, as indicated in panel D. 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. However, the large increase in university applications suggests that a change in signal structure favorable for this group of students may also have additionally encouraged university applications (and subsequent enrollment). Section VI. directly examines the change in signal for the marginal university entrant.

Why are some students, particularly the low-propensity high-income students, more likely to apply to university but no more likely to enroll? This may be the result of students aiming too high due to reduced information about their ability. In Panels B and D of

¹⁴We define low and high propensity as below or above median predicted propensity on the sample excluding the training sample.

Figure E4, we show 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. Thus, these students are more likely to apply to university, but do not secure offers. 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 E4. Although this suggests some “wasteful” university application attempts, this fact alone is insufficient to draw welfare conclusions for this group. These students were still shifted into A-levels, particularly STEM A-levels, which may have positive labor market treatment effects even in the absence of university attendance. Alternatively, given that these students are no more likely to enroll in university, they may have been better off taking non-STEM A-level courses or vocational courses. In Section V., we directly examine the labor market outcomes.

IV.D *Characterizing the marginal university entrant*

In Section IV.A, we documented negative selection into A-levels and STEM classes once a school stopped offering the pre-university exams. Is there similar negative selection into university? We examine how university graduation rates, degree-class, and dropout rates were affected among university students. Panels A, B, and C in Figure 7 present estimates from Equation 5 among students who began university between 2012 through 2017, comparing schools that stopped offering AS-levels to both control schools in England that continued requiring AS-levels through 2017 as well as high schools in Northern Ireland and Wales. Panel D in Figure 7 presents results from a staggered event study estimating Equation 1. We find that conditional on attending university, students in treated schools are 2pp less likely to graduate after AS-levels are no longer offered.

Turning to performance at university, students are 1pp less likely to graduate with first-class degrees and 2pp less likely to graduate with upper-second class (2:1) degrees. This suggests that the reform primarily affected students who were on the margin of graduating with an upper-second class (2:1) rather than those on the margin of graduating with a first-class degree, which is expected given that new entrants were likely marginal to applying to university altogether. 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 appears to increase over time, which may be partly due to the fact that an increasing number of subjects are reformed between 2017-2019. Event study

coefficients for these figures, along with additional specifications that control for university fixed effects, are presented in Appendix Table E3. 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 E1.

We run these event studies separately by schools that have high pre-reform teacher grade inflation fixed effects in 2017-reformed subjects and schools that have low teacher grade inflation fixed effects.¹⁵ Table E2 presents estimates separately by low- and high- grade inflation schools. 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, while the effects at schools with below-median teacher grade inflation are statistically indistinguishable from zero. Similarly, effects on graduating with a first and graduating with a first or upper-second are approximately twice as large at high teacher grade inflation schools. This is suggestive evidence that the decline in graduation is exacerbated at schools with historically higher grade inflation. This relies on the notion that schools that historically had inflated teacher grades even in the presence of AS-levels were more likely to assign teacher grades that would differ more from the AS-level grades students *would have* gotten in the presence of an exam. We therefore interpret these findings with caution as suggestive but not conclusive evidence that higher teacher grades may generate more academic mismatch at universities after the elimination of tests.

In order to directly examine the extent of negative selection, we use an IV framework (Imbens and Angrist, 1994; Angrist et al., 1996) to document the characteristics of *compliers* into university. Under this framework, compliers 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 ($D_i = 1$) otherwise (i.e. if $Z_i = 0$). Consistent with our event study results in Figure 5, 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 distribution within low-income compliers is:

¹⁵Schools are grouped 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”.

$$w_x = \frac{E[1\{X_i = x\}D_i|Z_i = 1] - E[1\{X_i = x\}D_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]} \quad (4)$$

Identification requires the assumptions of the LATE theorem outlined in Imbens and Angrist (1994) to hold, and in particular, that the instrument independent of the error term in the decision to attend university. The IV framework in our setting imposes more stringent assumptions than 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 4 using 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 also attend academically non-selective universities, with the average university rank of around 100.¹⁶ Consistent with our event studies, we find that 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. Next, we turn to directly examining the labor market returns to university attendance for these marginal entrants to university.

V. Labor Market Returns for Marginal Entrants

In Section IV., we documented that relaxing testing requirements increased university application and enrollment among low-income students, but that these marginal entrants are academically weaker than the typical low-income university student. Do these students succeed at university? And do they go on to pursue higher-earning careers? Some literature, typically focused on affirmative action, has examined the notion of *academic mismatch*, suggesting that students induced into university by certain expansionary policies may have been better off either not attending university or attending lower-ranked universities due to lack

¹⁶We construct university rankings using the average standardized A-level grades of university attendees in the 2011 cohort.

of adequate academic preparation to graduate successfully from these universities. In this section, we directly examine the effects in our setting on degree completion and employment outcomes measured three years after high school graduation at age 21-22. Regression tables for all figures can be found in Appendix C.

Due to limitations in our data, outcomes such as university graduation and 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:

$$Y_i = \alpha_{s(i)} + \gamma_{t(i)} + \sum_{\tau=2012}^{2017} \theta_\tau I(t(i) = \tau) \times I(T_{s(i)} = 2017) + \varepsilon_i \quad (5)$$

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 and schools in Northern Ireland and Wales, where data is available.

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

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

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. Our control group includes English schools that continued requiring AS-level exams in 2017.

V.A *Degree completion and early-career outcomes*

Because we observe age 21 outcomes only for students in the first treated cohort, our 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. We estimate Equation 5 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.¹⁷. 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 6 to allow for zeroes.

Figure 8 summarizes the difference-in-differences coefficients for outcomes and Table E4 reports event study coefficients to rule out pre-trends. We see that low-income students see statistically significant improvements in their university graduation and early-career outcomes. These students are 1pp more likely to have graduated with a university degree by age 21, are 1pp more likely to be employed, 2pp less likely to receive out-of-work benefits, and are 1pp less likely to be working at a non-degree firm or a low-paying firm (i.e. bottom 10% firm). However, we find a null effect on earnings at age 21. We also see 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 9 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

¹⁷We 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.

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 8 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 8. This leakage is consistent with some findings in the literature (Arcidiacono and Lovenheim, 2016; Black et al., 2023).

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

In the previous section, we presented results showing that, 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 the LATE theorem (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 employees¹⁸ 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*:

$$E[1\{Y_{1i}^{21} < c\}|D_{1i} > D_{0i}] = \frac{E[1\{Y_i^{21} < c\}D_i|Z_i = 1] - E[1\{Y_i^{21} < c\}D_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]} \quad (7)$$

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

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

We estimate Equations 7 and 8 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 10 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 left-tail firms (firms that do not require university degrees and are low-paying). 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 left-tail firms.

¹⁸In 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.

V.C Lifetime earnings gains for marginal entrants to university

To measure the long-run returns for students marginally shifted into university, we estimate the present discounted value of lifetime earnings gains for low-income marginal entrants. As before, we focus our attention on low-income marginal entrants for statistical power since we only observe first stage effects of the reform on university enrollment for low-income students. Additionally, we estimate 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 an expected completion in June 2020. First, we estimate the effect of university enrollment on earnings at ages 18 through 21 using two-stage least squares (2SLS) controlling for school and year fixed effects. We then adjust these estimates for three years of tuition, which is £9,250 per year during our sample years, paid at ages 18, 19, and 20. All earnings and tuition fees are CPI-adjusted to 2018 GBP.

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, using two related approaches to project age 27 earnings: (i) using the surrogate index method developed by Athey et al. (2019) and (ii) the experimental selection correction estimator Athey et al. (2025). 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:

$$Y_i^{27} = \beta X_i + \gamma_{firm(i)}^{21} + \gamma_{SIC(i)}^{21} + \sum_{a=18}^{21} (\delta_1^a Y_i^a + \delta_2^a E_i^a + \delta_3^a B_i^a) + \varepsilon_i \quad (9)$$

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_{firm}^{21} and $\gamma_{SIC(i)}^{21}$ are fixed effects for the firm ID and 2007 Standard Industrial Clas-

sification (SIC) codes for the firm of employment at age 21.¹⁹ 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:

$$Y_i^{27} = \beta X_i + \gamma X_{firm(i)}^{21} + \sum_{a=18}^{21} (\delta_1^a Y_i^a + \delta_2^a E_i^a + \delta_3^a B_i^a) + \varepsilon_i \quad (10)$$

where instead of firm and industry fixed effects, we control for age-21 firm characteristics. Concretely, $X_{firm(i)}^{21}$ 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.²⁰ We estimate Equations 9 and 10 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 D1.

Using these estimates, we then follow Athey et al. (2019) 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. Thus, 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 the second approach, we estimate the LATE of university attendance on “secondary outcomes”: earnings, employment, and unemployment benefits at ages 18 through 21 as well as age 21 firm characteristics. This is estimated using 2SLS with school and academic year fixed effects. Second, for all low-income students in the historical sample (individuals age 18 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, which is age

¹⁹We assign a common ID to small firms with fewer than 3 employees and draw on variation from the SIC 2007 codes instead.

²⁰As 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.

27 earnings in our context, 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 11 presents cumulative lifetime earnings gains for low-income students estimated using both the surrogate index and ESC projections at 3% and 5% discount rates. 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 a £50,000-£100,000 increase in lifetime earnings for students induced into attending university as a result of the 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. Informativeness and Calibration Bias of Standardized Tests

Reduced standardized testing requirements in our setting led to an increase in university enrollment among low-income students. Were these low-income students previously not attending university because they were screened out by a test that was miscalibrated to measure their academic potential? In this section, we empirically compare the predictive power of teacher-assigned grades to standardized exams for both low-income and high-income students and develop an empirical test for calibration bias in standardized testing for low-income students induced into university. We consider two sets of university outcomes—whether or not a university student graduates with a degree within 3 years (the standard duration of an undergraduate degree in the UK) and whether or not a graduating student gets a first-class degree (typically top 25-30% of graduating class).²¹ High school students in the UK who choose academic-track classes, called A-levels, receive three sets of signals for each class: (i) the teacher-assigned grades A-level students receive for each A-level subject before applying to university, (ii) intermediate A-level exam scores taken halfway through the class *before* university applications, and (iii) endline A-level exam scores taken *after* receiving university offers. We begin by comparing the predictive power and informativeness of (i) and (iii), observed for all university applicants both pre- and post-reform, on university performance

²¹Unlike US universities, universities in the UK do not report GPAs. Instead, students typically graduate with one of four degree classes, with the exception of medical students who do not receive degree classifications: first-class, upper-second (2:1), lower-second, third.

measures.

VI.A Standardized tests are more informative of future outcomes

We begin by comparing the predictive power and informativeness of (i) and (iii), observed for all university applicants both pre- and post-reform, on university performance measures.

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. We restrict our analysis to students who took large A-level subjects.²² Empirically, we bring this to the data using a three-pronged approach and focus on a pre-reform sample of university students pursuing undergraduate degrees who began university between 2012-2016 at large universities.²³ First, we estimate OLS regressions of each university outcome separately by income group to compare the explanatory power of teacher grades and test scores using the specification in Equation 11:

$$Y_i = \delta_g \sigma_i + \gamma_{umg} + \varepsilon_i \quad (11)$$

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. γ_{umg} is a major-university fixed effects. Table 3 presents the adjusted R^2 values each set of regression. Across both low-income and high-income students, we find that the A-level scores explain substantially more of the variance for both outcomes. Comparing columns 1 and 2 in Table 3, we find that 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%). Turning to columns 4 and 5, we see that tests have even more incremental explanatory power compared to teacher grades when it comes to the outcome of position within graduating university cohort. Tests explain 40% more variation in graduating with a first-class degree compared to teacher grades for low-income students and 33% more for high-income students. Columns 3 and 6 report adjusted R^2 values when using both teacher and test scores, and we see that across both income groups and university outcomes, adding in teacher scores add little explanatory power when test scores are already included.

²²We 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.

²³In 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.

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.²⁴ Because random forest classifiers generally perform best on continuous variables rather than fixed effects, instead of university fixed effects, we include a ranking of the average A-level score percentiles of matriculating students in 2010 at each university. Both models include either the average standardized A-level exam scores or the average standardized teacher grades. Figure 12 presents the precision-recall (PR) curves and the average precision (AP) measure for students matriculating between 2012-2016. The AP measure and the PR curves are standard diagnostics used in the literature (Cengiz et al., 2022) and allow us to compare the predictive power of the models.²⁵ Panel A in Figure 12 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 3, standardized tests have substantially higher out-of-sample predictive power for a student's position within the graduating cohort. Panels C and D in Figure 12 show that the AP of the PR curve increases from 0.241 to 0.260 for all students and from 0.201 to 0.233 for low-income students. Moreover, Panels C and D in Figure 12 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.²⁶

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 subjects²⁷, and the final A-level grade controlling for year fixed effects. 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

²⁴The 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.

²⁵We additionally report receiver operating characteristic (ROC) curves in Figure E5 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).

²⁶Intuitively, 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.

²⁷Core GCSE subjects are Mathematics, English Language, English Literature, and Science.

scores, such as subject-specific measures of ability recognized by teachers but not observed in standardized test scores. We estimate:

$$Y_{ig} = \alpha_{0g} + \delta_g \sigma_{ig} + \gamma_{sg} + \gamma_{tg} + \varepsilon_{ig} \quad (12)$$

where Y_{ig} is an outcome such as university graduation, g denotes the subject group (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. We repeat this regression for A-levels in STEM subjects (Mathematics, Further Mathematics, Physics, Biology, Chemistry) and non-STEM subjects. Table 4 shows that in both STEM and non-STEM subjects, a 1 s.d. 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 selective university. Although these residuals are strongly correlated with a student's university destination, they are not predictive of success.

In STEM subjects, a 1 s.d. increase in teacher-assigned grade residuals is associated with a 0.5pp *lower* probability of graduating from university and a 1pp lower probability of graduating with a first or second-class degree conditional on graduating. In non-STEM subjects, these residuals have small but positive correlations with the probability of graduating and getting a first or second-class degree conditional on graduating. However, even in non-STEM subjects, the 95% confidence interval of the coefficients on σ_{ig} lies below 1pp. Taken together, 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. Moreover, in STEM subjects, over-prediction is associated with *worse* outcomes, potentially due to academic mismatch.

Each of these methods have their own advantages and disadvantages, but together they paint a consistent picture that exams are informative signals that predict performance at university for both high-income and low-income university students. This echoes the statements from many UK universities in response to the reform. In particular, we find that the exams are much better at predicting whether or not a student graduates with a first-class degree, conditional on graduating, effectively capturing a student's position in their university cohort's distribution. This discrepancy in predictive power appears to be smaller when it comes to predicting whether or not a student graduates from university at all. All three of our comparisons suggest that post-reform, students and universities had access to a less informative measure of ability.

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

We now extend our analysis beyond explanatory power and out-of-sample predictions to examine *calibration* of tests and teachers in predicting university performance. 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)?

In order to empirically answer these questions, we first formalize 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 $E[Y|\hat{Y}(\zeta), \text{Low-income} = 1] \neq E[Y|\hat{Y}(\zeta), \text{Low-income} = 0]$.

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 13 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. Let $Z_i = 1$ if student i attends a school-by-cohort where AS testing was discontinued (instrument) and $D_i \in 0, 1$ indicate university attendance. Under the standard LATE assumption, this would generate a subpopulation of compliers among whom university attendance is conditionally random. We estimate marginal mean Y and \hat{Y} for low-income students in the holdout sample using 2SLS to control for school and year fixed effects:

$$E[Y_{1i}|D_{1i} > D_{0i}] = \frac{E[Y_{1i}D_i|Z_i = 1] - E[Y_{1i}D_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]} \quad (13)$$

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 13 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 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 and/or discrimination, 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”.²⁸ 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, 2023). 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.

²⁸Source: <https://www.bhasvic.ac.uk/parents-carers-dashboard/higher-education> as of February 12, 2025.

VII. Standardized Testing and Efficiency

Our findings suggest that the reform effectively increased student *attempts*, which in turn also increased *failures*. Students who were previously unlikely to take A-levels became 5pp more likely to take A-levels, specifically a complete course of A-levels (i.e. three or more). Students who were previously already likely to take A-levels were in turn shifted into STEM A-levels. Conditional on taking A-levels, students saw a 5pp increase in their likelihood of failing an A-level. This pattern of increased attempts and failures also emerges at the university level. More students apply to universities and enroll, but graduation rates conditional on attending university decline by 2pp. In the education literature, a decline in graduation rates conditional on attending is often described as “academic mismatch” and interpreted as a decrease in efficiency. However, graduation rates alone may be insufficient measures of welfare, which depend instead on the relative returns of a university degree and some amount of university without degree completion to 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. Our findings suggest that standardized testing may discourage disadvantaged students from attempting paths with large private returns, consistent with findings in the undermatching literature (Hoxby and Avery, 2012; Campbell et al., 2022).

These private returns also imply positive *social* returns. First, they exceed the average per-student cost to government to educate a typical student is £35,000. This means that the returns likely exceed the marginal government cost by an even larger amount, as the per-student average cost is typically higher than the marginal cost. Moreover, 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 B2.²⁹

²⁹Beginning 2012, enrollment caps in 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.

Thus, our results suggest positive social returns to university for marginal students suggests that relaxing testing requirements brought us closer to social optimum. Computing the marginal value of public funds (MVPF) using a marginal tax rate of 20%, we obtain between 2.5 to 7.1, putting it well above the median of estimated MVPFs of cost-effective university aid programs (Hendren and Sprung-Keyser, 2020). The net social benefit is also large and positive, at around £32,000 to £75,000 per student, net of the average per-student government cost.

Crucially, these social gains did not arise from the elimination of a miscalibrated or demographically *biased* exam that inefficiently *screened out* low-income students. Instead, our results suggest that one potentially overlooked downside of pre-university standardized testing is the direct disutility it brings to students who are on the margin of taking academic-track classes and attending university. Relaxing these pre-university exam requirements that were part of academic-track classes in the UK not only increased enrollment, but also increased the takeup of those classes even *prior* to the university application process. This shift into A-level classes and the eventual shift into university primarily occurred for low-income students with large labor-market returns to attending university. Under standardized testing regimes, students may inefficiently underapply to university or not take A-level classes. As a result, reduced standardized testing was likely efficiency enhancing from a social planner's standpoint in this particular context, despite decreases in academic performance and the absense of calibration bias in the exams.

VIII. Conclusion

Critics of standardized testing argue that they are inherently biased against low-income students and serve as a barrier to higher education access, inhibiting opportunity. This paper provides empirical evidence on the educational and labor-market consequences of reduced standardized testing, using the staggered elimination of pre-university standardized tests in the UK and administrative data on students outcomes from high school through university and postgraduate outcomes. Our results demonstrate that students are highly responsive to relaxed testing requirements—they update their educational choices during high school as well as their decisions to apply to and attend university.

We find that the reform substantially expanded A-level participation among disadvantaged, lower-ability students. Higher income students and students who were previously likely to take A-levels were also shifted toward STEM subjects. At the same time, we find that students became more likely to fail A-level classes, with the largest decreases in academic performance seen in STEM subjects. All students became more likely to apply

to university, but increases in enrollment and degree completion were 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. Average age 21 earnings, however, remained unchanged among both low and high-income students. These early career indicators at age 21 translate to large private returns: low-income students shifted into university see £50,000–£100,000 in lifetime earnings gains, net of tuition. These marginal university entrants reap large returns despite a low on-time degree completion rate of around 30–40%. Conditional on university matriculation, three-year degree completion declined by approximately 2pp, and the probability of graduating with at least upper-second honors fell similarly. Although the reform reduced conditional graduation rates, it expanded university access and degree completion on the extensive margin without depressing early-career earnings.

Moreover, we find that these effects are not the result of eliminating a test that was systematically biased against low-income students. Using the reform as an instrument for university attendance, we overcome selective labeling and test for calibration bias on 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 can discourage participation among disadvantaged students, and that reduced testing requirements may expand university access when teachers assign inflated grades. Reduced standardized testing improved outcomes at the lower end of the socioeconomic distribution with minimal losses on the upper end, but effectively increased student *attempts*. 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, Alberto**, “Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models,” *Journal of the American Statistical Association*, March 2002, 97 (457), 284–292. Publisher: ASA Website eprint: <https://doi.org/10.1198/016214502753479419>.
- , **Matthew M. Chingos, and Martin R. West**, “Endogenous Stratification in Randomized Experiments,” *The Review of Economics and Statistics*, October 2018, 100 (4), 567–580.
- Ahn, Tom, Peter Arcidiacono, Amy Hopson, and James Thomas**, “Equilibrium Grading Policies With Implications for Female Interest in STEM Courses,” *Econometrica*, 2024, 92 (3), 849–880. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA17876>.
- Allensworth, Elaine M. and Kallie Clark**, “High School GPAs and ACT Scores as Predictors of College Completion: Examining Assumptions About Consistency Across High Schools,” *Educational Researcher*, April 2020, 49 (3), 198–211.
- Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin**, “Identification of Causal Effects Using Instrumental Variables: Rejoinder,” *Journal of the American Statistical Association*, June 1996, 91 (434), 468.
- Angrist, Joshua, David Autor, and Amanda Pallais**, “Marginal Effects of Merit Aid for Low-Income Students,” *The Quarterly Journal of Economics*, April 2022, 137 (2), 1039–1090.
- Arcidiacono, Peter and Michael Lovenheim**, “Affirmative Action and the Quality–Fit Trade-off,” *Journal of Economic Literature*, March 2016, 54 (1), 3–51.
- , **Esteban M. Aucejo, Hanming Fang, and Kenneth I. Spenner**, “Does affirmative action lead to mismatch? A new test and evidence,” *Quantitative Economics*, November 2011, 2 (3), 303–333.
- Arnold, David, Will Dobbie, and Crystal S Yang**, “Racial Bias in Bail Decisions*,” *The Quarterly Journal of Economics*, November 2018, 133 (4), 1885–1932.
- , — , and **Peter Hull**, “Measuring Racial Discrimination in Bail Decisions,” *American Economic Review*, September 2022, 112 (9), 2992–3038.
- Astorne-Figari, Carmen and Jamin D. Speer**, “Are changes of major major changes? The roles of grades, gender, and preferences in college major switching,” *Economics of Education Review*, June 2019, 70, 75–93.
- Athey, Susan, Raj Chetty, and Guido Imbens**, “The Experimental Selection Correction Estimator: Using Experiments to Remove Biases in Observational Estimates,” *NBER Working Paper Series*, May 2025, p. w33817.

—, —, —, and Hyunseung Kang, “The Surrogate Index: Combining Short-Term Proxies to Estimate Long-Term Treatment Effects More Rapidly and Precisely,” *NBER Working Paper Series*, November 2019, p. w26463.

Autor, David H. and David Scarborough, “Does Job Testing Harm Minority Workers? Evidence from Retail Establishments*,” *Quarterly Journal of Economics*, February 2008, 123 (1), 219–277. Publisher: Oxford University Press (OUP).

Avery, Christopher, Lena Shi, and Preston Magouirk, “Test-Optional College Admissions: ACT and SAT scores, applications, and enrollment changes,” *NBER Working Paper Series*, September 2025.

Belasco, Andrew S., Kelly O. Rosinger, and James C. Hearn, “The Test-Optional Movement at America’s Selective Liberal Arts Colleges: A Boon for Equity or Something Else?,” *Educational Evaluation and Policy Analysis*, June 2015, 37 (2), 206–223.

Bennett, Christopher T., “Untested Admissions: Examining Changes in Application Behaviors and Student Demographics Under Test-Optional Policies,” *American Educational Research Journal*, February 2022, 59 (1), 180–216.

Black, Sandra E., Jeffrey T. Denning, and Jesse Rothstein, “Winners and Losers? The Effect of Gaining and Losing Access to Selective Colleges on Education and Labor Market Outcomes,” *American Economic Journal: Applied Economics*, January 2023, 15 (1), 26–67.

Bleemer, Zachary, “Affirmative Action, Mismatch, and Economic Mobility after California’s Proposition 209*,” *The Quarterly Journal of Economics*, February 2022, 137 (1), 115–160.

Bohren, J Aislinn, Peter Hull, and Alex Imas, “Systemic Discrimination: Theory and Measurement,” *The Quarterly Journal of Economics*, May 2025. Publisher: Oxford University Press (OUP).

Borghesan, Emilio, “The Heterogeneous Effects of Changing SAT Requirements in Admissions: An Equilibrium Evaluation,” 2023.

Britton, Jack, Laura van der Erve, Chris Belfield, Anna Vignoles, Matt Dickson, Yu Zhu, Ian Walker, Lorraine Dearden, Luke Sibieta, and Franz Buscha, “How much does degree choice matter?,” *Labour Economics*, December 2022, 79.

Campbell, Stuart, Lindsey Macmillan, Richard Murphy, and Gill Wyness, “Matching in the Dark? Inequalities in Student to Degree Match,” *Journal of Labor Economics*, October 2022, 40 (4), 807–850.

Carlana, Michela, “Implicit Stereotypes: Evidence from Teachers’ Gender Bias*,” *The Quarterly Journal of Economics*, August 2019, 134 (3), 1163–1224.

—, **Eliana La Ferrara, and Paolo Pinotti**, “Goals and Gaps: Educational Careers of Immigrant Children,” *Econometrica*, 2022, 90 (1), 1–29.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and David Zentler-Munro, “Seeing beyond the Trees: Using Machine Learning to Estimate the Impact of Minimum Wages on Labor Market Outcomes,” *Journal of Labor Economics*, April 2022, 40 (S1), S203–S247. Publisher: University of Chicago Press.

Chen, Jiafeng and Jonathan Roth, “Logs with Zeros? Some Problems and Solutions,” *The Quarterly Journal of Economics*, March 2024, 139 (2), 891–936. Publisher: Oxford University Press (OUP).

Chetty, Raj, David Deming, and John Friedman, “Diversifying Society’s Leaders? The Causal Effects of Admission to Highly Selective Private Colleges,” 2023.

— , **Nathaniel Hendren, Patrick Kline, and Emmanuel Saez**, “Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States *,” *The Quarterly Journal of Economics*, November 2014, 129 (4), 1553–1623. Publisher: Oxford University Press (OUP).

Creamer, Lisa, “Question 2 wins: Mass. voters end MCAS high school graduation requirement,” *wbur*, November 2024.

Dillon, Eleanor Wiske and Jeffrey Andrew Smith, “The Consequences of Academic Match between Students and Colleges,” *Journal of Human Resources*, 2020, 55 (3), 767–808.

Dilnot, Catherine, Lindsey Macmillan, and Gill Wyness, “The path increasingly travelled: Vocational entry qualifications, socioeconomic status and university outcomes,” *British Educational Research Journal*, December 2023, 49 (6), 1142–1160. Publisher: John Wiley & Sons, Ltd.

Friedman, John N, Bruce Sacerdote, Douglas O Staiger, and Michele Tine, “Standardized Test Scores and Academic Performance at Ivy-Plus Colleges,” *NBER Working Paper Series*, 2025.

Geiser, Saul and Maria Veronica Santelices, “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)*, 2007.

Gill, Tim, “Provision of level 3 qualifications in English schools 2008-2012,” Statistics Report 58, Research and Development Cambridge Assessment August 2013.

Haider, Steven and Gary Solon, “Life-Cycle Variation in the Association between Current and Lifetime Earnings,” *American Economic Review*, September 2006, 96 (4), 1308–1320. Publisher: American Economic Association.

Hendren, Nathaniel and Ben Sprung-Keyser, “A Unified Welfare Analysis of Government Policies*,” *The Quarterly Journal of Economics*, August 2020, 135 (3), 1209–1318.

Hillman, Nick, “A guide to the removal of student number controls,” *Higher Education Policy Institute Report* 69, 2014.

Hoxby, Caroline M. and Christopher Avery, “The Missing “One-Offs”: The Hidden Supply of High-Achieving, Low Income Students,” December 2012.

Imbens, Guido W. and Joshua D. Angrist, “Identification and Estimation of Local Average Treatment Effects,” *Econometrica*, March 1994, 62 (2), 467.

Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Ashesh Rambachan, “Algorithmic Fairness,” *AEA Papers and Proceedings*, May 2018, 108, 22–27.

—, **Sendhil Mullainathan, and Manish Raghavan**, “Inherent Trade-Offs in the Fair Determination of Risk Scores,” November 2016. arXiv:1609.05807 [cs].

Lakkaraju, Himabindu, Jon Kleinberg, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan, “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” ACM Halifax NS Canada August 2017, pp. 275–284.

Leckie, George and Konstantina Maragkou, “Student Sociodemographic and School Type Differences in Teacher-Predicted vs. Achieved Grades for University Admission,” September 2023.

McEwan, Patrick J., Sheridan Rogers, and Akila Weerapana, “Grade Sensitivity and the Economics Major at a Women’s College,” *AEA Papers and Proceedings*, May 2021, 111, 102–106.

Melrose, Karen and Rebecca Mead, “AS and A level decoupling: Implications for the maintenance of AS standards,” Technical Report Ofqual/18/6378/7 July 2018.

Murphy, Richard and Gill Wyness, “Minority report: the impact of predicted grades on university admissions of disadvantaged groups,” *Education Economics*, July 2020, 28 (4), 333–350.

Office for National Statistics, “Longitudinal Education Outcomes SRS Iteration 2 Standard Extract - England,” 2023.

Partington, Richard, “Predictive Effectiveness of Metrics in Admission to the University of Cambridge,” Technical Report February 2011.

Riehl, Evan, “Do less informative college admission exams reduce earnings inequality? Evidence from Colombia,” *Journal of Labor Economics*, April 2023.

Saboe, Matt and Sabrina Terrizzi, “SAT optional policies: Do they influence graduate quality, selectivity or diversity?,” *Economics Letters*, January 2019, 174, 13–17.

Sacerdote, Bruce, Douglas Staiger, and Michele Tine, “How Test Optional Policies in College Admissions Disproportionately Harm High Achieving Applicants from Disadvantaged Backgrounds,” Technical Report w33389, National Bureau of Economic Research, Cambridge, MA January 2025.

Sun, Liyang and Sarah Abraham, “Estimating dynamic treatment effects in event studies with heterogeneous treatment effects,” *Journal of Econometrics*, December 2021, 225 (2), 175–199.

UCAS, “Unpacking Qualification Reform: Results from the UCAS survey on A level reform,” Technical Report January 2015.

— , “UCAS A level Survey 2016 Update,” Technical Report January 2016.

— , “UCAS Qualification Provision Survey 2017,” Technical Report January 2017.

— , “UCAS Qualification Provision Survey 2018,” Technical Report June 2018.

US Department of Education, “Student Access to and Enrollment in Mathematics, Science, and Computer Science Courses and Academic Programs in U.S. Public Schools,” New Data Release May 2024.

Zimmerman, Seth D., “The Returns to College Admission for Academically Marginal Students,” *Journal of Labor Economics*, October 2014, 32 (4), 711–754.

Exhibits

Table 1: Summary statistics of GCSE students in England, 2007-2013

	Full Sample	3+ A-levels	<3 A-levels	Low-income	High-income
GCSE Percentiles					
Math	50	77	40	41	57
English Language	50	77	40	41	57
English Lit	50	72	38	41	55
Education Outcomes					
3+ 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
First-class Degree	0.09	0.23	0.03	0.05	0.12
First or 2:1 Degree	0.26	0.64	0.11	0.16	0.34
Gets Degree within 3 years	0.34	0.77	0.17	0.23	0.42
Observations	4,466,505	1,247,160	3,219,340	1,824,285	1,587,375

NOTE.—Summary statistics for students who completed GCSEs between 2007-2013 in England. Low-income students defined as students in the bottom two quintiles of neighborhood income (IMD 1 and 2) and high-income students defined as students in the bottom two quintiles of neighborhood income (IMD 4 and 5). GCSE percentiles are calculated within subject and qualification year.

Table 2: Characteristics of marginal university entrants

	Low-income compliers	Low-income university students
Characteristics		
White	64%	76%
Female	65%	56%
First-gen	64%	48%
Below-median Math GCSE	54%	28%
Avg. ranking of university attended	98	74
Performance		
Drops out	21%	13%
On-time degree completion	41%	66%

NOTE.— This table presents marginal mean characteristics for the complier population, i.e. low-income students induced into attending university by the reform. 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. Marginal means are estimated using 2SLS for students who finished high school between 2012–2017, controlling for school and year fixed effects.

Table 3: Adjusted R^2 in regressions of university outcomes

	Graduated (on-time)			Graduated with First		
	Teacher	Test	Teacher + Test	Teacher	Test	Teacher + Test
Low-income						
Adjusted R^2	0.092	0.111	0.112	0.066	0.092	0.094
Observations	75,535	75,535	75,535	75,535	75,535	75,535
High-income						
Adjusted R^2	0.117	0.126	0.127	0.069	0.092	0.094
Observations	89,230	89,230	89,230	89,230	89,230	89,230

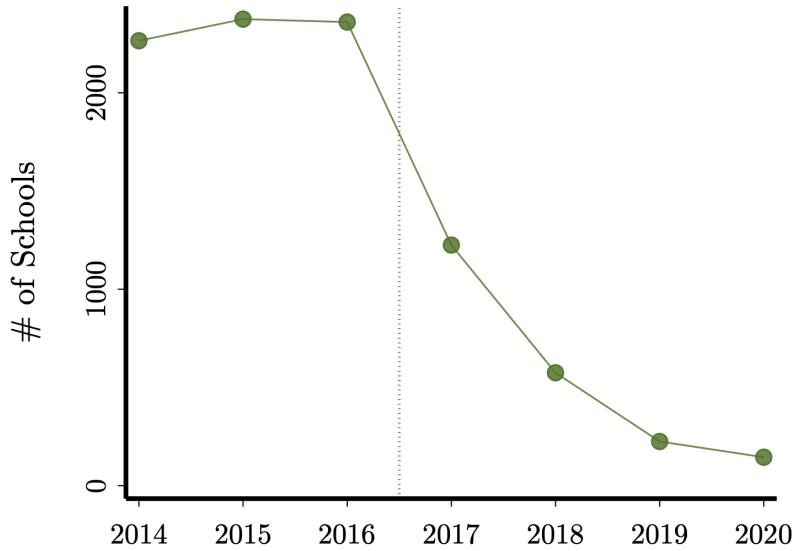
NOTE.— This table presents adjusted R^2 values for OLS regressions of university outcomes. Regressions in columns 1 and 4 under “Teacher” include only the average standardized teacher-assigned grades in A-level high school courses. Regressions in columns 2 and 5 under “Test” include only average standardized endline exam grades in A-level high school courses. Regressions in columns 3 and 6 under “Teacher + Test” include average standardized A-level 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 A-levels and began university between 2012–2016.

Table 4: Predictive power of teacher-assigned grades on university and post-graduate outcomes

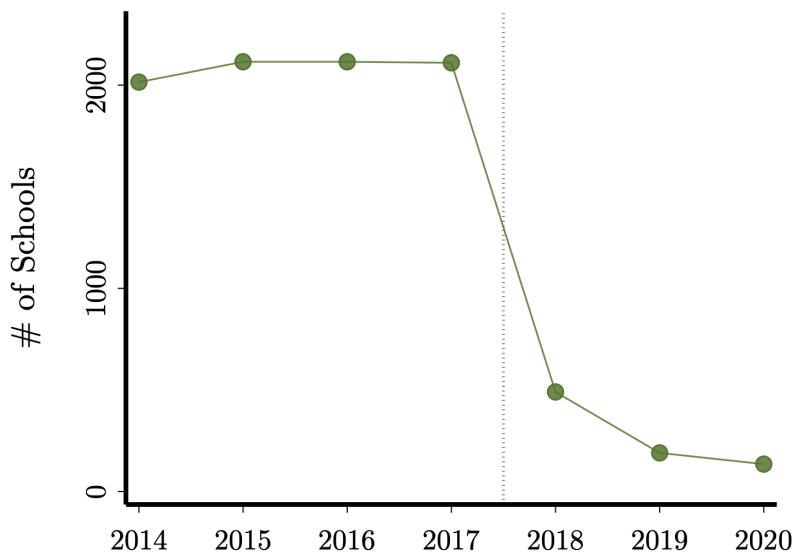
	Applied to Selective Uni	Attended Selective Uni	Graduated Uni	First-class Degree if graduated	Log Earnings 5-yrs
STEM Subjects					
Std Teacher-assigned resid	0.081*** (0.001)	0.052*** (0.001)	-0.004*** (0.001)	-0.004*** (0.001)	-0.016*** (0.002)
R^2	0.12	0.15	0.02	0.03	0.02
Mean Outcome	0.79	0.49	0.87	0.33	9.49
Observations	474,755	408,550	441,745	380.165	411,160
Non-STEM Subjects					
Std Teacher-assigned resid	0.090*** (0.001)	0.046*** (0.001)	0.005*** (0.001)	0.007*** (0.001)	-0.002 (0.001)
R^2	0.15	0.18	0.02	0.03	0.02
Mean Outcome	0.57	0.31	0.86	0.25	9.53
Observations	886,870	757,440	794,195	672,695	796,355

NOTE.— This table shows OLS coefficients from regressions on outcome variables separately by students taking STEM and non-STEM A-levels. Sample restricted to students in England taking A-levels between 2010-2015. All regressions include school and year fixed effects. Robust standard errors are reported in parentheses.

Figure 1: Number of schools requiring AS-levels in reformed subjects over time



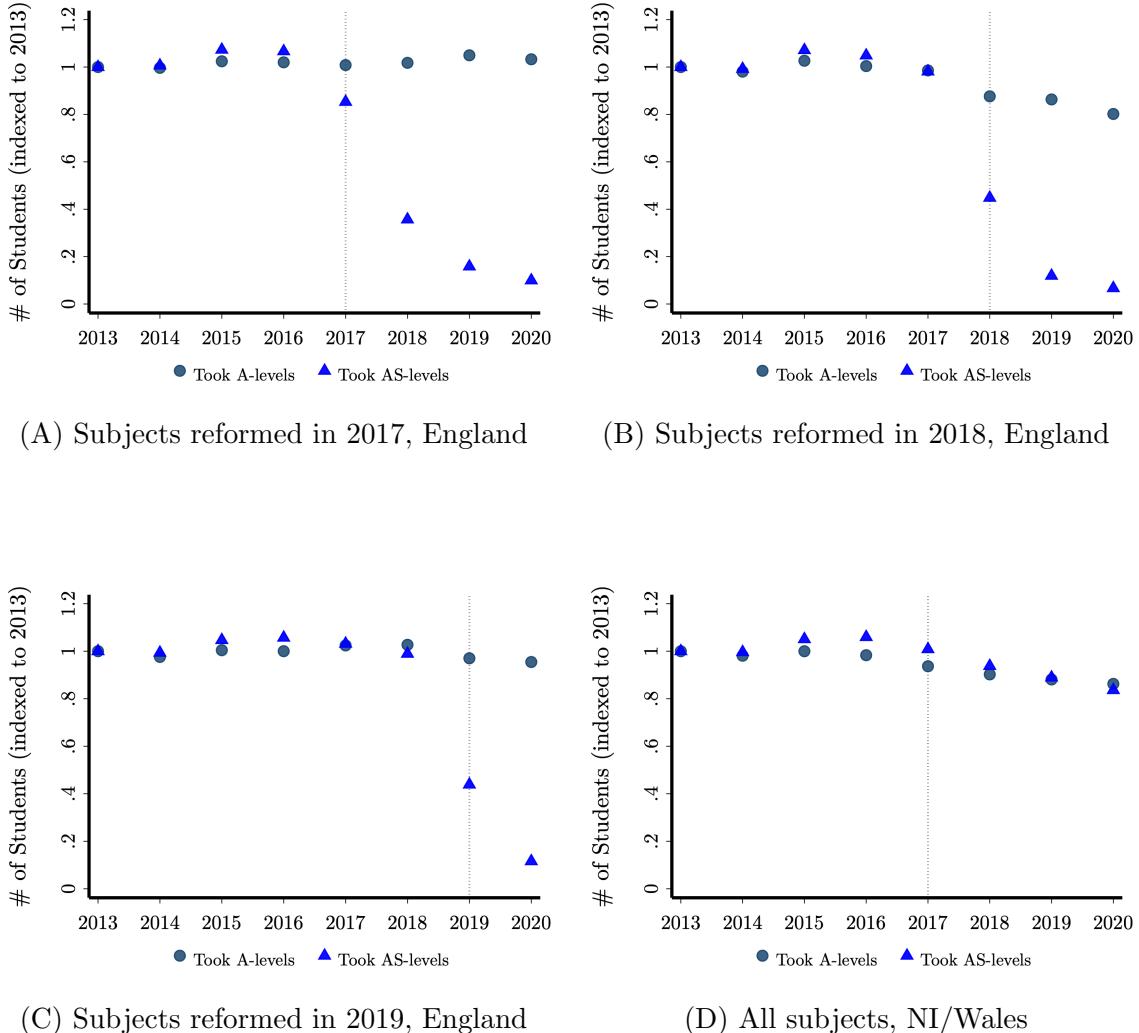
(A) Schools requiring AS-levels in 2017-reformed subjects



(B) Schools requiring AS-levels in 2018-reformed subjects

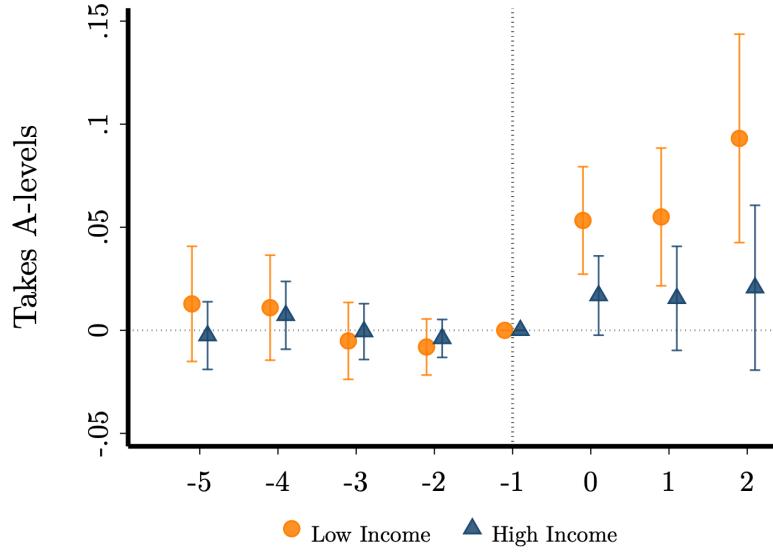
NOTE.– This figure shows the number of high schools in England that required AS-levels reformed subjects between 2014-2020. Panel A shows schools that required AS-levels in 2017-reformed subjects while Panel B shows schools that required AS-levels in 2018-reformed subjects. We define requiring AS-levels as observing more than 95% of the A-level students taking reformed subjects at the school who also take AS-level exams. Sample restricted to schools where more than 5 students a year took A-levels in 2017-reformed subjects for Panel A and 2018-reformed subjects for Panel B.

Figure 2: Time series of students taking A-level and AS-level exams by subject reform year

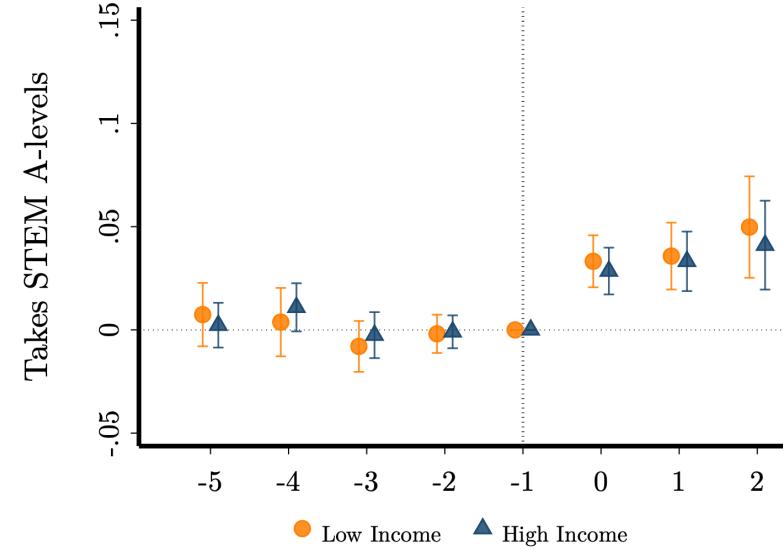


NOTE.— Time series of the total number of students A-level and AS-level exams taken indexed to 2013 numbers. Results presented separately by geographic region (England vs. Northern Ireland and Wales) and subject reform year. The complete list of subject titles and corresponding reform years can be found 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 Guilds, Eduqas, OCR, NCFE, Pearson and WJEC).

Figure 3: Shifts into A-levels and STEM, by income



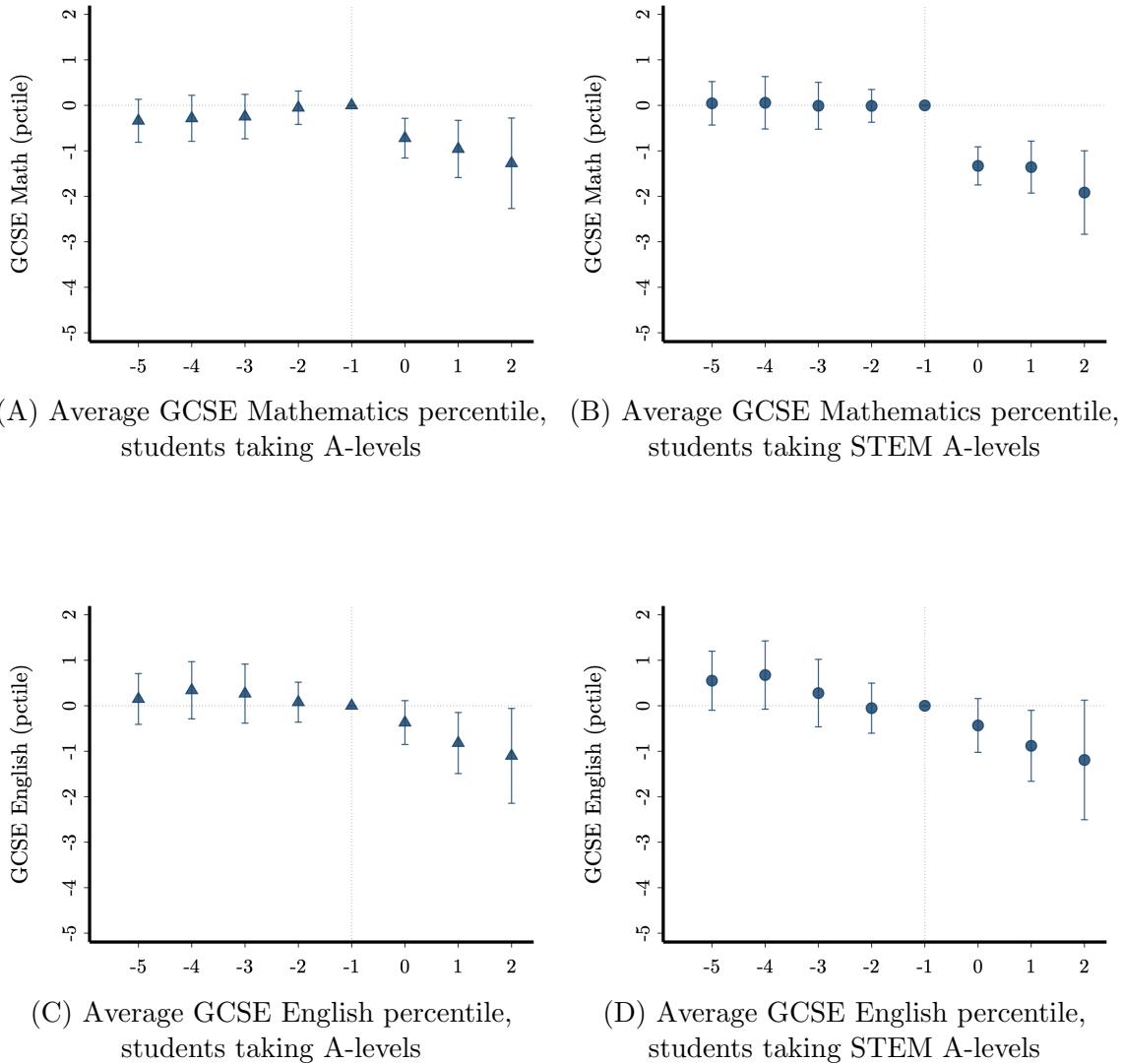
(A) Take-up of A-levels



(B) Take-up of STEM A-levels

NOTE.— Event study using specification in (1), with the outcome being an indicator for whether a student took at least three or more A-levels in Panel A and an indicator for whether a student took a STEM A-level (Biology, Chemistry, Physics) in Panel B. In each plot, the orange circular marker restricts to students in neighborhoods in the bottom two quintiles of the Index of Multiple Deprivation (IMD). The blue triangular marker restricts to students in neighborhoods in the top two quintiles of IMD. 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.

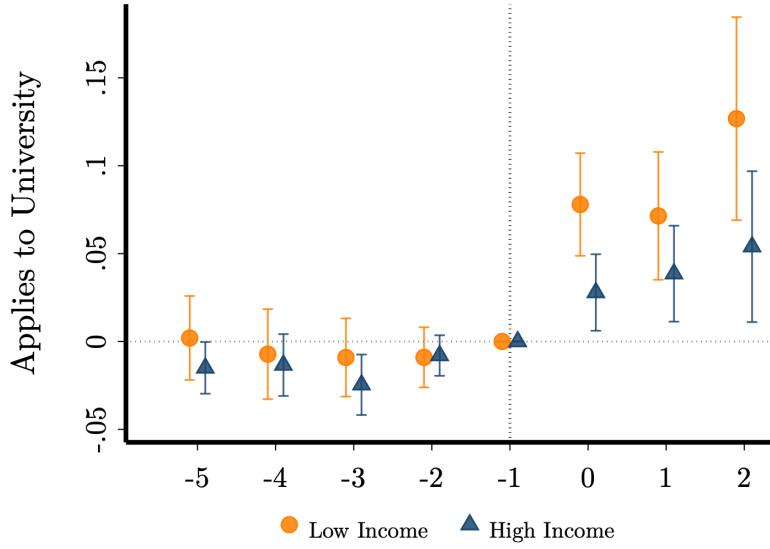
Figure 4: Negative academic selection in A-levels and STEM A-levels after the reform



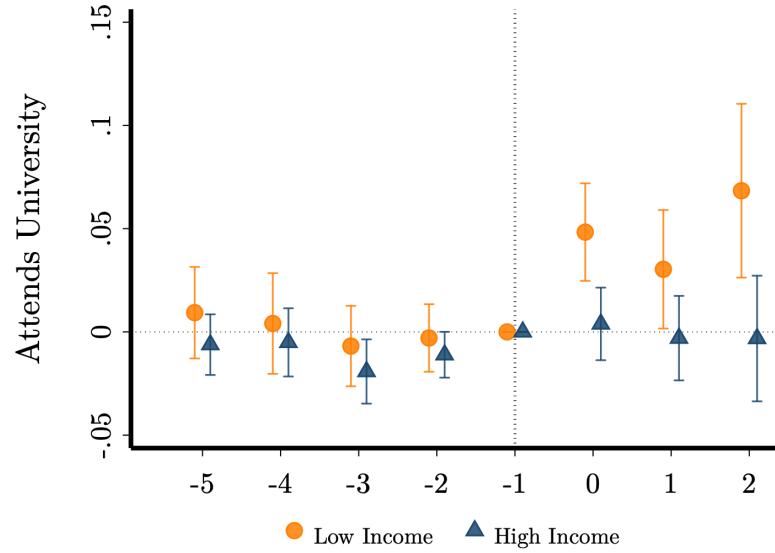
NOTE.— Event study using specification in (1) where the outcome is students' average GCSE Mathematics percentiles in Panels A and B and students' average GCSE English percentiles in Panels C and D.

Treatment is at the school level. Sample: Students age 18 at the end of high school in England between 2010-2019 taking A-levels in Panels A and B and taking STEM A-levels in Panels C and D.

Figure 5: Shifts into university, by income



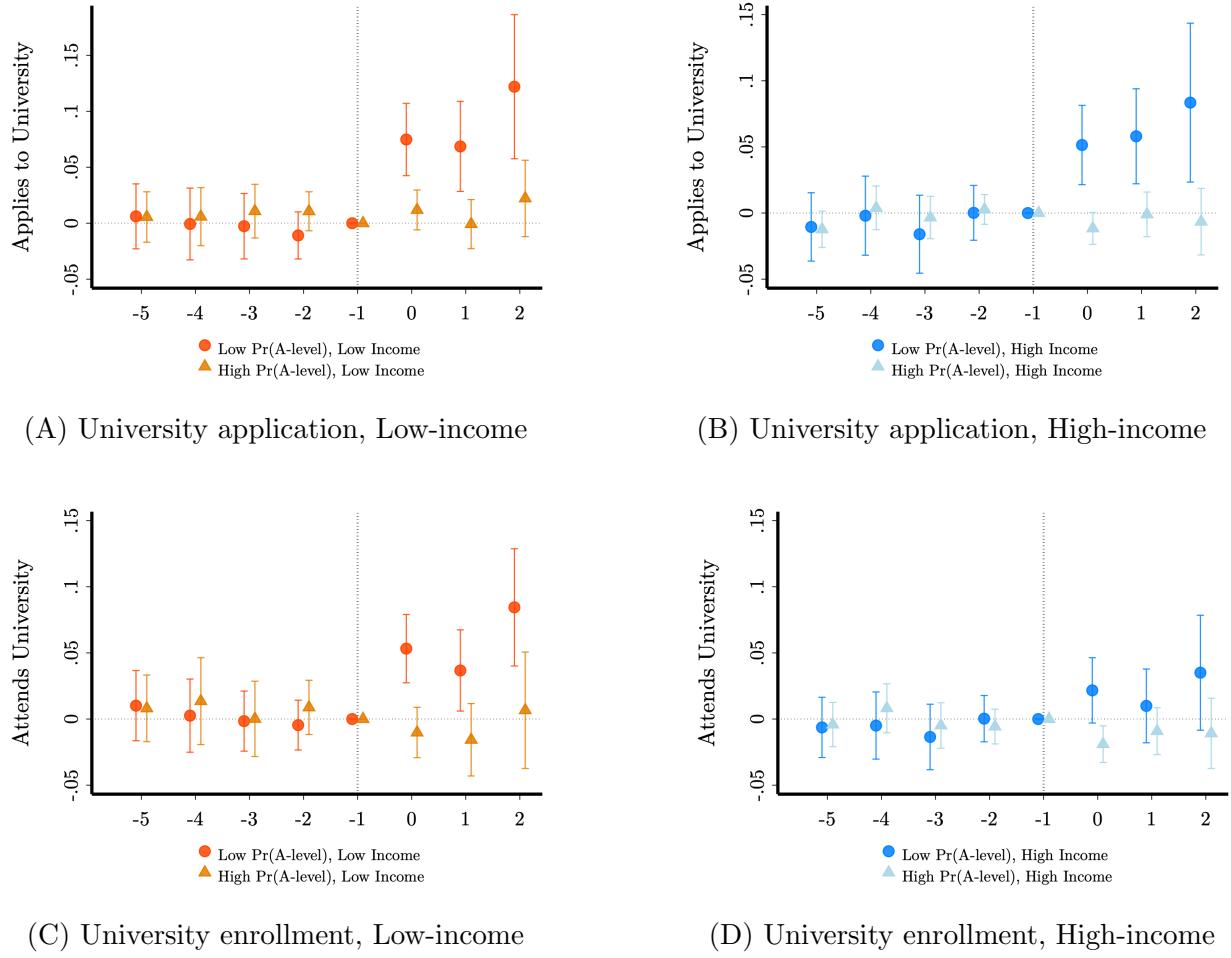
(A) University application



(B) University enrollment

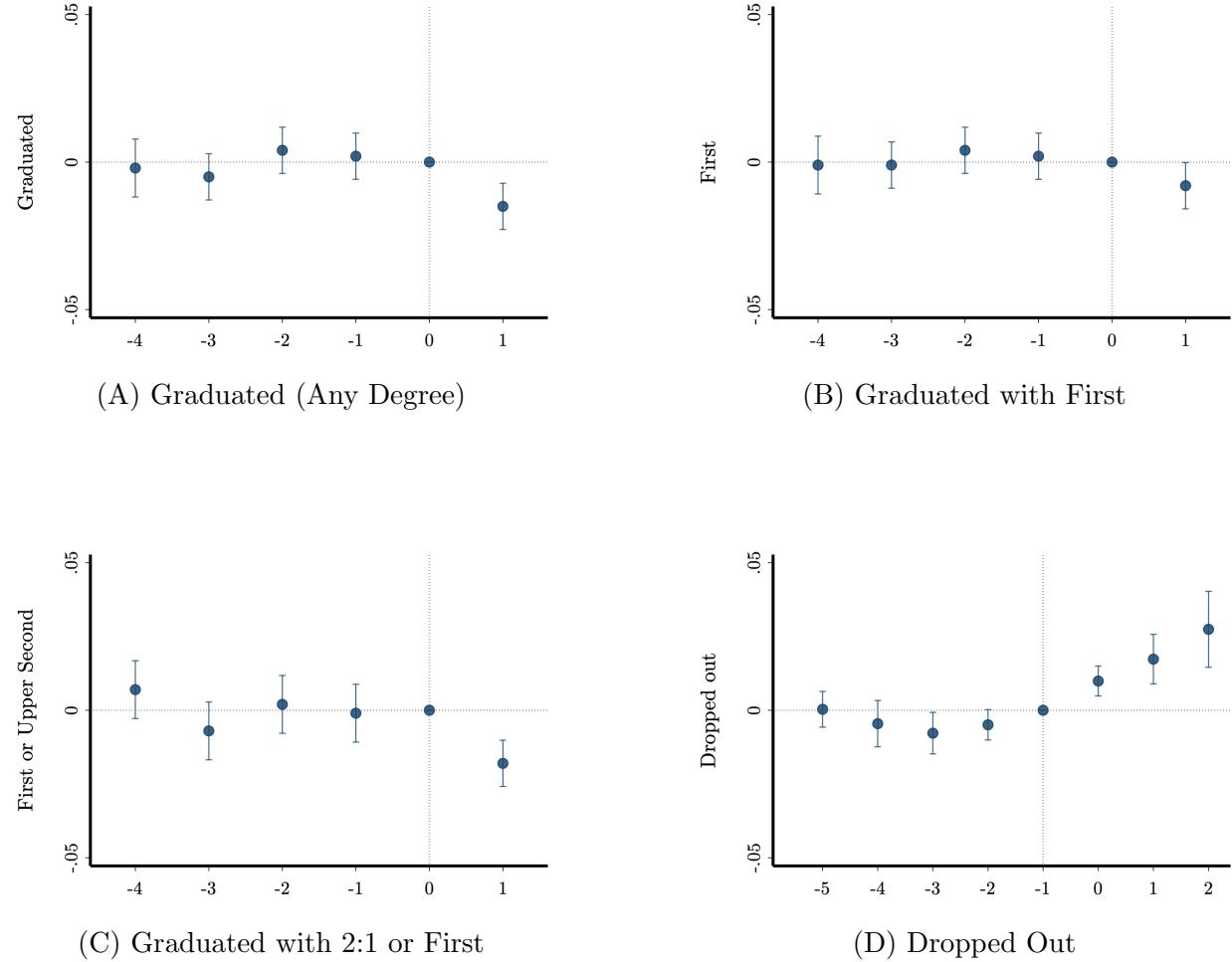
NOTE.— Event study estimates from Equation 1, with the outcome being an indicator for whether a student applied to university at age 18 in Panel A and an indicator for whether a student enrolled in university at age 18 in Panel B. In each plot, the orange circular marker restricts to students in neighborhoods in the bottom two quintiles of the Index of Multiple Deprivation (IMD). The blue triangular marker restricts to students in neighborhoods in the top two quintiles of IMD. 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 6: Shifts into university, by predicted propensity to take A-levels



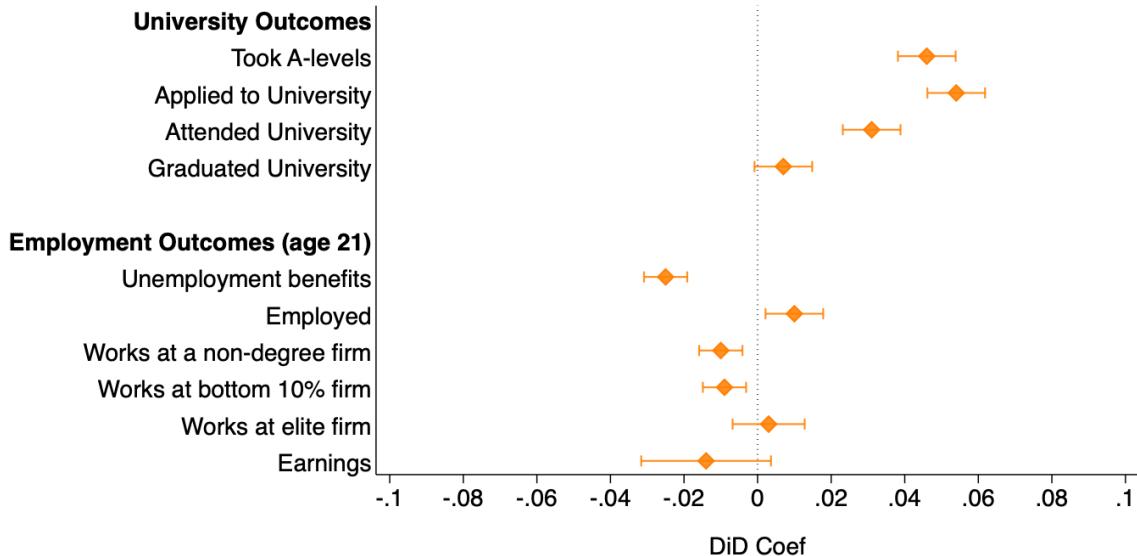
NOTE.— Event study estimates from Equation 1, with the outcome being an indicator for whether a student applied to university at age 18 in Panels A and B and an indicator for whether a student enrolled in university at age 18 in Panels C and D. In each plot, the orange markers restrict to low-income 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 7: Event study estimates on degree attainment conditional on attending university

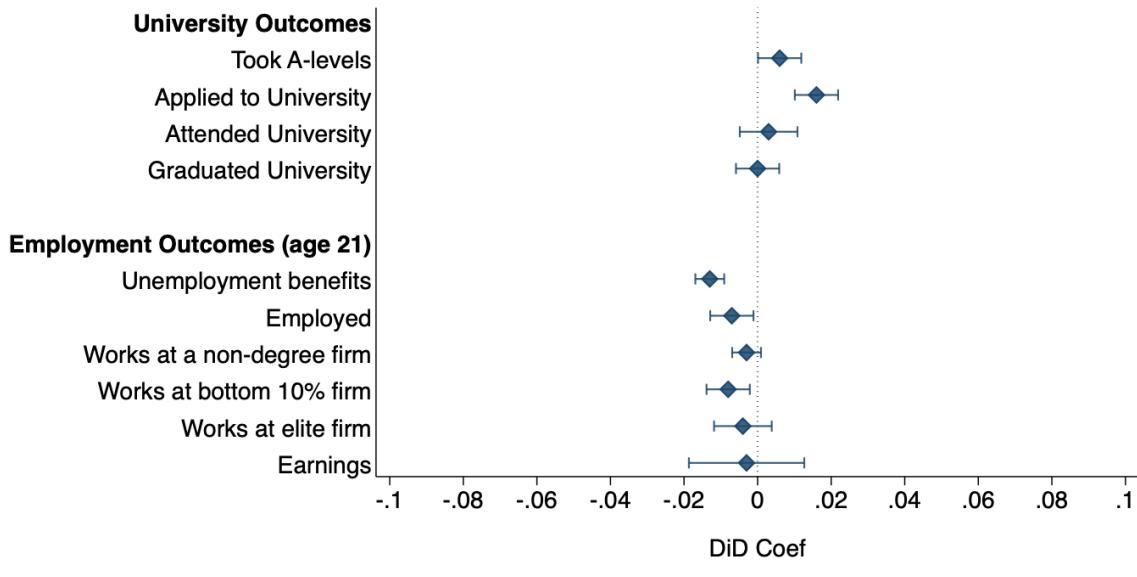


NOTE.—Panels A, B and C present OLS coefficients from event study regressions in Equation 5 on observed degree outcomes at age 21, comparing university students from schools that stopped offering AS-levels beginning in academic year 2016-17 to schools that still required all students to take AS-levels in reformed subjects in 2016-17 and schools in Northern Ireland and Wales. Panel D reports estimates from the staggered event-study specified in Equation 1, comparing 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. All regressions include university fixed effects. Sample (Panels A, B, C): university students who started an undergraduate program between 2012-2017. Sample (Panel D): university students who started an undergraduate program between 2012-2019. Event study coefficients presented in Table E3.

Figure 8: Summary of educational and post-graduate outcomes by income



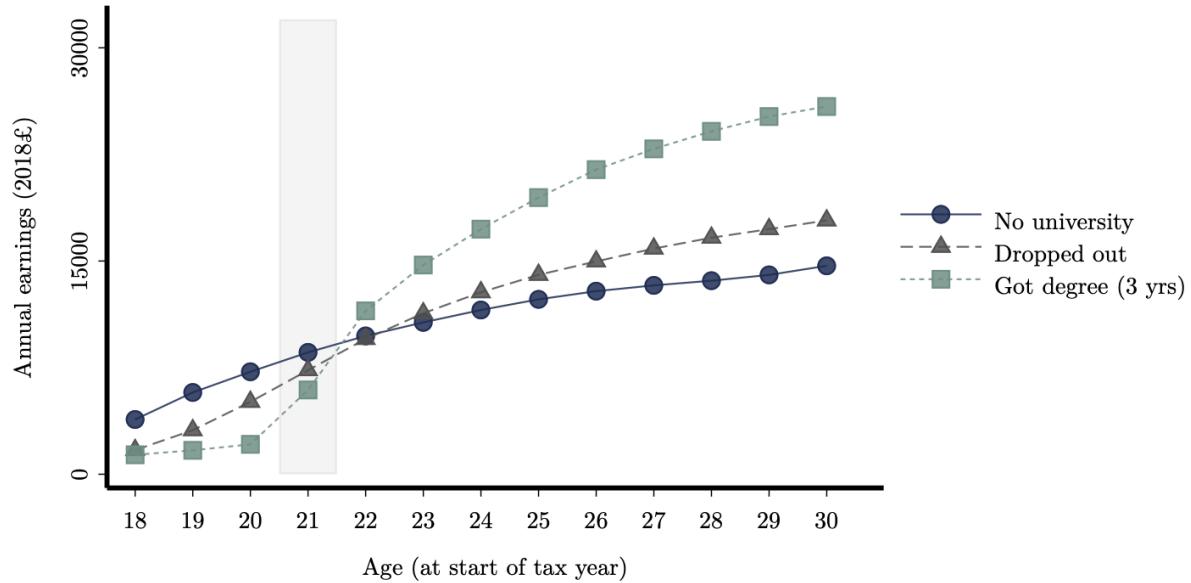
(A) Low-income students



(B) High-income students

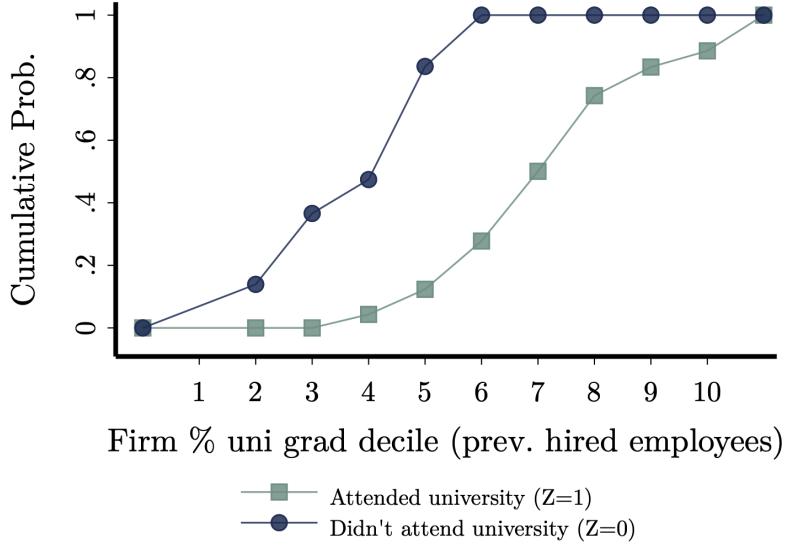
NOTE.— Difference-in-differences coefficients from Equation 5 for outcomes separately by low-income students and high-income students. Non-degree firms are defined as firms at which 0% of the employees hired prior to tax year 2010-11 had university degrees. Elite firms are defined as firms where > 50% 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. Predicted outcomes are based on age 27 earnings projected using observed employment outcomes between ages 18-21 using Equation 9. Sample: Students age 18 at the end of high school in England between 2010-2017 in the NPD.

Figure 9: Historical earnings profile by education

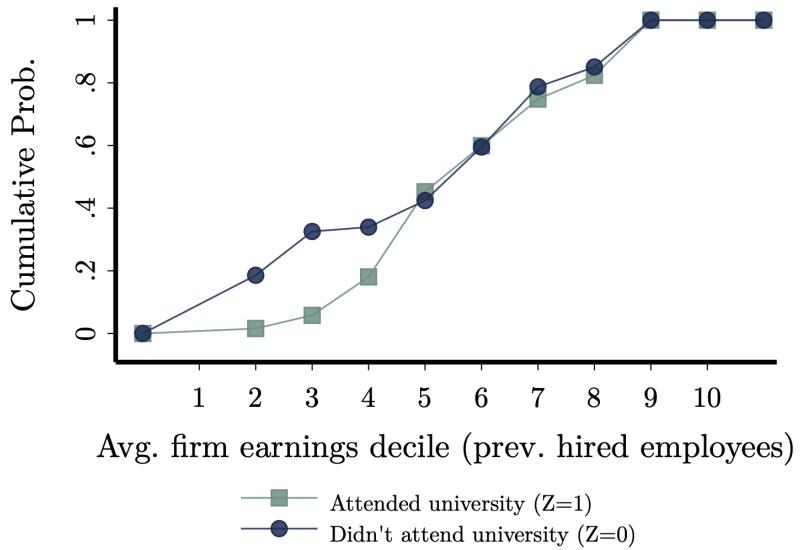


NOTE.— This figure shows earnings profiles for adults in England in the HMRC data between ages 18 through 30, separately by (i) workers who never attended university ; (ii) workers who attended university at 18 but never graduated; (iii) workers who attended university and completed a degree in the standard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

Figure 10: Distributional effects of age 21 firms, compliers into university



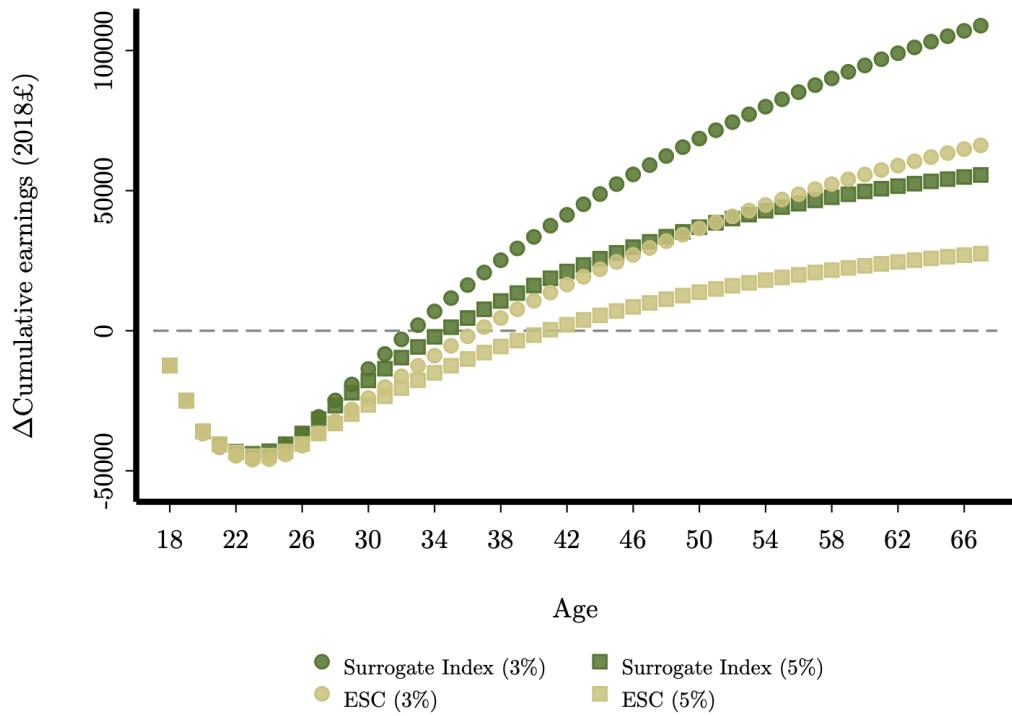
(A) CDF of age 21 firm education decile



(B) CDF of age 21 firm pay decile

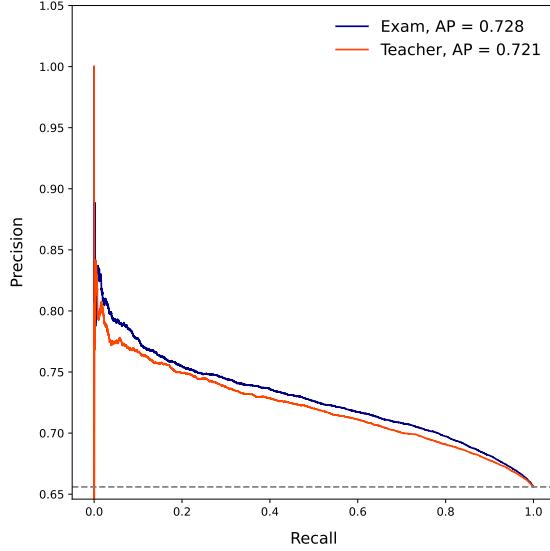
NOTE.— This figure compares potential outcome distributions of age 21 firm characteristics for compliers who are induced into attending university to compliers who are not induced into university. Distributions are estimated from Equations 7 and 8. Panel A presents distributions of firm education decile, defined by the share of employees hired in tax years 2007-08 through 2010-11 with university degrees. Panel AB presents distributions of firm earnings decile, defined by the average earnings of employees hired in tax years 2007-08 through 2010-11 with university degrees.

Figure 11: Cumulative lifetime earnings gains for low-income compliers into university

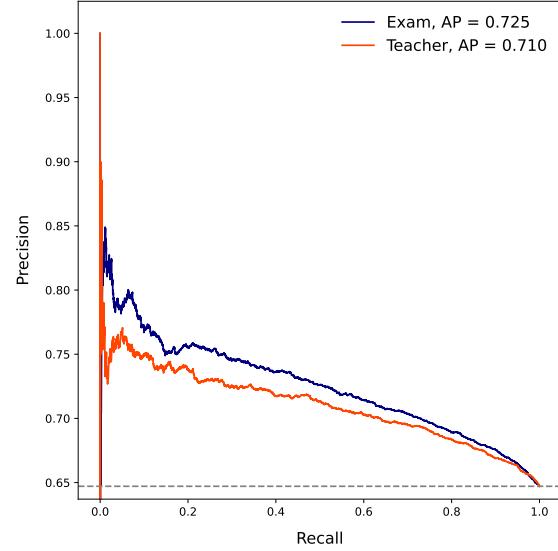


NOTE.— This figure presents cumulative lifetime earnings gains for low-income students estimated using both the surrogate index (Athey et al., 2019) and ESC (Athey et al., 2025) projections at 3% and 5% discount rates. We project age 27 earnings out until age 67. Earnings effects between age 22 to 27 are linearly interpolated. Earnings effects for ages 18 through 21 are estimated using 2SLS on the low-income quasi-experimental sample and are net of tuition priced at 3 x £9,250.

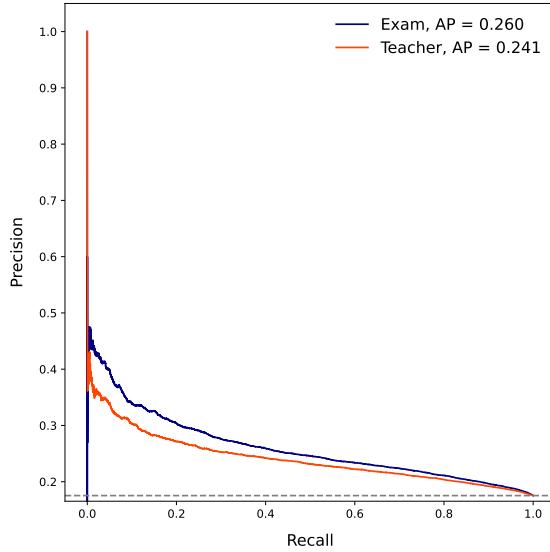
Figure 12: Out-of-sample predictive power of teacher-assigned grades vs. test grades



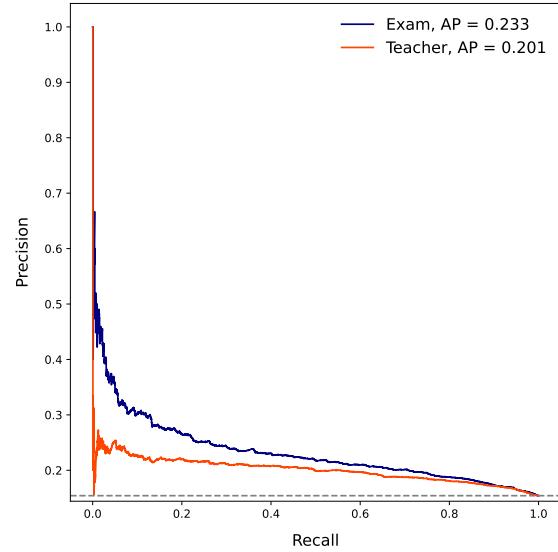
(A) Graduates (3yrs), All students



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



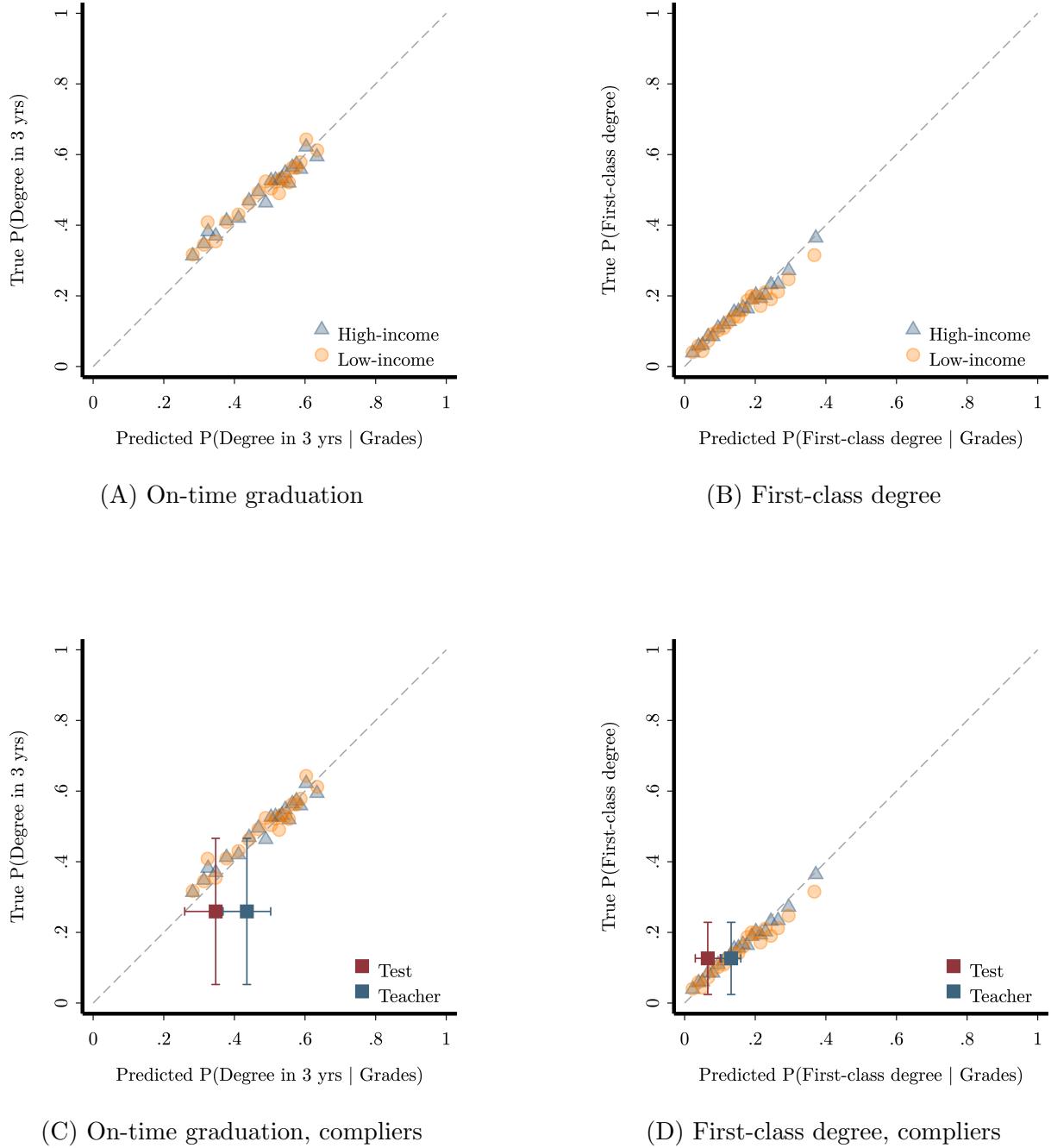
(C) Graduates with First, All students



(D) Graduates with First, Low-income

NOTE.– 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 20-25% of graduating class). Average precision (AP) metrics for each model are reported. Dashed gray lines represent benchmark for performance, the testing sample average positive class rate.

Figure 13: Calibration bias for low-income marginal entrants



NOTE.— This figure plots average university outcomes Y and predicted university outcomes based on test scores, \hat{Y} , for low-income and high-income university students (Panels A and B). Panels C and D overlay the marginal mean Y for low-income compliers induced into university, estimated in Equation 13 against the predicted \hat{Y}^{test} in red and predicted $\hat{Y}^{teacher}$ in blue. Sample: university students who begin university between 2012-2017 and compliers estimated in a 2SLS regression of 2012-2017 high school graduates in England.

Appendix

A Data

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).

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. All earnings were winsorized at the 99% and 1% levels within a tax year and inflation adjusted to 2015 GBP using the CPIH index. We restrict our analysis to exclude individuals who work at more than 10 firms in a tax year (< 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.³⁰

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. Next, 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.

³⁰We do not consider earlier years since the HESA dataset begins in 2005-06 so we are unable to observe university attendance.

B Additional background and historical context

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.³¹ BTEC grades were assigned based on coursework and were one of Unclassified, Near Pass, Pass, Merit, Distinction, and Distinction* in ascending order.

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

³¹ 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.

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 £3000 from 2006–2011 and increased to £9000 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 B1 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

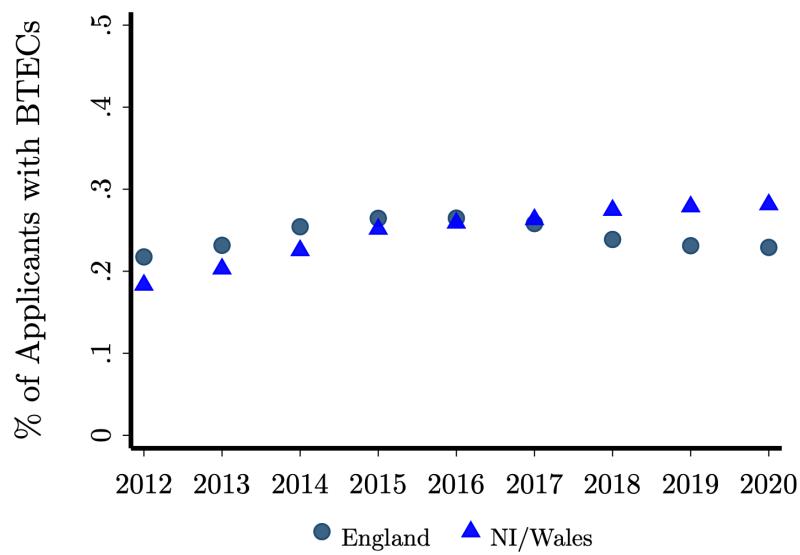
with a university through the main phase of applicants, with just over 60% attending their firm choice. Although around 1 in 4 students miss both of their offer conditions, only 15% of applicants are unmatched with universities at the end of each application cycle as 11-13% of applicants match through Clearing. During our sample period, the share of university applicants taking BTECs was increasing across England, Northern Ireland, and Wales prior to 2017. Figure B1 shows that after 2017, the rise in BTECs continues among applicants from Northern Ireland and Wales, but tapers off among English applicants.

Table B1: Summary statistics for university applicants

	2012–2016	2017–2019
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
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
University Outcomes		
Number of Offers	3.80	3.95
Share of Conditional Offers	0.99	0.94
Share of Unconditional Offers	0.01	0.06
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
Observations	1,112,608	674,662

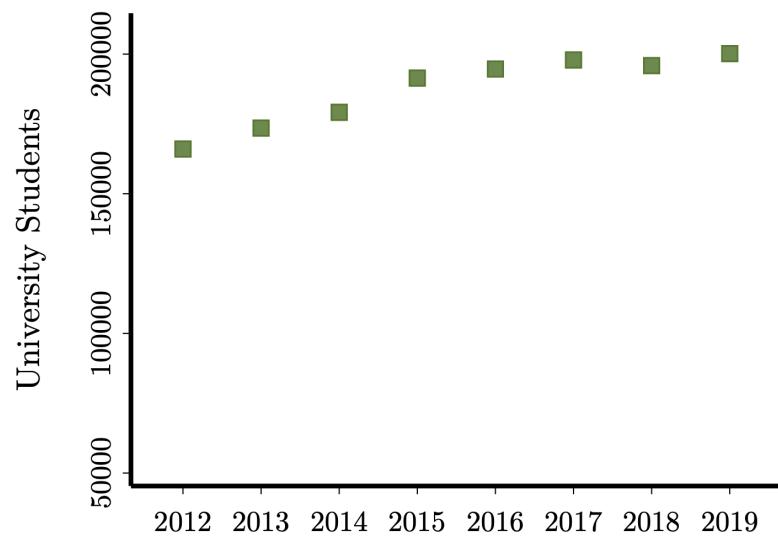
NOTE.– This table presents summary statistics for students in the UK (excl. Scotland) applying to university between 2012 through 2019.

Figure B1: Share of university applicants with BTECs



NOTE.– Time series illustrating the share of students in England, Northern Ireland, and Wales who applied to university with BTECs between 2012-2020.

Figure B2: University enrollment over time



NOTE.– Time series illustrating the number of UK university students over time by year of enrollment. Sample restricted to students 18 years of age at time of enrollment who applied from England.

C Supplementary tables and figures on AS-level reform

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).³² 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

³²Similarly, 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.

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: Timing of A-level reform (dropping AS-level requirement) 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

NOTE.– 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: Pre-reform summary statistics of high schools in England by adoption year

	Treated Schools (Switchers)			Control Schools	Excluded Schools
	2017	2018	2019	After 2019	Mixed Policies
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
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
# Schools	560	390	115	165	2105

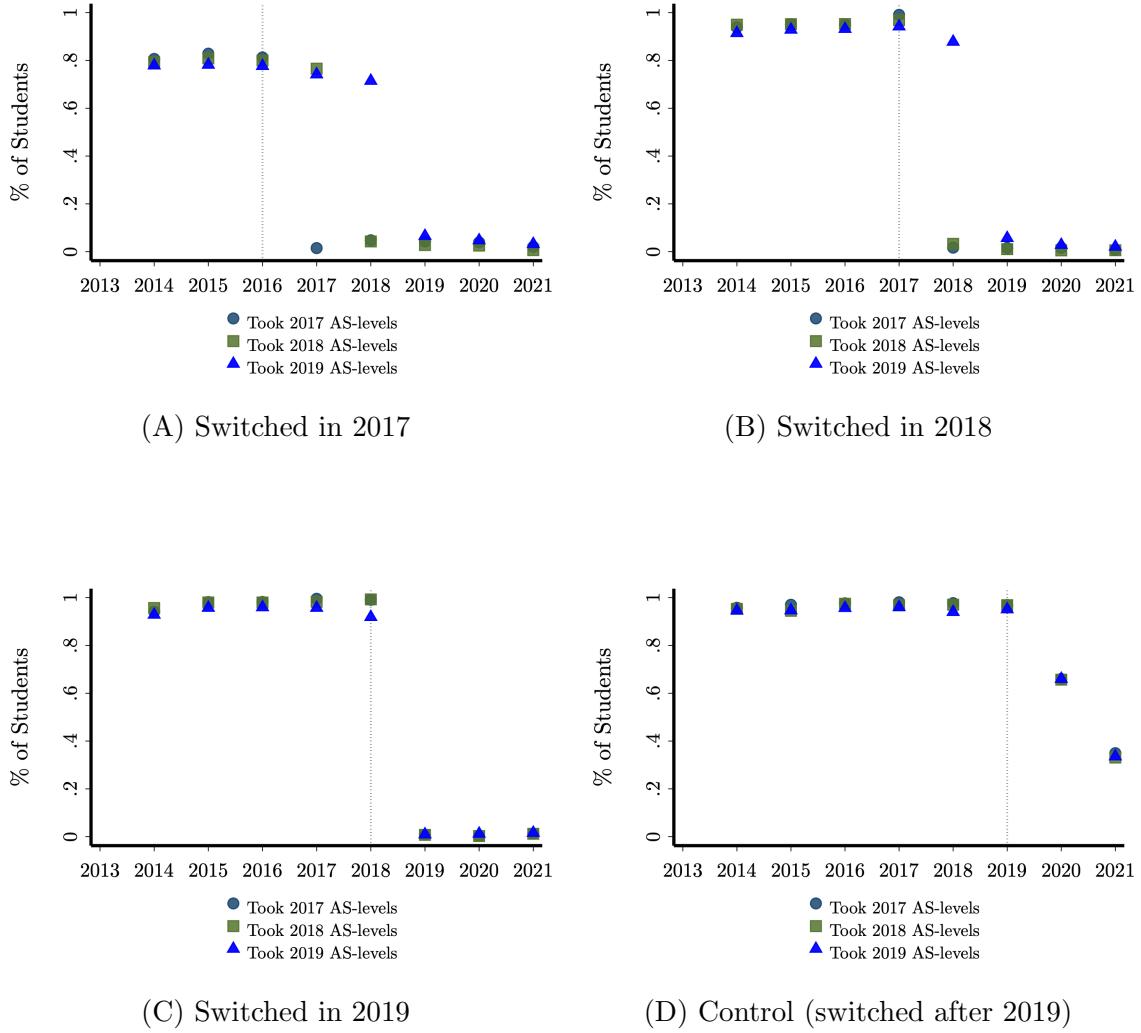
NOTE.— This table presents summary statistics in academic years 2011-12 through 2015-16 (pre-reform) for high schools in England by year adoption. The treated schools are schools that switched from having >90% of A-level students in reformed subjects taking AS-levels post-reform to <5% of A-level students 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 C1. 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 income-propensity cells

	Low Predicted P(A-levels)		High Predicted P(A-levels)	
	Low-income	High-income	Low-income	High-income
GCSE performance				
GCSE Maths percentile	27.87	31.36	68.13	70.68
GCSE English percentile	26.64	28.62	69.99	71.93
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
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
Observations	1,188,355	985,225	675,535	1,522,440

NOTE.– 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: Share of A-level students taking AS-levels by subject reform year and school adoption year



NOTE.— 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.

D Regression estimates used in projected earnings

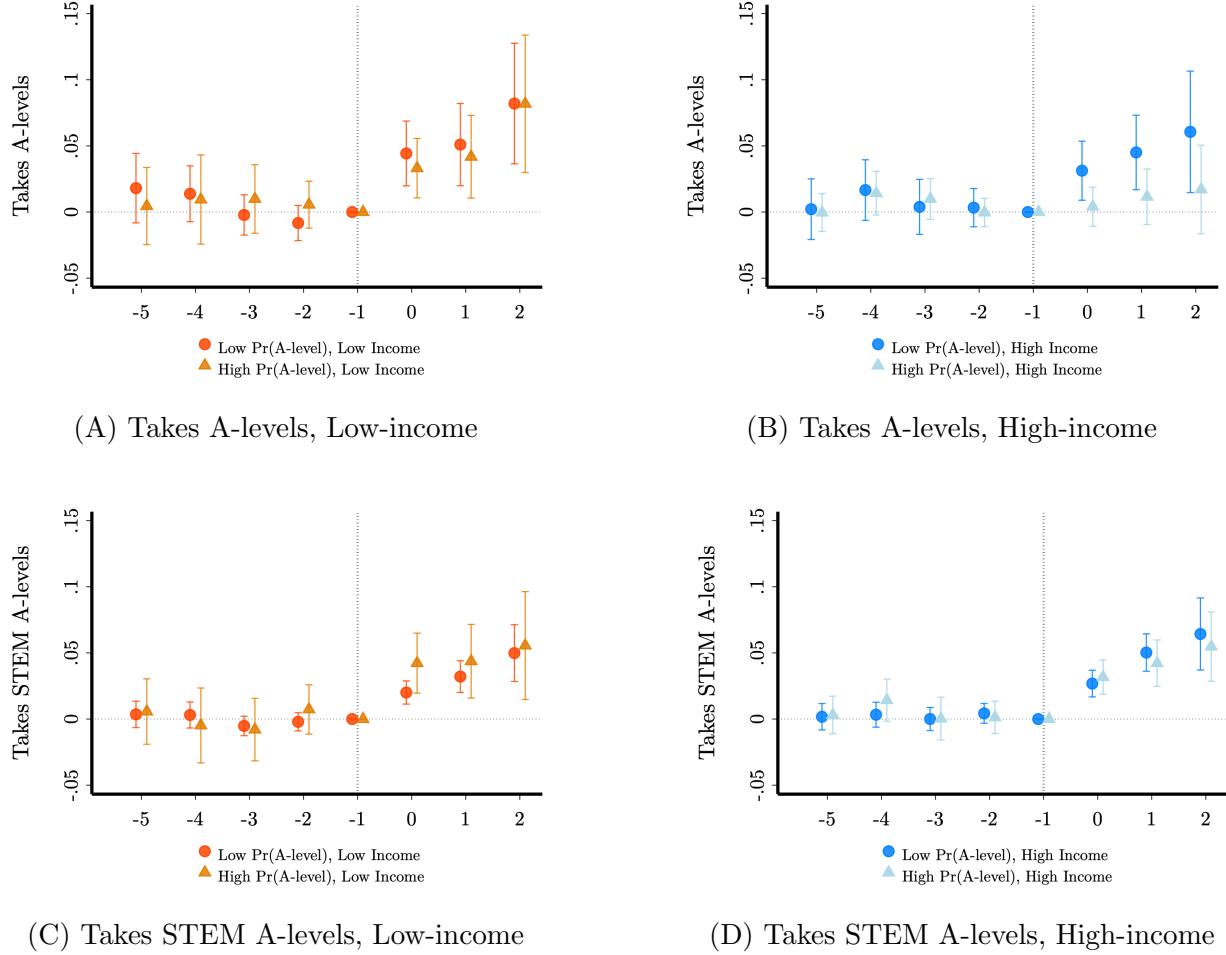
Table D1: Coefficients from Equations 9 and 10 used to project earnings

	Age 27 Earnings (GPB)	
Earnings (Age 21)	0.418*** (0.003)	0.280*** (0.003)
Earnings (Age 20)	0.022*** (0.004)	-0.023*** (0.004)
Earnings (Age 19)	-0.100*** (0.004)	-0.066*** (0.004)
Earnings (Age 18)	0.027*** (0.005)	0.032*** (0.005)
Unemployed (Age 21)	834.926*** (43.904)	-185.715 (694.018)
Unemployed (Age 20)	110.404*** (40.181)	-254.339*** (42.164)
Unemployed (Age 19)	-296.462*** (38.135)	-402.536*** (39.945)
Unemployed (Age 18)	-581.806*** (33.955)	-566.778*** (35.644)
Out-of-work Benefits (Age 21)	-741.278*** (39.864)	-1,429.815*** (42.151)
Out-of-work Benefits (Age 20)	-1,658.210*** (44.813)	-1,945.205*** (47.077)
Out-of-work Benefits (Age 19)	-1,689.799*** (42.763)	-1,646.122*** (44.843)
Out-of-work Benefits (Age 18)	-1,282.342*** (38.132)	-1,371.812*** (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)	
<i>R</i> ²	0.235	0.306
Observations	899,185	897,580
Outcome Mean	20,029.12	20,027.42
Firm, SIC2007 FE	N	Y
GCSE Math Decile, FE	Y	Y
Race, Gender, IMD FE	Y	Y

NOTE.– This table presents regression coefficients for Equations 9 and 10 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., 2019).

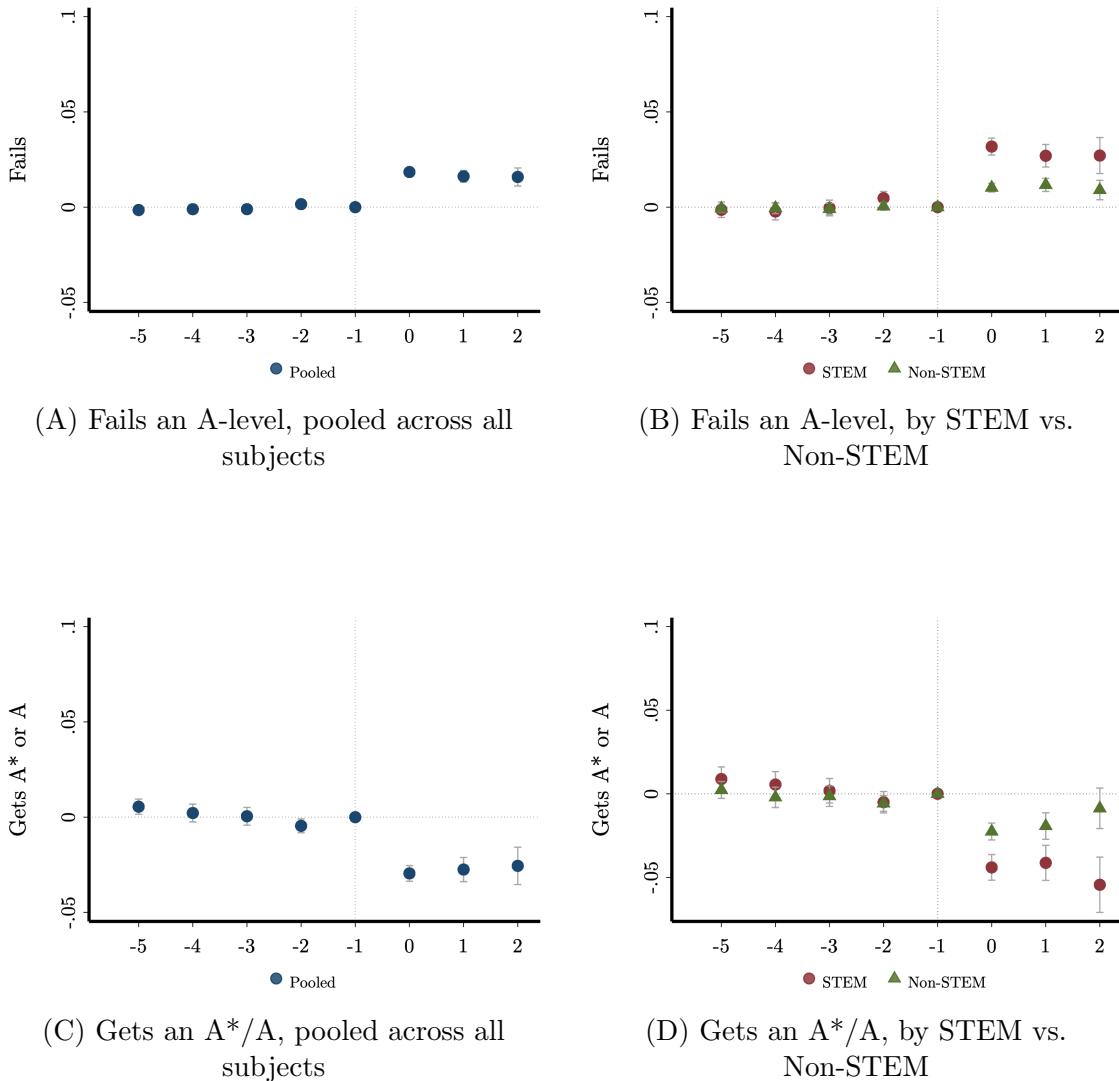
E Heterogeneity and robustness

Figure E1: Shifts into A-levels, by predicted propensity to take A-levels



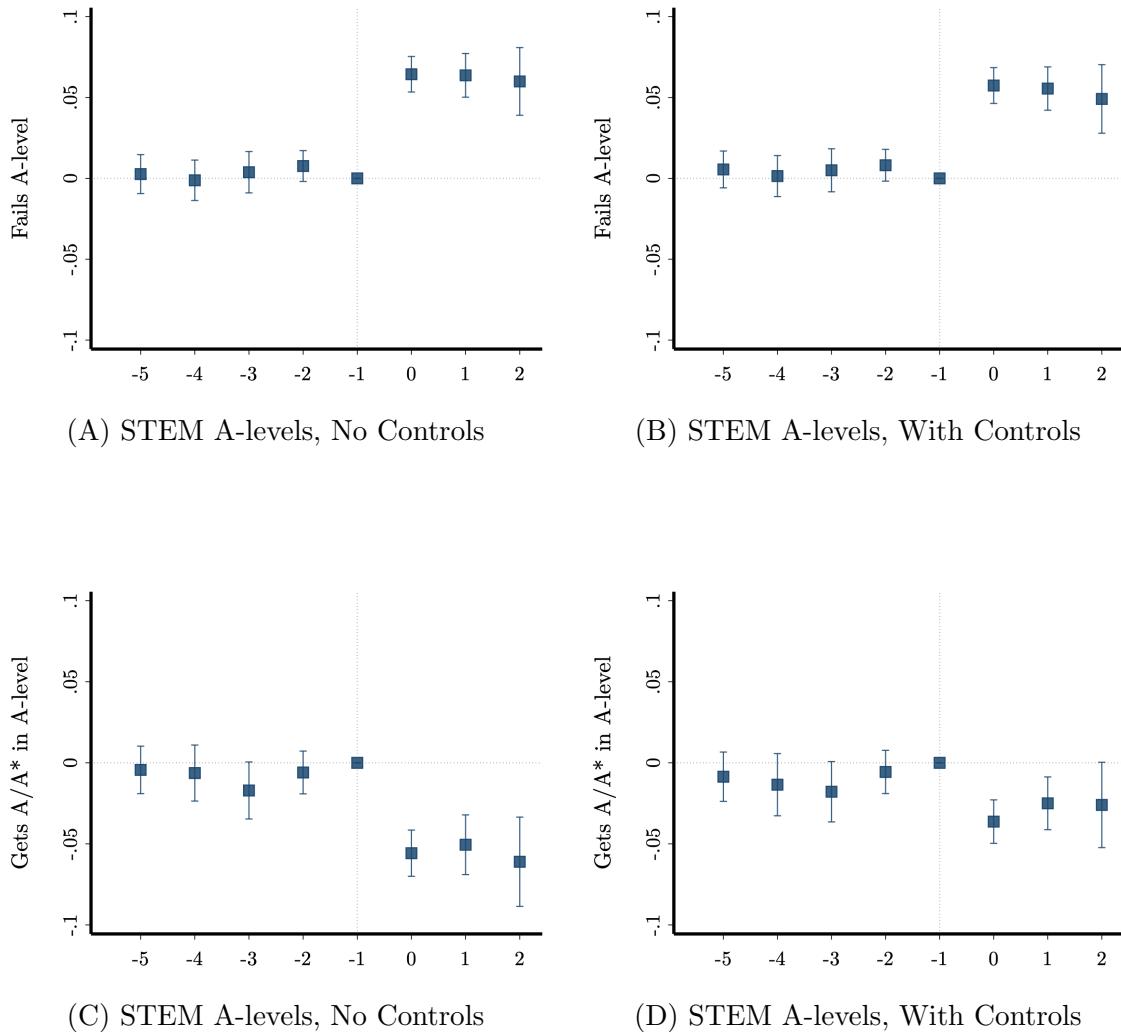
NOTE.– Event-study estimates from Equation 1, with the outcome being an indicator for whether a student took at least three or more A-levels in Panels A and B and an indicator for whether a student took a STEM A-level (Biology, Chemistry, Physics) in Panels C and D. In each plot, the orange markers restrict 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 E2: Changes in A-level performance



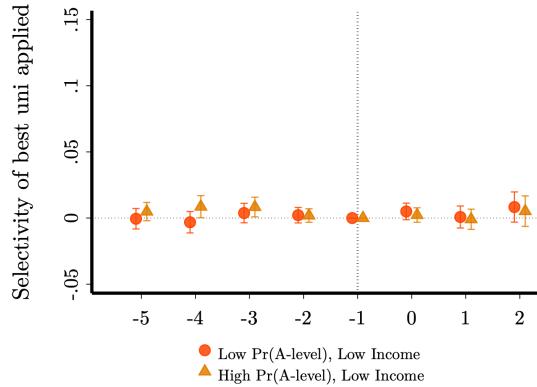
NOTE.– Event study using specification in (2), with the outcome being an indicator for whether a student failed an A-levels exam in Panels A and B and an indicator for whether a student got an A* or an A in Panels C and D. Treatment is at the school-subject level and fixed effects are included for student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

Figure E3: Changes in A-level performance, with controls

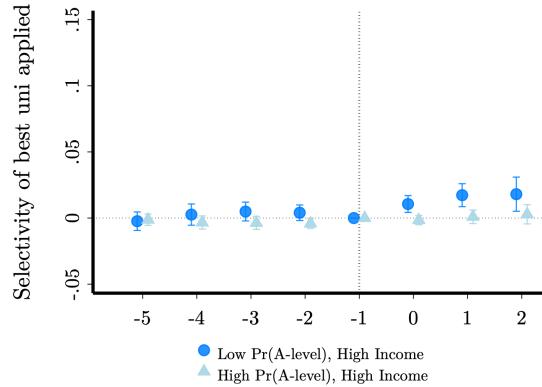


NOTE.— Event study using specification in (1), with the outcome being an indicator for whether a student taking STEM A-levels received a failing grade in Panels A and B and an indicator for whether a student taking STEM A-levels received a grade of A or A* in Panels C and D. Panels B and D include controls 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.

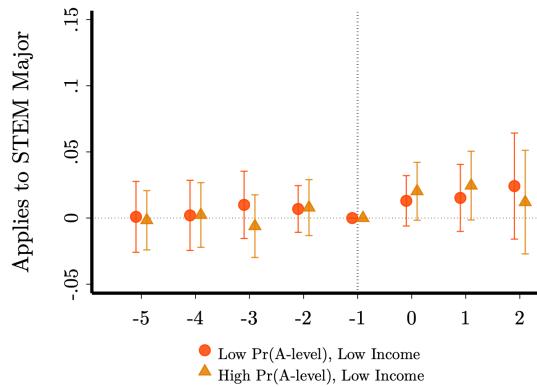
Figure E4: Changes in university application patterns conditional on applying



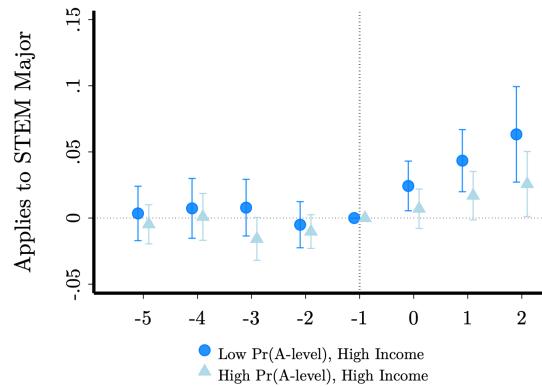
(A) Max university selectivity in application portfolio, Low-income



(B) Max university selectivity in application portfolio, High-income



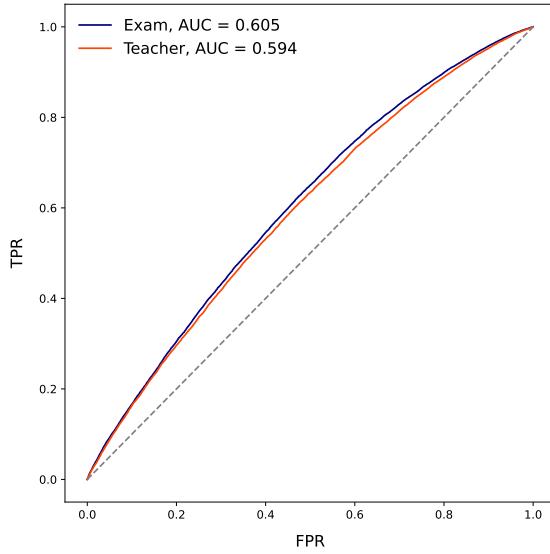
(C) Applies to STEM major, Low-income



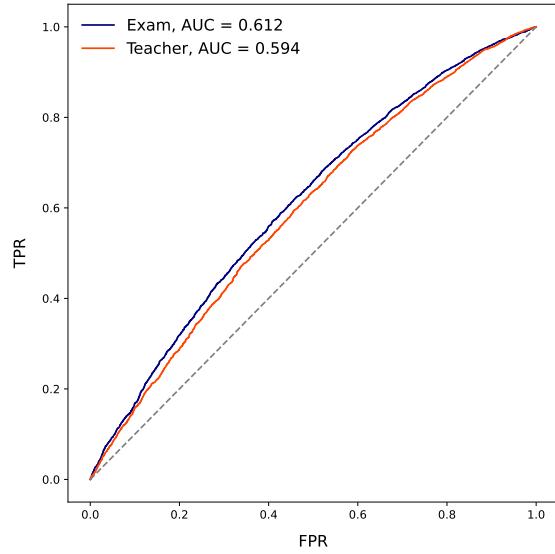
(D) Applies to STEM major, High-income

NOTE.— 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.

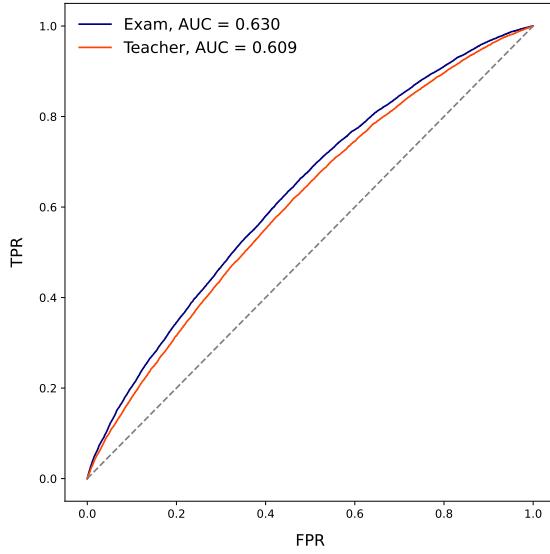
Figure E5: Out-of-sample predictive power of teachers vs. tests, ROC curves



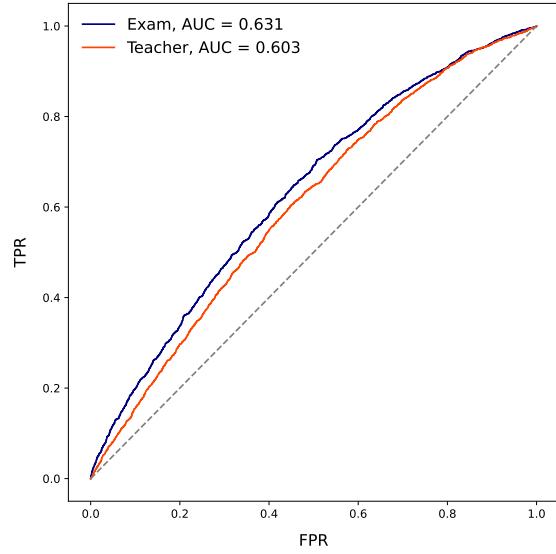
(A) Graduates (3yrs), All students



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



(C) Graduates with First, All students



(D) Graduates with First, Low-income

NOTE.– 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 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 E1: Robustness in event-study estimates on university degree outcomes

	Graduated University			Graduated with 2:1 or First		Graduated with First	
	NI/Wales	Within-England	Placebo, NI/Wales	NI/Wales	Within-England	NI/Wales	Within-England
$1\{t = -5\} \times 1\{\text{Treated}\}$	-0.012 (0.009)	-0.001 (0.005)	-0.013 (0.010)	0.003 (0.011)	0.007 (0.005)	-0.003 (0.008)	-0.001 (0.003)
$1\{t = -4\} \times 1\{\text{Treated}\}$	-0.007 (0.009)	-0.005 (0.004)	-0.002 (0.010)	-0.008 (0.010)	-0.007 (0.005)	0.001 (0.008)	-0.001 (0.003)
$1\{t = -3\} \times 1\{\text{Treated}\}$	-0.003 (0.009)	0.004 (0.004)	-0.007 (0.010)	-0.002 (0.010)	0.003 (0.005)	0.005 (0.008)	0.004 (0.003)
$1\{t = -2\} \times 1\{\text{Treated}\}$	-0.002 (0.009)	0.002 (0.004)	-0.005 (0.009)	0.002 (0.009)	-0.001 (0.004)	0.002 (0.008)	0.002 (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)	0.000 (0.000)
$1\{t = 0\} \times 1\{\text{Treated}\}$	-0.022*** (0.008)	-0.014*** (0.004)	-0.008 (0.008)	-0.037*** (0.009)	-0.016*** (0.005)	-0.018** (0.008)	-0.006* (0.003)
University FE	Y	Y	Y	Y	Y	Y	Y
Outcome Mean	0.71	0.72	0.71	0.57	0.58	0.21	0.21
Observations	215,680	455,542	289,901	215,680	456,790	215,680	456,790

NOTE.—This table presents robustness results for estimates from Equation 5. Columns 1, 4, and 6 compare university students from schools in England that eliminated AS-levels in 2017 to students from Wales/NI. Columns 2, 5, and 7 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. Column 3 presents placebo results comparing university students from schools in England that eliminated AS-levels in 2018 or later to students from Wales/NI. 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 E2: Heterogeneity in event-study estimates on university degree outcomes by previous school-level overprediction

	Graduated		Graduated with 2:1 or First		Graduated with First	
	Low-overpred	High-overpred	Low-overpred	High-overpred	Low-overpred	High-overpred
$1\{t=-5\} \times 1\{\text{Treated}\}$	-0.010 (0.010)	-0.009 (0.014)	0.000 (0.012)	-0.016 (0.012)	-0.005 (0.010)	-0.016* (0.010)
$1\{t=-4\} \times 1\{\text{Treated}\}$	-0.010 (0.010)	-0.003 (0.013)	-0.016 (0.011)	-0.013 (0.011)	0.001 (0.009)	-0.010 (0.009)
$1\{t=-3\} \times 1\{\text{Treated}\}$	-0.003 (0.009)	0.012 (0.013)	-0.003 (0.011)	-0.015 (0.011)	0.012 (0.009)	-0.015 (0.009)
$1\{t=-2\} \times 1\{\text{Treated}\}$	0.001 (0.009)	0.002 (0.012)	0.000 (0.011)	-0.008 (0.011)	0.010 (0.009)	-0.014 (0.009)
$1\{t=-1\} \times 1\{\text{Treated}\}$	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
$1\{t=0\} \times 1\{\text{Treated}\}$	-0.009 (0.009)	-0.030*** (0.012)	-0.024** (0.010)	-0.051*** (0.010)	-0.008 (0.009)	-0.025*** (0.008)
University FE	Y	Y	Y	Y	Y	Y
Outcome Mean	0.71	0.71	0.57	0.57	0.20	0.20
Observations	104,190	103,673	104,660	108,725	104,660	108,725

NOTE.— This table presents estimates from Equation 5 comparing university students from schools in England that eliminated AS-levels in 2017 to students from Wales/NI. Columns 1, 3, and 5 present estimates for schools in England with below-median overprediction for 2017-reformed subjects between 2010-2011, while columns 2, 4, and 6 present 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 E3: Event-study estimates of university graduation and dropout outcomes

	Graduated		Graduated with First		Graduated with 2:1 or First		Dropped Out	
$1\{t = -5\} \times 1\{\text{Treated}\}$	-0.003 (0.005)	-0.002 (0.005)	-0.003 (0.005)	-0.001 (0.005)	0.005 (0.005)	0.007 (0.005)	0.000 (0.003)	-0.001 (0.003)
$1\{t = -4\} \times 1\{\text{Treated}\}$	-0.006 (0.004)	-0.005 (0.004)	-0.002 (0.004)	-0.001 (0.004)	-0.009* (0.005)	-0.007 (0.005)	-0.005 (0.004)	-0.005 (0.004)
$1\{t = -3\} \times 1\{\text{Treated}\}$	0.003 (0.004)	0.004 (0.004)	0.002 (0.004)	0.004 (0.004)	0.001 (0.005)	0.002 (0.005)	-0.008** (0.004)	-0.008** (0.004)
$1\{t = -2\} \times 1\{\text{Treated}\}$	0.001 (0.004)	0.002 (0.004)	0.002 (0.004)	0.002 (0.004)	-0.002 (0.005)	-0.001 (0.005)	-0.005* (0.003)	-0.005** (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)	0.000 (0.000)	0.000 (0.000)
$1\{t = 0\} \times 1\{\text{Treated}\}$	-0.016*** (0.004)	-0.015*** (0.004)	-0.008** (0.004)	-0.008** (0.004)	-0.019*** (0.004)	-0.018*** (0.004)	0.010*** (0.003)	0.009*** (0.003)
$1\{t = 1\} \times 1\{\text{Treated}\}$	—	—	—	—	—	—	0.017*** (0.004)	0.017*** (0.004)
$1\{t = 2\} \times 1\{\text{Treated}\}$	—	—	—	—	—	—	0.027*** (0.007)	0.026*** (0.007)
University FE	N	Y	N	Y	N	Y	N	Y
Observations	485,875	485,875	485,875	485,875	485,875	485,875	652,275	652,275

NOTE.— This table shows OLS coefficients from event study regressions in Equation 5 on observed degree outcomes at age 21. Columns 7 and 8 report estimates from the staggered event-study specified in Equation 1 for dropouts. Columns 1 through 6 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 and schools in Northern Ireland and Wales. Sample: university students who started between 2012-2017 (columns 1-6) and university students who started between 2012-2019 (columns 7-8).

Table E4: Event-study estimates of graduation and early-career outcomes at age 21

	Graduated University		Employed		Earnings (GBP)		Out-of-work Benefits		Non-degree Firm	
	Low-inc	High-inc	Low-inc	High-inc	Low-inc	High-inc	Low-inc	High-inc	Low-inc	High-inc
1{t= -5} × 1{Treated}	-0.003 (0.006)	-0.003 (0.005)	0.005 (0.005)	-0.000 (0.004)	0.003 (0.014)	0.011 (0.012)	0.001 (0.005)	-0.003 (0.003)	-0.006 (0.004)	0.002 (0.003)
1{t= -4} × 1{Treated}	0.007 (0.005)	-0.005 (0.005)	0.000 (0.005)	0.002 (0.004)	0.012 (0.013)	-0.001 (0.011)	-0.003 (0.004)	0.004* (0.002)	-0.001 (0.004)	0.003 (0.003)
1{t= -3} × 1{Treated}	0.007 (0.005)	-0.007* (0.005)	0.007 (0.005)	-0.000 (0.004)	0.025** (0.013)	0.006 (0.011)	0.002 (0.004)	0.002 (0.002)	-0.001 (0.004)	0.002 (0.003)
1{t= -2} × 1{Treated}	0.005 (0.005)	-0.006 (0.004)	-0.000 (0.005)	-0.001 (0.004)	0.001 (0.012)	0.009 (0.011)	0.006 (0.004)	0.004* (0.002)	-0.004 (0.004)	0.002 (0.003)
1{t= -1} × 1{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)	0.000 (0.000)	0.000 (0.000)
1{t= 0} × 1{Treated}	0.010** (0.005)	-0.004 (0.004)	0.012*** (0.005)	-0.006* (0.004)	-0.006 (0.012)	0.003 (0.011)	-0.023*** (0.004)	-0.012*** (0.002)	-0.012*** (0.004)	-0.002 (0.003)
Spec	OLS	OLS	OLS	OLS	Poisson	Poisson	OLS	OLS	OLS	OLS
Observations	525,615	716,515	525,615	716,515	525,680	716,900	525,615	716,515	353,220	494,035

NOTE.— This table shows OLS coefficients from event study regressions in Equation 5 on observed graduation and post-graduate outcomes at age 21, separately by low-income and high-income students. Columns 5 and 6 report coefficients from the Poisson regression specified in Equation 6 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 2018 GBP and winsorized at 1 percent above and below.