Reshuffling University Access: Analyzing Grade Distributions in the Absence of Standardized Testing. *

Keitaro Ninomiya

Most recent version: `Click here`

Abstract

How do schools vary in assigning grades to students for university admissions relative to a standardized exam regime where teachers have no grading discretion? I exploit a natural experiment in the United Kingdom during the COVID-19 pandemic, when teacher-assigned grades replaced in-person standardized A-level exams. I provide evidence of differing grading policies across schools, shaped by school type and institutional quality. These policies followed persistent patterns across consecutive years, though schools with stricter standards in the first year revised them downward in the second. Within schools, grades were assigned differently by students' gender, race, and economic affluence, with patterns varying across school quality tiers. Despite the regressive tendencies between student background and grade improvements, the impact of grade inflation on application success into selective universities was limited for students from disadvantaged backgrounds. Overall, the findings highlight how non-standardized assessments shifted grade distributions but had limited effects on diversifying selective universities.

*I am indebted to Alex Bartik, Dan Bernhardt, Andy Garin, and Russell Weinstein for their guidance and support throughout the advancement of this project. This paper has benefited greatly from feedback and discussions with James Banks, Giovanni Facchini, Rachael Griffith, Katsushi Imai, Victor Lavy, Jack Mountjoy, Jesse Matheson, Robin Naylor, Elizabeth Powers, Joshua Shea, Jeremy Smith, as well as seminar and conference participants at ¬University of Illinois at Urbana-Champaign, University of Nottingham, University of Manchester, University of Bath, the University of Warwick CAGE - AMES seminar series, and RES 2025 Annual Conference. This work was undertaken in the Office for National Statistics Secure Research Service using data from ONS and other owners and does not imply the endorsement of the ONS or other data owners. All errors in this analysis are my own.

"After the government's announcement that 2021's A-levels and GCSEs [exams] would be cancelled, senior management at NLCS were openly excited at the possibility of obtaining 'our best grades yet' and the allure of this idea propelled them to make decisions that had little integrity, even though they appeared to be within the rules."

Anonymous tip from North London Collegiate School

1 Introduction

Policymakers' decisions over grading regimes shape how students sort across universities, carrying broader implications for the entire higher-education market. Each grading regime distinguishes itself by heterogeneity in grading standards, which in turn shapes how different groups of students gain access to universities of varying selectivity. Previous studies discuss how changes in students' final grades trigger responses from both students and universities: students may adjust their application behavior (Hoxby and Avery (2012), Dynarski et al. (2022)), while universities may revise their selection criteria (Bleemer (2023), Arcidiacono et al. (2023)). Furthermore, shifts in student allocation across universities spill over into adjacent markets, including secondary education (Cullen et al. (2013)) and the labor market (Chetty et al. (2020), Bleemer (2021), Black et al. (2023)). However, we know little about the behavioral responses of schools through their grading policies when governments transfer admission-deciding authority from central exam boards to schools.

I study how schools differ in their grading policies on qualifications used in university admissions by exploiting a natural experiment in the UK where teacher-assigned grades solely determined admissions outcomes. During the COVID-19 pandemic, the British government canceled the upcoming centrally graded standardized exams (e.g., A-levels) and relied on schools to provide replacement grades. With minimal guidance on grading standards or assignment procedures, schools exercised broad discretion in determining students' grades. The

TAG policy persisted for two consecutive years, but adjustments in applications were limited because the announcement came after students had completed their A-level coursework and submitted most university applications. Despite the government's effort to maintain stability in the admissions system, the grade distribution shifted sharply upward: the share of top grades rose by 12 percentage points in 2020 and by 20 percentage points in 2021 relative to 2019. This unintended change substantially and unevenly lowered admission barriers across the student population.

While the TAG policy transferred grading authority from central exam boards to schools, universities had limited control over admission decisions. In the British admissions system, students and universities agree in advance—prior to A-level exams—on the grade conditions required to secure a place.[1] Following the upward shift in the overall grade distribution, many applicants met their grade conditions, leading to an unprecedented inflow of students into universities (Figure A2). Admission officers had little ability to manage this surge, as pre-determined offer contracts required universities to accept all students who met their grade conditions.

Using administrative data on the universe of A-level entries from nearly 4,000 schools in the UK, I estimate the grading policies adopted by each school during the pandemic, and examine their effects on placement outcomes and the resulting compositional shift of students at universities. I identify grading policies by measuring the gap in final grades between pandemic cohorts and their immediate pre-pandemic counterparts. However, the difference in grades could reflect pre-existing trends rather than changes in grading methods. The historical *Comparable Standards* approach used by A-level exam boards precludes this concern, as A-level examiners set grade boundaries to preserve the historical relationship between students' prior attainment on the national GCSE exams—taken at age 16—and their subsequent A-level performance. This calibration keeps the relationship between GCSE performance and A-level outcomes stable over time, so deviations observed during the pandemic

[1]Students' A-level outcomes, combined with their pre-submitted ranking of university preferences, mechanically determine the final match between applicants and programs.

reflect changes caused by TAG.

I provide three sets of new results. First, I find substantial heterogeneity in how schools loosened their grading policies, with higher degrees of inflation concentrated in privately funded schools and schools with historically lower average performance on standardized A-level exams. I find a nearly 13-percentage-point gap in the assignment of top grades between schools in the bottom 10th percentile of quality and those in the top 10th percentile. However, this gap hides the real intensity of grade inflation across schools, since students at high-performing schools already had higher baseline chances of receiving top grades. After accounting for the limited headroom in raising grades across schools, inflation proved larger among high-performing schools than elsewhere. Additionally, I find that privately funded institutions (e.g. Independent schools) exhibited greater grade inflation across all levels of school quality by significantly inflating grades in the second year, suggesting that private institutions rapidly learned about and took advantage of the new environment. Most schools maintained similar levels of grading policies across the two consecutive years, but I observe a downward revision of grading standards in 2021 among schools that had stricter grading policies in the first year.

Secondly, I find that schools' grading policies incorporated students' academic preparedness, demographic characteristics, and household affluence, with the magnitude of these effects differing across subject areas. I find that cross-schools differences in grading policies were largely explained by grading patterns in non-quantitative subject areas, where the subject content naturally relied on graders' discretion for assessment. Turning to grading patterns within a course in 2020, I find that academically well-prepared students received higher grades, as did whites and females students, and those with parents in higher-paying occupations.[2] Overall, the expansion of grades widened attainment gaps within classrooms, but the extent of this widening depended on students' demographic and economic advan-

[2]I also find similar results on student affluence and demographics for 2021, however such results could be influenced by the learning loss due to school closures and remote learning during COVID.

tage—factors that teachers, consciously or not, incorporated into their grading.

Thirdly, I show that grading policies increased students' likelihood of admission and shifted the composition of university cohorts away from their pre-pandemic mix. The variation in grading leniency across schools generated uneven flows of students from schools into universities: students from more leniently graded schools were significantly more likely to secure places at their top-choice universities than those from stricter schools. In particular, grading policies at independent schools disproportionately improved the placement rate of their students compared with those of academically similar students at publicly funded schools. Turning to university composition, looser grading policies reduced the average academic preparedness, parental income level, and other demographic indicators of incoming cohorts. A 10 percent increase in cohort size lowered the average GCSE Mathematics and English Literature scores of entrants by 0.5 and 0.4 standard deviations, respectively, and decreased the average parental income score by 0.18 standard deviations. These effects were strongest at selective universities, where larger and more heterogeneous applicant pools amplified the compositional shifts among admitted students.

My results contribute to the broader literature on how admission policies shape the composition of students across universities. Previous studies have examined how external interventions can expand access for students from underrepresented backgrounds to more selective institutions (Dynarski (2003),Arcidiacono (2005),Epple et al. (2006), Bettinger et al. (2012), Hoxby and Avery (2012), Bleemer (2021), Dynarski et al. (2022),Chetty et al. (2023), Dessein et al. (2025)). The COVID-19 pandemic created an unintended opportunity to examine how the removal of standardized testing changed university composition. My results provide the first documentation of this episode. I find that, at least in the UK, the policy narrowed the qualification gap between students from contrasting backgrounds, but had limited effects on the upward mobility of disadvantaged students into more selective universities. The grading regime change mainly benefited students with lower baseline probabilities of admission who nonetheless applied to selective institutions and were accepted. Consequently, the average academic preparedness of entrants declined at selective institutions, while diversity increased along other dimensions such as socioeconomic status and parental income.

My results also contribute to the literature on grade inflation in teacher-assessed (non-standardized) evaluations. Many educational institutions show a persistent upward drift in students' average grades, which weakens employers' and admissions officers' ability to distinguish candidates by academic competence (Bar et al. (2009), Bleemer (2020), Denning et al. (2023), Denning et al. (2022)). I document a deterioration in aggregate grading standards over the two consecutive pandemic years, consistent with theories suggesting that graders strategically adjust to others' grading decisions, thereby depreciating the informational value of grades as a signal (Chan et al. (2007), Rojstaczer and Healy (2012), Bleemer (2020)). I extend this literature by showing that inflation in the first year shaped inflation in the next: schools that graded more strictly initially tended to relax their standards, while those with looser grading maintained their approach.

Finally, my results contribute to the large empirical literature on grading biases. Previous literature has documented grading bias across demographic groups, including gender (Lavy (2009), Carlana (2019)), ethnicity (Burgess and Greaves (2013), Botelho et al. (2015), Alesina et al. (2024)), economic affluence (Hanna and Linden (2012)), and classroom behavior (Ferman and Fontes (2022), Diamond and Persson (2016)). My results support most of the findings, as schools' grading policies favored female, white, and economically affluent students more than others. My paper provides empirical evidence that grading biases across student groups vary across schools according to their institutional features. Earlier work on school governance and misreporting by Figlio and Lucas (2004) and Neal and Schanzenbach (2008) shows that private institutions and schools near accountability thresholds tend to misreport student grades. My paper highlights how institutional features of schools shape grading biases across different student groups, deepening our understanding of how such biases arise.

The paper is organized as follows: Section 2 provides the institutional background of the higher education system in the UK as well as the details of the policy set in place during the pandemic. Section 3 describes the data source I use in the analysis. Section 4 presents the empirical model and discuss the validity of the estimates along with the institutional details. Section 5 presents the results. Section 6 discuss the results with the literature. Section 7 concludes with policy discussions.

2 UK's higher education system and the TAG policy

2.1 A-Levels

A-levels are the terminal, standardized qualifications taken at the end of upper-secondary education (age 16–18) and form the principal basis for university admission (Figure 1). Applicants may apply to up to five courses via UCAS, typically by a January deadline; because the examinations are held after applications close, applications are supported by teacher-predicted grades. Universities then issue conditional offers (usually by May) that specify the grades required for admission. Applicants nominate a firm and an insurance choice; these choices establish the binding order in which offers are accepted.

Final placement is determined by the centrally awarded A-level grades. A-level scripts are marked by examination boards: markers are blind to candidates' identities and to application information, and letter grades are assigned according to standardised grade boundaries. Once actual results are released in August, UCAS assigns students according to the ranking they submitted and the conditions of the conditional offers (the firm choice is accepted if its conditions are met; otherwise the insurance choice may be used; failing both, the student can enter Clearing). Because grading and placement follow this centralised, rule-based procedure, teachers and schools have no formal discretion over the final marks or over the contractual placement outcome.

A-level scripts are marked centrally by examination boards; markers are blind to candidates' identities and demographic attributes. Letter grades are awarded according to grade boundaries that examiners set annually. Those boundaries are calibrated with reference to historical performance and other standardisation procedures so that the difficulty of achieving a given grade is broadly comparable across years. We confirm this property empirically by predicting A-level attainment from pre-treatment observables (here, GCSE mathematics and English) and comparing the conditional expectation of A-level outcomes across non-COVID years (Figure 2): visual inspection of binned conditional expectation functions, pooled regressions with year × predicted-score interactions, and formal equivalence tests within narrow predicted-score bins reveal no meaningful shifts in the conditional means. Complementary checks — subject-by-exam-board splits — produce the same conclusion, supporting the claim that, conditional on prior attainment, A-level outcomes are stable across non-COVID cohorts.

Subject choice and course provision vary across the system. There are well over 100 distinct A-level subjects, and schools and sixth-form colleges decide which subjects to offer according to local demand and resources. Some pupils remain at their secondary school for sixth-form study, while others transfer to separate sixth-form or further-education colleges; this spatial and institutional dispersion affects which subjects pupils can take. Universities often specify required subjects for particular courses, and highly selective institutions tend to favour so-called "facilitating" subjects; independent schools frequently encourage uptake of these subjects and invest in resources (e.g. specialised equipment and teaching) to support them.

2.2 University application system in the UK

The UK uses a centralised admissions system (UCAS) to match applicants to university courses. Applicants may apply to up to five courses through UCAS; universities reply—typically before students sit their A-levels—with acceptances, rejections, or conditional offers that specify the grades required for admission. After receiving offers, applicants must choose two to keep: a *firm* choice and an *insurance* choice, and withdraw any remaining offers. These ranked choices determine allocation once actual A-level results are released: if the firm offer's conditions are met the firm place is confirmed, otherwise the insurance offer may be used; failing both, the applicant may enter Clearing to seek an alternative place.

2.3 The British Education Sector for Higher Education

The UK has on average 3,300 university programs per year, matching with on average 570,000 students per a year. 3,500 secondary schools provide students with education to prepare for exams.

British secondary schools are mostly run publicly. 93% of students participate in public schools and only 7% in private schools ((Table A1)). Private schools, equally known as independent schools, run mainly on tuition collected from their pupils' parents. Past literature documents that attendants of independent schools are often children with rich parents. Henseke et al (2021) finds that 92 percent of the students attending independent schools belong to households in the top 10th decile group of the income distribution. 90% of the students in independent schools proceed to higher education after graduation, with 4% attending Oxford of Cambridge and 52% attending one of the top 25th ranked universities. Students from independent schools apply to better universities more than students from public schools. Almost 85% of the students apply to at least one university of the Russell group, while 45% of the students from public schools apply. Independent school students apply less to lower-tier schools than students from public schools. Almost 40% of the students apply to lower-tier schools, while 80% of the students from public schools apply to at least one lower tier school. Independent school students receive more offer of admission from Russell Group schools than public schools. 80% of students from independent schools receive at least one offer from a Russell Group school.

There are multiple types of public school in the UK higher education sector (Table A1). The most common type of schools are academy schools that consists of 35% of the total number of schools. Academy schools provide education to pupils living within their local authority jurisdiction. As an exception, grammar schools select students based on their performance on a previous national exam (*year 11*).

2.4 COVID-19 pandemic

The COVID-19 pandemic led the UK authorities to cancel in-person A-level examinations and substitute centrally-marked exams with teacher-based assessments. For the 2020 cohort this change was announced on 18 March 2020, and for the 2021 cohort on 6 January 2021. In place of externally marked scripts, teachers were asked to submit centre-assessed grades for their pupils, drawing on their knowledge of each student and the student's prior classroom performance and exam history. Exam boards instructed centres to review and, where necessary, lower implausibly high submissions before finalising their returns; nonetheless, the grades submitted by schools frequently stood as the final awards and were used in place of the usual centrally-marked A-level marks.

Crucially, both announcements came after applications had been submitted and universities had issued responses (acceptances, rejections, conditional offers or unconditional offers). The teacher-based grades therefore stood in for the usual, centrally-awarded A-level marks used to satisfy conditional offers and determine final placements. At that point universities had little practical means to control the incoming cohort — their main lever was to tighten conditional offers (raise grade requirements) but many institutions were reluctant to do so for fear of deterring applicants during the pandemic. Importantly, students had already nominated a firm and an insurance choice before teachers assigned centre-assessed grades, so final placements followed the pre-submitted ranking together with the teacher-submitted grades rather than any subsequent adjustment by schools or universities.

2.5 Consequence of the method change

The immediate consequence of the method change was substantial grade inflation (Figure 1). The share of students awarded A* or A rose by roughly 12 percentage points in 2020 and by about 20 percentage points in 2021 relative to pre-pandemic cohorts. Importantly, the shift was not confined to the top tail: the entire grade distribution moved (see Figure 1), with noticeable changes across middle and lower bands (C–E) as well as at A/A*. In short, teacher-based assessment redistributed mass across multiple grade bands rather than merely increasing top grades.

Those inflated grades translated directly into larger incoming cohorts because universities were obliged to honour conditional offers for students who met the stated requirements. Figures A2 and A2a document this over-acceptance in both 2020 and 2021. The expansion was concentrated among more selective programmes: the most competitive programmes—those with low pre-pandemic acceptance rates—experienced the largest proportional increases in cohort size, with some programmes almost doubling intake despite their typically small pre-pandemic capacities.

Figure A2a shows that a substantial part of the expansion came from students being placed at their firm choice. In particular, programmes that lay below the 30th percentile of pre-pandemic acceptance rates increased their intake by more than twice the pre-pandemic average, indicating that the grade shifts interacted with the existing selectivity hierarchy to produce outsized growth at the top end of the market.

3 Data

I use administrative data on university application in the UK (UCAS) to recover the grade distribution of students with detailed demographic information of the students and basic attributes of schools.

The UCAS data includes the A-level grade that applicants submitted to the university, as well as the student background information. As UCAS is the single centralized clearing house of university applications in the UK, UCAS handles the content of the application, including the A-level grades as well the other qualification that the applicant submits including their GCSE scores, their AS level scores, and third party qualifications as the International Baccalaureate. In addition to the qualifications, applicants also submit their background demographics information such as their gender, race, socio economic status, and the income deprivation index scores of their residential area. Students also submit their school of attendance in addition to the governance type of the school. My dataset covers 2007 to 2021.

The UCAS dataset covers a wide range of qualifications, including the grades that universities use for filtering applicants into acceptance, as well as the record of the applicants past qualifications. Many universities in the UK request basic academic certifications from their applicants despite not using the results of the past qualification for deciding the reply to the applicants.

Despite some universities and students match outside of the UCAS system, a majority of students use the UCAS system to apply to universities. The UCAS system also manages the match between foreign students and British universities. As many British universities accept the use of internationally recognized standardized exam as certificates for acceptance, the UCAS data also includes applications from foreign applicants as well as their demographics and their grades in the foreign exams.

Lastly, the UCAS data also includes application related information, namely their firm/insurance choice, as well as their acceptance.

Although the UCAS data covers the entire university applicant that used UCAS to apply to university, their coverage is not comprehensive of the entire A-levels exam outcome as a subset of students do not apply to university, therefore the students grade is missing from the UCAS data.

4 Empirical Model

4.1 Empirical Specification and Estimation

I estimate the changes in grade assignment patterns between the pandemic and non-pandemic years through a binary choice model. Namely, I estimate

$$Y_{i,j,k,t} \;=\; \mathbf{1}\{Y^*_{i,j,k,t} \geq 0\}.$$

$$Y^*_{i,j,k,t} \;=\; \alpha_j \;+\; X_i\eta \;+\; \gamma_k \;+\; \sum_{\tau\in\{2020,2021\}} D_\tau \times (\Delta\alpha_{j,\tau} \;+\; X_i\Delta\eta_\tau \;+\; \Delta\gamma_{k,\tau})\,, \qquad (1)$$

where i denotes a student, j denotes the school, k denotes the exam subject, and t denotes a year. The outcome variable $Y_{i,j,k,t}$ is a binary variable for student i's exam outcome at year t measuring if he/she received a top grade (A or A*). α_j is the school fixed effects. X_i is a vector of student level controls (Previous scores in a national standardized exam (GCSE), gender, race, and socio-economic group). D_τ is a dummy variable for the treatment indicator, which distinguishes between both 2020 and 2021. $\Delta\alpha_{j,\tau}, \Delta\eta_\tau, \Delta\gamma_{k,\tau}$ measures the coefficient on the interaction term between the school fixed effect and the treatment year indicator.

Equation (1) identifies the causal effect of the grading method across various student backgrounds by measuring the difference in the probability in receiving a top grade between academically and demographically similar students that both attended the same school but varied only in their year in taking the A-levels. For regression of Equation (1) to identify the causal effect, one must assure that no unobserved factors are influencing the difference in the final grade between the pair of students. A primary concern is differences in underlying trends in the academic performance of students a school, which would confound the treatment effect with the unobserved trend ($E[\ell_{i,j,k,t}|X_{t<2020}] = E[\ell_{i,j,k,t'}|X_{t'<2020}]$)).

I control for the unobserved trends across schools and various student backgrounds in their performance in the A-level before the pandemic by leveraging the deterministic grading system of the A-level in the pre-pandemic periods. To avoid fluctuating grading standards across cohorts, the grade boundaries for the A-levels are centrally determined by a statistical relationship between the test-takers grades in their national standardized test results that all UK students take when entering secondary school (*the GCSE*). Namely, the central graders set the grade boundaries so that the expected probability of a student to receive a certain grades is preserved to be constant in any years.

To establish the stationarity of the overall grade distribution of the A-levels before the pandemic years, I conduct a series of stationarity tests. I first estimate the conditional mean probability of a student receiving an top grade (A or A*) by estimating the following equation.

$$Y_{i,t} = \beta_0 + \sum_{\substack{\tau=t_{\min} \\ \tau \neq t_0}}^{t_{\max}} \beta_\tau \mathbf{1}\{t = \tau\} + X_i' \gamma + \varepsilon_{i,t},$$

β denotes the conditional average probability of a student to achieve an A or A* in their A-level exams.

Figure 2 plots the conditional average probabilities from 2015 to 2021. The difference in the conditional average probabilities is statistically insignificant. The difference between mean probability of obtaining top grades (A/A*) between 2015 to 2018 are statistically insignificant from 2019, but increases to 0.5 in the logit model coefficient in in 2020 and 7.0 in 2021[3]. I show how the average probability in receiving top grade is constant up until 2019 for each school types that composes the overall estimate (Figure 3) are all similar to the probability to the predicted levels in 2019.

Another potential threat to identification is how the student difference in student composition between the grading method change and before is not affected by the announcement of the policy. Namely, students changing school of attendance just for the pandemic years would confound the treatment effects with difference in compositions of students within a school across years ($E[\ell_{i,j,k,t}|X_{t<2020}] = E[\ell_{i,j,k,t'}|X_{t'<2020}]$). Any sorting of students across

[3]The slight increase in the probability of receiving top grades in 2019 is driven by the rescalling of the GCSE scores that was conducted in 2017.

schools due to the pandemic is strongly implausible due to the unpredictability of the pandemic, as the students had already made their decisions on which schools to attend to prepare for their A-levels. Table A1 tests for the difference across cohort characteristics between a pandemic year and a non-pandemic year and finds no significant difference between cohort characteristics for both years.

Although Equation (1) is robust to underlying pre-trends and unobserved differences in students characteristics across the treatment dummy, the final concern is that teachers may take into account how students suffered in their learning experiences during the pandemic into their grading. Past literature (Diamond and Persson (2016) , Ferman and Fontes (2022)) documents how teachers may take into account the exam taking conditions of the students into the students final grades. The learning disruption is less of a concern for the 2020 samples in my setup as these students had already finished most of their preparation for the exams, and teachers had immense experience with teaching the students before the pandemic occurred. To establish that the estimated changes in grades for 2020 is largely due to grading method changes, I conduct a series of place-bo tests on the results of standardized foreign exams conducted during the pandemic years. Many UK higher education institutions allow students to use standardized foreign exams outside the A-level qualifications to allow foreign students to study at their universities. **??** reports the difference in the average grades of the non-A-level standardized tests for 2020 against 2019, and it confirms that the learning loss for the 2020 year cohort was minimal.

However, I can not strongly reject the possibility that the learning loss's for the 2021 cohorts may confound the effect of grading change with disruptions in the learning environments due to the shutdown during the pandemic year. Especially, the cross school differences in grading policies may reflect the differences in learning environment disruptions, as schools may differed in their ability to compensate for the lack of in-person classes by providing online courses. Despite the caveats for the school level difference in grading policies, I claim that the student level estimates are less prone to differences in learning losses across demographic groups. Equation (1) estimates the differences across demographic groups by comparing across students sharing the same subject course.

Estimation I estimate the parameters by searching for the set of estimates that maximizes the joint likelihood function of the observed dependent variable (e.g. students achieving A or A*). The joint likelihood function is expressed as the following;

$$\mathcal{L}(\theta) = \prod_{i,j,k,t} \left[F(z_{i,j,k,t})^{Y_{i,j,k,t}} \; [1 - F(z_{i,j,k,t})]^{1-Y_{i,j,k,t}} \right],$$

where,

$$z_{i,j,k,t} \; = \; \alpha_j + X_i\eta + \gamma_k + \sum_{\tau \in \{2020,2021\}} D_\tau \left(\Delta\alpha_{j,\tau} + X_i \Delta\eta_\tau + \Delta\gamma_{k,\tau} \right),$$

and F denotes the cumulative density function of the logistic function. I estimate the parameters that maximize the logged function;

$$\ell(\theta) = \sum_{i,j,k,t} Y_{i,j,k,t} \log F(z_{i,j,k,t}) + (1 - Y_{i,j,k,t}) \log \left[1 - F(z_{i,j,k,t}) \right].$$

I take an empirical Bayes approach (Walters (2024), **?**) to correct for the estimation noise related to the school fixed effects $\alpha_j, \Delta\alpha_j$. I assume that the school effects are drawn from a common distribution with separate variance parameters.

$$\alpha_j \sim \mathcal{N}(0, \sigma^2_\alpha), \;\; \Delta\alpha_{j,\tau} \sim \mathcal{N}(0, \sigma^2_{\Delta\alpha,\tau})$$

$\sigma^2_\alpha, \sigma^2_{\Delta\alpha,\tau}$) measures the variance of the common distribution. Following **?** and **?** I correct the school level estimators with the following formula;

$$\widehat{\alpha}_j^{EB} = \frac{\widehat{\sigma}^2_\alpha}{\widehat{\sigma}^2_\alpha + \widehat{\text{SE}}(\widehat{\alpha}_j)^2} \, \widehat{\alpha}_j, \quad \widehat{\Delta\alpha}_{j,\tau}^{EB} = \frac{\widehat{\sigma}^2_{\Delta\alpha,\tau}}{\widehat{\sigma}^2_{\Delta\alpha,\tau} + \widehat{\text{SE}}(\widehat{\Delta\alpha}_{j,\tau})^2} \, \widehat{\Delta\alpha}_{j,\tau},$$

$\widehat{\sigma}^2_{\Delta\alpha,\tau}$ denotes the estimate of the variance of the common distribution, which I calculate as the following.

$$\widehat{\sigma}^2_{\Delta\alpha,\tau} = \max\left\{0,\ \mathrm{Var}_j\left(\widehat{\Delta\alpha}_{j,\tau}\right) - \overline{\mathrm{SE}\left(\widehat{\Delta\alpha}_{j,\tau}\right)^2}\right\}.$$

$\widehat{\mathrm{SE}}(\widehat{\Delta\alpha}_{j,\tau})^2$ measures the estimated sampling variance of the school fixed effects[4].

School Specific grading pattern of students Equation (1) estimates the grading pattens across demographic groups by estimating the average difference in students tendency to receiving a top grade over schools. I extend the model into Equation (2) so that I estimate grading patterns across demographic groups seperatly for each schools. Namely I augment the interaction term $X_i\Delta\eta_\tau$ so that the treatment response to observed demographics is school-specific. Let $\mathcal{G}$ index a set of demographic groups (e.g. gender, race, SES bins) and denote by $d_{i,g}$ the indicator (or normalized score) for student i belonging to group $g \in \mathcal{G}$. I replace the common demographic treatment vector $\Delta\eta_\tau$ with school-by-group parameters $\{\Delta\eta_{g,j,\tau} : g \in \mathcal{G},\ j \in \mathcal{J}\}$ and write the latent index as

$$z_{i,j,k,t} = \alpha_j + X_i\eta + \gamma_k + \sum_{\tau\in\{2020,2021\}} D_\tau\Big(\Delta\alpha_{j,\tau}\ +\ \sum_{g\in\mathcal{G}} d_{i,g}\,\Delta\eta_{g,j,\tau}\ +\ \Delta\gamma_{k,\tau}\Big), \tag{2}$$

so that the treatment effect on students in demographic group g at school j in year τ is captured by $\Delta\eta_{g,j,\tau}$. The binary outcome and likelihood remain as in the main specification with F the logistic CDF.

Because the number of school–group parameters can be large and some school–group cells have few observations, I impose a hierarchical (partial-pooling) prior on the school–specific demographic effects and estimate this empirically (empirical Bayes). Concretely,

$$\Delta\eta_{g,j,\tau}\ \sim\ \mathcal{N}\big(\mu_{g,\tau},\ \sigma^2_{g,\tau}\big), \qquad \mu_{g,\tau} \in \mathbb{R},\ \sigma^2_{g,\tau} \geq 0,$$

[4]I solve for the estimates that maximizes the likelihood function with a quasi-Newton approach. As the sample size and the number of parameters is large, a standard Newton-Raphson solver does not scale well, as the calculation of the Hessian matrix takes time. Instead the L-BFGS method, which is a quasi-Newton approach, approximates the Hessian using past information that reduces the computation time significantly (?). I derive the standard errors of the estimates by clustering at the school year level.

so that estimates for small or noisy cells are shrunk toward the group-level mean $\mu_{g,\tau}$. The empirical Bayes shrinkage estimator has the closed form (posterior mean under Gaussian approximations)

$$\widehat{\Delta\eta}^{EB}_{g,j,\tau} = w_{g,j,\tau}\,\widehat{\Delta\eta}_{g,j,\tau} \;+\; \big(1 - w_{g,j,\tau}\big)\,\widehat{\mu}_{g,\tau}, \qquad w_{g,j,\tau} = \frac{\widehat{\sigma}^2_{g,\tau}}{\widehat{\sigma}^2_{g,\tau} + \widehat{\mathrm{SE}}\big(\widehat{\Delta\eta}_{g,j,\tau}\big)^2},$$

where $\widehat{\Delta\eta}_{g,j,\tau}$ and $\widehat{\mathrm{SE}}(\widehat{\Delta\eta}_{g,j,\tau})$ are the raw (maximum likelihood) estimates and their sampling standard errors, and $\widehat{\mu}_{g,\tau}, \widehat{\sigma}^2_{g,\tau}$ are estimated from the cross-school empirical distribution. As before, I estimate the variance component by the usual nonnegative deconvolution

$$\widehat{\sigma}^2_{g,\tau} = \max\Big\{0,\ \mathrm{Var}_j\big(\widehat{\Delta\eta}_{g,j,\tau}\big) - \overline{\mathrm{SE}\big(\widehat{\Delta\eta}_{g,j,\tau}\big)^2}\Big\}.$$

4.2 Discussion on the empirical model

My model is similar to the approach used by Jacob and Levitt (2003) Neal and Schanzenbach (2008) and Dee et al. (2019) in which researcher leverage the sudden change in grading policies and use the discrepancy between the predicted grades and the actual grades to measure the grading bias[5]. The magnitudes in the grade discrepancies are averaged over students with different demographics groups and the differences in their magnitudes are used to assess the demographic group that the teacher favors more than others.

I extend the pre-existing framework for detecting grading biases to schools, which has been left undone in the literature. Schools often do not share common standardized tests that allows researchers to compare the grading bias across schools. As the number of schools in the UK is nearly 4000, I use the methods from the large scale inference literature (Efron (2010)Walters (2024)) to obtain precise estimates about the grading bias of schools. The high-dimension statistics literature documents how maximum-likelihood estimators in such setting often fails to precisely estimate each parameters due to the large number of statistical

[5]Other papers uses grading systems in which teachers provide grades on the students along side a externally graded exams (Lavy (2009)Lavy and Sand (2018), Ferman and Fontes (2022)).

test a single regression conducts, and suggests improving the statistical power of individual tests by combining the multiple tests to improve the statistical power of each individual test[6]. I use the shrinkage models suggested in the high-dimensional statistics literature to improve the precision of each school level grading leniency estimate[7].

I estimate the grade assignment process with a binary choice model rather than using a linear probability model (Dee et al. (2019)) or an OLS that standardizes the dependent variable (Lavy (2009)), as my sample varies substantially in their underlying academic abilities. Namely a linear probability model fails to identify the marginal effect of a parameter when the baseline probability for the sample is near 0 or 1 (Wooldridge (2001)). Students with a high GCSE score of 9 or students attending elite schools are likely to achieve a high grade, and a linear probability would fail to identify the effect of the grading method change on such populations. A non-linear probability model consistently estimates the effect of the grading policy on the students grades by taking into account the underlying difference in the success rate for a student to achieve a high grade in the standardized exams. To avoiding confounding the estimates from the estimation noise from other parameters, I remove schools with less than 30 students in either 2020 and 2021.

5 Results

5.1 Inflation effect by schools

Figure A3 shows that the degree of grade inflation was higher in lower-quality schools. The top and bottom panels respectively report the extent of grade inflation for every school in the UK in 2020 and 2021, by school quality; school quality is measured by the probability that a nationally median student at the school achieves a top grade. In both 2020 and 2021, grade inflation was greater in lower-quality schools. A one-unit increase in school quality

[6] Applications include **?**, Chetty et al. (2020), and surveyed in Walters (2024)

[7]

— measured by the nationally median student's probability of receiving a top grade — was associated with 0.45 and 0.47 percentage-point smaller degrees of grade inflation in 2020 and 2021, respectively (Table A5).

The negative relationship between school quality and the extent of inflation holds after controlling for students' underlying probability of receiving a top grade. Because grades are capped at the top, Equation (1) may fail to identify grade inflation for schools that already have high probabilities of students receiving top grades: most students in those schools would obtain top A-level grades regardless of the grading method. To separate compositional effects from true grade-inflation effects, I re-estimate the inflation using A* as the dependent variable and additionally interact students' GCSE scores with the school indicators. Columns 3 and 4 ofTable A5 report the correlation between school quality and the extent of inflation for A* and B, respectively. The smaller correlation coefficient for A* reflects how rarely low-to-mid school assigned A*, which flattens the estimated effect. By contrast, the coefficient for B is larger in magnitude than that for A, because lowering the grade threshold increases the pool of schools included in the regression.

I find that schools loosened their grading standards in subjects with less quantitative content. Figure A15a and Figure A15b report the relationship between school quality and the extent of grade inflation by subject type. On average, a one-unit increase in school quality was associated with a 1.5 percentage-point lower probability of receiving a top grade in non-quantitative subjects and a 0.5 percentage-point lower probability in quantitative subjects (Table 2). Columns 3 to 6 of Table 2 report regression coefficients using A* and B as dependent variables. The results show that the negative correlation between school quality and inflation is concentrated in non-quantitative subjects; no statistically significant relationship appears in quantitative subjects. These findings underscore that subjects which leave more grading discretion to teachers were the main channels schools used to raise students' grades.

I provide evidence that privately run schools inflated their pupils' grades more than other school types. Section 7 shows a bin-scatter plot comparing major publicly funded schools (e.g., academies and state schools) with privately run (independent) schools. Table 1 reports differences in grading patterns across school types and decomposes those estimates by school quality. Across the school-quality distribution, independent schools were more likely than publicly funded schools to assign top grades. Independent schools also adapted fastest to the new grading regime, increasing the rate of top-grade assignments in 2021 more than any other school type.

Notably, the difference in inflation between independent and other public schools does not appear to be driven by selectivity: I find no difference in inflation between grammar schools (public selective schools) and other public schools (Figure A4). I also find that sixth-form colleges—schools focused on preparing students for A-levels—experienced lower degrees of grade inflation in both 2020 and 2021. These differences are not driven by subject composition or student composition: the degree of grade inflation at sixth-form colleges was lower than at academies and state schools.

5.2 Persistence and Feedback of Grade Inflation

I estimate persistence in grade inflation across years by regressing 2021 inflation levels on 2020 inflation levels within the same school. Namely, I estimate

$$\Delta\alpha_{j,2021} = \rho\,\Delta\alpha_{j,2020} + \varepsilon_j. \tag{3}$$

I further decompose persistence by estimating subject-group–level grade inflation. Let

$$Y_{i,j,k,t} = \mathbf{1}\{Y^*_{i,j,k,t} \geq 0\}, \tag{4}$$

and model the latent index as

$$Y^*_{i,j,k,t} = \alpha_j + X_i\eta + \gamma_k + \sum_{\tau\in\{2020,2021\}} D_\tau \times \left(\sum_{g\in\mathcal{G}\setminus\{g_0\}} G_{k,g}\, \Delta\alpha_{j,\tau,g} + X_i\Delta\eta_\tau + \Delta\gamma_{k,\tau} \right), \tag{5}$$

where $G_{k,g}$ is an indicator for whether subject k belongs to group g (e.g., quantitative vs. non-quantitative).

To study within-school persistence by subject group, I estimate

$$\Delta\alpha_{j,g,2021} = \rho_{\text{same}}\, \Delta\alpha_{j,g,2020} + \rho_{\text{other}}\, \bar{\Delta}\alpha_{j,-g,2020} + \varepsilon_{j,g}, \tag{6}$$

where $\bar{\Delta}\alpha_{j,-g,2020}$ denotes the average 2020 inflation in school j across subject groups other than g.

The coefficients reported are descriptive and should not be given a causal interpretation. These estimates may confound unobserved factors that changed across the two pandemic years. The dynamic-panel literature (e.g., **?**) also discusses potential biases from estimating dynamics with specifications like (6). Nonetheless, the estimates are informative about the persistence of grade inflation.

Figure 7 reports the dynamic pattern of grade inflation across the two consecutive years. The y-axis measures the average increase in a student's probability of achieving a top grade in 2021, and the x-axis measures that increase in 2020. Schools with inflation levels near zero continued to inflate their students' grades at similar levels in the second year. By contrast, schools with low inflation in the first year increased their rate of grade inflation in 2021, while schools with high initial inflation slightly reduced their degree of inflation in the second year. This pattern is reflected in the estimates from Equation (6): a one-unit difference in the degree of inflation in 2020 is associated with a 0.6-unit difference in 2021. As Figure 7 shows, the slope is driven by adjustments among schools that did not inflate much in the initial year. Moreover, the extent of assigning B-or-higher grades did not vary across years (column 3 of Panel A in Table 3).

Figure A15 reports school-level inflation separately by subject group (quantitative vs. non-quantitative). The figure plots the average inflation levels across subject groups over the two consecutive years. As Table 3 shows, a one-unit difference in the degree of inflation in 2020 for non-quantitative subjects was associated with a 0.5-unit difference in the same subject group in 2021. By contrast, a one-unit difference in grade inflation for quantitative subjects in 2020 was associated with a 0.2-unit difference in that subject group in 2021.

Subfigure A of Figure A15 indicates that persistence within subject groups is pronounced for non-quantitative subjects. Schools' inflation levels in non-quantitative subjects remained similar to their initial-year levels, although schools that inflated less in the first year tended to increase their inflation in the second year. Subfigure D shows that quantitative subjects exhibit a weaker correlation across years: a positive relationship exists, but schools with high initial inflation tended to reduce their inflation in the second year, whereas schools with low initial inflation tended to increase it.

Subfigures B and D in Figure A15 report cross-subject inflation levels between neighboring subject groups over the two years. As the columns for cross-subject combinations in Table 3 confirm, I find little evidence of spillovers or feedbacks of grade inflation across neighboring subjects. Columns (1)–(4) show that a one-unit increase in grade inflation in a given subject group is not associated with a statistically significant increase in inflation in the neighboring subject in the following year.

However, the cross-subject coefficient is statistically significant for the outcome "B-or-higher." A one-unit difference in 2020 grade inflation in quantitative (non-quantitative) subjects is associated with a 0.3-unit difference in 2021 grade inflation for non-quantitative (quantitative) subjects, respectively. Notably, these cross-subject coefficients are smaller than the within-group coefficients.

5.3 Institutional differences in grading biases within classrooms

How were grades reassigned to students within a classroom? Figure A6 shows how the degree of grade inflation varied across students with different prior academic scores, measured on the national GCSE scale[8]. I plot the average grade-inflation level across score bins for the two mandatory subjects (Mathematics and English Literature), where 9 is the highest and 3 is the lowest observed grade in the sample.

In both subjects, the ordering of students' probabilities of receiving a top grade across score bins remained unchanged. However, I find differences in how top grades were assigned within classrooms by students' baseline GCSE scores. Specifically, in 2020 top A-level grades in Mathematics were concentrated among students with GCSE scores of 7–9. By contrast, the increase in the probability of receiving a top grade in English Literature was similar across students with different GCSE scores.

This pattern is not solely driven by baseline differences in students' propensity to receive a top grade. Although students with higher GCSE scores already have higher baseline probabilities of obtaining a top A-level grade, the observed inflation reflects larger probability increases for high-GCSE students: top GCSE students experienced bigger increases in their chances of receiving a top grade than lower-GCSE students, even after accounting for baseline differences. Figure A10 presents the same relationship using log-odds differences in a student's probability of receiving a top grade. Inflation remains concentrated among students with high GCSE Mathematics scores in both 2020 and 2021.

Figure A11 shows how grading patterns across GCSE score bins vary by school quality. The figure plots the change in the probability of receiving a top A-level grade across GCSE score bins, standardized to the mean GCSE score (GCSE = 6). It also shows aggregated mean differences in inflation by school-quality groups; school quality is measured using the school fixed effects estimated from Equation (1) and then split into three bins.

[8] The GCSE is a nationally standardized exam that all students take before starting their secondary-school (16–18) education; mathematics, English Language, and English Literature are mandatory.

I find that lower-quality schools — defined as those below the 30th percentile of the school-quality measure — tended to inflate the grades of students with top GCSE scores in both Mathematics and English Literature more than higher-quality schools. On the GCSE Mathematics scale, the increase in the probability of receiving a top A-level grade is roughly proportional to the student's GCSE score, and this proportional pattern is similar across school-quality bins. On the English Literature scale, the increases are more uniform across score bins, but low-quality schools show relatively larger increases for students scoring 8 or 9 on the GCSE English Literature exam. These patterns are consistent in both 2020 and 2021.

Demographic differences and students' economic affluence were also taken into account in grading. I find that female students had a higher probability of receiving a top grade than male students with similar academic and socioeconomic backgrounds (Section 7). I also find that White students had a slightly higher probability of receiving a top grade than ethnic-minority students. Female students were about one percentage point more likely to receive a top grade than male students.

Figure 8a plots how the gender grading bias varied by school-quality bins across the two pandemic years. Higher-quality schools were more likely to inflate female students' grades, whereas schools below the low-to-mid range of the quality distribution showed little gender-based difference. Schools in the middle of the quality distribution increased their gender grading bias in the second year, while low-quality schools did not show notable increases in gender differences.

Teachers and schools awarded higher grades to students from more affluent backgrounds. Section 7 reports grading patterns using income-related measures. Panel A shows how grading varies by parental occupational class: students whose parents hold higher-income occupations (for example, managers or professionals) were more likely to receive a top grade than students whose parents work in elementary occupations. The difference in the probability of receiving a top grade between students with parents in elementary occupations and those with parents in occupations with only slightly higher average incomes (for example, caring or sales) is small and not statistically significant.

The monotonic relationship between parental occupational income and the degree of inflation is stronger in lower-quality schools (Figure A8). I plot the average difference across parental occupational classes by school-quality bins; the bin-scatter shows that the correlation between parental income class and inflation is concentrated in low-quality schools. I do not observe a statistically significant relationship between parental income class and grade inflation in higher-quality schools.

These patterns are confirmed using alternative measures of parental wealth, such as the average university-progression rate in a student's home area: students from regions with higher progression rates were assigned higher grades at low-quality schools, but not at high-quality schools.

5.4 Effect on student placement at universities

Did the extent of grade inflation differ for similar students with different application portfolios? To assess whether schools adjusted inflation depending on the universities students applied to, I test whether grade inflation differentially affected students' placement after conditioning on student and school attributes, but without controlling for application patterns.

First I estimate whether student-level inflation predicts placement into a given course/program:

$$Y^*_{i,j,s} = X_i\beta \;+\; \gamma\,\mathrm{Infl}_i \;+\; \alpha_s \;+\; \kappa_j \;+\; u_{i,j,s}, \tag{7}$$

$$Y_{i,j,s} = \mathbf{1}\{Y^*_{i,j,s} > 0\}, \tag{8}$$

where $Y_{i,j,s}$ indicates that student i from school s is placed into (or accepts) course/program j; X_i contains the student's predicted grade from a standardized exam and other observables; Infl_i is the student-level assigned-grade inflation (observed grade minus predicted grade); α_s are school fixed effects; and κ_j are course/program fixed effects.

Next I allow the effect of inflation to vary by school to test whether some schools tailor inflation to students' target universities:

$$Y^*_{i,j,s} = X_i\beta \; + \; \gamma\,\mathrm{Infl}_i \; + \; \delta_s\big(\mathrm{Infl}_i \times D_s\big) \; + \; \alpha_s \; + \; \kappa_j \; + \; u_{i,j,s}, \tag{9}$$

$$Y_{i,j,s} = \mathbf{1}\{Y^*_{i,j,s} > 0\}, \tag{10}$$

where D_s denotes a (school) indicator used to recover school-specific interaction effects and δ_s captures how a marginal increase in Infl_i at school s mechanically alters the student's probability of meeting the grade requirement for (or being placed into) higher-quality universities, above the average effect γ.

Identification of δ_s exploits cross-student variation in Infl_i conditional on covariates X_i, school fixed effects α_s, and course fixed effects κ_j. In practice I recover δ_s by interacting Infl_i with school indicators, so comparisons are made between students with different estimated inflation but otherwise similar observables within and across schools.

Because I do not observe each student's counterfactual grade, I construct a predicted grade $gr\hat{a}de_{i,2020}$ for each student using an elastic-net trained on pre-2020 outcomes, and define inflation as the realized minus predicted grade. I then regress this measure of inflation on students' observed tariff scores.[9]

$$\widehat{\mathrm{Infl}}_{i,2020} = grade_{i,2020} - gr\hat{a}de_{i,2020}, \qquad \widehat{\mathrm{Infl}}_{i,2020} = \beta\,\mathrm{Tariff}_i + \varepsilon_{i,2020}.$$

Here Tariff_i denotes the student's observed tariff score (a summary of prior attainment). The coefficient β captures how inflation varies with students' position in the prior academic

[9] The prediction step $gr\hat{a}de_{i,2020} = \mathcal{F}(X_i)$ uses a penalized regression,

$$\hat{\beta}_\lambda = \arg\min_{\delta} \; \frac{1}{2n} \sum_{i \in \mathrm{pre}} \big(y_i - X_i^\top \delta\big)^2 \; + \; \lambda\left[\frac{1-\xi}{2}\,\|\delta\|_2^2 + \xi\,\|\delta\|_1\right],$$

where $\lambda > 0$ is the tuning parameter and $\xi \in [0,1]$ controls the ℓ_1–ℓ_2 mix. This design isolates how much of the observed inflation can be attributed to students' position in the academic distribution while allowing the prediction step to flexibly incorporate a rich set of covariates X_i.

distribution.

I find that schools differed in how grade inflation affected students' chances of being placed at their firm-choice university. Figure A13 plots the marginal effect of a one-unit increase in assigned-grade inflation on the probability that a student from a given school is placed at their firm choice. In 2020, these marginal effects are largest for schools near the 20th, 55th, and 85th percentiles of the school-quality distribution. A simple linear regression of the marginal effect on school quality yields a positive coefficient in 2020, indicating that higher-quality schools extracted larger placement gains from inflation. By 2021, however, this positive correlation disappears (Table A13).

Furthermore, I decompose placement improvements by two grade-inflation channels: (i) school-specific inflation and (ii) grade changes induced by students' GCSE scores (and by the selectivity of the universities they apply to). Figure A14 presents bin-scatter plots of the marginal effect of a one-unit increase in assigned-grade inflation (separately for 2020 and 2021) on the probability of securing a place at a student's firm-choice university, decomposed into variation attributable to GCSE Mathematics scores and to the school's overall inflation level.

The channels operate differently across years. In 2020, GCSE-induced grade increases translated into better placements mainly for schools in the middle of the quality distribution; the marginal effect for top-ranked schools was similar to that for schools at the bottom of the distribution despite differences in pupil composition. In 2021, however, the placement gains from GCSE-induced inflation rose for high-quality schools: the marginal probability of placement for high-quality schools was about 8 percentage points higher than for lower-quality schools. By contrast, the school-specific inflation channel lost relative influence in 2021 — its contribution to placement success declined in the second year.

Regression results in Table A9 corroborate these patterns. A one–percentage-point difference in school quality is associated with a decline in the placement payoff from school-level inflation, while within-school re-grading based on pupils' GCSE grades increases the probability of placement at higher-quality universities. In both years, the interaction coefficients on school quality are statistically significant, indicating that the effectiveness of each inflation channel systematically varies with school quality.

Additionally, I decompose placement effects by the selectivity of the universities applicants target. Figure A13 plots the marginal effect of a one-unit increase in assigned-grade inflation on the probability of securing a place at a selective program and at a non-selective program; Panel A shows 2020 and Panel B shows 2021. In 2020 the placement impact of inflation is largest for schools near the 20th–30th percentile of the school-quality distribution. In 2021, the group of schools that gains the most shifts toward the middle of the distribution (roughly the 50th–60th percentiles): mid-rank schools led other schools in converting inflation into placements at students' firm-choice programs.

Table A8 reports regression estimates of the probability of placing into a firm-choice program by the selectivity level of the target university. The point estimates show a weakly negative relationship between placement gains and school quality, but these coefficients are not statistically significant. Taken together with the bin-scatter in Figure A13, the evidence suggests that mid-quality schools took greater advantage of grading discretion to improve their pupils' placement outcomes. The conversion of grades into placements therefore does not follow a monotonic pattern: lower-quality schools generally did not succeed in pushing pupils into better programs, while mid-quality schools were most effective at using grade inflation to secure students' firm-choice placements.

6 Discussion

My results qualitatively aligns with **?** on how schools with worse quality inflates grade more than good schools. I find that a one unit difference in school quality was associated with a 10 percentage point difference in the extent of inflating grades. **?** and Denning et al. (2023) similarly find concentration of grade manipulation in schools with worse quality. **?**

finds that a one standard deviation in students achievements is associated with a 0.4 percent point probability in manipulating the students scores. Denning et al. (2023) finds that a one percentage point increase in shares of students with free lunch was associated with 4.8 more students getting manipulated scores.

I find evidence on the positive evolution of grade inflation levels consistent with Dee et al. (2019) and Bleemer (2020). I find that experiencing a year of grade discretionary grading scheme is associated with on average a 10 percent increase in the average grade inflation level in the following year. This results aligns withDee et al. (2019) which finds that Dartmouths incoming student's average GPA increased by an annual rate of 40 percent, while Bleemer (2020) finds the number to be 10 percent.

My paper provides a novel evidence on the dynamics of grade inflation by decomposition the extent of the grading policy update by their school types and also their inflation levels in the initial year.I find that the private schools schools increased their inflate rate much faster than other type of schools. Private schools increased their inflation rate by 4 folds of the previous year, while difference in inflation levels during the two years were similar for public schools. My findings matches with Bleemer (2020) who finds that schools with students with higher income parents exhibit a faster rate of inflation then other types of schools. However, the persistence of grade inflation is confined with subject groups with weaker effect to spreading to less related subjects. Table 3 reports that the correlation between same subject groups across the two consecutive years are significantly stronger than the correlation between neighboring subject groups. My findings support how grades are contagious within schools (Chan et al. (2007)), but also highlights how the spread of the grading policy is varies within a school.

My results support how students economic affluence influences teachers grading behavior, supporting the findings by Burgess and Greaves (2013) and Botelho et al. (2015), while contradicting the findings in Hanna and Linden (2012). My results point to how teachers, especially at worse quality schools, tended to assign higher grades to students with lower economic affluence, measured by their parents occupational class and the income deprivation index of their permanent residence. Hanna and Linden (2012) finds that lower caste students were assigned at most 0.9 standard deviation lower grades than a student from a higher caste.

I provide evidence that grade inflation is positively correlated with the academic achievements of the student, measured by their GCSE scores. However, the monotone pattern is observed mainly in low quality schools. The extent of the grade inflation is exponentially larger for students with higher GCSE math scores. The results matches Cornwell et al. (2013) and Botelho et al. (2015), who finds similar effect in primary school children. Cornwell et al. (2013) find that an one standard deviation increase in math scores increases grade assigned by nearly 0.2 standard deviation points in their grade measure. However, my paper highlights that this pattern is only hold for a subpopulation, namely the low quality schools. For mid-quality schools, the schools/teachers do not seem to assign grades by the academic score of their pupils, and in high quality school grade inflation is larger for the students with lower academic scores.

I find evidence of gender bias against males. The findings support the findings of the grade bias literature (Lavy (2009), Lavy and Sand (2018), Cornwell et al. (2013)). Female students received top grades by 6 percentage points higher than male students. The results are at odds with the findings of Hanna and Linden (2012) and Diamond and Persson (2016), that does not find any biases against males when teachers have grading discretion. Diamond and Persson (2016) attempted to identify gender biases by the size of the discontinuity near the cutoff grades for a top grade which is different from my approach with compares final grades between a predicted scores in the standardized regime against the observed teacher-assigned grade.

I provided a decomposition exercise to explain the changes in the overall grading distribution during the COVID 19 pandemic into the grading biases from the student attributes and the schools grading policies. I find rich heterogeneity in how students attributes lead to reshuffling of top grades within a school with their patterns varying by the schools quality.

At low quality schools, students with higher academic achievements increase their chances of receiving a top grade, thereby exacerbating the achievement inequality across students with varying academic ability. However, low quality schools also provides leniency towards students from socio-economically disadvantaged backgrounds, by assigning better grades towards to the disadvantaged student when comparing academically similar students. However, the mean-spreading property of academic achievements of students weaken in mid-high quality schools, as mid-tier schools show lower tendency to change the ordering of students by academic or demographic characteristics. High quality school inflates the grades for lower academically capable students. However the results for the high quality schools needs to be carefully interpreted as students in high quality schools already have a high chance of receiving a top grade regardless of the grading regime.

Despite the non-standardized grading regime increased students chances of receiving a top grade, only a limited population of students gained from the opportunity to use the opportunity to secure a placement at a selective university that they had little chance of acceptance. The findings are similar to Dee et al. (2019) and Phillips and Reber (2022) in which they show how improving access to universities don't necessary lead to better placements. My results show that most under-represented students in 2020 did not place into selective universities despite higher grades, even in 2021 although students could have applied to better universities if they successfully predicted that their teachers will grant them higher grades then they deserve.

7 Conclusion

This paper demonstrates how replacing grading discretion from a central examiner to the the local educators of students reduces the information content of grades. The new grades reflected the quality of the grade rewarding institution, with lower quality school adding higher grades to than higher quality schools. The grading standards rapidly evolves upwards with educational standards and the information content deteriorating in an increasing rate, with profit driven schools among the schools with fastest adjustments to the new grading regime. The loosening of grading standards are also accompanied by a pattern for teachers/schools to reward grades that satisfies the acceptance requirement provided by the universities. The grade matching adds to another source of information deterioration of the rewarded grade, as more institutions with stricter grade requirement are subject to a worse deterioration of grade standards as schools/teachers adjust for the universities responses.

Despite the loosening of grading standard was larger in lower quality institutions the impact of the jump in grade were limited to placement improvements for students from disadvantaged socio-economic background. Additionally the improvements in representations at selective institutions were limited to students above the average in the pre-pandemic grade distribution. The non-monotone relationship between grade and placement mobility were driven by differences in application choices between students with varying baseline backgrounds, as disadvantaged students do not apply to selective institutions despite their rewarded grades facilitating their placements into selective institutions than students from more advantaged backgrounds.

My results points to the dangers of accommodating test-optional policies into university admission systems. Proponents of test-optional policies advocate for a holistic admission system to diversify students background at selective universities, as well as using schools and teachers as a method for extracting information on students quality and potential that would be missed out in traditional standardized exams in which the examiner has no-knowledge of the students background when they are grading the tests. The grading patterns in the UK demonstrates the vulnerability of such decentralized grading regimes to potential gaming of grades from teachers, as well as their limited effect on diversifying selective universities with students socio-economic backgrounds. The speed of the adaptation were fastest for private schools, which teaches the students from the wealthiest background of parents. I suggest that simply lowering the barriers to selective universities will not improve the diversity of selective universities. Even among the students that apply to universities, students from socio-economics advantaged vary with their attitudes towards selective institutions, such as the composition of students or their parents attitudes or knowledge on elite institutions. I propose that the lowering the barriers for universities to diversity their knowledge to students from diverse backgrounds could be met by reducing frictions in the benefits of higher-education but not on the application system.

A important caveat of the policy implication is whether policy makers can take advantage of the regressive grade inflation from non-standardized testing while assuring that the increase in grades for socio-economically disadvantaged students are converted to improvements into placements into selective institutions. As my results uses the changes in grades due to COVID19 pandemic, the application patterns that I use for assessing placement pattern may be influenced by the non-normal situation accruing from the pandemic chaos. For example, students from disadvantaged backgrounds may refrain from applying ambitiously despite receiving high grades, and also students may prefer to study at universities closer to their parents to ease the civil unrest during the pandemic. A feasible extension of my results would be to take advantage of the growing literature on estimating of university application portfolio choice (Ali and Shorrer (2025)) and test whether placement patterns would change if students had changed their application choice by foreseeing the changes in the grading scheme.

References

Alesina, A., Carlana, M., La Ferrara, E., and Pinotti, P. (2024). Revealing stereotypes: Evidence from immigrants in schools. *American Economic Review*, 114(7):1916–48.

Ali, S. N. and Shorrer, R. I. (2025). Hedging when applying: Simultaneous search with correlation. *American Economic Review*, 115(2):571–98.

Arcidiacono, P. (2005). Affirmative action in higher education: How do admission and financial aid rules affect future earnings? *Econometrica*, 73(5):1477–1524.

Arcidiacono, P., Kinsler, J., and Ransom, T. (2023). What the students for fair admissions cases reveal about racial preferences. *Journal of Political Economy Microeconomics*, 1(2).

Avery, C., Shi, L., and Magouirk, P. (2025). Test-optional college admissions: Act and sat scores, applications, and enrollment changes. Working Paper 34260, National Bureau of Economic Research.

Bar, T., Kadiyali, V., and Zussman, A. (2009). Grade information and grade inflation: The cornell experiment. *Journal of Economic Perspectives*, 23(3):93–108.

Becker, G. S. (1971). *The Economics of Discrimination.* University of Chicago Press.

Bennett, C. T. (2022). Untested admissions: Examining changes in application behaviors and student demographics under test-optional policies. *American Educational Research Journal*, 59(1):180–216.

Bettinger, E. P., Long, B. T., Oreopoulos, P., and Sanbonmatsu, L. (2012). The role of application assistance and information in college decisions: Results from the h&r block fafsa experiment*. *The Quarterly Journal of Economics*, 127(3):1205–1242.

Betts, J. R. and Grogger, J. (2003). The role of grading standards in education policy. *American Economic Review*, 93(2):94–99.

Black, S. E., Denning, J. T., and Rothstein, J. (2023). Winners and losers? the effect of gaining and losing access to selective colleges on education and labor market outcomes. *American Economic Journal: Applied Economics*, 15(1):26–67.

Bleemer, Z. (2020). Grade inflation at more- and less-affluent high schools. Technical report, UC-CHP (UC Center for Health Policy) Policy Brief 2020.1.

Bleemer, Z. (2021). Affirmative action, mismatch, and economic mobility after california's proposition 209*. *The Quarterly Journal of Economics*, 137(1):115–160.

Bleemer, Z. (2023). Affirmative action and its race-neutral alternatives. *Journal of Public Economics*, 220:104839.

Borusyak, K. and Hull, P. (2023). Nonrandom exposure to exogenous shocks. *Econometrica*, 91(6):2155–2185.

Botelho, F., Madeira, R. A., and Rangel, M. A. (2015). Racial discrimination in grading: Evidence from brazil. *American Economic Journal: Applied Economics*, 7(4):37–52.

Burgess, S. and Greaves, E. (2013). Test scores, subjective assessment, and stereotyping of ethnic minorities. *Journal of Labor Economics*, 31(3):535–576.

Carlana, M. (2019). Implicit stereotypes: Evidence from teachers' gender bias*. *The Quarterly Journal of Economics*, 134(3):1163–1224.

Chan, W., Hao, L., and Suen, W. (2007). A signaling theory of grade inflation. *International Economic Review*, 48(3):1065–1090.

Chetty, R., Deming, D. J., and Friedman, J. N. (2023). Diversifying society's leaders? the determinants and causal effects of admission to highly selective private colleges. Working Paper 31492, National Bureau of Economic Research.

Chetty, R., Friedman, J. N., Saez, E., Turner, N., and Yagan, D. (2020). Income segregation and intergenerational mobility across colleges in the united states*. *The Quarterly Journal of Economics*, 135(3):1567–1633.

Cornwell, C., Mustard, D. B., and Van Parys, J. (2013). Non-cognitive skills and the gender disparities in test scores and teacher assessments: Evidence from primary school. *Journal of Human Resources*, 48(1):236–264.

Cullen, J. B., Long, M. C., and Reback, R. (2013). Jockeying for position: Strategic high school choice under texas' top ten percent plan. *Journal of Public Economics*, 97:32–48.

Dee, T. S., Dobbie, W., Jacob, B. A., and Rockoff, J. E. (2019). The causes and consequences of test score manipulation: Evidence from the new york regents examinations. *American Economic Journal: Applied Economics*, 11(3):339–365.

Denning, J. T., Eide, E. R., Mumford, K. J., Patterson, R. W., and Warnick, M. (2022). Why have college completion rates increased? *American Economic Journal: Applied Economics*, 14(3):1–29.

Denning, J. T., Nesbit, R., Pope, N., and Warnick, M. (2023). The distributional and long-term effects of grade inflation. Technical report, NBER Working Paper No. 31123.

Dessein, W., Frankel, A., and Kartik, N. (2025). Test-optional admissions. *American Economic Review*, 115(9):3130–70.

Diamond, R. and Persson, P. (2016). The long-term consequences of teacher discretion in grading of high-stakes tests. Working Paper 22207, National Bureau of Economic Research.

Dynarski, S., Nurshatayeva, A., Page, L. C., and Scott-Clayton, J. (2022). Addressing non-financial barriers to college access and success: Evidence and policy implications. NBER Working Paper 30054, National Bureau of Economic Research.

Dynarski, S. M. (2003). Does aid matter? measuring the effect of student aid on college attendance and completion. *American Economic Review*, 93(1):279–288.

Efron, B. (2010). *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*. Institute of Mathematical Statistics Monographs. Cambridge University Press.

Epple, D., Romano, R., and Sieg, H. (2006). Admission, tuition, and financial aid policies in the market for higher education. *Econometrica*, 74(4):885–928.

Ferman, B. and Fontes, L. F. (2022). Assessing knowledge or classroom behavior? evidence of teachers' grading bias. *Journal of Public Economics*, 216:104773.

Fernández-Val, I. and Weidner, M. (2016). Individual and time effects in nonlinear panel models with large n, t. *Journal of Econometrics*, 192(1):291–312.

Figlio, D. N. and Getzler, L. S. (2006). Accountability, ability, and disability: Gaming the system? *Advances in Applied Microeconomics*, 14:35–49.

Figlio, D. N. and Lucas, M. E. (2004). Do high grading standards affect student performance? *Journal of Public Economics*, 88(9):1815–1834.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis*. Chapman and Hall/CRC, 3rd edition.

Hanna, R. N. and Linden, L. L. (2012). Discrimination in grading. *American Economic Journal: Economic Policy*, 4(4):146–168.

Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature*, 24(3):1141–1177.

Hoxby, C. M. and Avery, C. (2012). The missing one-offs: The hidden supply of high-achieving, low income students. Working Paper 18586, National Bureau of Economic Research.

Jacob, B. A. and Levitt, S. D. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating*. *The Quarterly Journal of Economics*, 118(3):843–877.

Lavy, V. (2009). Performance pay and teachers' effort, productivity, and grading ethics. *American Economic Review*, 99(5):1979–2011.

Lavy, V. and Sand, E. (2018). On the origins of gender gaps in human capital: Short- and long-term consequences of teachers' biases. *Journal of Public Economics*, 167:263–279.

Manski, C. F. and Wise, D. A. (1983). *College Choice in America*. Harvard University Press, Cambridge, MA.

Neal, D. and Schanzenbach, D. W. (2008). Left behind by design: Proficiency counts and test-based accountability. *Review of Economics and Statistics*, 92(2):263–283.

Pattison, E., Grodsky, E., and Muller, C. (2013). Is the sky falling? grade inflation and the signaling power of grades. *Educational Researcher*, 42(5):259–265.

Phillips, M. and Reber, S. (2022). Does virtual advising increase college enrollment? evidence from a random-assignment college access field experiment. *American Economic Journal: Economic Policy*, 14(3):198â€“234.

Rojstaczer, S. and Healy, C. (2012). Where a is ordinary: The evolution of american college and university grading, 1940-2009. *Teachers College Record*, 114(7):1–23.

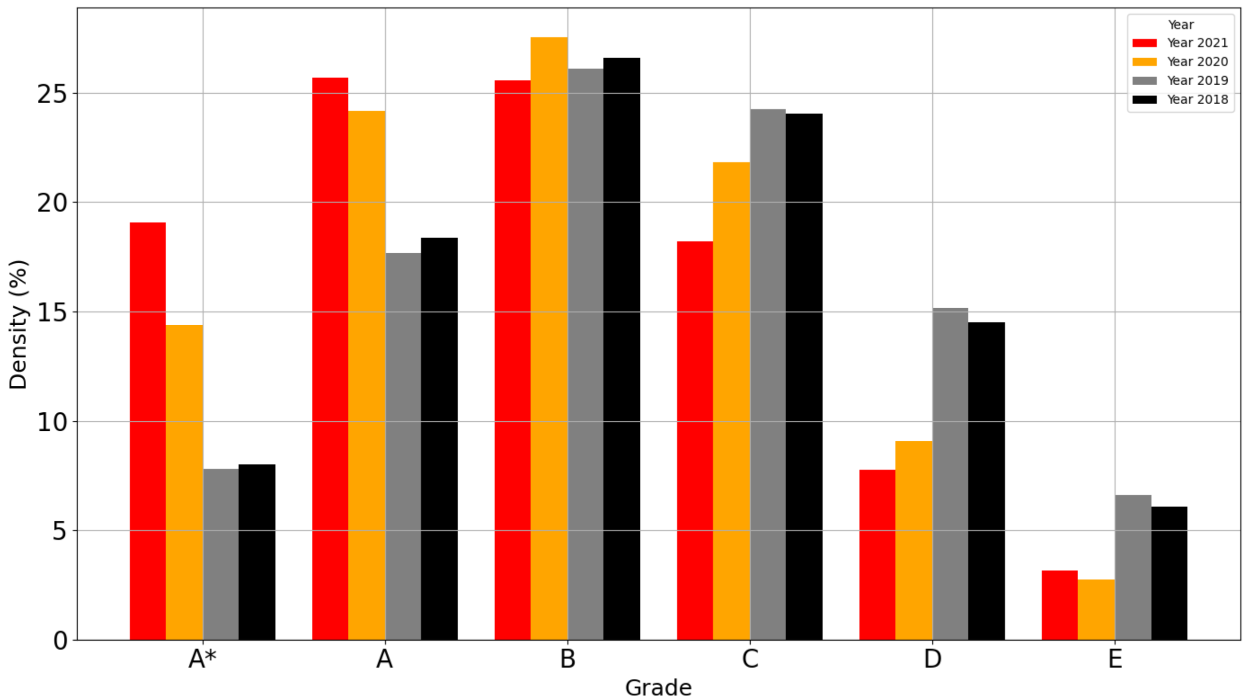
Sacerdote, B., Staiger, D. O., and Tine, M. (2025). How test optional policies in college admissions disproportionately harm high achieving applicants from disadvantaged backgrounds. Working Paper 33389, National Bureau of Economic Research.

Tyner, A. and Gershenson, S. (2020). Conceptualizing grade inflation. *Economics of Education Review*, 78:102037.

Walters, C. (2024). Chapter 3 - empirical bayes methods in labor economics. volume 5 of *Handbook of Labor Economics*, pages 183–260. Elsevier.

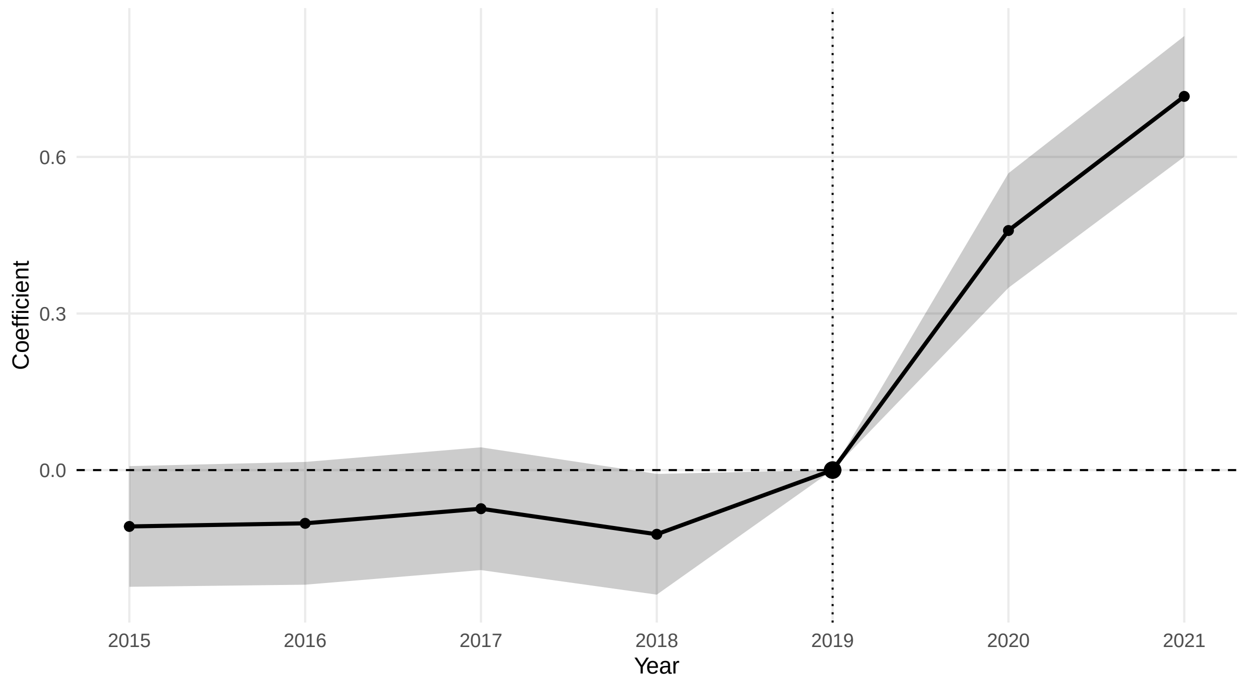
Wooldridge, J. M. (2001). *Econometric Analysis of Cross Section and Panel Data*, volume 1 of *MIT Press Books*. The MIT Press, 1 edition.

Figure 1: Grade distribution



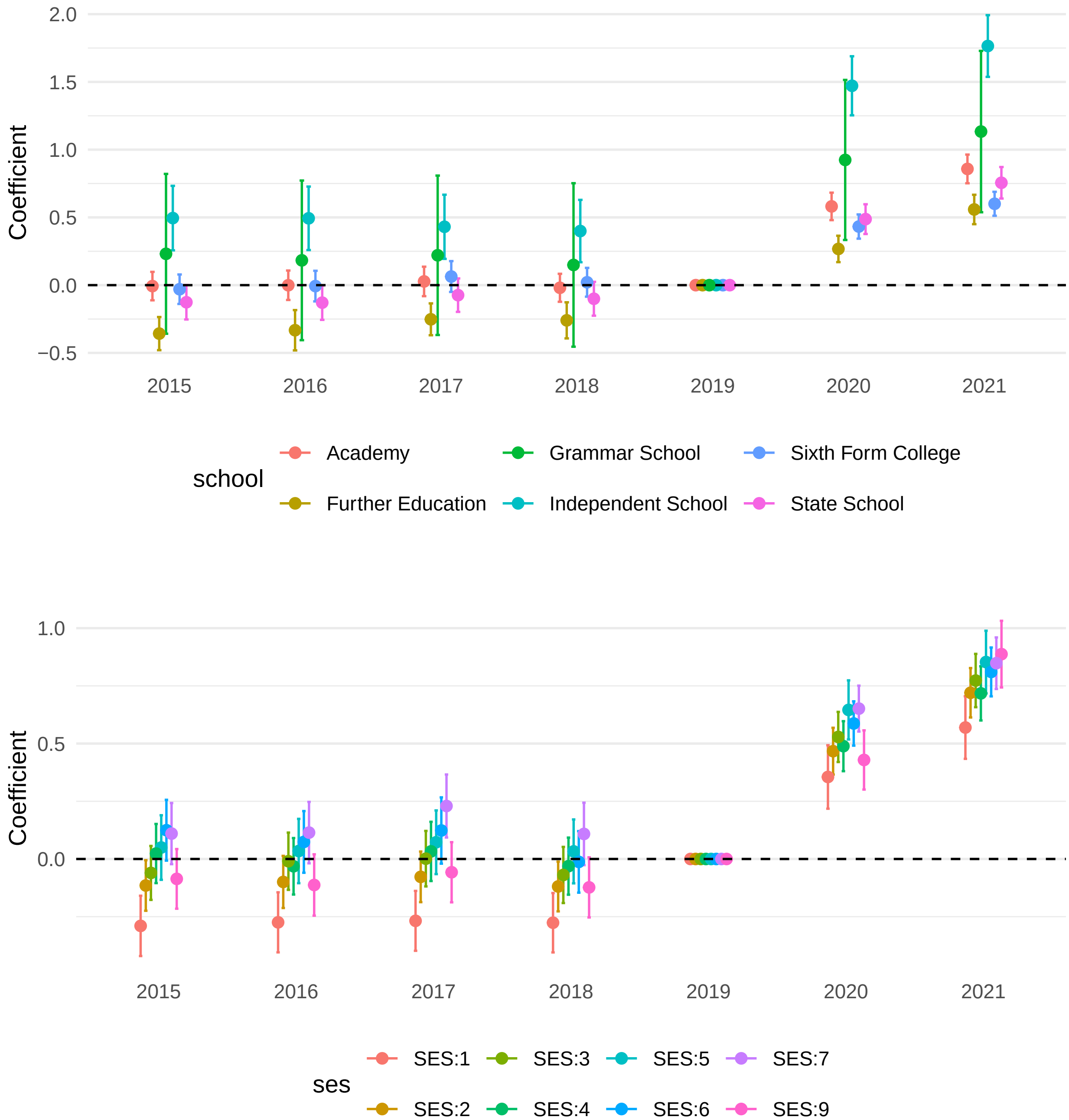
Note: The histogram displays the share of letter grades issued for each grade bins in the A-levels between 2018 and 2021. A*,A denotes the highest grade assigned to exam results scoring above the top 90th and 80th percentile respectively. Sample is the universe of test takers in each year. Data is derived from grade reports from the Joint Councils of Qualifications.

Figure 2: Results on stationarity test comparing average grade across years



Note: The figure displays the exam level year-to-year differences in probability to receive a top grade (A or A*) between the reference year (2019) and other years from 2015 to 2021. The sample is the universe of A-level exams used for admissions at British universities through the centralized administrative system (N = 2,802,651). Year-specific coefficients and 95% confidence intervals are from a logistic regression that regresses a success dummy indicating the student receiving a top grade (A or A*) on a series of time indicator dummy variables while controlling for students' achievement score in a previous standardized national exam (GCSE) in mandatory subjects (mathematics and English literature) and a dummy variable for the subject course of the exam.. Standard errors are clustered at the school-subject course level.

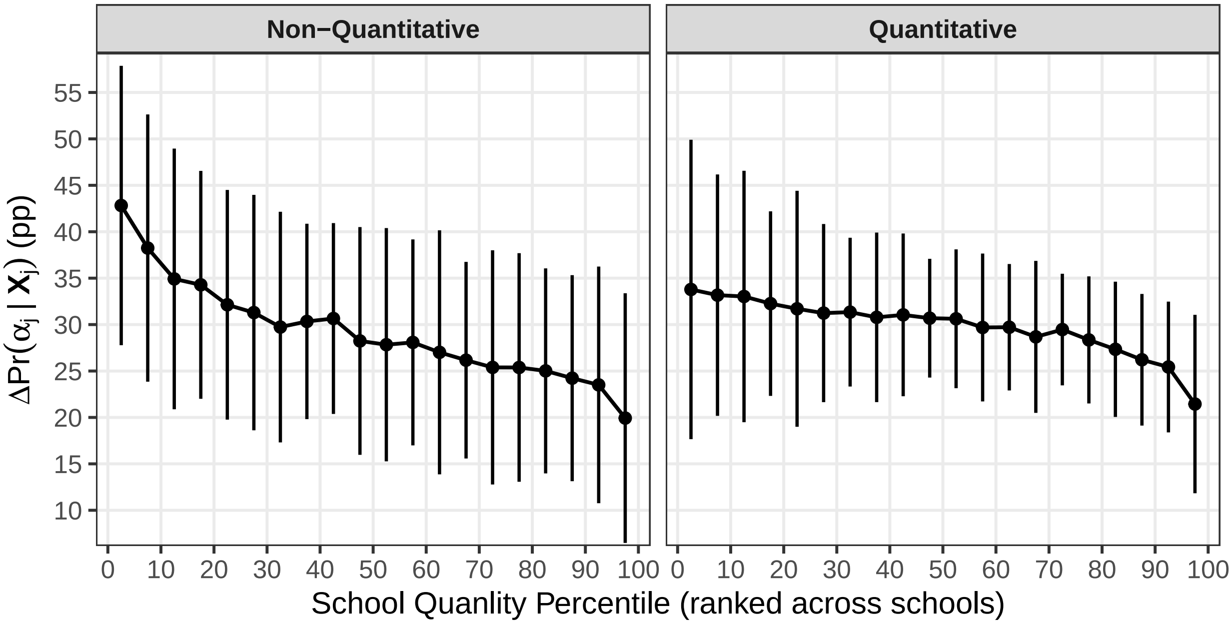
Figure 3: Stationatity test by (a) school types and (b) parental occupation class



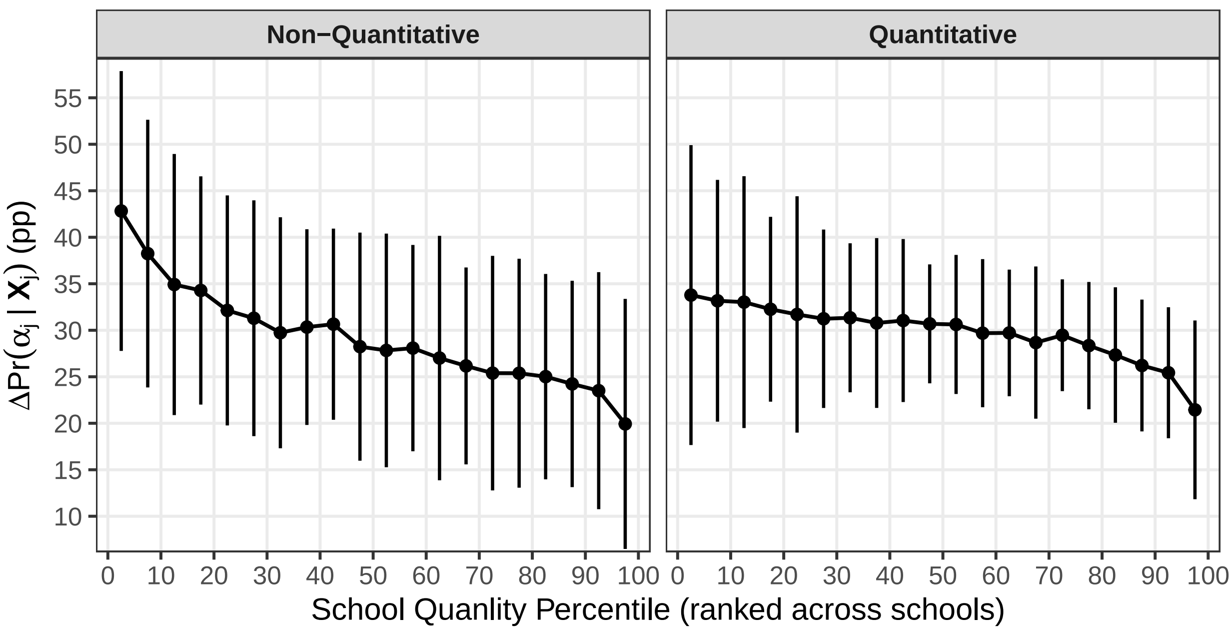
Note: The figure displays the exam level year-to-year differences in probability to receive a top grade (A or A*) between the reference year (2019) and other years from 2015 to 2021 after sub-sampling the sample by the test takers' (a) school types or (b) parents' parental occupation class. The sample is the universe of A-level exams used for admissions at British universities through the centralized administrative system (N = 2,802,651). Year-specific coefficients and 95% confidence intervals are from a logistic regression that regresses a success dummy indicating the student receiving a top grade (A or A*) on a series of time indicator dummy variables while controlling for students' achievement score in a previous standardized national exam (GCSE) in mandatory subjects (mathematics and English literature) and a dummy variable for the subject course of the exam. Coefficients are estimated jointly by interacting the times indicator dummy with either a (a) school type indicator or a (b) parental occupation indicator. Standard errors are clustered at the school-subject course level.

Figure 4: Grade inflation by school quality and subject group.

(a) 2020



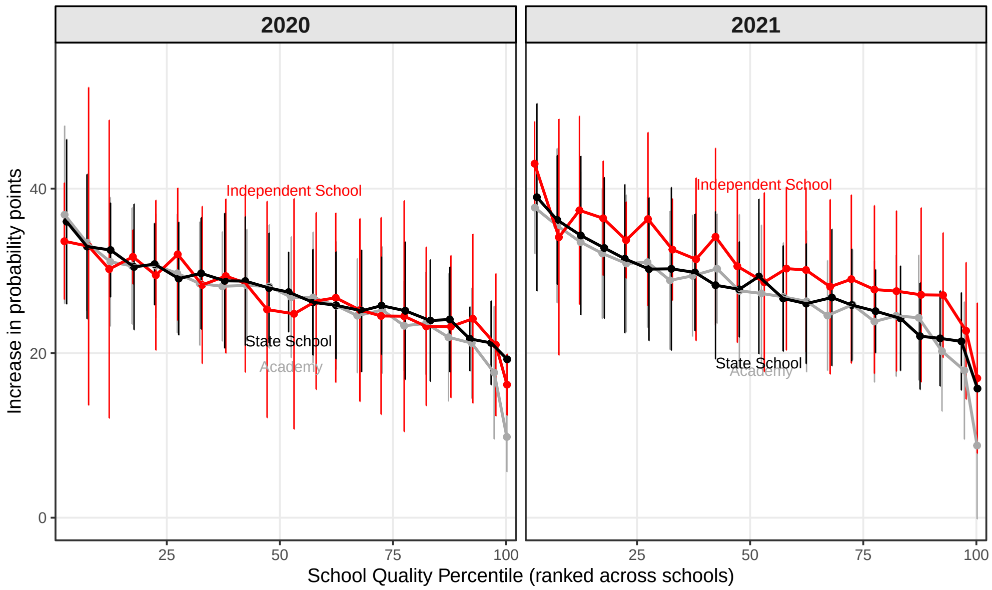
(b) 2021



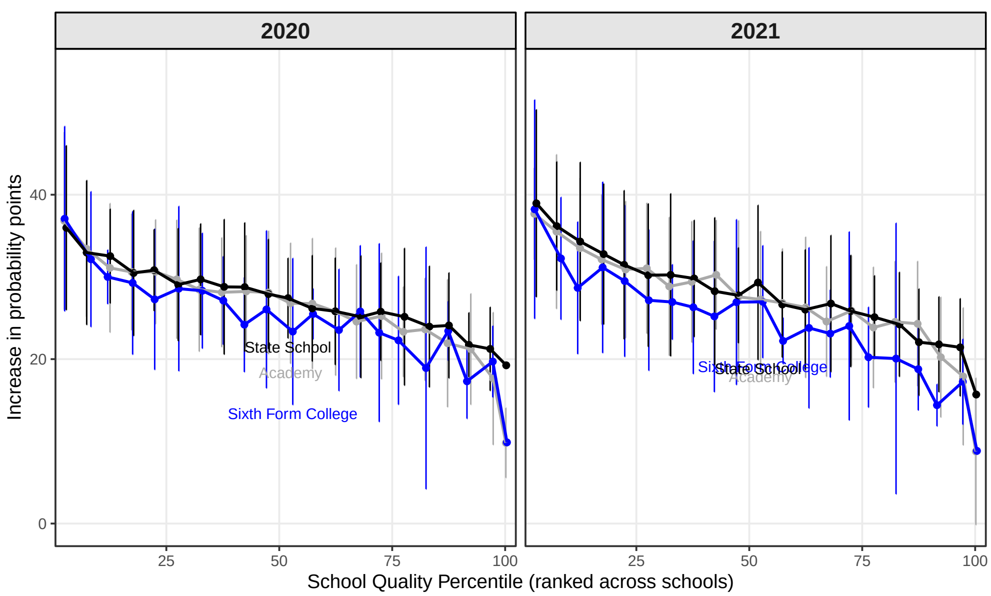
Note: This bin-scatter displays the subject group level increments in grade improvements defined as a student receiving a top grade (A or A^*) in (a) 2020 and (b) 2021. The sample is the universe of A-level exams used for admissions at British universities through the centralized administrative system (N = 2,802,651). Year-specific coefficients and 95% confidence intervals are from a logistic regression that regresses a success dummy indicating the student receiving a top grade (A or A^*) on the interaction between school subject group dummy and a time indicator for both pandemic years. The regression controls for students' achievement score in a previous standardized national exam (GCSE) in mandatory subjects (mathematics and English literature) and a dummy variable for the subject course of the exam. Standard errors are clustered at the school-subject course level.

Figure 5: Grade inflation by school type

(a)

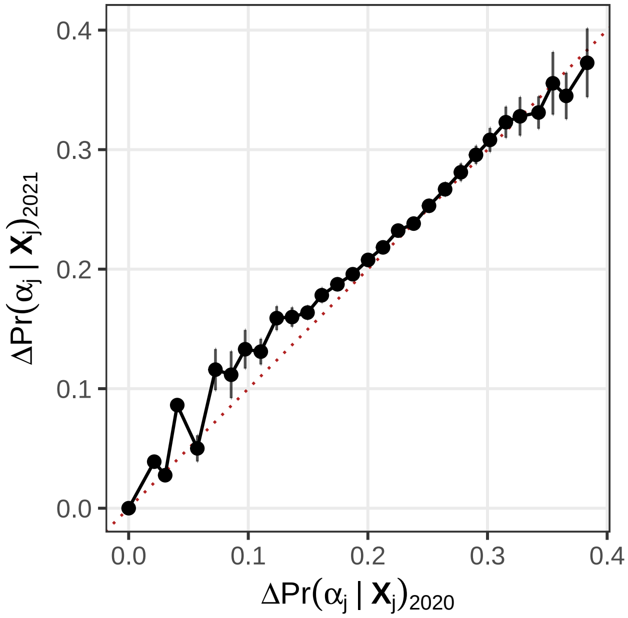


(b)



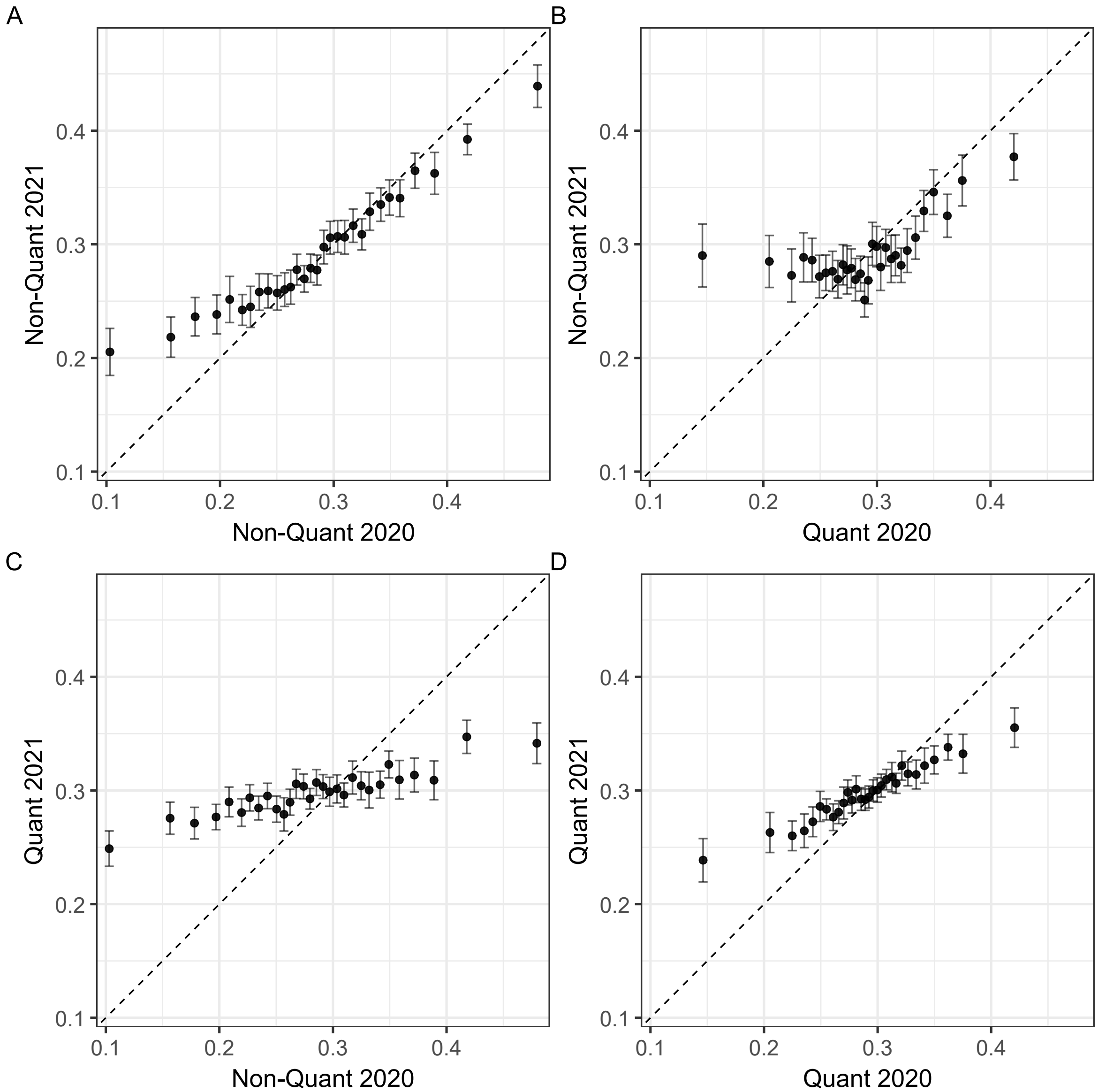
Note: The bin-scatter displays the school level increments in grade improvements defined as a student receiving a top grade (A or A*) in (a) 2020 and (b) 2021. The sample is the universe of A-level exams used for admissions at British universities through the centralized administrative system (N = 2,802,651). Year-specific coefficients and 95% confidence intervals are from a logistic regression that regresses a success dummy indicating the student receiving a top grade (A or A*) on the interaction between school subject group dummy and a time indicator for both pandemic years. The regression controls for students' achievement score in a previous standardized national exam (GCSE) in mandatory subjects (mathematics and English literature) and a dummy variable for the subject course of the exam. Standard errors are clustered at the school-subject course level.

Figure 6: Aggregate dynamics



Note: The bin-scatter displays the school level grade improvements in 2021 relative the degree of grade improvements in 2020 for the same school. The sample is the universe of A-level exams used for admissions at British universities through the centralized administrative system (N = 2,802,651). Year-specific coefficients and 95% confidence intervals are from a logistic regression that regresses a success dummy indicating the student receiving a top grade (A or A*) on the interaction between school subject group dummy and a time indicator for both pandemic years. The regression controls for students' achievement score in a previous standardized national exam (GCSE) in mandatory subjects (mathematics and English literature) and a dummy variable for the subject course of the exam. Standard errors are clustered at the school-subject course level.

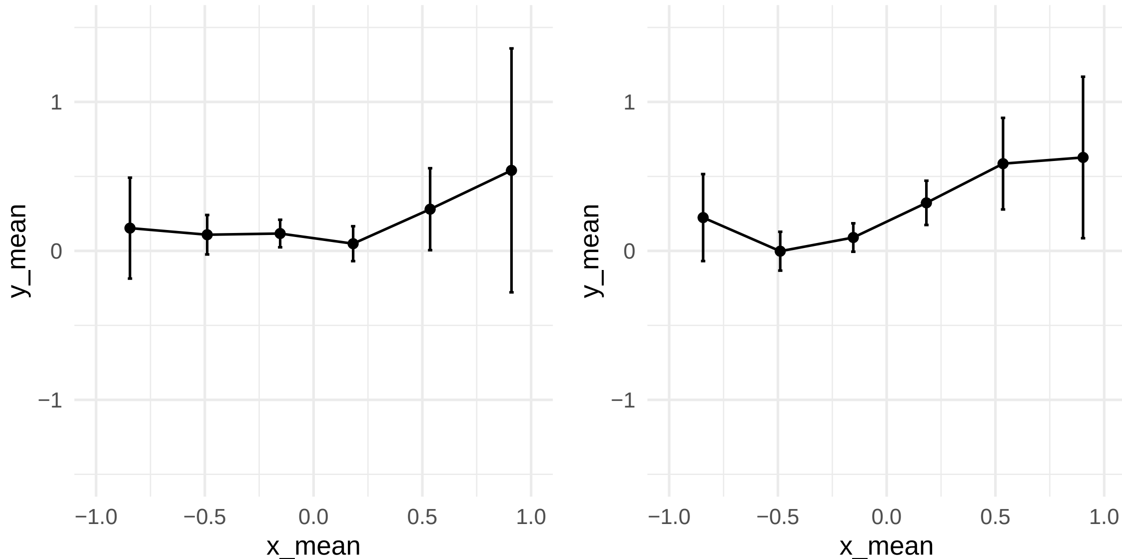
Figure 7: By subject groups



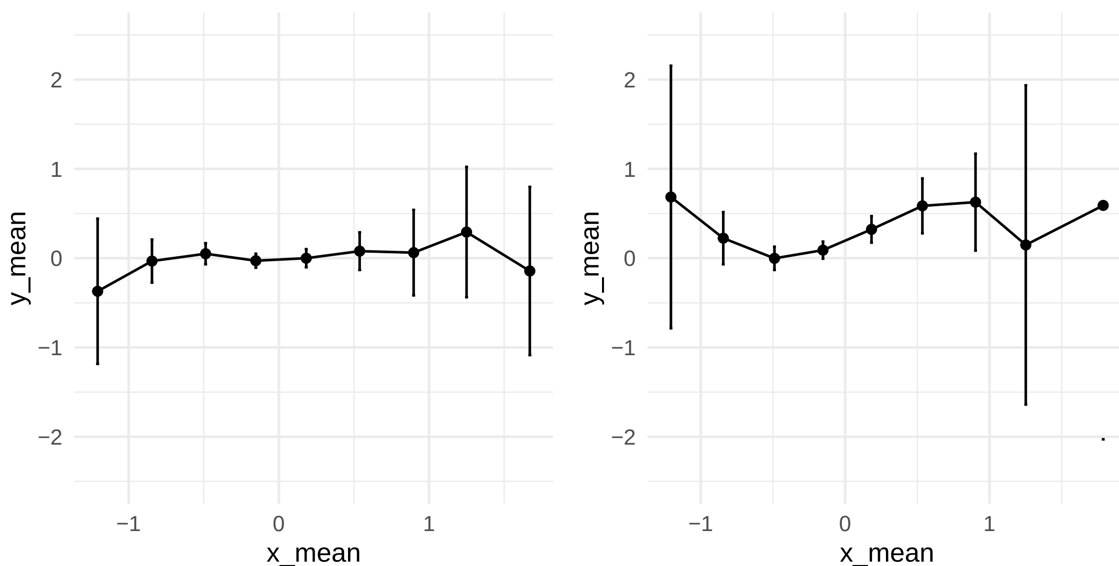
Note: The bin-scatter displays the school-subject-group level grade improvements in 2021 relative the degree of grade improvements in 2020 for the same school. The sample is the universe of A-level exams used for admissions at British universities through the centralized administrative system (N = 2,802,651). Year-specific coefficients and 95% confidence intervals are from a logistic regression that regresses a success dummy indicating the student receiving a top grade (A or A*) on the interaction between school subject group dummy and a time indicator for both pandemic years. The regression controls for students' achievement score in a previous standardized national exam (GCSE) in mandatory subjects (mathematics and English literature) and a dummy variable for the subject course of the exam. Standard errors are clustered at the school-subject course level.

Figure 8: Grade inflation by GCSE score by exam subject.

(a) Gender gap in grade inflation by school quality

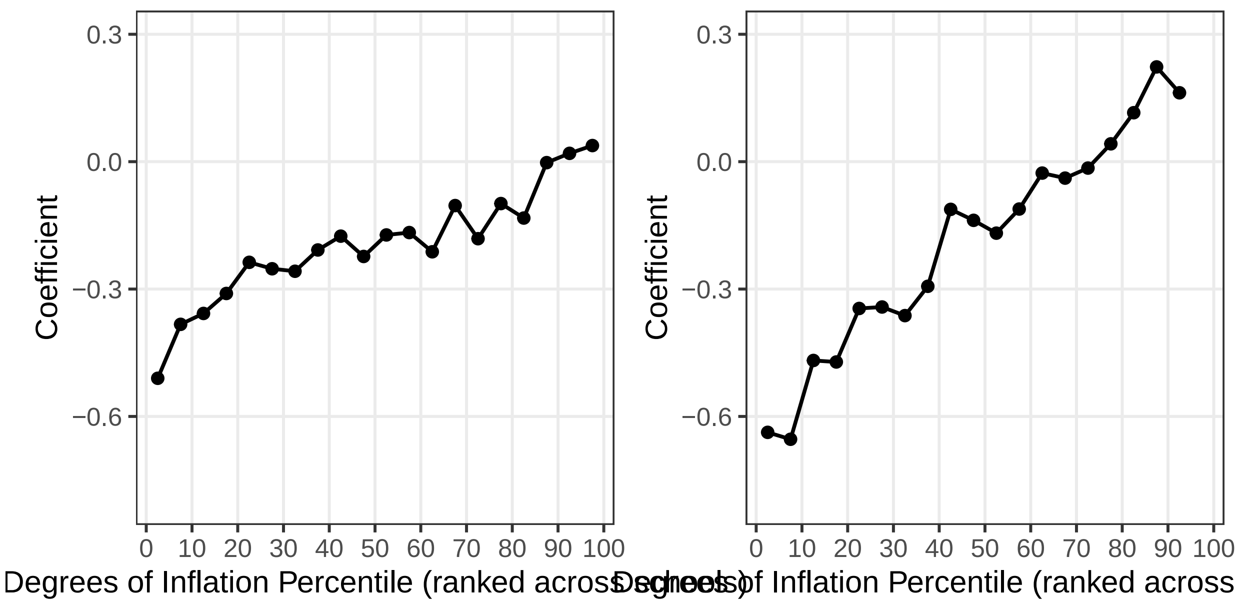


(b) Ethnicity gap in grade inflation by school quality



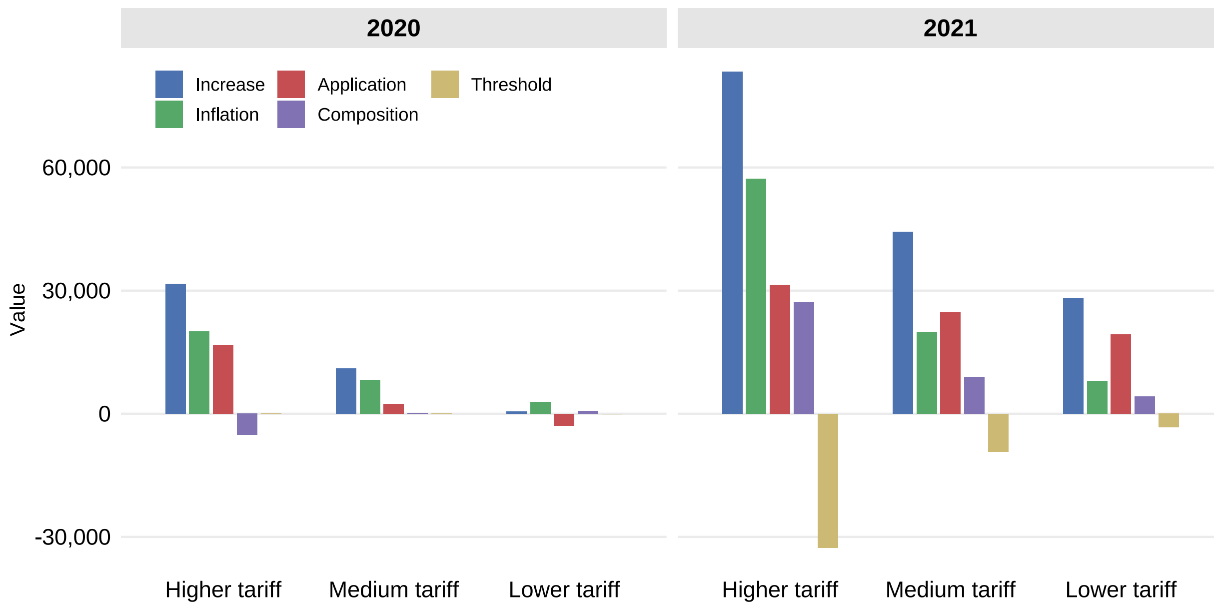
Note: The bin-scatter displays the school level difference in grade improvements between (a)female and male students and (b) while and non-white students. The sample is the universe of A-level exams used for admissions at British universities through the centralized administrative system (N = 2,802,651). Year-specific coefficients and 95% confidence intervals are from a logistic regression that regresses a success dummy indicating the student receiving a top grade (A or A*) on the interaction between school subject group dummy and a time indicator for both pandemic years. The regression controls for students' achievement score in a previous standardized national exam (GCSE) in mandatory subjects (mathematics and English literature) and a dummy variable for the subject course of the exam. Standard errors are clustered at the school-subject course level.

Figure 9: Reduced form result of application success on grade inflation



Note: The figure displays school level improvements in success rate of securing a placement at a students top choice university program versus the extent of grade inflation at the school. The sample is the universe of A-level exams used for admissions at British universities through the centralized administrative system (N = 2,802,651). The y-axis coefficients and 95% confidence intervals are from a student level logistic regression that regresses a success of student application on the interaction between school subject group dummy and a time indicator for both pandemic years. The x-axis coefficients and 95% confidence intervals are from a student level logistic regression that regresses a dummy variable indicating whether a student received a top grade (A or A*) in their exam on the interaction between school subject group dummy and a time indicator for both pandemic years. Both regressions control for students' achievement score in a previous standardized national exam (GCSE) in mandatory subjects (mathematics and English literature) and a dummy variable for the subject course of the exam. Standard errors are clustered at the school-subject course level.

Figure 10: Decomposition of university composition by source



Note: The bin plot displays a decomposition of the increase in the number of incoming student at each university tier group by separate channels. The sample is the universe of A-level exams used for admissions at British universities through the centralized administrative system (N = 2,802,651). Tariff groups denotes the degree of selectivity of the university program, with high tariff indicating the most selective university tier group and the lower tariff indicating the least selective group.

Table 1: Grade inflation and school type

	A		A*		B	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
School Quality	-0.570***	-0.618***	-0.289***	-0.307***	-0.289***	-0.307***
	(0.031)	(0.034)	(0.027)	(0.029)	(0.027)	(0.029)
Further Education	0.139*	-0.136	-0.094***	-0.040	-0.094***	-0.040
	(0.077)	(0.084)	(0.026)	(0.028)	(0.026)	(0.028)
Grammar School	-0.062	0.013	-0.047	0.007	-0.047	0.007
	(0.049)	(0.053)	(0.047)	(0.050)	(0.047)	(0.050)
Independent School	0.082***	0.344***	-0.065***	-0.006	-0.065***	-0.006
	(0.022)	(0.024)	(0.017)	(0.018)	(0.017)	(0.018)
Other	-0.501	-0.124	0.118	-0.567**	0.118	-0.567**
	(0.502)	(0.546)	(0.209)	(0.223)	(0.209)	(0.223)
Sixth Form College	0.026	-0.210***	-0.068***	-0.047*	-0.068***	-0.047*
	(0.036)	(0.039)	(0.021)	(0.023)	(0.021)	(0.023)
State School	0.023	0.024	-0.048***	-0.019	-0.048***	-0.019
	(0.023)	(0.025)	(0.015)	(0.016)	(0.015)	(0.016)
School Quality × school type						
School Quality × Further Education	-0.045	-0.179	0.249*	0.108	0.249*	0.108
	(0.087)	(0.095)	(0.103)	(0.109)	(0.103)	(0.109)
School Quality × Grammar School	-0.087	0.135	0.085	-0.031	0.085	-0.031
	(0.107)	(0.118)	(0.140)	(0.149)	(0.140)	(0.149)
School Quality × Independent School	0.301***	0.222***	0.146**	0.148**	0.146**	0.148**
	(0.055)	(0.060)	(0.049)	(0.053)	(0.049)	(0.053)
School Quality × Other	-1.261	1.237	-0.210	2.075*	-0.210	2.075*
	(2.023)	(2.206)	(0.811)	(0.863)	(0.811)	(0.863)
School Quality × Sixth Form College	0.043	-0.098	0.141*	0.014	0.141*	0.014
	(0.073)	(0.080)	(0.069)	(0.074)	(0.069)	(0.074)
School Quality × State School	0.123*	-0.040	0.154***	0.074	0.154***	0.074
	(0.055)	(0.060)	(0.046)	(0.049)	(0.046)	(0.049)
Year	2020	2021	2020	2021	2020	2021
Num. Obs.	1858	1858	1852	1852	1852	1852
R^2	0.257	0.318	0.099	0.150	0.099	0.150

Notes: The table reports the regression coefficient that regresses the inflation measure defined in Equation (1) on the measure of school quality, the school type dummy variable, and the interaction of the two variables. Degree of inflation is measured as the average marginal effect (AME) at the school level, in which I calculates the average marginal effect of the school inflation for all students within the school. Namely, the latent scores of each students without the inflation effects are calculated by mapping the student attributes on the fixed effect coefficient in Equation (1). Before the AME is calculated, I apply the shrinkage correction method by **?**.School quality is defined as the fixed effect derived from Equation (1). The school fixed effect is evaluated by calculating the probability of obtaining A or A* for the nationally representative student, which I define as the student with median values in their continuous or categorical variables. School fixed effects are corrected for the incidental parameter bias by using methods by Fernández-Val and Weidner (2016) . Schools without less than 30 students in both 2020 and 2021 taking A-level exams are dropped from the regression. Column 1 and 2 uses A, Column 3 and 4 uses A*, and Column 5 and 6 uses B as the binary dependent variable for estimating Equation (1). Bootstrapped standard errors at obtained by blocking at the school level. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 2: Regression results: Treatment estimates by subject type

Non-Quantitative subjects

	A		A*		B	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
School Quality	-0.598***	-0.729***	-0.523***	-0.538***	-0.062**	-0.111***
	(0.032)	(0.033)	(0.027)	(0.031)	(0.021)	(0.021)
Year	2020	2021	2020	2021	2020	2021
Num. Obs.	1887	1889	1887	1889	1879	1876
R^2	0.155	0.208	0.163	0.141	0.004	0.014

Quantitative subjects

	A		A*		B	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
School Quality	-0.200***	-0.220***	0.018	0.006	-0.001	-0.071***
	(0.027)	(0.034)	(0.023)	(0.022)	(0.021)	(0.021)
Year	2020	2021	2020	2021	2020	2021
Num. Obs.	1887	1889	1879	1876	1879	1876
R^2	0.029	0.022	0.000	0.000	0.000	0.006

Notes: The table reports the regression coefficient that regresses the inflation measure defined in Equation (1) on the measure of school quality, the school type dummy variable, and the interaction of the two variables. Degree of inflation is measured as the average marginal effect (AME) at the school level, in which I calculates the average marginal effect of the school inflation for all students within the school. Namely, the latent scores of each students without the inflation effects are calculated by mapping the student attributes on the fixed effect coefficient in Equation (1). Before the AME is calculated, I apply the shrinkage correction method by **?**.School quality is defined as the fixed effect derived from Equation (1). The school fixed effect is evaluated by calculating the probability of obtaining A or A* for the nationally representative student, which I define as the student with median values in their continuous or categorical variables. School fixed effects are corrected for the incidental parameter bias by using methods by Fernández-Val and Weidner (2016) . Schools without less than 30 students in both 2020 and 2021 taking A-level exams are dropped from the regression. Column 1 and 2 uses A, Column 3 and 4 uses A*, and Column 5 and 6 uses B as the binary dependent variable for estimating Equation (1). Bootstrapped standard errors at obtained by blocking at the school level. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3: Descriptive regression on persistence of grade inflation

Panel A: Dynamics between 2021 and 2020

	A	**A***	**B**
	(1)	**(2)**	**(3)**
Inflation in 2020	0.624***	0.454***	1.018***
	(0.022)	(0.024)	(0.009)
Num. Obs.	1890	1869	1869
R^2	0.302	0.157	0.875

Panel B: Decomposition by subject type

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
	Q	**NQ**	**Q**	**NQ**	**Q**	**NQ**
Inflation in 2020 + Non-Quant	0.036**	0.507***	-0.023	0.296***	0.287***	0.537***
	(0.016)	(0.023)	(0.033)	(0.023)	(0.022)	(0.018)
Inflation in 2020 + Quant	0.225***	-0.020	0.393***	-0.007	0.744***	0.311***
	(0.023)	(0.034)	(0.025)	(0.017)	(0.019)	(0.016)
Num. Obs.	1863	1869	1846	1852	1852	1846
R^2	0.047	0.210	0.100	0.141	0.695	0.649

Notes: The table reports the regression coefficient that regresses the inflation measure defined in Equation (1) on the measure of school quality, the school type dummy variable, and the interaction of the two variables. Degree of inflation is measured as the average marginal effect (AME) at the school level, in which I calculates the average marginal effect of the school inflation for all students within the school. Namely, the latent scores of each students without the inflation effects are calculated by mapping the student attributes on the fixed effect coefficient in Equation (1). Before the AME is calculated, I apply the shrinkage correction method by **?**.School quality is defined as the fixed effect derived from Equation (1). The school fixed effect is evaluated by calculating the probability of obtaining A or A* for the nationally representative student, which I define as the student with median values in their continuous or categorical variables. School fixed effects are corrected for the incidental parameter bias by using methods by Fernández-Val and Weidner (2016) . Schools without less than 30 students in both 2020 and 2021 taking A-level exams are dropped from the regression. Column 1 and 2 uses A, Column 3 and 4 uses A*, and Column 5 and 6 uses B as the binary dependent variable for estimating Equation (1). Bootstrapped standard errors at obtained by blocking at the school level. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 4: Grade inflation by demographic groups

	A		A*		B	
	(1)	(2)	(3)	(4)	(5)	(6)
SES (ref. Elementary occupations.)						
Managers, directors and senior officials	0.190***	0.011	0.171***	0.149***	0.177***	0.150***
	(0.013)	(0.012)	(0.012)	(0.012)	(0.014)	(0.017)
Professional occupations	0.173***	0.024	0.134***	0.105***	0.181***	0.128***
	(0.012)	(0.012)	(0.016)	(0.018)	(0.023)	(0.029)
Associate professional and technical	0.190***	0.032	0.179***	0.135***	0.267***	0.168***
	(0.012)	(0.016)	(0.022)	(0.022)	(0.034)	(0.035)
Administrative and secretarial	0.161***	-0.016	0.166***	0.087***	0.178***	0.028
	(0.019)	(0.020)	(0.028)	(0.013)	(0.042)	(0.015)
Skilled trades occupations	0.264***	0.028	0.075***	0.054***	0.064***	0.069***
	(0.023)	(0.024)	(0.012)	(0.016)	(0.016)	(0.023)
Caring, leisure and other service occupations	0.175***	-0.029	0.021	0.046*	0.039	0.048
	(0.023)	(0.023)	(0.019)	(0.023)	(0.028)	(0.034)
Sales and customer service occupations	0.206***	-0.019	0.002	0.029	0.017	0.040
	(0.029)	(0.028)	(0.022)	(0.028)	(0.033)	(0.039)
Year	2020	2021	2020	2021	2020	2021
Number of Obs.	2,802,651	2,802,651	2,802,651	2,802,651	2,802,651	2,802,651

Notes: The table reports the regression coefficient that regresses the inflation measure defined in Equation (1) on the measure of school quality, the school type dummy variable, and the interaction of the two variables. Degree of inflation is measured as the average marginal effect (AME) at the school level, in which I calculates the average marginal effect of the school inflation for all students within the school. Namely, the latent scores of each students without the inflation effects are calculated by mapping the student attributes on the fixed effect coefficient in Equation (1). Before the AME is calculated, I apply the shrinkage correction method by **?**.School quality is defined as the fixed effect derived from Equation (1). The school fixed effect is evaluated by calculating the probability of obtaining A or A* for the nationally representative student, which I define as the student with median values in their continuous or categorical variables. School fixed effects are corrected for the incidental parameter bias by using methods by Fernández-Val and Weidner (2016) . Schools without less than 30 students in both 2020 and 2021 taking A-level exams are dropped from the regression. Column 1 and 2 uses A, Column 3 and 4 uses A*, and Column 5 and 6 uses B as the binary dependent variable for estimating Equation (1). Bootstrapped standard errors at obtained by blocking at the school level. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 5: Treatment interaction estimates by demographic groups

	A		A*		B	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel B: POLAR Quintiles (ref: Quintile 5)						
Quintile 1	0.019	-0.003	0.064^{***}	0.032^{*}	0.030	0.003
	(0.018)	(0.019)	(0.017)	(0.018)	(0.027)	(0.028)
Quintile 2	0.036^{**}	-0.003	0.066^{***}	0.020	0.033	0.004
	(0.016)	(0.016)	(0.015)	(0.016)	(0.021)	(0.022)
Quintile 3	–	0.021	0.037^{***}	0.023	0.041^{**}	0.013
		(0.015)	(0.014)	(0.015)	(0.019)	(0.020)
Quintile 4	-0.019	0.015	0.018	0.032^{*}	-0.020	0.001
	(0.017)	(0.017)	(0.018)	(0.018)	(0.022)	(0.022)
Panel C: IMD Quintiles (ref: Quintile 5)						
Quintile 1	0.013	-0.081	-0.014	-0.091	0.041	-0.046
	(0.015)	(0.016)	(0.014)	(0.015)	(0.022)	(0.024)
Quintile 2	0.008	-0.066	-0.026	-0.058	0.061	-0.033
	(0.014)	(0.014)	(0.014)	(0.014)	(0.019)	(0.020)
Quintile 3	0.018	-0.040	-0.024	-0.033	0.039	-0.020
	(0.013)	(0.014)	(0.014)	(0.014)	(0.018)	(0.018)
Quintile 4	0.010	-0.035	-0.011	-0.049	0.021	-0.007
	(0.016)	(0.016)	(0.017)	(0.017)	(0.020)	(0.020)
Panel A: Gender						
Female × Treat	0.055^{***}	0.069^{***}	0.041^{***}	0.062^{***}	0.063^{***}	0.079^{***}
	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.003)
Panel B: Ethnicity (ref: White)						
Asian	-0.037^{***}	0.040^{***}	-0.082^{***}	0.031	-0.034^{**}	0.005
	(0.013)	(0.014)	(0.018)	(0.019)	(0.013)	(0.014)
Black	-0.004	-0.024	-0.018	-0.026	-0.067^{***}	-0.073^{***}
	(0.020)	(0.021)	(0.032)	(0.034)	(0.018)	(0.019)
Mixed	-0.029	-0.063^{***}	-0.063^{**}	-0.066^{**}	-0.048^{**}	-0.077^{***}
	(0.019)	(0.020)	(0.026)	(0.026)	(0.020)	(0.020)
Other	0.005	0.053	0.048	0.127^{***}	-0.015	0.006
	(0.031)	(0.033)	(0.045)	(0.047)	(0.030)	(0.031)
Unknown / Prefer not to say	0.069^{*}	-0.065^{*}	0.243^{***}	0.109^{**}	-0.037	-0.082^{*}
	(0.037)	(0.037)	(0.043)	(0.043)	(0.041)	(0.042)
Year	2020	2021	2020	2021	2020	2021
Number of Obs.	2,802,651	2,802,651	2,802,651	2,802,651	2,802,651	2,802,651

Notes: Columns (1)-(2)=A, (3)-(4)=AS, (5)-(6)=B. Entries are coefficients; standard errors are in parentheses. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Baseline level and fixed effects omitted.

A Appendix

A.0.1 Application-Adjusted Assignment with Demographic-Specific Preferences

The second approach allows students to adjust their application choices in response to grade changes. Preferences over universities vary systematically by observable demographic characteristics. These preferences are estimated using pre-COVID application data within a multinomial logit (MNL) framework. Specifically, the probability that student i applies to and attends university j is given by:

$$P_{ij} = \frac{e^{V_{ij}}}{\sum_{m \in \mathcal{J}_i} e^{V_{im}}}, \quad V_{ij} = \beta \times S_i \times \overline{S}_{j,2019} + \iota_i + \delta_j,$$

where δ_j is a university fixed effect, ι_i is a student fixed effect, $S_i = \sum_k w_k G_{ik}$ is the weighted sum of grades for student i, and $\overline{S}_{j,2019}$ is the mean S_i among students admitted to j in 2019. The choice set $\mathcal{J}_i$ consists of all universities to which student i could plausibly apply.

The student–university ranking under this approach is

$$Rank(X_{i,j}, u) = F\big(G(X_{i,j}) \,\big|\, G(X_{-i,j}), u\big),$$

with assignments determined by

$$\mathcal{U}_u = \left\{ X_{i,j} \,\middle|\, \sum_{Q_k > Q_u} C_k \leq Rank(X_{i,j}, u) \leq \sum_{Q_k > Q_u} C_k + C_u \right\}.$$

Endogenous Application Mechanism. Under this model, the effect of grade changes on university composition arises through two channels: (i) the direct reordering of students in the admissions ranking due to changes in $G(X_{i,j})$, and (ii) the induced change in application probabilities P_{ij} through the MNL utility specification. By estimating β, ι_i, and δ_j using pre-COVID data, I recover the structural relationship between student characteristics, relative grade position, and application behavior. I then simulate counterfactual application sets $\mathcal{J}_i$ under alternative grade distributions, re-solving the assignment problem to capture both ranking and choice adjustments.

A.1 Estimating counterfactual grades for the 2020–21 cohorts

I obtain counterfactual grades for students in the pandemic years by training a prediction rule on the *pre-pandemic* cohorts and then applying that rule to the pandemic cohorts. The idea is simple: estimate the relationship between observable student features (GCSEs, prior attainment, demographics, subject and school indicators, etc.) and realised grades before the pandemic, and then use that relationship to predict what each pandemic-year student would have received had the pre-pandemic mapping continued to hold.

A plain ordinary least squares estimator (a maximum likelihood estimator for a linear model) performs poorly in this setting for two reasons: (i) the predictor set is large (many dummies, interactions and polynomial terms), which makes a plain regression prone to *overfitting*, and (ii) many predictors are highly correlated, producing unstable coefficient estimates. To address both issues I use *elastic net* regularisation, which shrinks coefficients and — when helpful — sets some coefficients exactly to zero.

The elastic-net estimator is

$$\hat{\beta}_{\lambda,\xi} = \arg\min_{\beta}; \frac{1}{2n_{\text{pre}}} \sum_{i\in\text{pre}} \big(y_i - X_i^\top \beta\big)^2$$

$\lambda\Big(\frac{1-\xi}{2}, |\beta|_2^2 + \xi, |\beta|_1\Big)$, (11) where y_i is the (numeric) grade for pre-pandemic student i, X_i is the vector of predictors, $\lambda > 0$ controls the overall strength of regularisation (larger λ produces stronger shrinkage), and $\xi \in [0,1]$ mixes between ridge ($\xi = 0$) and lasso ($\xi = 1$) behaviour.

I select the tuning parameter λ by K-fold cross-validation. Split the pre-pandemic sample into K folds, estimate the penalised problem on $K-1$ folds and measure prediction error on the held-out fold; repeat for each fold and choose the λ that minimises the average held-out mean-squared error:

$$\lambda^\star = \arg\min_{\lambda}; \frac{1}{K} \sum_{k=1}^{K} \frac{1}{n_k} \sum_{i\in\text{fold}k} \big(y_i - X_i^\top \hat{\beta}\lambda, -k\big)^2, \tag{12}$$

where $\hat{\beta}_{\lambda,-k}$ is the estimator obtained without fold k.

With $\lambda^\star$ fixed, the counterfactual continuous score for a pandemic-year student with covariates X_i is

$$\widehat{y}i^{\text{CF}} = X_i^\top \hat{\beta}\lambda^\star, \xi. \tag{13}$$

If desired, this continuous prediction can be mapped back to discrete grade categories using empirically estimated cutoffs (for example, the observed thresholds between letter grades in the pre-pandemic data). Alternatively, one may estimate a multinomial or ordinal penalised model directly and obtain predicted grade probabilities; the two approaches are broadly equivalent in spirit, and the choice depends on whether you want direct probability estimates or a continuous latent score.

Practical implementation notes

- **Standardisation.** Elastic-net penalties are scale-dependent. Standardise continuous predictors (zero mean and unit variance) before estimation (most software does this automatically).
- **Categorical variables.** Convert factors to dummies (one-hot encoding) but be mindful of huge design matrices (e.g. subject×school interactions). Elastic net handles large p but runtime and memory still matter.
- **Cross-validation folds.** If observations within schools are correlated, form CV folds that respect clustering (e.g. fold by school) so the CV error is not overly optimistic.
- **Choosing ξ.** Try a few values (e.g. $\xi \in 0, 0.5, 1$) and report sensitivity. In practice $\xi \in (0,1)$ often combines the stability of ridge with the sparsity of lasso.

A.2 Estimating school-level effects with Empirical Bayes

With many schools and uneven sample sizes, raw MLEs for school effects are noisy in small samples. Empirical Bayes (EB) "borrows strength" across schools by shrinking imprecise estimates toward a common mean, while leaving precise estimates largely unchanged. The amount of shrinkage is data-driven: noisier schools shrink more.

Setup. Let $\widehat{\alpha}_j$ be the MLE of the school effect α_j from Equation (1) with estimated standard error $s_j \equiv \widehat{\mathrm{SE}}(\widehat{\alpha}_j)$. Under standard regularity conditions,

$$\widehat{\alpha}_j \mid \alpha_j \approx \mathcal{N}(\alpha_j,\, s_j^2).$$

EB posits a *working* prior (random-effects distribution) for the true school effects

$$\alpha_j \overset{\text{iid}}{\sim} \mathcal{N}(\mu,\, \tau^2),$$

with unknown hyperparameters (μ, τ^2) estimated from the collection $\{\widehat{\alpha}_j, s_j^2\}_{j=1}^J$.

Hyperparameter estimation (marginal MLE). Integrating out α_j yields the marginal sampling model

$$\widehat{\alpha}_j \sim \mathcal{N}\big(\mu,\, \tau^2 + s_j^2\big).$$

The marginal log-likelihood is

$$\ell(\mu, \tau^2) = -\tfrac{1}{2} \sum_{j=1}^{J} \Big\{ \log(\tau^2 + s_j^2) + \tfrac{(\widehat{\alpha}_j - \mu)^2}{\tau^2 + s_j^2} \Big\}.$$

The maximizer satisfies

$$\widehat{\mu} = \frac{\sum_j w_j\, \widehat{\alpha}_j}{\sum_j w_j}, \qquad w_j \equiv \frac{1}{\widehat{\tau}^2 + s_j^2},$$

and $\widehat{\tau}^2$ solves

$$\sum_{j=1}^{J} \frac{(\widehat{\alpha}_j - \widehat{\mu})^2}{(\widehat{\tau}^2 + s_j^2)^2} = \sum_{j=1}^{J} \frac{1}{\widehat{\tau}^2 + s_j^2},$$

which can be obtained by a short 1D Newton or grid search.[10]

EB shrinkage estimator (posterior mean). Given $(\widehat{\mu}, \widehat{\tau}^2)$, the EB estimate of α_j is the posterior mean under the normal–normal model:

$$\widehat{\alpha}_j^{EB} = (1 - B_j)\, \widehat{\alpha}_j \; + \; B_j\, \widehat{\mu}, \qquad B_j \equiv \frac{s_j^2}{s_j^2 + \widehat{\tau}^2}.$$

Equivalently,

$$\widehat{\alpha}_j^{EB} = \frac{\widehat{\tau}^2}{\widehat{\tau}^2 + s_j^2}\, \widehat{\alpha}_j \; + \; \frac{s_j^2}{\widehat{\tau}^2 + s_j^2}\, \widehat{\mu}.$$

[10] A nonnegative constraint on τ^2 is imposed; if the MLE hits $\widehat{\tau}^2 = 0$, no cross-school variation is detected and EB reduces to pooling at $\widehat{\mu}$.

Schools with large s_j^2 (imprecise MLEs) have B_j closer to 1 and hence shrink more toward $\widehat{\mu}$; precise schools (small s_j^2) are shrunk little.

The corresponding EB posterior variance (useful for uncertainty display) is

$$\widehat{\mathrm{Var}}(\alpha_j \mid \widehat{\alpha}_j) = \frac{s_j^2\,\widehat{\tau}^2}{s_j^2 + \widehat{\tau}^2}.$$

Treatment-year school effects. For treatment-year deviations $\Delta\alpha_{j,\tau}$ with MLEs $\widehat{\Delta\alpha}_{j,\tau}$ and errors $s_{j,\tau}$, apply the same steps with a potentially different hyper-variance $\tau^2_{\Delta,\tau}$ and mean $\mu_{\Delta,\tau}$:

$$\widehat{\Delta\alpha}^{EB}_{j,\tau} = (1 - B_{j,\tau})\,\widehat{\Delta\alpha}_{j,\tau} \; + \; B_{j,\tau}\,\widehat{\mu}_{\Delta,\tau}, \quad B_{j,\tau} \equiv \frac{s^2_{j,\tau}}{s^2_{j,\tau} + \widehat{\tau}^2_{\Delta,\tau}}.$$

Remarks. (i) EB uses plug-in $(\widehat{\mu}, \widehat{\tau}^2)$, so intervals based solely on the posterior variance above understate hyperparameter uncertainty; in large J this is typically minor. (ii) Centering at $\widehat{\mu}$ (rather than 0) avoids implicit assumptions that the grand mean effect is zero. (iii) The gain is variance reduction for noisy schools with limited bias, which is crucial with many parameters and heterogeneous cell sizes.

A.3 Computation of variance in full model

Computation of variances for the full model. Computing standard errors in our high-dimensional, sparse likelihood required estimating the diagonal of the observed-information matrix H^{-1} (the parameter variances). A direct approach (form H^{-1} or compute a full Cholesky factor) proved infeasible: sparse Cholesky factorization produced excessive fill-in and exhausted memory for our problem size. To avoid factorization and the attendant memory costs we combine an iterative linear solver (Conjugate Gradient, CG) with the Hutchinson trace estimator. Concretely, for m independent Rademacher probe vectors $z^{(r)} \in \{\pm 1\}^n$ we solve $Hx^{(r)} = z^{(r)}$ by CG and accumulate the elementwise products $z^{(r)} \odot x^{(r)}$; the sample average is an unbiased estimator of $\mathrm{diag}(H^{-1})$. This approach only requires sparse matrix–vector products and a small working set of vectors, so its memory use scales with the number of nonzeros of H (and not with the fill-in of a factor). For reproducibility we implement the CG solver using Eigen's sparse CG via `RcppEigen`, run $m = 200$ probes, set the solver tolerance to $\mathrm{tol} = 10^{-8}$ and cap iterations at $\mathrm{maxit} = 2000$. Practical diagnostics include: (i) verify H is stored as a `dgCMatrix` and is numerically symmetric/positive definite (force symmetry and add a tiny ridge if needed); (ii) monitor CG residuals for each probe and increase tol or add an incomplete-Cholesky preconditioner if convergence is slow; (iii) choose m by trading off accuracy and CPU (typical $m \in [100, 500]$); and (iv) validate the estimator on a small index set by comparing Hutchinson estimates to exact variances computed by direct solves for those indices when feasible. These choices deliver accurate, scalable variance estimates without ever forming H^{-1} or a full factor, and the implementation details and minimal code used in this paper are provided below for reproducibility.[11]

[11] A compact description of this procedure used for computing the appendix variances is given in the original implementation notes.

Figure A1: Timeline for university application in the UK

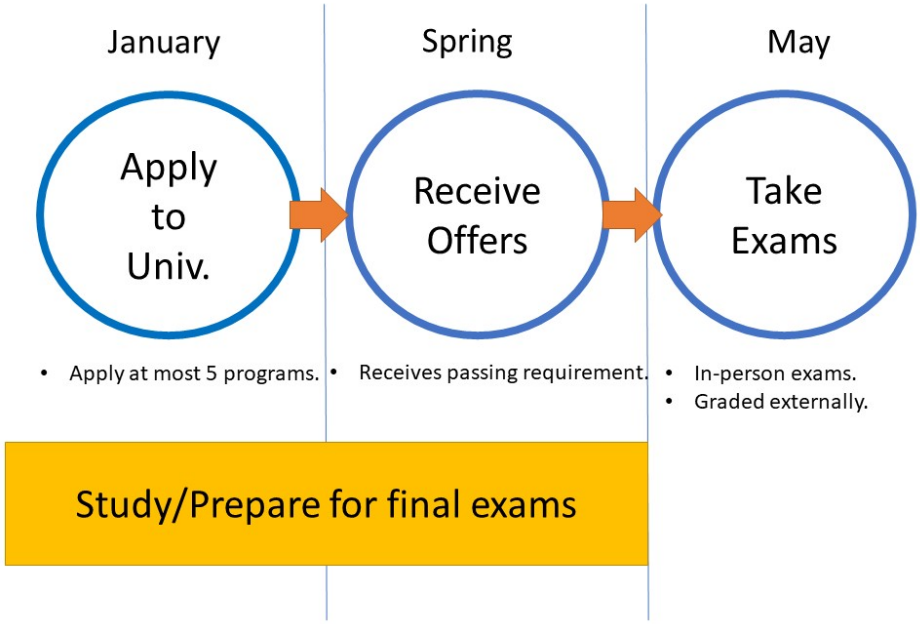
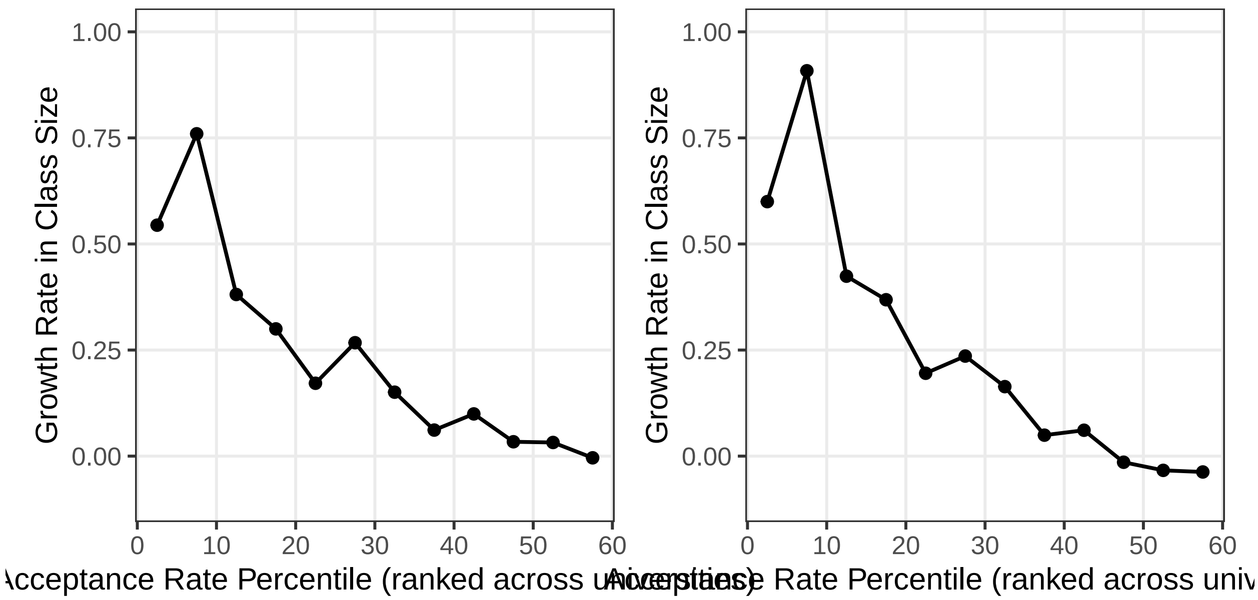


Figure A2: Cohort Size Growth



(a) Cohort Size Firm Growth

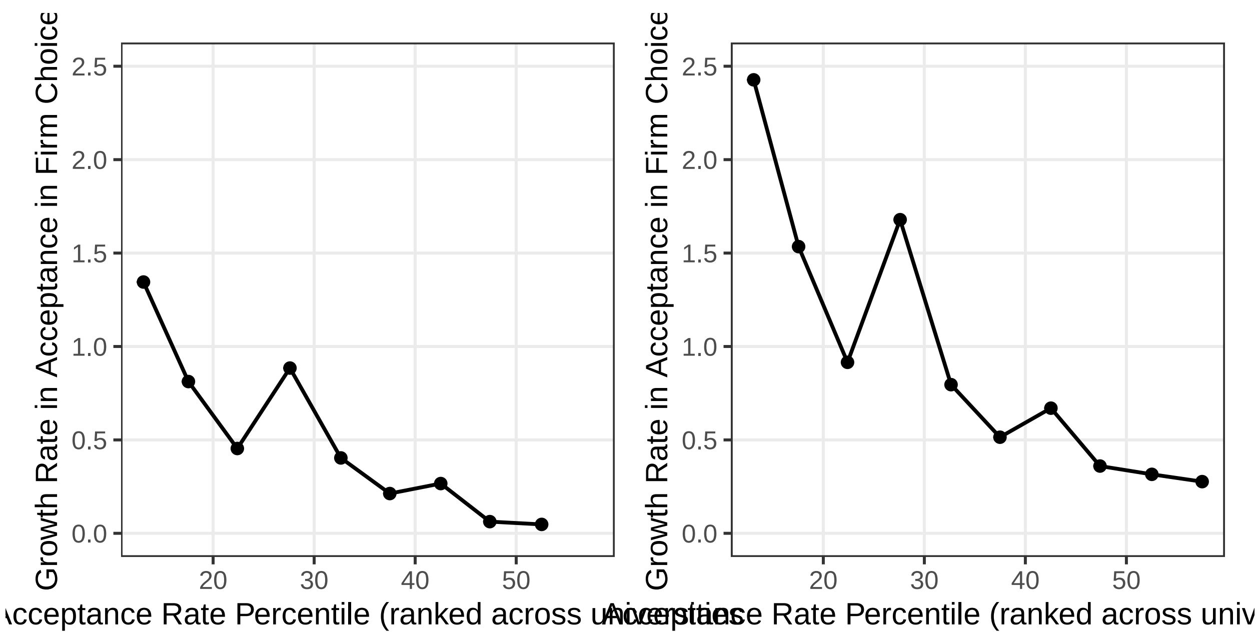
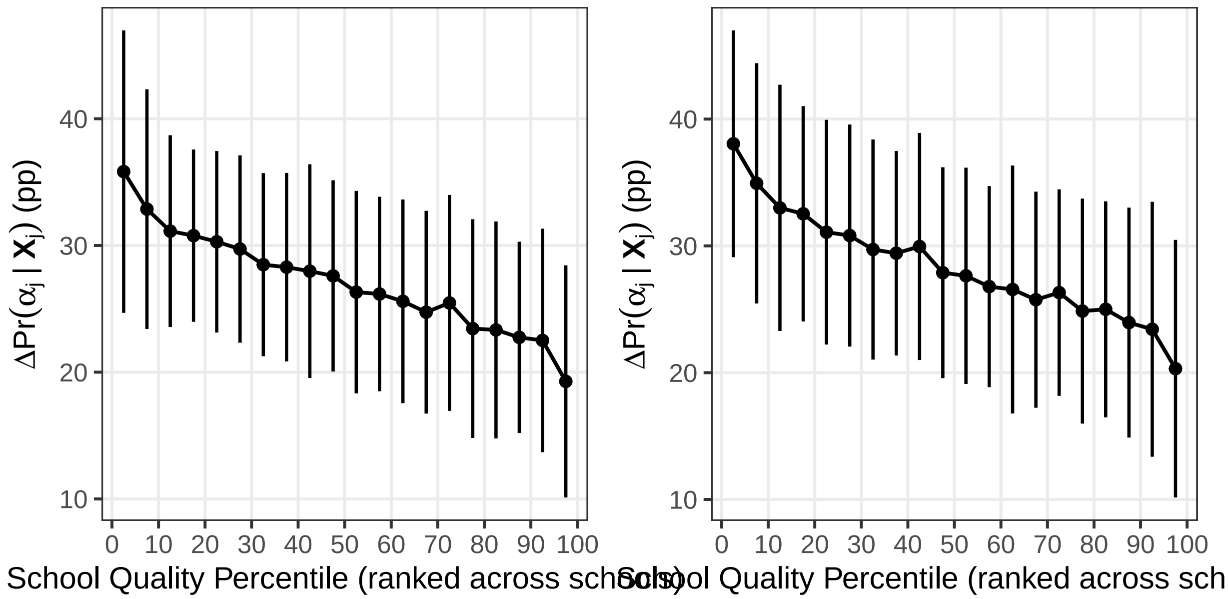


Figure A3: Grade inflation by school quality (2020, 2021).



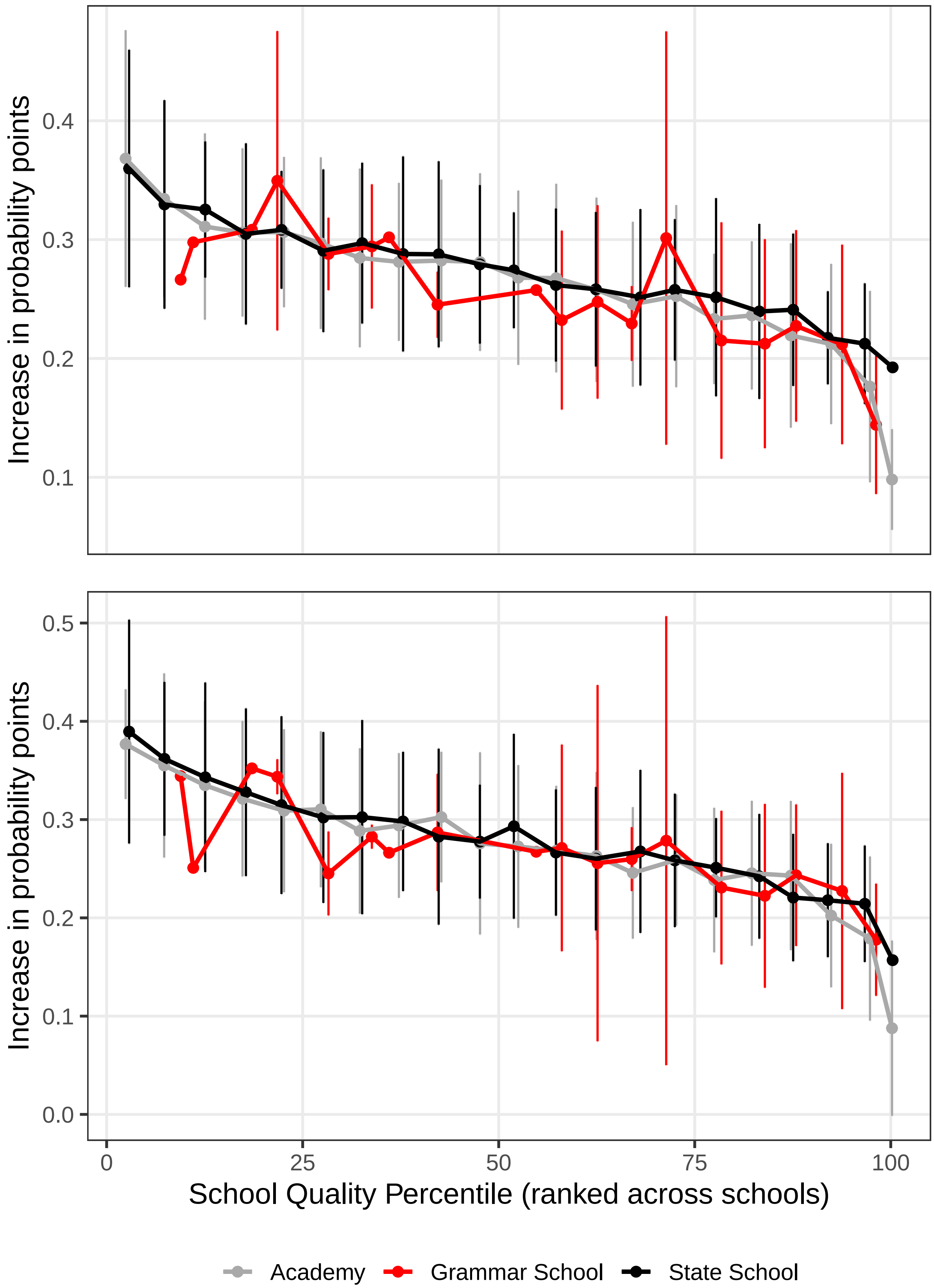


Figure A4: Grade inflation at grammar schools

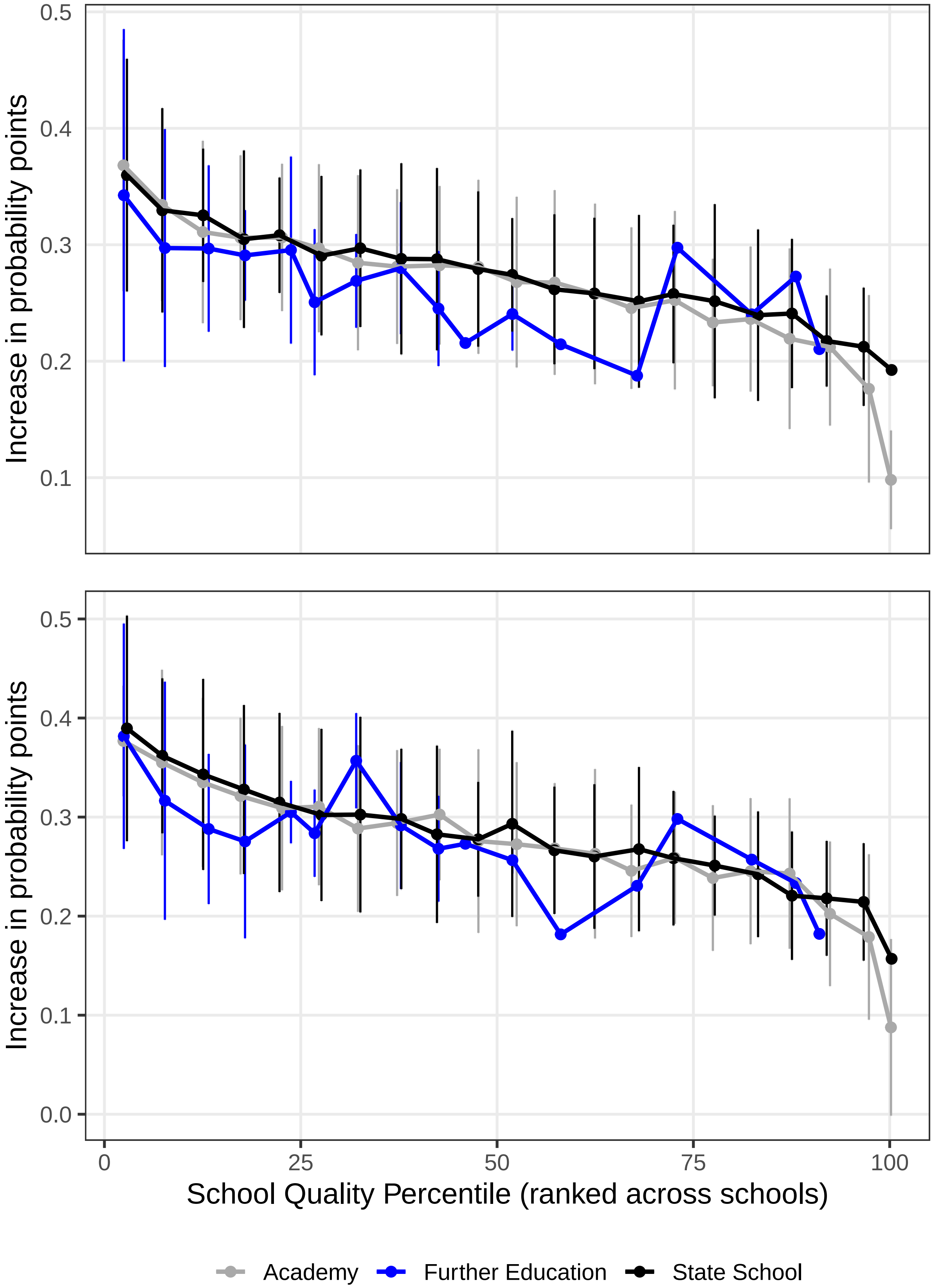
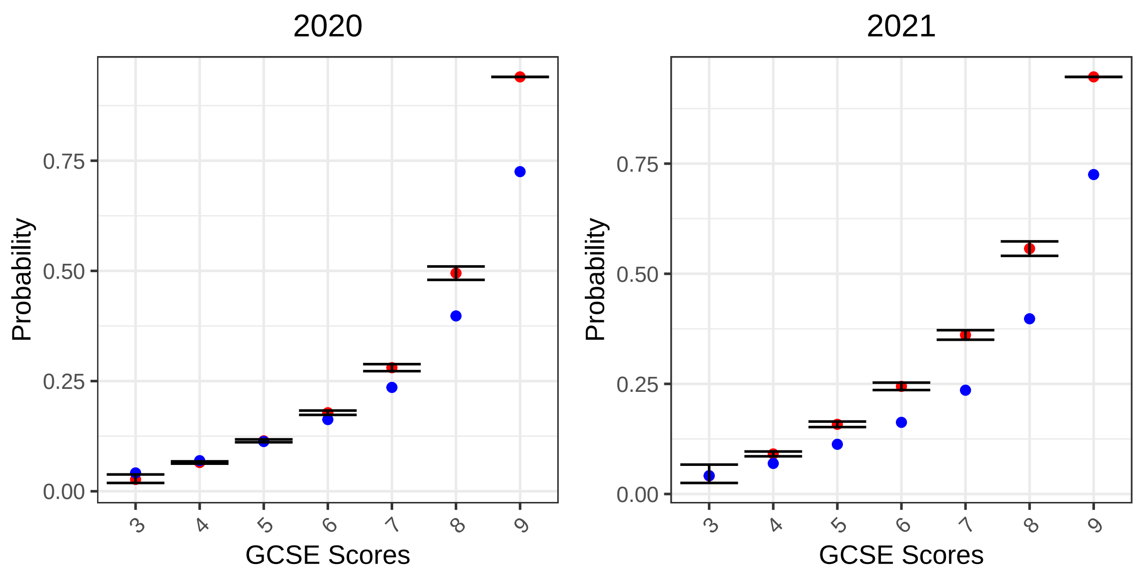
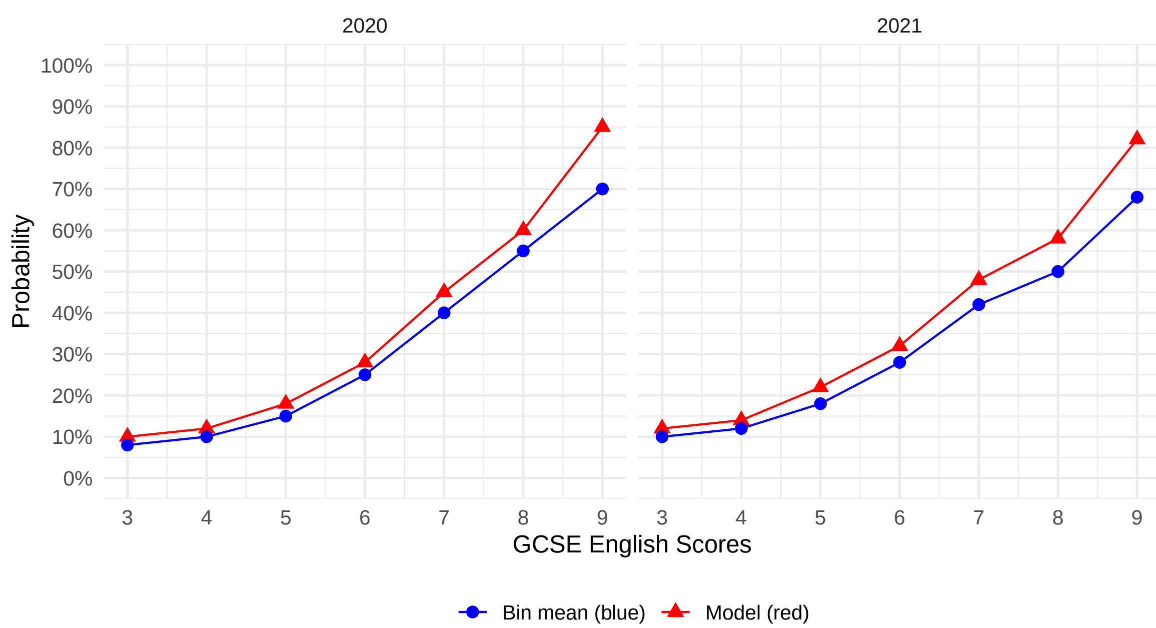


Figure A5: Grade inflation at further education schools

Figure A6: Grade inflation by GCSE score by exam subject.

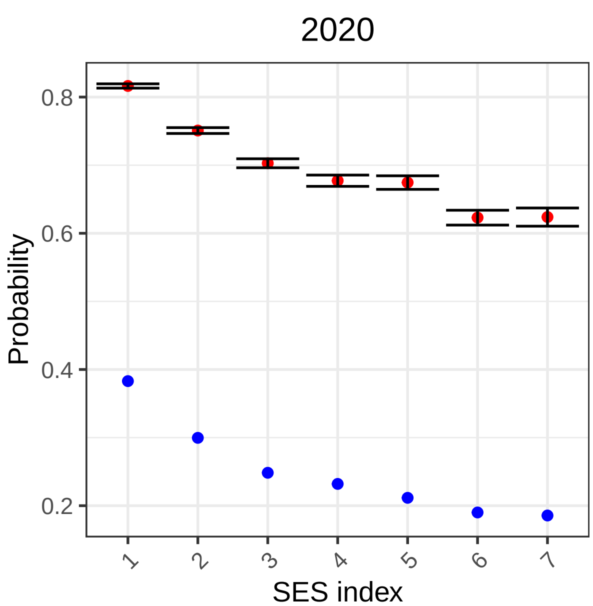


(a) Grade inflation by GCSE score in Mathematics



(b) Grade inflation by GCSE score in English Literature.

Figure A7: Inflation by parental occupation class



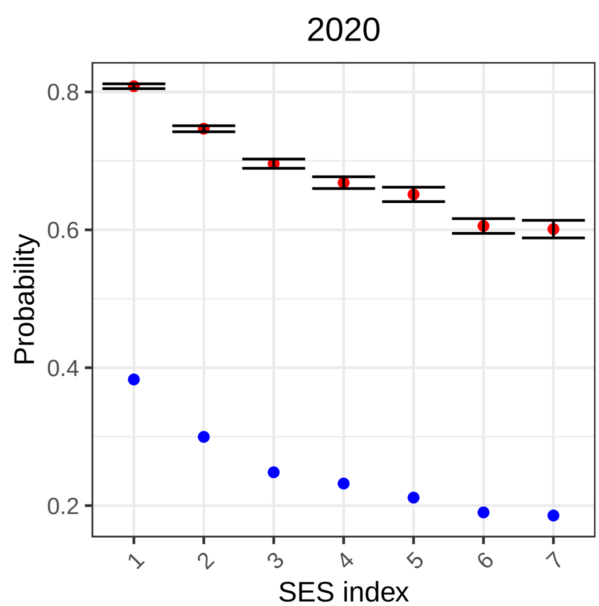
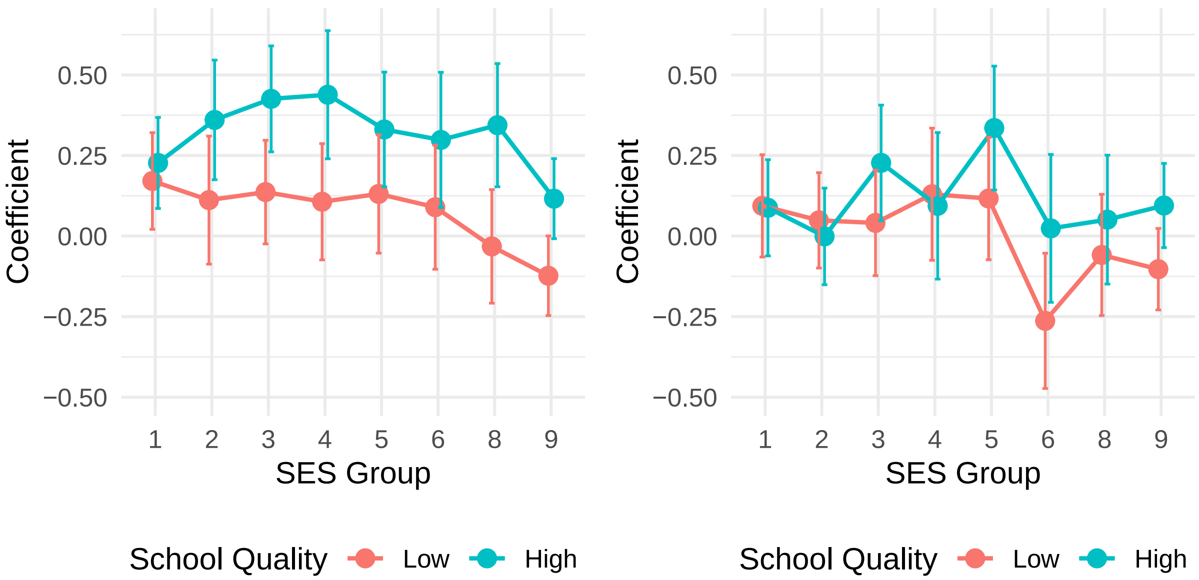
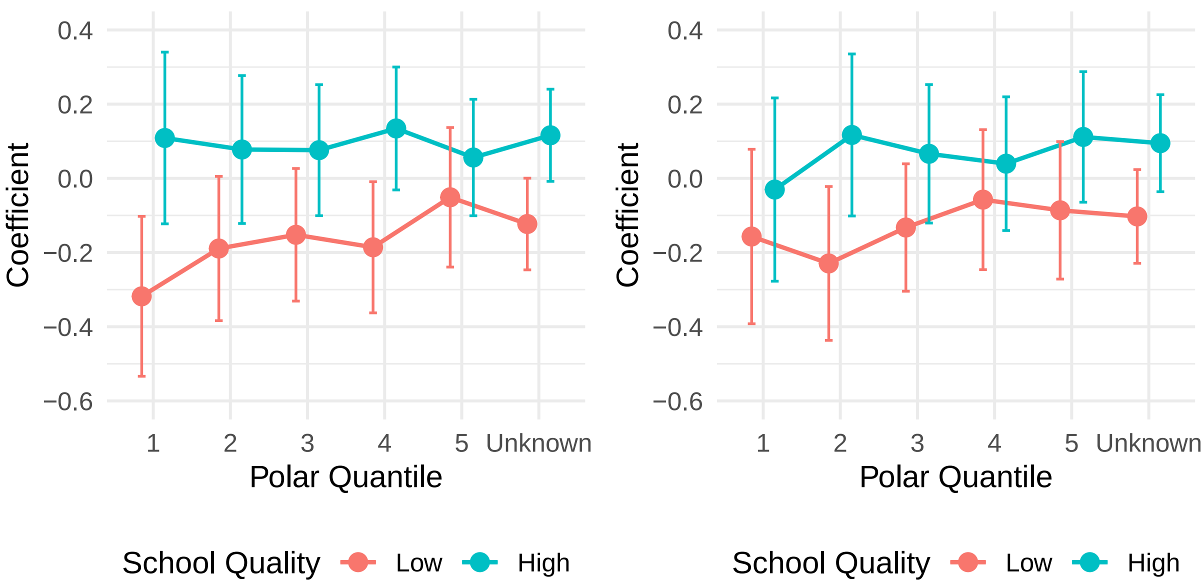


Figure A8: Grade inflation by (a) Parental occupation class, and (b) Univ. progression rate

(a)



(b)



Note: The bin-scatter displays student level difference in grade improvements across (a) parental occupational class and (b) average university progression rate of the students' residential address. The sample is the universe of A-level exams used for admissions at British universities through the centralized administrative system (N = 2,802,651). Year-specific coefficients and 95% confidence intervals are from a logistic regression that regresses a success dummy indicating the student receiving a top grade (A or A*) on the interaction between school subject group dummy and a time indicator for both pandemic years. The regression controls for students' achievement score in a previous standardized national exam (GCSE) in mandatory subjects (mathematics and English literature) and a dummy variable for the subject course of the exam. Standard errors are clustered at the school-subject course level.

Figure A9: Inflation by poverty score score

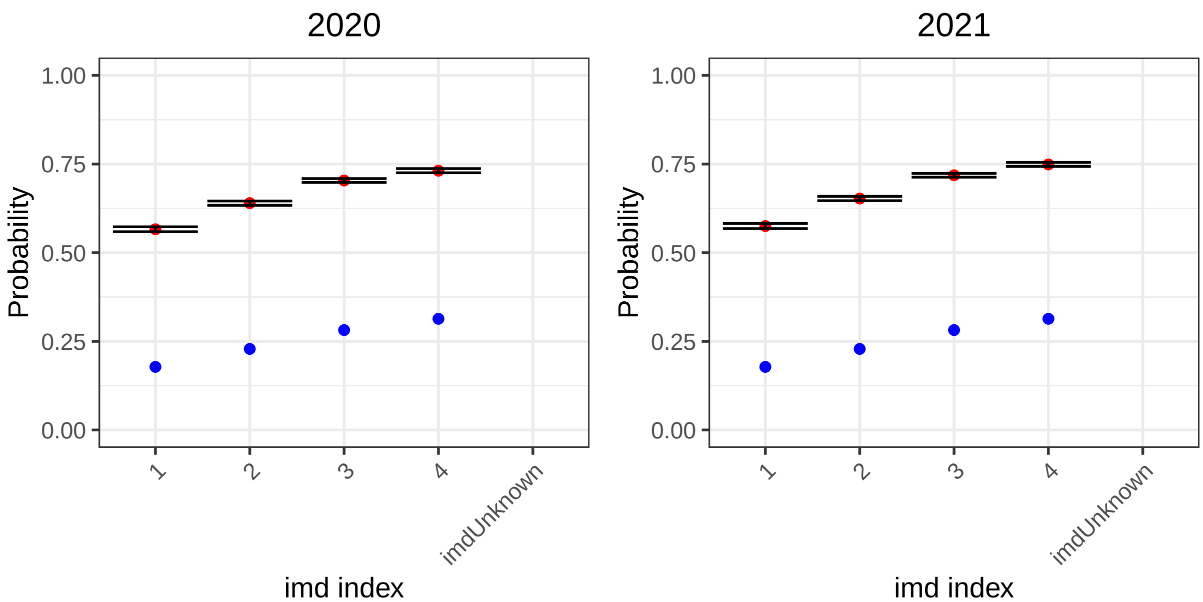


Figure A10: Inflation by GCSE scores measured by log-odds

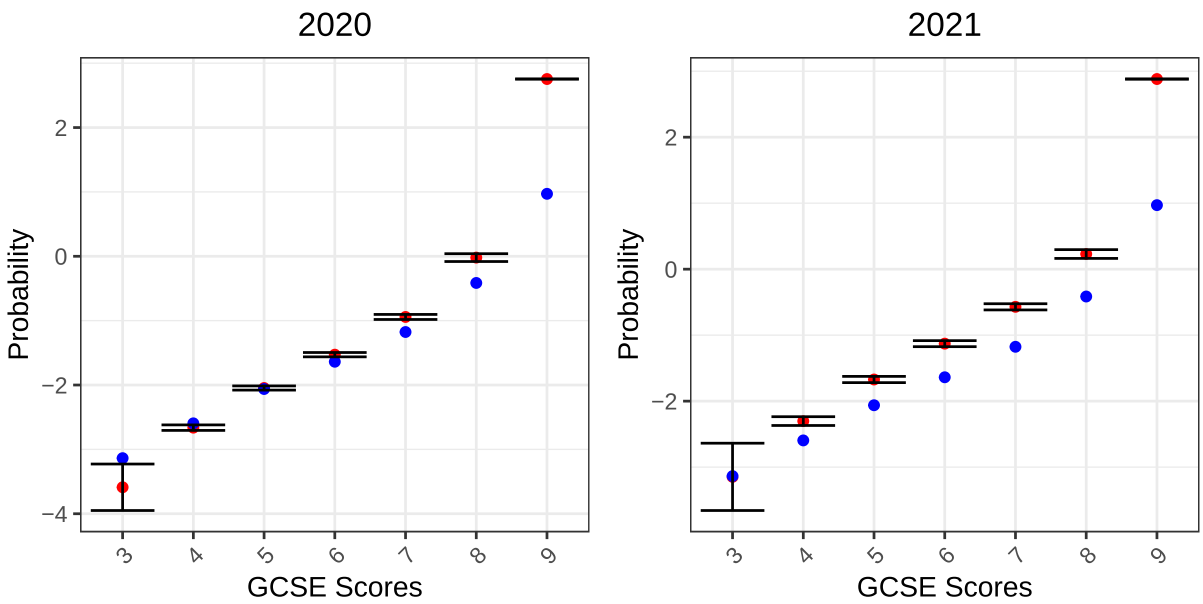


Figure A11: Inflation by GCSE scores by school quality

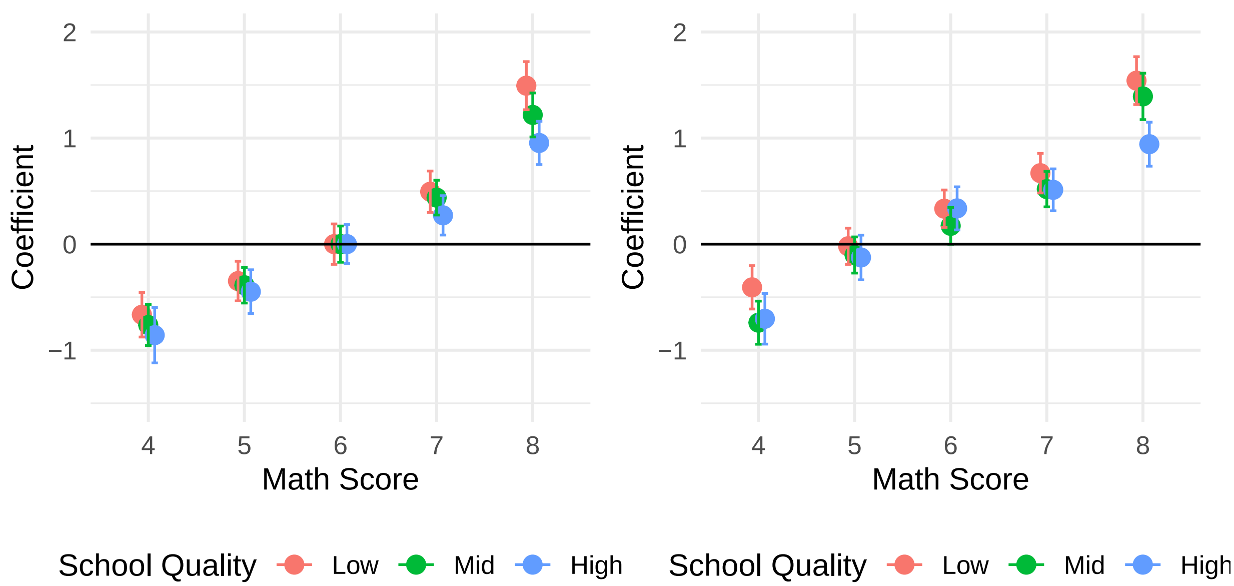


Figure A12: Inflation by GCSE scores and schools

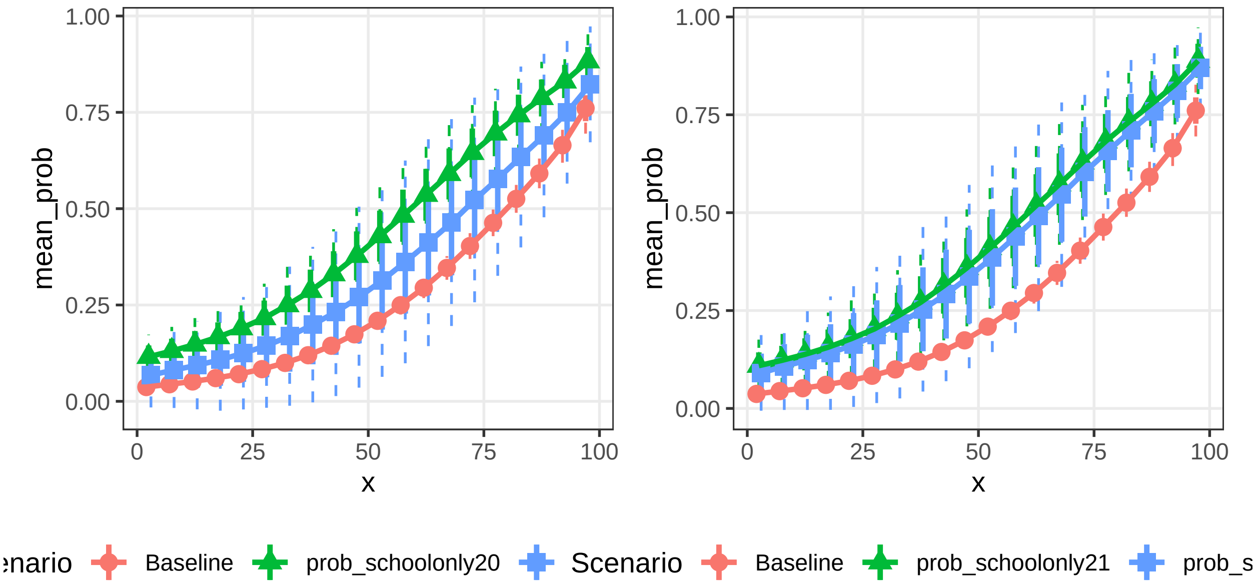


Figure A13: Effect of inflation on placements in 2020

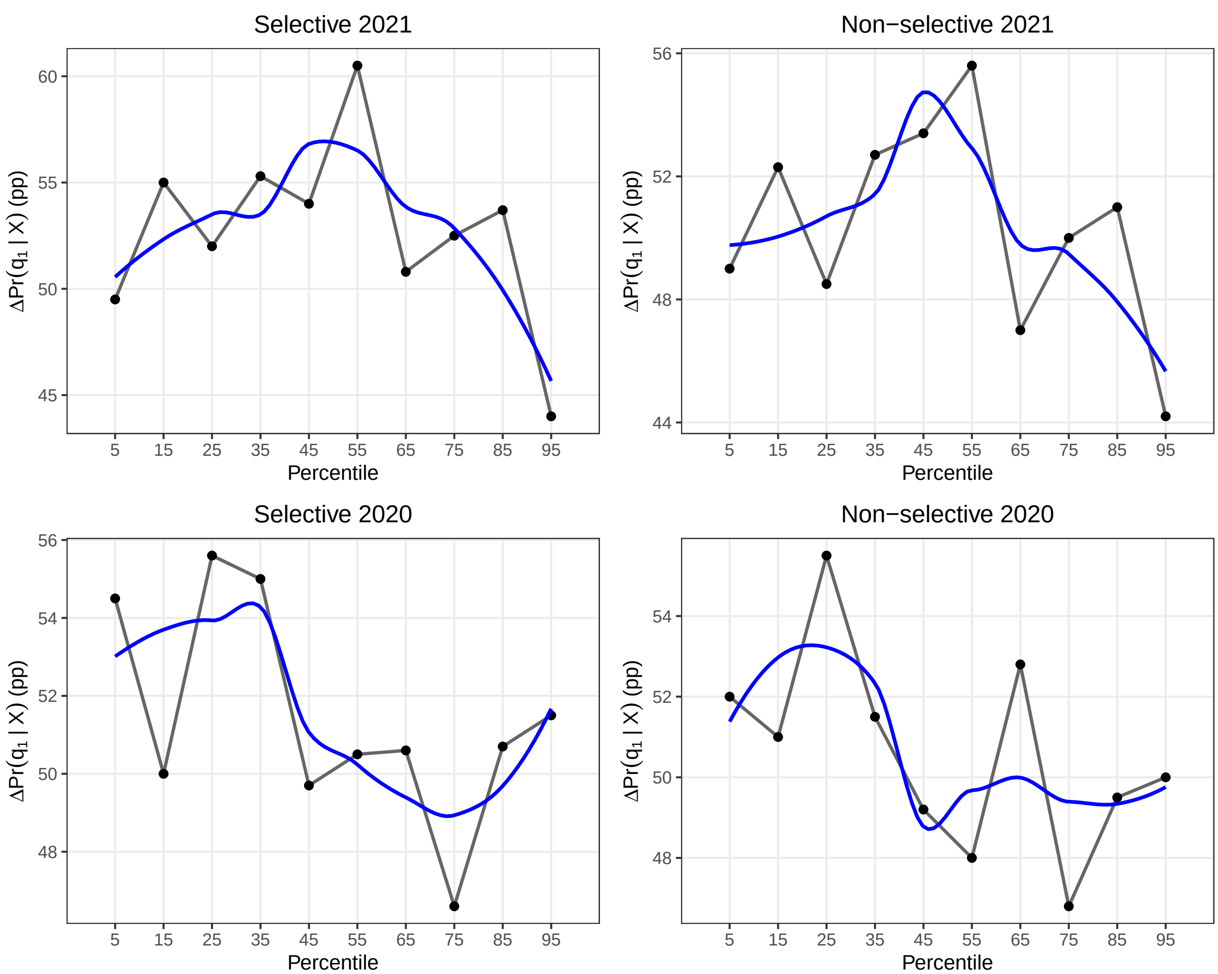


Figure A14: Effect of inflation on placements in 2021

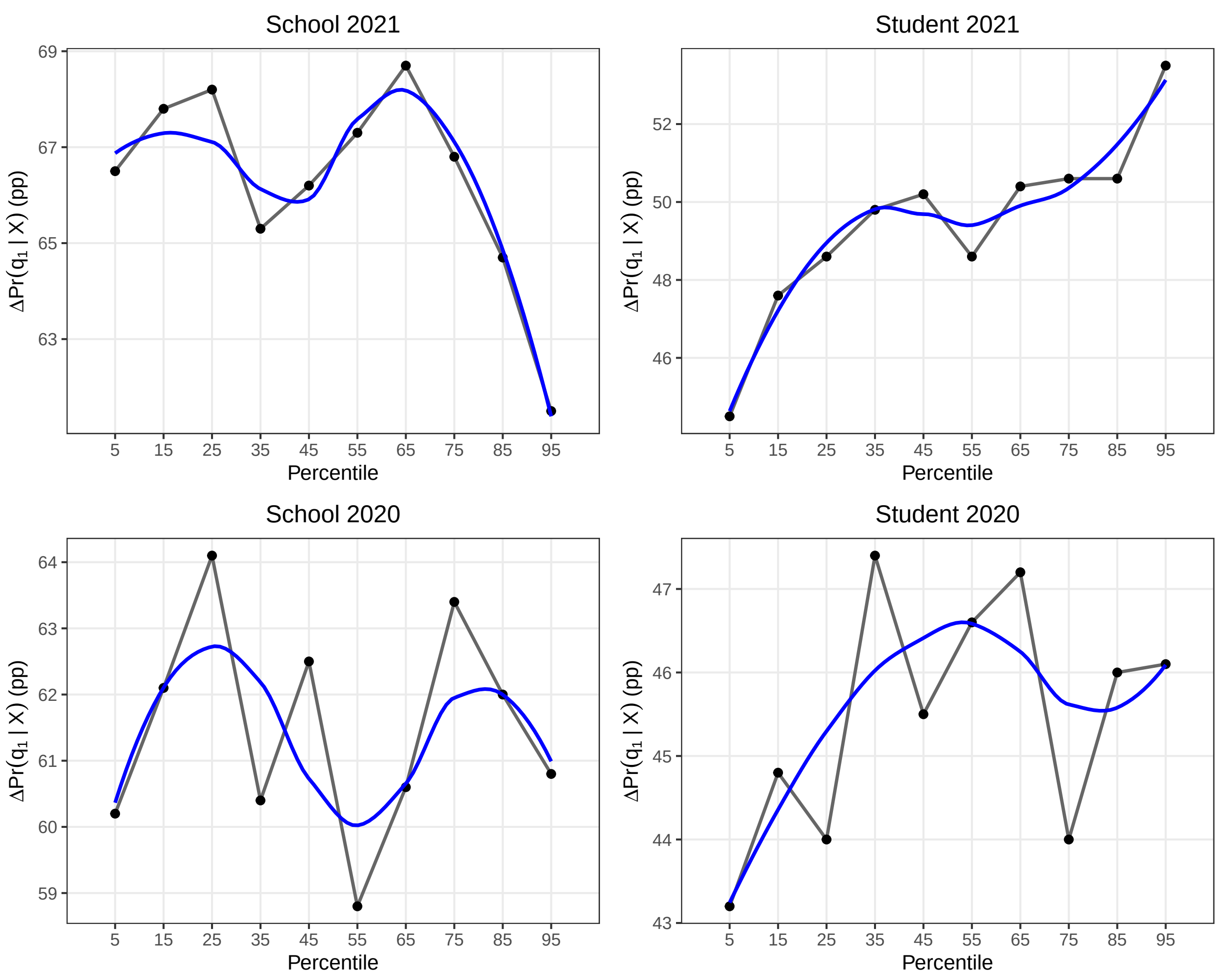
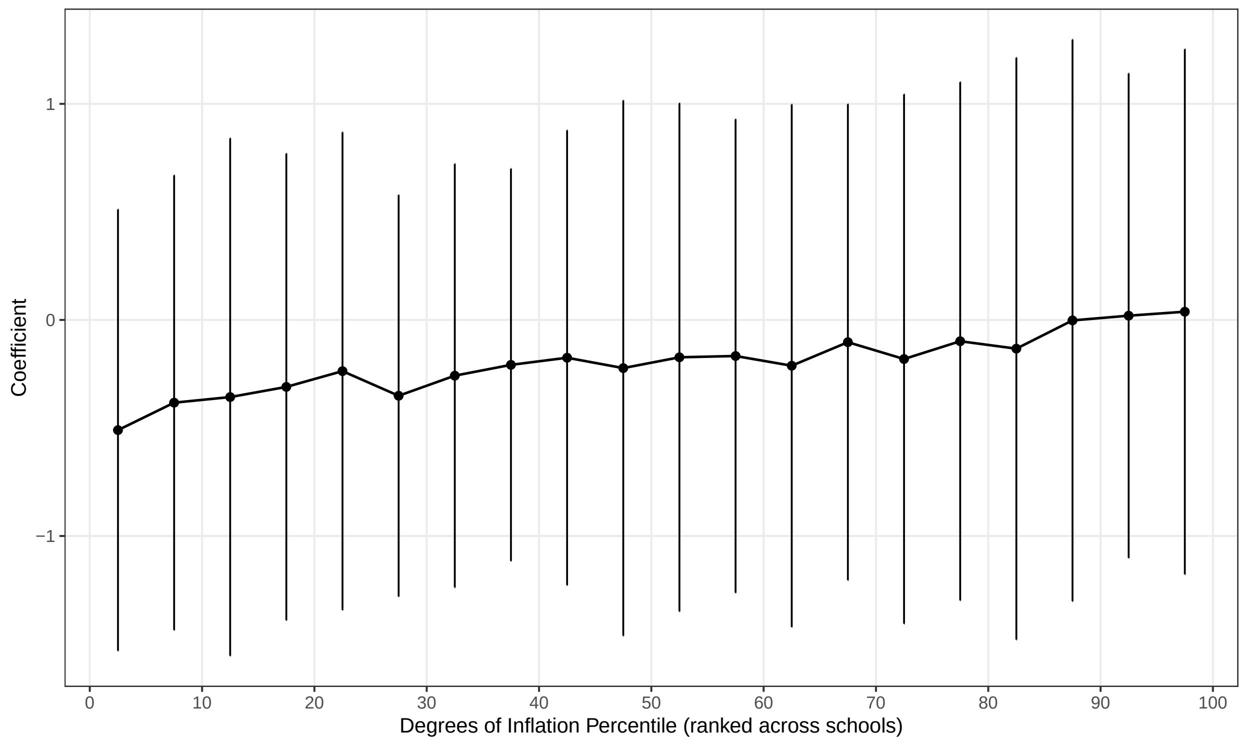


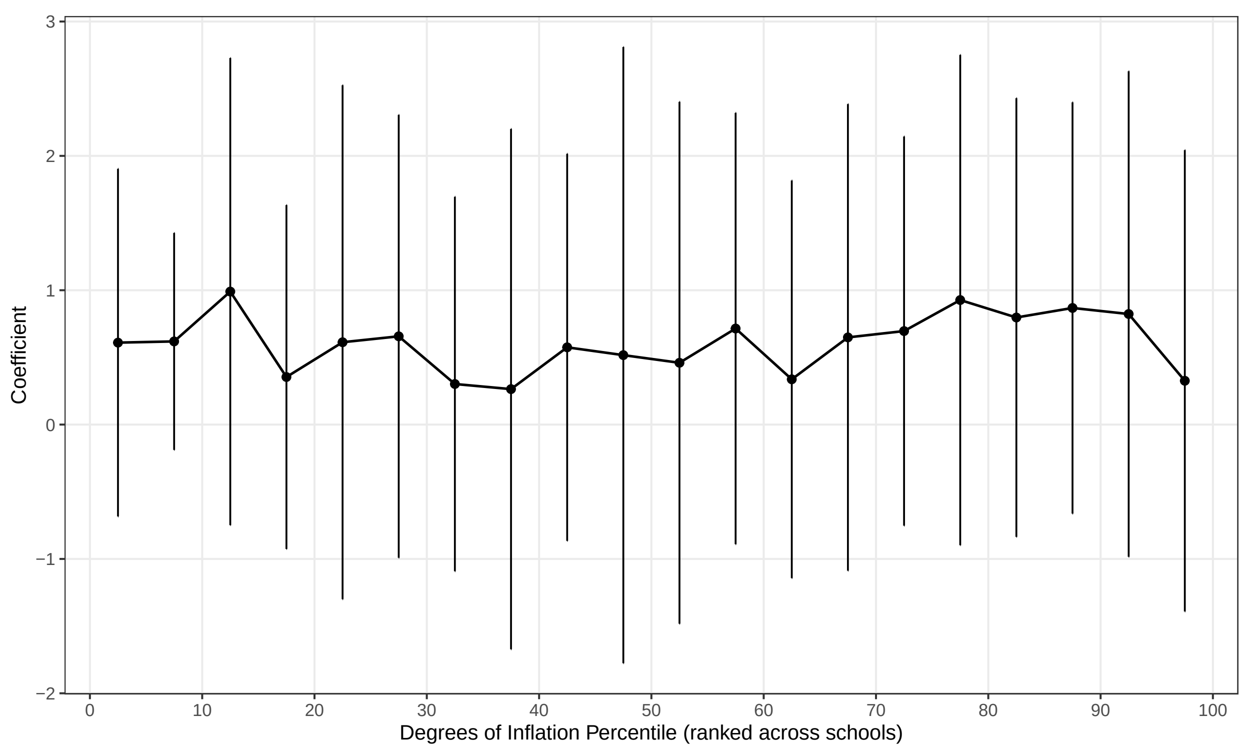
Table A1: Balance Table of Applicant Characteristics

Characteristic	$N = 1,192,935$ (2020)	$N = 219,929$ (2021)	$N = 234,294$ (2020 + 2021)	$N = 1,647,158$ (2015-2019)
Gender				
Men	522,795 (44%)	95,289 (43%)	102,174 (44%)	720,258 (44%)
Women	670,140 (56%)	124,640 (57%)	132,120 (56%)	926,900 (56%)
School Type				
Academy	411,969 (35%)	84,186 (38%)	90,521 (39%)	586,626 (36%)
Further Education	62,864 (5.3%)	9,840 (4.5%)	10,614 (4.5%)	83,318 (5.1%)
Grammar School	55,818 (4.7%)	5,897 (2.7%)	6,252 (2.7%)	67,973 (4.1%)
Independent	167,995 (14%)	24,869 (11%)	26,301 (11%)	223,561 (14%)
Sixth Form	237,426 (20%)	46,177 (21%)	48,868 (21%)	332,471 (20%)
State School	256,863 (22%)	48,960 (22%)	51,738 (22%)	344,140 (21%)
Ethnic Group				
Asian	156,871 (13%)	29,111 (13%)	30,869 (13%)	235,067 (14%)
Black	61,526 (5.2%)	11,964 (5.4%)	12,568 (5.4%)	86,058 (5.2%)
Mixed	53,751 (4.5%)	11,936 (4.5%)	13,335 (5.7%)	79,022 (4.8%)
Other	51,823 (4.3%)	8,774 (4.0%)	9,207 (3.9%)	69,804 (4.2%)
Unknown/Prefer NS	41,923 (3.5%)	7,936 (3.6%)	7,817 (3.3%)	57,676 (3.5%)
White	857,238 (72%)	144,645 (66%)	152,763 (65%)	1,154,617 (70%)
Domicile				
England	1,070,768 (90%)	210,220 (96%)	224,699 (96%)	1,505,687 (91%)
EU (excl. UK)	36,002 (3.0%)	6173 (2.8%)	6456 (2.8%)	48,631 (3.0%)
Northern Ireland	46,319 (3.9%)	1764 (0.8%)	1698 (0.7%)	49,781 (3.0%)
Not EU	30,252 (2.5%)	5,535 (2.5%)	5,283 (2.3%)	41,070 (2.5%)
Scotland	2,503 (0.2%)	1,477 (0.7%)	2,310 (1.0%)	3,928 (0.2%)
Wales	39,491 (3.3%)	1352 (0.6%)	1588 (0.7%)	42,431 (2.6%)
POLAR4 Quintile				
Q1	113,733 (9.5%)	21,763 (9.9%)	23,175 (9.9%)	158,671 (9.6%)
Q2	163,860 (14%)	29,981 (14%)	32,260 (14%)	226,101 (14%)
Q3	211,895 (18%)	41,877 (17%)	44,111 (19%)	291,461 (18%)
Q4	268,260 (22%)	49,360 (22%)	52,248 (22%)	369,868 (22%)
Q5	400,492 (34%)	74,065 (34%)	79,426 (34%)	553,918 (34%)
Unknown	34,758 (2.9%)	6345 (2.9%)	6036 (2.6%)	47,139 (2.9%)
IMD Quintile				
Q1	151,778 (13%)	30,905 (14%)	32,784 (14%)	215,467 (13%)
Q2	186,674 (16%)	36,096 (16%)	38,057 (16%)	260,827 (16%)
Q3	220,488 (18%)	39,982 (18%)	42,878 (18%)	303,348 (18%)
Q4	162,748 (22%)	46,752 (21%)	47,873 (20%)	310,122 (19%)
Q5	336,024 (28%)	59,774 (27%)	64,598 (28%)	460,396 (28%)
Unknown	35,223 (3.0%)	6420 (2.9%)	6100 (2.6%)	47,743 (2.9%)

Figure A15: Impact of grade inflation into placement success rates by school.



(a) Grade inflation by subject group, 2020.



(b) Grade inflation by subject group, 2021.

Table A2: Distribution by Social Class, School Background, and Ethnic Group across University Types

Category	Russell Group	Other Old	New
Social class origin			
Higher professional/managerial	35	23	42
Lower professional/managerial	25	22	53
Routine non-manual	20	20	60
Manual class	13	17	70
School background			
Private	53	24	23
State	20	20	60
Ethnic group			
White	24	20	56
Black Caribbean/African	6	17	77
Pakistani/Bangladeshi	12	23	65
Indian	18	21	61
Chinese	33	19	48
Mixed/Other	21	21	58
All	22	20	58

Table A3: Selectivity and Mean demographics

Panel A. Female Share

	(1)	(2)	(3)
AcceptRate19	0.046*	0.002	0.041*
	(0.019)	(0.019)	(0.019)
Num. Obs.	6443	6026	5708
R^2	0.001	0.000	0.001

Panel B. High Math score share

	(1)	(2)	(3)
AcceptRate19	0.325***	0.170***	0.197***
	(0.018)	(0.019)	(0.020)
Num. Obs.	6443	6026	5708
R^2	0.050	0.014	0.017

Panel C. High income parent

	(1)	(2)	(3)
AcceptRate19	0.039***	-0.004	0.000
	(0.005)	(0.005)	(0.005)
Num. Obs.	18901	16981	15982
R^2	0.004	0.000	0.000

Panel D. Independent school share

	(1)	(2)	(3)
AcceptRate19	0.202***	0.149***	0.149***
	(0.007)	(0.007)	(0.007)
Num. Obs.	18901	16981	15982
R^2	0.047	0.026	0.026

Notes: Standard errors in parentheses. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A4: Applications by School Type and Tariff Group

Characteristic	State School $N = 112{,}415$	Academy $N = 231{,}948$	Grammar $N = 18{,}181$	Independent $N = 91{,}663$	Sixth Form $N = 120{,}562$	Further $N = 26{,}053$	Overall $N = 600{,}822$
High Tariff							
0	33,441 (30%)	61,637 (27%)	4,041 (22%)	9,708 (11%)	33,849 (28%)	8,313 (32%)	150,989 (25%)
1	19,026 (17%)	34,927 (15%)	2,101 (12%)	6,720 (7.3%)	20,213 (17%)	4,826 (19%)	87,813 (15%)
2	15,898 (14%)	31,084 (13%)	2,067 (11%)	7,492 (8.2%)	16,906 (14%)	4,426 (17%)	77,475 (13%)
3	14,311 (13%)	30,611 (13%)	2,562 (14%)	10,660 (12%)	16,673 (14%)	2,988 (11%)	77,805 (13%)
4	14,627 (13%)	34,386 (15%)	2,871 (16%)	16,613 (18%)	17,505 (15%)	2,897 (11%)	90,796 (15%)
5	13,748 (12%)	36,017 (16%)	3,737 (21%)	34,909 (38%)	15,475 (13%)	2,470 (9.5%)	106,356 (18%)
6	1,364 (1.2%)	3,286 (1.4%)	412 (2.3%)	3,839 (4.2%)	1,487 (1.2%)	229 (0.9%)	10,617 (1.8%)
7	0 (0%)	0 (0%)	0 (0%)	3 (<0.1%)	0 (0%)	0 (0%)	3 (<0.1%)
Mid Tariff							
0	26,335 (23%)	56,366 (24%)	4,942 (27%)	41,324 (45%)	36,986 (31%)	6,721 (26%)	172,674 (29%)
1	29,573 (26%)	60,002 (26%)	4,638 (26%)	21,987 (24%)	32,455 (27%)	6,852 (26%)	155,507 (26%)
2	26,555 (24%)	53,062 (23%)	4,054 (22%)	13,828 (15%)	25,234 (21%)	5,991 (23%)	128,724 (21%)
3	18,636 (17%)	37,851 (16%)	2,747 (15%)	8,698 (9.5%)	16,450 (14%)	3,943 (15%)	88,325 (15%)
4	8,652 (7.7%)	18,937 (8.2%)	1,296 (7.1%)	4,462 (4.9%)	7,319 (6.1%)	1,998 (7.7%)	42,664 (7.1%)
5	2,454 (2.2%)	5,352 (2.3%)	477 (2.6%)	1,261 (1.4%)	1,991 (1.6%)	511 (2.0%)	12,046 (2.0%)
6	210 (0.2%)	378 (0.2%)	27 (0.1%)	100 (0.1%)	126 (0.1%)	37 (0.1%)	878 (0.1%)
7	0 (0%)	0 (0%)	0 (0%)	0 (<0.1%)	0 (0%)	0 (0%)	0 (0%)
Low Tariff							
0	48,878 (43%)	118,550 (51%)	11,156 (61%)	72,696 (79%)	51,397 (43%)	9,770 (38%)	312,447 (52%)
1	22,914 (20%)	44,555 (19%)	2,983 (17%)	9,715 (23%)	23,636 (20%)	6,300 (24%)	109,403 (18%)
2	16,745 (15%)	30,257 (13%)	1,954 (11%)	4,928 (5.4%)	19,756 (16%)	2,287 (8.8%)	75,927 (13%)
3	12,415 (11%)	21,074 (9.1%)	1,247 (6.9%)	2,649 (2.9%)	13,910 (12%)	3,358 (13%)	54,653 (9.1%)
4	7,700 (6.8%)	12,005 (5.2%)	585 (3.2%)	1,221 (1.3%)	9,078 (7.5%)	2,011 (7.7%)	32,600 (5.4%)
5	3,445 (3.1%)	5,901 (2.5%)	256 (1.4%)	414 (0.5%)	4,430 (3.7%)	906 (3.5%)	14,542 (2.4%)
6	318 (0.3%)	416 (0.2%)	11 (<0.1%)	40 (<0.1%)	354 (0.3%)	104 (0.4%)	1,243 (0.2%)
7	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)

Table A5: Grade inflation and school quality

	A		A*		B	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
School Quality	-0.452***	-0.473***	-0.212***	-0.212***	-0.591***	-0.591***
	(0.019)	(0.022)	(0.018)	(0.019)	(0.006)	(0.006)
Year	2020	2021	2020	2021	2020	2021
Num. Obs.	1890	1890	1869	1869	1869	1869
R^2	0.227	0.193	0.072	0.061	0.818	0.820

Notes: The table reports the regression coefficient that regresses the inflation measure defined in Equation (1) on the measure of school quality. Degree of inflation is measured as the average marginal effect (AME) at the school level, in which I calculates the average marginal effect of the school inflation for all students within the school. Namely, the latent scores of each students without the inflation effects are calculated by mapping the student attributes on the fixed effect coefficient in Equation (1). Before the AME is calculated, I apply the shrinkage correction method by **?**.School quality is defined as the fixed effect derived from Equation (1). The school fixed effect is evaluated by calculating the probability of obtaining A or A* for the nationally representative student, which I define as the student with median values in their continuous or categorical variables. School fixed effects are corrected for the incidental parameter bias by using methods by Fernández-Val and Weidner (2016) . Schools without less than 30 students in both 2020 and 2021 taking A-level exams are dropped from the regression. Column 1 and 2 uses A, Column 3 and 4 uses A*, and Column 5 and 6 uses B as the binary dependent variable for estimating Equation (1). Bootstrapped standard errors at obtained by blocking at the school level. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A6: Grading bias across demographics by school quality

	(1)	(2)
Gender ×SQ	-0.006	0.281^{***}
	(0.078)	(0.082)
Num.Obs.	1818	1824

Ethnic ×SQ	0.062	-0.175^{+}
	(0.064)	(0.105)
Num.Obs.	1838	1855

SES Class2 ×SQ	-0.084	-0.034
	(0.193)	(0.130)
SES Class3 ×SQ	0.155	0.190
	(0.155)	(0.156)
SES Class4 ×SQ	0.165	0.202
	(0.174)	(0.185)
SES Class5 ×SQ	0.399^{*}	0.264
	(0.180)	(0.199)
SES Class6 ×SQ	0.367^{*}	0.182
	(0.162)	(0.189)
SES Class7 ×SQ	0.331^{+}	0.579^{**}
	(0.196)	(0.218)
SES Class9 ×SQ	0.033	0.204
	(0.128)	(0.126)
Num.Obs.	13997	14006

IMD Quantile 1 ×SQ	0.001	0.355^{*}
	(0.154)	(0.157)
IMD Quantile 2 ×SQ	0.286^{*}	0.040
	(0.136)	(0.131)
IMD Quantile 3 ×SQ	0.208^{+}	0.286^{*}
	(0.113)	(0.119)
IMD Quantile 4 ×SQ	0.317^{**}	0.175^{+}
	(0.115)	(0.103)
Num.Obs.	8719	8759

POLAR Quantile 1 ×SQ	0.452^{*}	-0.040
	(0.199)	(0.199)
POLAR Quantile 2 ×SQ	0.163	0.148
	(0.172)	(0.204)
POLAR Quantile 3 ×SQ	-0.074	-0.172
	(0.149)	(0.171)
POLAR Quantile 4 ×SQ	-0.068	-0.162
	(0.151)	(0.156)
Num.Obs.	8061	8162

Notes: standard errors in parentheses. $+p < 0.1$, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

Table A7: Decomposing Inflation by school and student channel

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Inflation	-0.437	-0.429	0.199	0.229	-0.100	-0.409	-0.409	-0.437
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Year	2020	2021	2020	2021	2020	2021	2020	2021
Num. Obs.	522667	568060	522667	568060	522667	568060	522667	568060
R^2	0.191	0.184	0.082	0.072	0.207	0.044	0.126	0.255

Table A8: Regression results: Selectivity slopes

	plogis(slope_20)	**plogis(slope_21)**	**plogis(slope_selective_20)**	**plogis(s**
(Intercept)	0.541***	0.566***	0.562***	
	(0.040)	(0.038)	(0.047)	
plogis(Base_Estimate)	-0.073	-0.128	-0.101	
	(0.087)	(0.082)	(0.101)	
Num. Obs.	1878	1881	1878	
R^2	0.000	0.001	0.001	

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A9: Regression results: School vs. Student channels

	(School Channel)	**(School Channel)**	**(Student Channel)**	**(Student C**
(Intercept)	0.618***	0.713***	0.420***	0.396*
	(0.022)	(0.023)	(0.020)	(0.02
plogis(Base_Estimate)	-0.009	-0.105*	0.078*	0.222*
	(0.047)	(0.049)	(0.043)	(0.04
Year	2020	2021	2020	202
Num. Obs.	1878	1881	1878	188
R^2	0.000	0.002	0.002	0.01
R^2 Adj.	0.000	0.002	0.001	0.01
AIC	-999.2	-812.4	-1372.0	-979
BIC	-982.6	-795.8	-1355.4	-963
Log Lik	502.625	409.210	689.024	492.9
F	0.039	2.479	3.524	22.15
RMSE	0.19	0.19	0.17	0.19

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A10: Performance of predicting tariff score

Table A11: Regression Results: Acceptance and Rejection

Dep.	#Accept		#Rejection	
Treat	-0.0025	-0.0017	-0.0279***	-0.0348**
	(0.0016)	(0.0024)	(0.0078)	(0.0115)
(x) Further Education		-0.0084		0.0736.
		(0.0071)		(0.0432)
(x) Grammar School		-0.0096		0.0493
		(0.0086)		(0.0371)
(x) Independent School		-0.0072		-0.0424
		(0.0090)		(0.0362)
(x) Sixth Form College		0.0005		0.0103
		(0.0040)		(0.0186)
(x) State School		-0.0004		0.0012
		(0.0043)		(0.0222)
Obs.	371,477	371,477	262,595	262,595
R2	0.44417	0.44418	0.36535	0.36541

Table A12: Regression Results: High, Medium, and Low Tariff

Dep.	#H Tariff		#M Tariff		#L Tariff
Treat	0.0161***	0.0192**	0.0161***	-0.0224**	-0.0078
	(0.0044)	(0.0065)	(0.0044)	(0.0069)	(0.0055)
(x) Further Education		-0.0081		0.0068	
		(0.0274)		(0.0175)	
(x) Grammar School		-0.0582**		0.0040	
		(0.0214)		(0.0286)	
(x) Independent School		-0.0516***		0.0619*	
		(0.0156)		(0.0307)	
(x) Sixth Form College		0.0038		-0.0033	
		(0.0110)		(0.0119)	
(x) State School		0.0069		-0.0005	
		(0.0120)		(0.0122)	
Obs.	306,983	306,983	306,983	326,639	264,154
R2	0.58502	0.58508	0.58502	0.51234	0.52139

Table A13: Regression results: Improvements in placements by school quality

	(1)	(2)	(3)	(4)
School Quality	-0.067*	-0.171***	0.086*	-0.049
	(0.040)	(0.036)	(0.052)	(0.048)
Year	2020	2021	2020	2021
Num. Obs.	1,878	1,881	1,878	1,881
Uni FE			X	X
School FE			X	X
Tariff			X	X

Notes: Standard errors in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. δ is the (estimated) logit coefficient and SQ denotes the coefficient for the school fixed effect from a seperate grade regression.

References

Alesina, A., Carlana, M., La Ferrara, E., and Pinotti, P. (2024). Revealing stereotypes: Evidence from immigrants in schools. *American Economic Review*, 114(7):1916–48.

Ali, S. N. and Shorrer, R. I. (2025). Hedging when applying: Simultaneous search with correlation. *American Economic Review*, 115(2):571–98.

Arcidiacono, P. (2005). Affirmative action in higher education: How do admission and financial aid rules affect future earnings? *Econometrica*, 73(5):1477–1524.

Arcidiacono, P., Kinsler, J., and Ransom, T. (2023). What the students for fair admissions cases reveal about racial preferences. *Journal of Political Economy Microeconomics*, 1(2).

Avery, C., Shi, L., and Magouirk, P. (2025). Test-optional college admissions: Act and sat scores, applications, and enrollment changes. Working Paper 34260, National Bureau of Economic Research.

Bar, T., Kadiyali, V., and Zussman, A. (2009). Grade information and grade inflation: The cornell experiment. *Journal of Economic Perspectives*, 23(3):93–108.

Becker, G. S. (1971). *The Economics of Discrimination*. University of Chicago Press.

Bennett, C. T. (2022). Untested admissions: Examining changes in application behaviors and student demographics under test-optional policies. *American Educational Research Journal*, 59(1):180–216.

Bettinger, E. P., Long, B. T., Oreopoulos, P., and Sanbonmatsu, L. (2012). The role of application assistance and information in college decisions: Results from the h&r block fafsa experiment*. *The Quarterly Journal of Economics*, 127(3):1205–1242.

Betts, J. R. and Grogger, J. (2003). The role of grading standards in education policy. *American Economic Review*, 93(2):94–99.

Black, S. E., Denning, J. T., and Rothstein, J. (2023). Winners and losers? the effect of gaining and losing access to selective colleges on education and labor market outcomes. *American Economic Journal: Applied Economics*, 15(1):26–67.

Bleemer, Z. (2020). Grade inflation at more- and less-affluent high schools. Technical report, UC-CHP (UC Center for Health Policy) Policy Brief 2020.1.

Bleemer, Z. (2021). Affirmative action, mismatch, and economic mobility after california's proposition 209*. *The Quarterly Journal of Economics*, 137(1):115–160.

Bleemer, Z. (2023). Affirmative action and its race-neutral alternatives. *Journal of Public Economics*, 220:104839.

Borusyak, K. and Hull, P. (2023). Nonrandom exposure to exogenous shocks. *Econometrica*, 91(6):2155–2185.

Botelho, F., Madeira, R. A., and Rangel, M. A. (2015). Racial discrimination in grading: Evidence from brazil. *American Economic Journal: Applied Economics*, 7(4):37–52.

Burgess, S. and Greaves, E. (2013). Test scores, subjective assessment, and stereotyping of ethnic minorities. *Journal of Labor Economics*, 31(3):535–576.

Carlana, M. (2019). Implicit stereotypes: Evidence from teachers' gender bias*. *The Quarterly Journal of Economics*, 134(3):1163–1224.

Chan, W., Hao, L., and Suen, W. (2007). A signaling theory of grade inflation. *International Economic Review*, 48(3):1065–1090.

Chetty, R., Deming, D. J., and Friedman, J. N. (2023). Diversifying society's leaders? the determinants and causal effects of admission to highly selective private colleges. Working Paper 31492, National Bureau of Economic Research.

Chetty, R., Friedman, J. N., Saez, E., Turner, N., and Yagan, D. (2020). Income segregation and intergenerational mobility across colleges in the united states*. *The Quarterly Journal of Economics*, 135(3):1567–1633.

Cornwell, C., Mustard, D. B., and Van Parys, J. (2013). Non-cognitive skills and the gender disparities in test scores and teacher assessments: Evidence from primary school. *Journal of Human Resources*, 48(1):236–264.

Cullen, J. B., Long, M. C., and Reback, R. (2013). Jockeying for position: Strategic high school choice under texas' top ten percent plan. *Journal of Public Economics*, 97:32–48.

Dee, T. S., Dobbie, W., Jacob, B. A., and Rockoff, J. E. (2019). The causes and consequences of test score manipulation: Evidence from the new york regents examinations. *American Economic Journal: Applied Economics*, 11(3):339–365.

Denning, J. T., Eide, E. R., Mumford, K. J., Patterson, R. W., and Warnick, M. (2022). Why have college completion rates increased? *American Economic Journal: Applied Economics*, 14(3):1–29.

Denning, J. T., Nesbit, R., Pope, N., and Warnick, M. (2023). The distributional and long-term effects of grade inflation. Technical report, NBER Working Paper No. 31123.

Dessein, W., Frankel, A., and Kartik, N. (2025). Test-optional admissions. *American Economic Review*, 115(9):3130–70.

Diamond, R. and Persson, P. (2016). The long-term consequences of teacher discretion in grading of high-stakes tests. Working Paper 22207, National Bureau of Economic Research.

Dynarski, S., Nurshatayeva, A., Page, L. C., and Scott-Clayton, J. (2022). Addressing non-financial barriers to college access and success: Evidence and policy implications. NBER Working Paper 30054, National Bureau of Economic Research.

Dynarski, S. M. (2003). Does aid matter? measuring the effect of student aid on college attendance and completion. *American Economic Review*, 93(1):279–288.

Efron, B. (2010). *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*. Institute of Mathematical Statistics Monographs. Cambridge University Press.

Epple, D., Romano, R., and Sieg, H. (2006). Admission, tuition, and financial aid policies in the market for higher education. *Econometrica*, 74(4):885–928.

Ferman, B. and Fontes, L. F. (2022). Assessing knowledge or classroom behavior? evidence of teachers' grading bias. *Journal of Public Economics*, 216:104773.

Fernández-Val, I. and Weidner, M. (2016). Individual and time effects in nonlinear panel models with large n, t. *Journal of Econometrics*, 192(1):291–312.

Figlio, D. N. and Getzler, L. S. (2006). Accountability, ability, and disability: Gaming the system? *Advances in Applied Microeconomics*, 14:35–49.

Figlio, D. N. and Lucas, M. E. (2004). Do high grading standards affect student performance? *Journal of Public Economics*, 88(9):1815–1834.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis*. Chapman and Hall/CRC, 3rd edition.

Hanna, R. N. and Linden, L. L. (2012). Discrimination in grading. *American Economic Journal: Economic Policy*, 4(4):146–168.

Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature*, 24(3):1141–1177.

Hoxby, C. M. and Avery, C. (2012). The missing one-offs: The hidden supply of high-achieving, low income students. Working Paper 18586, National Bureau of Economic Research.

Jacob, B. A. and Levitt, S. D. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating*. *The Quarterly Journal of Economics*, 118(3):843–877.

Lavy, V. (2009). Performance pay and teachers' effort, productivity, and grading ethics. *American Economic Review*, 99(5):1979–2011.

Lavy, V. and Sand, E. (2018). On the origins of gender gaps in human capital: Short- and long-term consequences of teachers' biases. *Journal of Public Economics*, 167:263–279.

Manski, C. F. and Wise, D. A. (1983). *College Choice in America*. Harvard University Press, Cambridge, MA.

Neal, D. and Schanzenbach, D. W. (2008). Left behind by design: Proficiency counts and test-based accountability. *Review of Economics and Statistics*, 92(2):263–283.

Pattison, E., Grodsky, E., and Muller, C. (2013). Is the sky falling? grade inflation and the signaling power of grades. *Educational Researcher*, 42(5):259–265.

Phillips, M. and Reber, S. (2022). Does virtual advising increase college enrollment? evidence from a random-assignment college access field experiment. *American Economic Journal: Economic Policy*, 14(3):198â€“234.

Rojstaczer, S. and Healy, C. (2012). Where a is ordinary: The evolution of american college and university grading, 1940-2009. *Teachers College Record*, 114(7):1–23.

Sacerdote, B., Staiger, D. O., and Tine, M. (2025). How test optional policies in college admissions disproportionately harm high achieving applicants from disadvantaged backgrounds. Working Paper 33389, National Bureau of Economic Research.

Tyner, A. and Gershenson, S. (2020). Conceptualizing grade inflation. *Economics of Education Review*, 78:102037.

Walters, C. (2024). Chapter 3 - empirical bayes methods in labor economics. volume 5 of *Handbook of Labor Economics*, pages 183–260. Elsevier.

Wooldridge, J. M. (2001). *Econometric Analysis of Cross Section and Panel Data*, volume 1 of *MIT Press Books*. The MIT Press, 1 edition.