# Educational policy and skill formation of students: Evidence from UK

University of Antwerp

April 19, 2024

**Abstract**

In this article, we exploit the introduction of a policy reform in England focussed on improving educational quality, not introduced in neighboring regions in the United Kingdom, to estimate the impact on PISA study-scores by UK students. We find that the policy has had an overall positive effect on mathematics scores. This result is robust to the introduction of student- and school-level control variables. We investigate heterogeneity issues and find that the increase in mathematics scores is most pronounced for disadvantaged children and females. Reading scores however did not improve for children from low socioeconomics background, while they dropped for female students. Our results suggest that the policy reform has had a positive effect on educational quality overall, benefitting socially disadvantaged groups most.

*JEL Classification: I21, I26*
*Keywords: Education, testing scores, Difference-in-Difference*

# 1 Introduction and Background

Individual skill development starts at a very early stage in life among other channels, largely through the years spent at school. Skills are important because they have significant returns on the labour market in terms of wage earnings. Furthermore, better cognitive skills are related to greater economic growth (Hanushek and Woessmann, 2012) and higher employment rates because these result in more education and employment opportunities (Hanushek and Woessmann, 2012; OECD, 2010).

When considering skill formation in terms of an educational production function, several inputs are imperative for increasing skill levels. One of these inputs is the quality of education, which can be influenced through many mechanisms such as the quality of the curriculum. However, simply increasing the budget spending on education is not an effective way of increasing quality Hanushek (1995) writes. Resources are spent in inefficient ways that sparingly contribute to the quality of education, decreasing the cost effectiveness of such spending. A more efficient way of tackling quality is by increasing the incentive to perform better (Hanushek, 1995, 2020).

In this paper, we consider a policy reform that fundamentally changed the testing of General Certificate of Secondary Education (GCSE) for students in England. By increasing the difficulty of the exam for example, schools are incentivized to increase educational quality such that students may pass on the one hand. On the other hand, students are incentivized to study harder such that they pass the exam. By exploiting that the policy was only implemented in England, and not in the rest of the UK, we try to indicate causal effects of this policy.

Ever since the 1960s (Carroll, 1963), education economists have tried to define the determinants of educational achievement. The education production function literature production function of education economics uses inputs such as student, family, and school resources to produce output like test scores (Worthington, 2001). In an ideal scenario, information from both the past and present resources and innate ability are present to develop this education production function (Todd and Wolpin, 2003). A common type of empirical analysis is conducted through standardized test scores which measure children's cognitive skills (Hanushek, 2020).

In recent years, much interest has been directed toward PISA scores as they measure the real-life skills of future generations. The decline in literacy skills is worrying as reading proficiency and mathematical skills at a young age correlate with socioeconomic status at age 42 (Ritchie and Bates, 2013). Additionally, testing for PISA underwent changes in testing for 2022. The focus in this wave was more on mathematics than previous years.

In our analysis, we leverage the GCSE policy reform to set up a Difference in Difference analysis, with the England being the only treated region in the United Kingdom. The introduction of the policy reform over time, combined with the fact that it only impacts England, makes it suitable for a DiD analysis. Our results indicate that the policy reform has statistically significantly increased overall mathematics performance in England compared to other regions in the UK. Reading scores might have suffered, but after controlling for student- and school-level covariates, this result is insignificant. Furthermore, we find clear evidence of heterogeneity in responses, with socially disadvantaged students benefitting more from the introduction of the GCSE policy reform.

## 2 Education System

### 2.1 Education System in the UK and the GCSE

Since the start of PISA in 2000, the UK has performed higher than average. But in the UK there is also the General Certificate of Secondary Education (GCSE), an academic qualification evaluation, usually taken at age 16 in England, Wales and Northern Ireland, which signals the completion of compulsory secondary education. With the GCSE qualification students can enter training, apprenticeship and continue higher education. Therefore, the GCSE is regarded as having high stakes. In 2015, the UK introduced reforms to the GCSEs to enrich the curriculum and prepare students better for study and employment. The new GSCE certifications post-reform were awarded in 2017. The first subjects to be reformed were English, English literature, and mathematics. Under the new GCSE, most subjects are assessed by exam alone, compared to previous form of assessment which was subject to the school (which included non-exam assessment) and not to the board of education. However, some non-exam assessment exists and are decided subject by subject.

The changes to English literature aimed to better prepare students for A-levels. The reforms introduced wide range of texts at a higher level than the previous ones. This included 19th century novels, Romantic poetry, Shakespeare and fiction or drama from British Isles from 1914 onwards. The exams are closed book. The reforms for the math GCSE introduced more demanding topics especially those that were taught at A-levels and emphasised on problem-solving.

The reformed GCSE have the following intentions: a) To show students' achievements in challenging and demanding content. b) To establish a solid basis for further academic, and vocational studies, and employment. c) To offer a foundation for schools and colleges to be accountable for their students' performance. (UK.GOV, 2018)

In the UK, schools usually decide the tiers that students are introduced into. The foundational tier focuses on core understanding and basic skills. Usually, students placed in tier C cannot obtain more than grade C even if they perform well. With the reforms, the qualifications of students are un-tiered so all students will take the same exam. Tiering however only exists for subjects such as Math where one exam cannot evaluate students across their full range of abilities such as understanding, knowledge and skills in the subject. The GSCS reforms had considered the potential impact of the changes, on students who share the protected characteristics defined in the Equality Act 2010. This Equality Act 2010 in the UK is established against someone because of characteristics like their age, disability, gender reassignment, marriage and civil partnership, race, religion, gender and sexual orientation (Equality and Human Rights Commission, 2021).

### 2.2 Global Education Reform Movement and GCSE reforms

The GERM is an ideology of a new form of the education system. It was launched in the 1980s by neo-liberalism think tanks in the United States and the United Kingdom during the regimes of Reagan and Thatcher (Fuller and Stevenson, 2019). The main factors of this movement are summarized as follows (Sahlberg, 2012): (1) Standardization: setting up standardized high-stakes testing to make international comparisons and install detailed curricula. (2) Focus on specific topics and basic skills tested in international assessments such as PISA and PIRLS. (3) Search for low-risk ways to reach learning goals, resulting in fewer experiments. (4) Test-based accountability policies: competition between students and schools results in awarding or punishing schools and teachers based on standardized tests.

## 3 Descriptive Statistics

This section introduces descriptive statistics for the education system in the United Kingdom. We will describe test scores and show how they evolve over time, by region and by socioeconomic background. Afterwards, we will look at individual characteristics of students in the United Kingdom.

### 3.1 Test scores

Tables 1 and 2 introduce test scores in different regions in the UK, reported separately by wave. The dataset has the most observations for England. In every wave, we observe that test scores for reading and mathematics are the highest in England. Wales consistently performs worst on reading and mathematics. Within each region, students typically perform better on reading scores than on mathematics scores, Wales being the only exception. Test scores also show evolution over time, where most scores increase going from 2015 to 2018. In 2022, test scores generally decrease, which might be due to the impact of the Covid-19 pandemic.

| **Reading** | **Wave** | **N** | **Mean** | **Std. Dev.** | **Median** |
|---|---|---|---|---|---|
| England | 2015 | 5,194 | 509.2 | 94.37 | 513.3 |
| | 2018 | 5,242 | 504.8 | 97.13 | 508.5 |
| | 2022 | 4,763 | 500.5 | 100 | 503.2 |
| Northern Ireland | 2015 | 2,401 | 496.2 | 78.6 | 498.4 |
| | 2018 | 2,413 | 503.6 | 93.48 | 510.8 |
| | 2022 | 2,384 | 494.2 | 95.02 | 499.7 |
| Wales | 2015 | 3,451 | 476.9 | 79.14 | 476.9 |
| | 2018 | 3,165 | 485 | 93.64 | 486.5 |
| | 2022 | 2,568 | 465.6 | 95.02 | 467.2 |
| Scotland | 2015 | 3,111 | 491 | 84.94 | 495.3 |
| | 2018 | 2,998 | 502.7 | 90.51 | 503.8 |
| | 2022 | 3,257 | 492.7 | 92.27 | 493.8 |

**Table 1:** Summary Statistics by Region and Wave for Reading test-scores

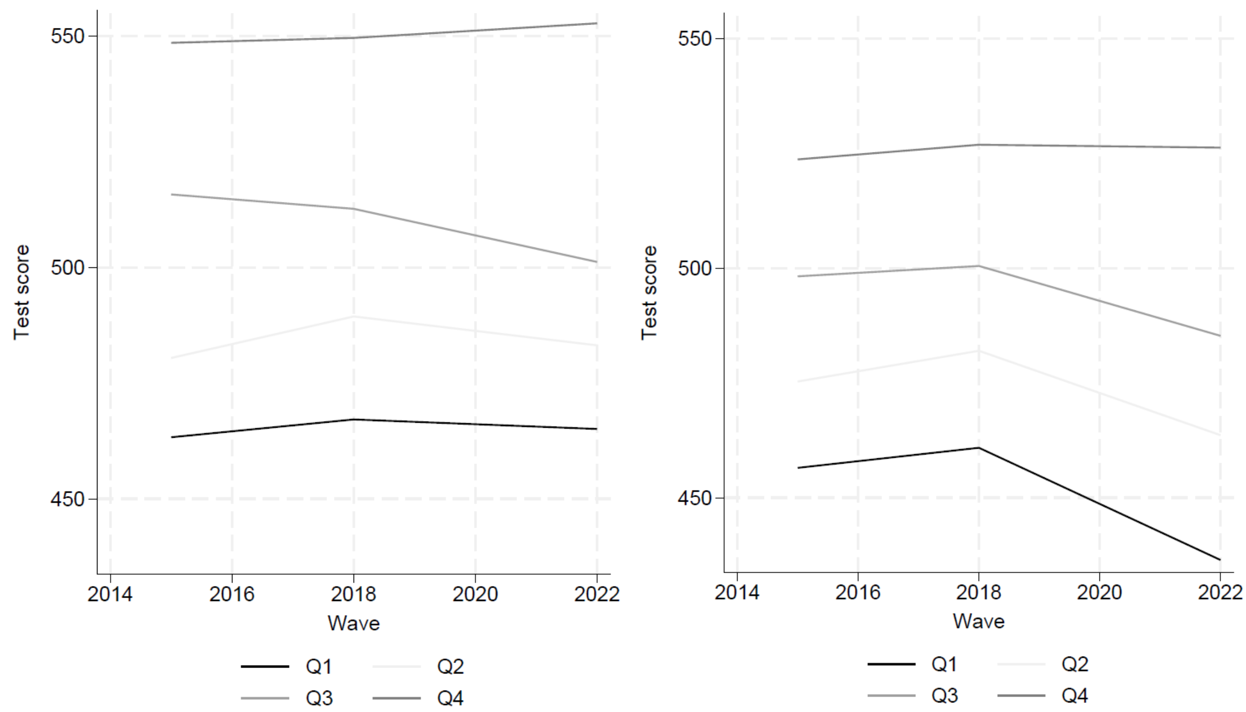
| Mathematics | Year | N | Mean | Std. Dev. | Median |
|---|---|---|---|---|---|
| England | 2015 | 5,194 | 502 | 90.36 | 507 |
| | 2018 | 5,242 | 503.8 | 86.84 | 506.1 |
| | 2022 | 4,763 | 496.5 | 94.66 | 496 |
| Northern Ireland | 2015 | 2,401 | 491.9 | 72.18 | 495.8 |
| | 2018 | 2,413 | 494.4 | 79.18 | 498.6 |
| | 2022 | 2,384 | 484.8 | 89.63 | 486.9 |
| Wales | 2015 | 3,451 | 477.7 | 72.35 | 477.5 |
| | 2018 | 3,165 | 488.2 | 76.12 | 488.3 |
| | 2022 | 2,568 | 466.7 | 86.87 | 463.6 |
| Scotland | 2015 | 3,111 | 488.9 | 77.8 | 489.2 |
| | 2018 | 2,998 | 488 | 79.67 | 486.8 |
| | 2022 | 3,257 | 470.1 | 88.28 | 468 |

**Table 2:** Summary Statistics by Region and Wave for Mathematics test-scores

### 3.2 Socioeconomic status

Test scores not only evolve over time and region, there is also significant difference in test scores for students from different socioeconomic background. Figure 1 shows the test scores for students from different socioeconomic backgrounds. The categorisation is done per wave and region, such that for every wave the 25% most socially disadvantaged students are in the first quantile and the 25% most advantaged students are in the fourth quantile. Clearly, test scores increase with socioeconomic status. Socially disadvantaged students lag behind in their PISA math scores, a critical issue in education policy. Furthermore, for every increase in quantile of socioeconomic status, the math performance increases. This observation is valid for both England and the other regions of the United Kingdom.

**Figure 1:** Math scores across time




England

Scotland, Northern Ireland, Wales

Note: this Figure shows the average math scores across time, represented for different students with socioeconomic statuses. The quartiles in the legend refer to quartile of socioeconomic status, where Q1 refers to the 25% of students most socially disadvantaged while Q4 refer to the 25% most socially advantaged students.

### 3.3 Covariates

This section considers the covariates, all of which are measured at the student level. *Teachers fully certified* is an index that indicates the proportion of teachers fully certified by the appropriate authority within the school of the student. The *school type* is represented by three dummy variables: private independent, private government-dependent, and public school. *Class size* measures the average class size in the school of the student. *School size* indicate the number of students in the student's school. *Parental education* is a variable that displays the highest level of educational attainment of either parent in years of schooling. *Immigration* status is dummy consisting of two different categories, specifically native students (0), first-generation and second-generation student (1). The *socioeconomic status* (SES) is an index of economic, social and cultural status that is based on household characteristics. *Female* is a dummy for whether the student is female. *ICT-Resources* is a variable indicating the ICT resources that the child has at home. Home possession is variable describing the material situation of a household for which the data is collected through a survey.

| Variable | Control Mean (SD) | Treatment Mean (SD) | Difference (p-value) |
|---|---|---|---|
| Proportion cert. teachers | 0.96 | 0.93 | -0.03*** |
| | (0.17) | (0.20) | (0.00) |
| Public School | 0.82 | 0.24 | -0.57*** |
| | (0.39) | (0.43) | (0.00) |
| Private (Indep.) | 0.03 | 0.07 | 0.05*** |
| | (0.17) | (0.26) | (0.00) |
| Private (Gov. Dep.) | 0.03 | 0.52 | 0.49*** |
| | (0.17) | (0.50) | (0.00) |
| Class Size | 24.93 | 24.97 | 0.04 |
| | (3.80) | (4.05) | (0.39) |
| School Size (Sum) | 959.45 | 1097.41 | 137.96*** |
| | (375.12) | (384.96) | (0.00) |
| Parental education | 14.15 | 14.24 | 0.09*** |
| | (2.09) | (2.17) | (0.00) |
| Immigration | 0.10 | 0.21 | 0.11*** |
| | (0.30) | (0.41) | (0.00) |
| Socio-economic status | 0.17 | 0.25 | 0.09*** |
| | (0.87) | (0.89) | (0.00) |
| Female | 0.50 | 0.49 | -0.00 |
| | (0.50) | (0.50) | (0.52) |
| ICT resources at home | 0.44 | 0.52 | 0.08*** |
| | (1.01) | (1.04) | (0.00) |
| Home possessions | 0.21 | 0.29 | 0.08*** |
| | (0.94) | (0.99) | (0.00) |
| Observations | 25,748 | 15,199 | 40,947 |

**Table 3:** Summary Statistics of Treatment Effects

Note: This table shows the difference in student characteristics for the treated and non-treated regions, pre- and post-treatment. To test for statistical significance, we use a heteroskedasticity robust t-test on the mean.

# 4 Empirical Strategy

## 4.1 Identification Strategy

In this study we examine the effect of educational quality on test scores. To resolve for potential endogeneity problems in estimating this effect, we will examine the effects of a policy change implemented in the United Kingdom. Because of stagnant results in mathematics and reading proficiency, the administrative region of England has implemented changes to the General Certificate of Secondary Education (GCSE) qualifications, as described above (OECD, 2019). The changes in the GCSE qualifications only apply to the administrative region of England, and do not apply to other regions of the the United Kingdom (Northern Ireland, Scotland and Wales). Furthermore, since the GCSE qualifications were implemented as of 2015, students who took the PISA-test in 2015 are not affected by the implementation of the new regulation, whereas the students taking the test in 2018 and 2022 are affected. Naturally, therefore, the implementation of GCSE lends itself to a Difference-in-Difference (DiD) estimation.

Ideally, the DiD regression would look as follows:

$$y_{i,t} = \alpha + \beta_1 \times \text{Treat}_{i,t} + \beta_2 \times \text{Post}_{i,t} + \beta_3 \times \text{Treat}_{i,t} \times \text{Post}_{i,t} + \delta_i + \varepsilon_{i,t}. \tag{1}$$

In this specification $Treat_{i,t}$ is a dummy variable indicating if a region is treated by changes in GCSE regulation. It takes a value of one for England, and takes a value of zero for all other regions in the United Kingdom. $Post_{i,t}$ is also a dummy variable, indicating if the observation is post-treatment. This variable takes a value of zero for wave 2015 and a value of one for waves 2018 and 2022. The variable $\delta_i$ represents individual fixed effects and $\varepsilon_{i,t}$ is the idiosyncratic error term.

In the above, we would control for individual-specific covariates to remove all possible fixed effects at the level of the individual student and school. The PISA-data, however, do not allow such a construction, since it makes use of a repeated cross-section. Therefore, instead of leveraging fixed-effects estimation to get rid of individual-specific effects, we aim to largely control for school and student characteristics by adding a vector of control variables:

$$y_{i,t} = \alpha + \beta_1 \times \text{Treat}_{i,t} + \beta_2 \times \text{Post}_{i,t} + \beta_3 \times \text{Treat}_{i,t} \times \text{Post}_{i,t} + \boldsymbol{\delta} \times \boldsymbol{x}'_{i,t} + \varepsilon_{i,t}. \tag{2}$$

This specification includes the vector of control variables, $x_{i,t}$, containing variables pertaining to the school (e.g. teaching hours, teacher quality, class size, public versus private school) and variables relating to the individual student background (e.g. socioeconomic status, migration background, parental education). The validity of a Difference-in-Difference estimation depends on the assumption of parallel trends. Because of data availability, we cannot explicitly test for parallel trends prior to 2015. We can, however, control for heterogeneity in composition of treatment and control groups to attribute most credibly the effect to the treatment and not to unobservables.

### 4.2 Evolution of Test Scores

There is an underlying reason for applying a DiD-analysis that becomes apparent from the dataset. Table 4 shows the evolution over time of math and reading scores. The table makes a distinction between score pre-treatment (2015 wave) and post-treatment (2022 wave), reported separately for the treated group (England) and the control group (the other regions, i.e. Scotland, Northern Ireland and Wales).

We test for statistical significance in test scores using a heteroscedasticity robust t-test on the mean. We observe that test scores have reduced over time in all groups and both for reading and math. These differences in test scores pre and post treatment are also statistically significant for both groups. This statistical significance indicates a natural evolution over time. For example, the reduction in test scores observed for math and reading post treatment can be explained by the effects of the Covid-19 pandemic.

It is, however, a different observation that is more striking, and justifies a DiD-analysis, which is the size of the difference. The size of the difference (pre and post treatment) differs remarkably when comparing England with the other regions. While math scores only decreased by 5.54 on average in England, the decrease in the other regions is much more pronounced. For reading, the observed effect is the other way around, with more pronounced reductions in England. This evolution might be the result of policy changes, which we aim to test using the DiD.

**Table 4:** Balance test pre/post treatment

| Region | Variable | Pre treatment Mean | Post treatment Mean | Difference (p-value) |
|---|---|---|---|---|
| England | math | 502.04 | 496.50 | -5.54*** |
| | | | | (0.00) |
| | read | 509.20 | 500.48 | -8.72*** |
| | | | | (0.00) |
| Observations | 5,194 | 4,763 | 15,199 | |
| Other regions | math | 485.40 | 473.31 | -12.09*** |
| | | | | (0.00) |
| | read | 486.97 | 484.63 | -2.34* |
| | | | | (0.08) |
| Observations | 8,963 | 8,209 | 25,748 | |

Note: This table shows the difference in mean math and reading scores for the treated and non-treated regions, pre- and post-treatment. To test for statistical significance, we use a heteroskedasticity robust t-test on the mean.

## 5 Results

### 5.1 Base Regression

Table 5 presents the results for our baseline DiD regression, with math scores as an outcome variable. The coefficient on England represents the region fixed-effects for being in England, as compared to the other regions. The coefficients for 2018 and 2022 are time-fixed effects. The coefficients of (causal) interest are the interaction term between being in England and being post treatment. These are indicated with 'England 2018' and 'England 2022'. Furthermore, column 1 presents the results without any further control variables. Column 2 adds school-specific control variables, while column 3 adds student-specific covariates. Our most complete specification is presented in column 4, with both school- and student-specific controls.

Moving to the interpretation of the results, we observe a statistically significant increase in mathematics scores for following education England after the introduction of the GCSE qualifications policy changes. For the cohort of 2018 we do not see that math performance has statistically significantly changed: although the policy was announced in 2015, the complete rollout of the policy took up to 2017. Therefore, the 2018 cohort might not fully be affected by the policy change. However, in 2022, the policy was in place for multiple years, and we observe statistically significant increase in math scores. This observation holds in all possible specifications, and remains in a similar range with and without the inclusion of control variables. For reading scores, the results are already significant in 2018. However, in the specification with student and school-level control variables the decrease loses its significance in 2022.

**Table 5:** Regression Results: DiD - math scores

| | (1) Base | (2) Controls | (3) Controls | (4) Controls |
|---|---|---|---|---|
| England | 16.643*** | 5.084* | 14.982*** | 7.361*** |
| | (11.24) | (2.53) | (10.42) | (3.72) |
| 2018 | 4.505*** | 2.240 | 4.523*** | 3.527* |
| | (3.90) | (1.59) | (3.93) | (2.49) |
| 2022 | -12.089*** | -15.708*** | -4.491*** | -7.528*** |
| | (-9.63) | (-9.91) | (-3.65) | (-4.79) |
| England x 2018 | -2.737 | -1.581 | -3.567 | -3.131 |
| | (-1.31) | (-0.61) | (-1.74) | (-1.23) |
| England x 2022 | 6.546** | 6.542* | 7.608*** | 7.930** |
| | (2.92) | (2.37) | (3.37) | (2.82) |
| Constant | 485.397*** | 537.467*** | 537.093*** | 552.318*** |
| | (616.86) | (132.76) | (107.11) | (75.49) |
| Student controls | No | No | Yes | Yes |
| School controls | No | Yes | No | Yes |
| N | 40947 | 25960 | 34868 | 22235 |
| F | 113.988 | 261.412 | 531.030 | 328.509 |
| r2 | 0.015 | 0.078 | 0.165 | 0.197 |

Note: this table shows the regression results for the responses to the GCSE reform, with PISA math scores as the outcome variable. Column 1 is the base specification without control variables, column 2 includes school level controls, column 3 includes student specific controls and column 4 includes both types of control variables. The rows England 2018 and England 2022 represent the coefficients of (causal) interest, expressing the effect of being a student in England after the policy reform. t-statistics between parenthesis, significance stars: *** 0.001, ** 0.01, * 0.05

**Table 6:** Regression Results: DiD - reading scores

| | (1) Base | (2) Controls | (3) Controls | (4) Controls |
|---|---|---|---|---|
| England | 22.235*** | 10.809*** | 21.878*** | 14.329*** |
| | (14.19) | (5.09) | (14.29) | (6.85) |
| 2018 | 9.453*** | 4.820** | 9.939*** | 6.916*** |
| | (7.15) | (2.97) | (7.49) | (4.23) |
| 2022 | -2.340 | -4.461** | 5.991*** | 4.118* |
| | (-1.73) | (-2.61) | (4.46) | (2.42) |
| England x 2018 | -13.887*** | -10.815*** | -15.487*** | -12.829*** |
| | (-6.05) | (-3.84) | (-6.81) | (-4.56) |
| England x 2022 | -6.383** | -7.041* | -5.522* | -5.390 |
| | (-2.69) | (-2.43) | (-2.28) | (-1.82) |
| Constant | 486.968*** | 538.418*** | 516.880*** | 534.272*** |
| | (565.91) | (121.11) | (92.83) | (65.50) |
| Student controls | No | No | Yes | Yes |
| School controls | No | Yes | No | Yes |
| N | 40947 | 25960 | 34868 | 22235 |
| F | 74.204 | 191.530 | 437.090 | 260.161 |
| r2 | 0.009 | 0.057 | 0.135 | 0.159 |

Note: this table shows the regression results for the responses to the GCSE reform, with PISA reading scores as the outcome variable. Column 1 is the base specification without control variables, column 2 includes school level controls, column 3 includes student specific controls and column 4 includes both types of control variables. The rows England 2018 and England 2022 represent the coefficients of (causal) interest, expressing the effect of being a student in England after the policy reform. t-statistics between parenthesis, significance stars: *** 0.001, ** 0.01, * 0.05

### 5.2 Robustness Checks

If a true treatment effect exists for the implementation of GCSE requirement changes in England, the only region that should be affected is England. The implemenation of GCSE requirements should not have an effect on the test scores for Scotland, Wales and Northern Ireland, since these regions were not affected by the policy change. To verify whether this requirement holds, Tables 7 and 9 replace the true treated region with the placebo regions Northern Ireland, Wales and Scotland for our most complete specification including all control variables. Since the policy has not affected these regions, we should not observe interaction effects of being in this region, posterior to the implementation of the treatment. Tables 7 and 9 largely confirm this intuition, with most of the interaction effects being insignificant. Furthermore, there is also no clear trend or sign for the interaction effects. This indicates that there are not persistent effects of the policy in the untreated regions of the United Kingdom, justifying our use of DiD regression. As a critical note, however, we must add here that there might have been educational policies implemented that were specific to Northern Ireland, Wales or Scotland. These can then reduce a clean identification, but are difficult to control for using the available data.

**Table 7:** Robustness: DiD - math scores

| | (1) | (2) | (3) |
|---|---|---|---|
| | Northern Ireland | Wales | Scotland |
| Region x 2018 | 4.093 | 6.414* | -5.574 |
| | (1.43) | (2.48) | (-1.94) |
| Region x 2022 | 3.716 | 1.934 | -16.062*** |
| | (1.16) | (0.58) | (-5.56) |
| Constant | 552.737*** | 555.803*** | 557.518*** |
| | (75.53) | (76.40) | (76.77) |
| Student controls | Yes | Yes | Yes |
| School controls | Yes | Yes | Yes |
| N | 22235 | 22235 | 22235 |
| F | 329.800 | 329.326 | 328.617 |
| r2 | 0.197 | 0.198 | 0.196 |

Note: this table shows the robustness check results for the responses to the GCSE reform, with PISA math scores as the outcome variable. The columns indicate the placebo region used for 'hypothetical' treatment. All specifications contain school and student controls. We only show the coefficients of interest. t-statistics between parenthesis, significance stars: *** 0.001, ** 0.01, * 0.05

**Table 8:** Robustness: DiD - math scores

| | (1) | (2) | (3) |
|---|---|---|---|
| | Northern Ireland | Wales | Scotland |
| Region x 2018 | 9.542** | 4.381 | 5.403 |
| | (2.92) | (1.46) | (1.67) |
| Region x 2022 | -1.581 | 1.392 | 0.752 |
| | (-0.46) | (0.38) | (0.24) |
| Constant | 536.905*** | 539.328*** | 541.408*** |
| | (65.62) | (66.59) | (66.69) |
| Student controls | Yes | Yes | Yes |
| School controls | Yes | Yes | Yes |
| N | 22235 | 22235 | 22235 |
| F | 260.568 | 269.237 | 253.473 |
| r2 | 0.159 | 0.164 | 0.157 |

Note: this table shows the robustness check results for the responses to the GCSE reform, with PISA reading scores as the outcome variable. The columns indicate the placebo region used for 'hypothetical' treatment. All specifications contain school and student controls. We only show the coefficients of interest. t-statistics between parenthesis, significance stars: *** 0.001, ** 0.01, * 0.05

## 5.3 Heterogeneity

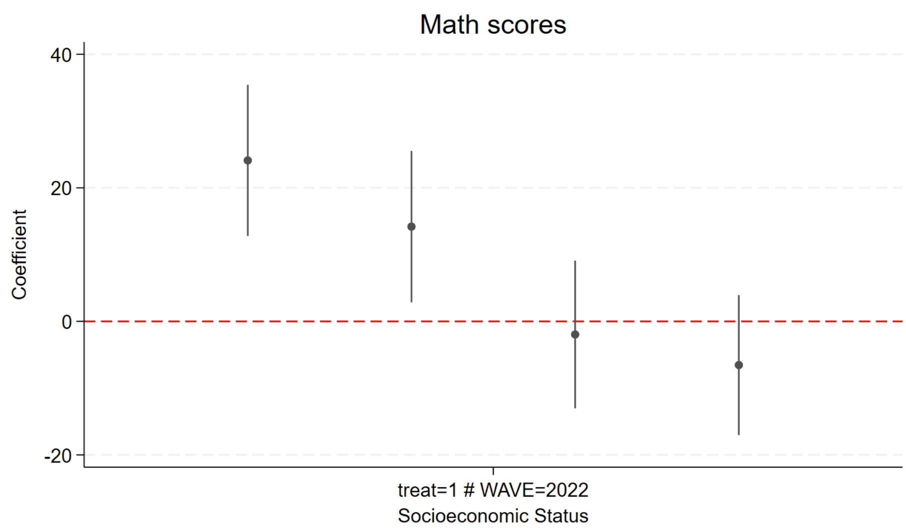
It is important to consider heterogeneity in student responses. Factors like socioeconomic status, parental education or gender might affect the results. In previous analysis, we included these covariates by controlling for them. In this section, we will explicitly disentangle heterogeneity in student performance by examining subgroups separately.

### 5.3.1 Socioeconomic status

We start by examining the differential impact of the GCSE reform on students with different socioeconomic status. We carefully divided the whole population of students into four quartiles based on their socioeconomic status. The categorisation is done per wave and region, such that for every wave the 25% most socially disadvantaged students are in the first quantile and the 25% most advantaged students are in the fourth quantile. Afterwards, we estimated Equation 2 separately for each subgroup of students. Full regression results are printed in the Appendix in Tables 11 and 12.

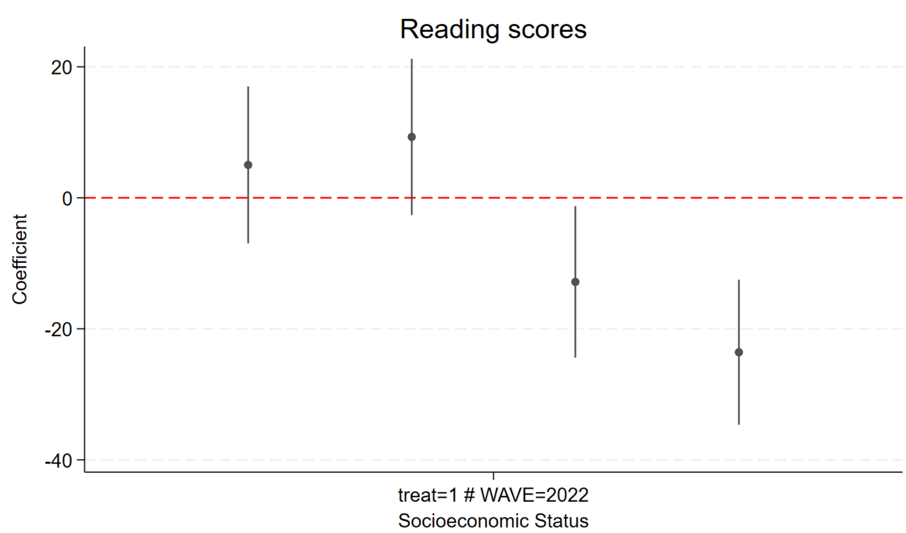
This section only shows the coefficients of causal interest, which are represented in Figures 4 and 5. The coefficients in these figures represent the interaction between being in the treatment group (England) post-treatment (2022) for each quartile of socioeconomic status. From left to right, socioeconomic status increases. Clearly, the effect of the GCSE reform differs by socioeconomic background. The most socially disadvantaged students tend to benefit from the reform most. That is, two lower quartiles of the population, the GCSE reform has a positive and significant impact on their math scores, and does not affect reading scores significantly. Students from in the higher socioeconomic classes, however, see their test scores decrease after the implementation of the new policy, where this decrease is most pronounced for reading.

**Figure 4:** Evolution of treatment effect: Math scores




Note: this Figure shows the beta-coefficients and corresponding 95% confidence interval for the DiD regression with math scores as an outcome variable. The coefficients represent the interaction between being in the treatment group (England) post-treatment (2022) for each quartile of socioeconomic status. From left to right, socioeconomic status increases.

**Figure 5:** Evolution of treatment effect: Reading scores




Note: this Figure shows the beta-coefficients and corresponding 95% confidence interval for the DiD regression with reading scores as an outcome variable. The coefficients represent the interaction between being in the treatment group (England) post-treatment (2022) for each quartile of socioeconomic status. From left to right, socioeconomic status increases.

### 5.3.2 Parental Education

Apart from socioeconomic status, parental education is also documented to affect student performance (Tamayo Martinez et al., 2022). Therefore, it might be that PISA reading and math scores also change differently for students with parents with different education background. Table 9 displays the results, for subgroups based on the parental education. The columns indicate the subgroups and outcome variable. In column headers, math and read refer to PISA scores on mathematics and reading, while high and low refer to higher or lower education of the parents. In line with what we found for socioeconomic background, we again find evidence pointing to the fact that disadvantaged students are beneficially treated by the GCSE reform. More specifically, for students of which the parents have low education, the implementation of the reform in England has increased their performance on mathematics in 2022, relative to the regions in which no reform was introduces. For this subgroup, reading scores have remained unchanged. Furthermore, for students of which the parents have followed higher education, their reading scores decreased in 2022 with no statistically significant effect increase mathematics score.

**Table 9:** Heterogeneity: Parental education

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | High (math) | Low (math) | High (read) | Low (read) |
| Region | 7.677** | 7.004* | 15.376*** | 12.451*** |
| | (3.11) | (2.14) | (5.91) | (3.59) |
| 2018 | 0.670 | 7.590** | 5.784** | 7.914** |
| | (0.38) | (3.27) | (2.81) | (2.96) |
| 2022 | -4.247* | -14.959*** | 6.536** | -1.848 |
| | (-2.09) | (-6.03) | (2.97) | (-0.69) |
| England x 2018 | -2.130 | -4.632 | -13.082*** | -12.130** |
| | (-0.67) | (-1.10) | (-3.72) | (-2.62) |
| England x 2022 | 1.524 | 20.337*** | -10.077** | 3.846 |
| | (0.44) | (4.20) | (-2.75) | (0.76) |
| Constant | 548.072*** | 460.777*** | 560.176*** | 421.186*** |
| | (29.39) | (34.94) | (27.19) | (28.37) |
| Student controls | Yes | Yes | Yes | Yes |
| School controls | Yes | Yes | Yes | Yes |
| N | 14381 | 7854 | 14381 | 7854 |
| F | 220.751 | 72.732 | 185.742 | 52.927 |
| r2 | 0.199 | 0.140 | 0.169 | 0.108 |

Note: this table shows the regression results for the heterogeneity in responses to the GCSE reform. The columns indicate the subgroups and outcome variable. Math and read refer to PISA scores on mathematics and reading, while high and low refer to higher or lower education of the parents. Parents are identified as higher educated when they followed at least 14 years of education. t-statistics between parenthesis, significance stars: *** 0.001, ** 0.01, * 0.05

### 5.3.3 Gender

Finally, we consider gender as a potential driver of school performance. A recent meta-study confirmed how female students in recent years outperform their male collegues in almost all domains and at various levels of the educational ladder (Wrigley-Asante et al., 2023). Table 10 shows the our results for the distinction between male and female school performance using the subsamples of male or female students in the United Kingdom. This table follows the same structure as the previous ones, with columns indicating the specific subsample under study. The coefficients of interest are the coefficients indicating the effects of following school in England after the introduction. Our results are less clear for this variable, with female students in England having higher math scores and reduced reading scores after the implementation of the policy, compared to students in other regions. There is, thus, no general trend in the performance of female students after the policy changed. For the subgroup of male students, the 2022 coefficients are ofverall insignificant.

**Table 10:** Heterogeneity: Parental education

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Female (math) | Male (math) | Female (read) | Male (read) |
| Region | 2.803 | 12.113*** | 18.472*** | 10.163*** |
| | (1.06) | (4.12) | (6.54) | (3.29) |
| 2018 | 5.705** | 1.174 | 11.767*** | 1.637 |
| | (3.03) | (0.55) | (5.40) | (0.67) |
| 2022 | -9.256*** | -5.401* | 7.724*** | 0.528 |
| | (-4.46) | (-2.27) | (3.42) | (0.21) |
| England x 2018 | -7.557* | 1.794 | -22.519*** | -2.751 |
| | (-2.18) | (0.48) | (-5.84) | (-0.67) |
| England x 2022 | 8.981* | 6.429 | -12.435** | 1.538 |
| | (2.37) | (1.55) | (-3.10) | (0.35) |
| Constant | 545.470*** | 544.500*** | 563.340*** | 520.544*** |
| | (55.99) | (50.21) | (51.53) | (43.40) |
| Student controls | Yes | Yes | Yes | Yes |
| School controls | Yes | Yes | Yes | Yes |
| N | 11437 | 10798 | 11437 | 10798 |
| F | 190.131 | 157.191 | 157.857 | 111.407 |
| r2 | 0.203 | 0.183 | 0.169 | 0.138 |

Note: this table shows the regression results for the heterogeneity in responses to the GCSE reform. The columns indicate the subgroups and outcome variable. Math and read refer to PISA scores on mathematics and reading, while male and female refer to higher or lower education of the parents. t-statistics between parenthesis, significance stars: *** 0.001, ** 0.01, * 0.05

## 6 Discussion, policy recommendation and conclusion

Schools and teachers matter for developing individual skills. The reforms made to educational institutions such as changes to assessments and examination techniques thus play a vital role in improving skills of 15 year old students. In our analysis, we empirically examined if such a reform has an impact on student grades in the UK. To do so we used three waves of the PISA-2015, 2018, 2022. We focused our analysis on the UK because the data was sufficiently balanced for most variables, but also because there was within country regional variation. We employ a Difference-in-Difference strategy to uncover the different effects of the reforms to the GCSE academic qualification in the UK. Our findings show that, in the post reform period (mostly for 2022), we see a positive significant effect for the math scores in England as against Northern Ireland, Wales and Scotland. This positive impact also holds after including a vector of student and school specific controls. We do not see any such significant results for reading scores robust to all controls. We account for heterogeneity, by classifying four quartiles for socio-economic status in increasing order. We observe that the effect of the math score is pronounced for the students with lower socio-economic status.

For students of low socio-economic backgrounds, the policy recommendation would be to focus on lowering the gap in performance between the students. The main policy implication of our analysis is that teaching in schools should not just be for testing purposes but rather for strengthening the knowledge and skills. And if teaching is for testing purposes, then a very good policy would focus on assistance to students in the lower and middle tiers of schools. This means that more resources are directed towards students who come from disagvantaged backgrounds so that they can gradually improve their progress into the middle and high tiers.

Such an educational policy should be designed with a focus on equality and contribute towards strengthening academic qualification skills for students with disadvantaged socio-economic backgrounds.

## References


Carroll, J. B. (1963). A model of school learning. *Teachers college record*, *64*(8), 1–9.

Equality and Human Rights Commission. (2021). *Protected characteristics*. https://www.equalityhumanrights.com/equality/equality-act-2010/protected-characteristics

Fuller, K., & Stevenson, H. (2019). Global education reform: Understanding the movement.

Hanushek, E. A. (1995). Interpreting recent research on schooling in developing countries. *The world bank research observer*, *10*(2), 227–246.

Hanushek, E. A. (2020). Education production functions. In *The economics of education* (pp. 161–170). Elsevier.

Hanushek, E. A., & Woessmann, L. (2012). Do better schools lead to more growth? cognitive skills, economic outcomes, and causation. *Journal of economic growth*, *17*, 267–321.

OECD. (2010). The high cost of low educational performance. *The long run economic impact of improving PISA outcomes. Programme for International Student Assessment OECD.*

OECD. (2019, September). *Education policy outlook 2019: Working together to help students achieve their potential*. https://doi.org/10.1787/2b8ad56e-en

Ritchie, S. J., & Bates, T. C. (2013). Enduring links from childhood mathematics and reading achievement to adult socioeconomic status. *Psychological science*, *24*(7), 1301–1308.

Sahlberg, P. (2012). Global educational reform movement is here. *Blog. Online: https://pasisahlberg.com/global-educationalreform-movement-is-here [Stand: 2022-12-27].*

Tamayo Martinez, N., Xerxa, Y., Law, J., Serdarevic, F., Jansen, P. W., & Tiemeier, H. (2022). Double advantage of parental education for child educational achievement: The role of parenting and child intelligence. *European Journal of Public Health*, *32*(5), 690–695. https://doi.org/10.1093/eurpub/ckac044

Todd, P. E., & Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. *The Economic Journal*, *113*(485), F3–F33.

UK.GOV. (2018). *Get the facts: Gcse reform*. https://www.gov.uk/government/publications/get-the-facts-gcse-and-a-level-reform/get-the-facts-gcse-reform

Worthington, A. C. (2001). An empirical survey of frontier efficiency measurement techniques in education. *Education economics*, *9*(3), 245–268.

Wrigley-Asante, C., Ackah, C. G., & Frimpong, L. K. (2023). Gender differences in academic performance of students studying science technology engineering and mathematics (stem) subjects at the university of ghana. *SN Social Sciences*, *3*(1). https://doi.org/10.1007/s43545-023-00608-8

# Appendix

### Full regression results: Socioeconomic status

**Table 11:** Heterogeneity: socioeconomic status - math

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Q1 | Q2 | Q3 | Q4 |
| Region | -0.290 | -1.619 | 13.733*** | 15.883*** |
| | (-0.07) | (-0.38) | (3.53) | (4.37) |
| 2018 | 5.892* | 6.717* | 7.556** | 1.346 |
| | (2.05) | (2.40) | (2.69) | (0.48) |
| 2022 | -17.378*** | -11.398*** | -10.102** | 1.909 |
| | (-5.69) | (-3.65) | (-3.15) | (0.58) |
| England x 2018 | -2.756 | 1.184 | -8.220 | -4.287 |
| | (-0.52) | (0.22) | (-1.62) | (-0.91) |
| England x 2022 | 24.106*** | 14.188* | -1.964 | -6.550 |
| | (4.17) | (2.45) | (-0.35) | (-1.22) |
| Constant | 484.015*** | 543.022*** | 549.264*** | 400.054*** |
| | (28.05) | (37.73) | (35.69) | (14.08) |
| Student controls | Yes | Yes | Yes | Yes |
| School controls | Yes | Yes | Yes | Yes |
| N | 5241 | 5428 | 5564 | 6002 |
| F | 30.084 | 30.346 | 32.665 | 45.378 |
| r2 | 0.089 | 0.080 | 0.081 | 0.096 |

Note: this table shows the regression results for the responses to the GCSE reform, with PISA math scores as the outcome variable. Each column represents a subgroup of socioeconomic status, with Q1 representing the lowest status. The rows England 2018 and England 2022 represent the coefficients of (causal) interest, expressing the effect of being a student in England after the policy reform. t-statistics between parenthesis, significance stars: *** 0.001, ** 0.01, * 0.05

Table 12: Heterogeneity: socioeconomic status - reading

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Q1 | Q2 | Q3 | Q4 |
| Region | 8.513 | 0.372 | 20.873*** | 25.322*** |
| | (1.95) | (0.09) | (5.08) | (6.47) |
| 2018 | 9.866** | 6.180 | 12.177*** | 6.876* |
| | (3.01) | (1.91) | (3.74) | (2.09) |
| 2022 | -1.953 | -1.537 | 1.104 | 12.426*** |
| | (-0.59) | (-0.45) | (0.32) | (3.52) |
| England x 2018 | -14.443* | -0.595 | -18.126** | -19.001*** |
| | (-2.49) | (-0.10) | (-3.24) | (-3.57) |
| England x 2022 | 5.027 | 9.291 | -12.841* | -23.575*** |
| | (0.82) | (1.53) | (-2.18) | (-4.17) |
| Constant | 437.776*** | 524.468*** | 550.784*** | 412.425*** |
| | (23.24) | (34.01) | (32.80) | (12.39) |
| Student controls | Yes | Yes | Yes | Yes |
| School controls | Yes | Yes | Yes | Yes |
| N | 5241 | 5428 | 5564 | 6002 |
| F | 24.184 | 26.147 | 28.298 | 41.547 |
| r2 | 0.072 | 0.070 | 0.073 | 0.088 |

Note: this table shows the regression results for the responses to the GCSE reform, with PISA math scores as the outcome variable. Each column represents a subgroup of socioeconomic status, with Q1 representing the lowest status. The rows England 2018 and England 2022 represent the coefficients of (causal) interest, expressing the effect of being a student in England after the policy reform. t-statistics between parenthesis, significance stars: *** 0.001, ** 0.01, * 0.05