Teacher Coaching, Educational Triage, and Student Achievement: Evidence from a Mathematics Intervention in Jamaica*

Patrice Anderson[†] Augustine Denteh[‡] Nicholas A. Wright[§]

Abstract

Policymakers are increasingly using teacher coaching programs to improve student achievement in developing countries. This paper evaluates a nationwide intervention that deployed specialized math coaches to train secondary school teachers in Jamaica. We use administrative data on all high schools from 2009 to 2016 and various difference-in-differences methods to estimate the causal effect of this intervention on student achievement. We find that the program mitigated educational triage by increasing the number of math test takers. However, the intervention also significantly worsened student math performance, with the largest effects concentrated at higher-quality schools with better teachers and greater resources.

Keywords: Teacher Coaching, Educational Triage, Education Quality, Mathematics

JEL No: I20, I21, I28, J24

**Date:* The current draft is as of December 9, 2022. We would like to thank Katy Ann Bergstrom, Joshua Goodman, Adrienne Lucas, Isaac Mbiti, Christine Moser, Tavneet Suri, and participants at the Association of Education Finance and Policy and Florida International University research seminars for their insightful comments. We are equally grateful to the officials at the Ministry of Education in Jamaica for data and institutional support. The views expressed in this paper do not reflect the opinions of the organization from which the data was collected. Any remaining errors are our own.

[†]Department of Economics, Western Michigan University, Kalamazoo, MI.

[‡]Department of Economics, Tulane University, New Orleans, LA.

[§]Corresponding Author: Department of Economics, Florida International University, Miami, FL 33199, USA. Email: nicwrigh@fiu.edu.

1. Introduction

As access to schooling in developing countries improves, policymakers are increasingly focused on enhancing educational quality and attainment (World Bank, 2018). As such, governments are devoting more public resources to professional development (PD) programs that provide various on-the-job training activities for teachers (Filmer and Rogers, 2018; OECD, 2021).[1] Teacher PD programs are appealing because teachers play a critical role in improving students' academic and later-life outcomes (Rockoff, 2004; Hanushek and Rivkin, 2010; Chetty et al., 2011, 2014). Moreover, educators in developing economies are often not sufficiently trained or equipped to teach effectively.[2] Despite the proliferation of teacher PD programs, the empirical evidence of their effectiveness is limited and mixed (Popova et al., 2022).

This paper evaluates a national professional development program that targeted high school math teachers in Jamaica. In 2014, the government of Jamaica implemented a large-scale intervention they dubbed the "Operation Turn Around" (OTA) program to address the perennial poor performance in the standardized high school math exit examination. The OTA intervention deployed specialists to provide in-person instructional training for math teachers with the aim of improving student achievement in the Caribbean Secondary Education Certificate (CSEC) examination. The OTA program involved a series of intensive training activities in the teachers' schools throughout the academic year. The math coaches also provided targeted feedback to improve the teaching styles and approaches of existing teachers.

Jamaica is an interesting case study for examining the impact of teacher-specific PD programs because high school teachers typically serve as gatekeepers in determining access to the exit exam. A unique feature of the Jamaican high school educational system is that only *recommended* grade 11 students can take the standardized exit exam (Evans, 2001). That is, among the eligible graduating cohort, teachers nominate students they deem prepared to take the exit exam in each subject. This student recommendation system ensures that high schools maintain their positions in national rankings and unprepared students do not pay for examinations they are likely to fail. As such, the

[1] The Indian government spent $1.2 billion on teacher PD programs between 2012 and 2017 (Popova, 2021). Similarly, sixteen sub-Saharan African countries had a PD program for secondary school teachers in 2019, while teachers in Mexico devote an average of 23 days to in-service training each year (Taylor et al., 2019; Popova, 2021).

[2] For instance, about 7 to 51 percent of Grade 4 teachers lack the minimum content knowledge in Mathematics across seven African countries (Bold et al., 2017).

educational system provides incentives for teachers to engage in educational triage, where teachers discourage academically weaker students from taking the exit exam (Gilligan et al., 2022). Therefore, this paper provides new evidence on the effectiveness of a national teacher PD program in a developing country where teachers can influence secondary school completion rates.[3]

Using administrative data on the universe of high schools from 2009 to 2016 and the doubly robust difference-in-differences (DID) method, we estimate the causal effect of the OTA program on access to the math exit exam and students' subsequent performance. The doubly robust DID methods provide more credible estimates of the treatment effect because they do not make functional form assumptions that restrict effect heterogeneity (Sant'Anna and Zhao, 2020). Such DID methods also account for covariate imbalances between the treatment and control groups related to the outcomes of interest. We provide strong supporting evidence that the treated and untreated schools followed a similar path before the OTA policy, and these trends would have continued in the absence of the intervention. In addition, we demonstrate the robustness of our main results to alternative outcome measures, estimation approaches, level of analysis (school vs. student level), the composition of the control group, and a randomization inference test.

Several interesting findings emerge from our analysis. First, we find positive unintended effects of the OTA intervention on student access to the CSEC math exam as measured by the number of students taking the exam. Our results show that the intervention significantly increased the number of math test takers by about 14 students, representing 12.7 percent of the pre-intervention average number of test takers. We also find a limited or practically small effect of the intervention on the number of absentees on the exam day. As such, while the government designed the OTA program to improve student performance, the policy mitigated educational triage by encouraging teachers to recommend more students to take the math exam. Our findings are consistent with the results in Gilligan et al. (2022), where a pay-for-percentile incentive scheme for teachers in Uganda

[3] Despite increased access to schooling, low completion rates hinder educational attainment. Previous studies have suggested other reasons for low educational attainment or completion rates, including limited school resources, low teacher availability and quality, lack of teacher accountability, and inadequate instructional time (Chaudhury et al., 2006; Hanushek and Rivkin, 2007; Bruns et al., 2011; Glewwe and Muralidharan, 2016; World Bank, 2018; Glewwe et al., 2011).

reduced dropout rates of primary school students. Our results suggest that when policymakers are designing teacher PD programs in developing countries, they should factor in the unintended effects on access, especially when teachers play a crucial role in determining which students can attempt the exit exams.

Second, in terms of student achievement, we find that the OTA intervention significantly worsened math performance. Our results show that the intervention reduced the share of students obtaining a passing score by 3.6 percentage points (10.6 percent). The negative effects of the OTA program are concentrated in the upper portions of the grade distribution, where we find that the program reduced the share of students obtaining a high passing score by 22 percent. In addition, when we examine the number of students receiving specific exam grades as an outcome measure, we find that the intervention reduced the number of students obtaining a high passing score by about 3 students (13.5 percent). As such, the program harmed the academic achievement of higher-ability students. However, our results suggest that the adverse effects of the intervention did not spill over to non-math subjects.

Third, when we focus on the lower end of the distribution, we find that the program increased the number of students who failed the math exam (i.e., the number obtaining the bottom three grades) by almost 14. Remarkably, the magnitude of the effect on the number of failures is nearly identical to the overall number of students who were induced into sitting the exam, as discussed above. Our results are consistent with the hypothesis that the OTA intervention led teachers to recommend lower-ability students to take the math exam, partly leading to the negative effect on performance.

Finally, we also investigate whether the OTA intervention had a differential impact on student performance by school quality and resources. We find that the OTA intervention negatively affected higher-achieving students in higher-quality schools. This pattern of results suggests that by asking higher-quality teachers in relatively better schools to modify their proven teaching methods, the teacher coaching intervention negatively impacted student learning and performance in those schools. Together, our results suggest that the increased access to the math exam only partially explains the negative effects on student performance. On the one hand, our results indicate that the program worsened performance in the schools that experienced an increase in test takers regardless of whether the school had historically low or high math performance. On the other hand, in most subgroups where the intervention increased the number of test takers, we do not find evidence of any changes in student performance. As such, the heterogeneous effects provide a nuanced interpretation of our findings because the effects of the intervention interact with school and teacher quality and are not entirely driven by the increased number of math test takers on the exit exam.

The main contribution of this paper is that it is the first to evaluate the OTA intervention in Jamaica and one of a few evaluations of a national-level teacher PD program. Popova et al. (2022) reports that there are no rigorous empirical evaluations of 139 large-scale, government-funded PD programs that have been implemented across 14 countries.[4] In lower-income countries, the evaluation of teacher PD programs is sparse and has yielded mixed results. For example, small-scale evaluations of primary school literacy PD interventions in Liberia and Uganda find large positive impacts on student reading achievement (Piper and Korda, 2011; Kerwin et al., 2015). In contrast, a randomized evaluation of a large-scale national teacher training policy in China found no impact on average performance but a large negative effect on the achievement of students who were taught by more qualified teachers (Loyalka et al., 2019).[5] One key innovation of this paper relative to previous studies is that we use administrative data on the academic performance of all high school students who took a national high-stakes exit exam. Other studies measure performance using a researcher-designed assessment (e.g., Loyalka et al., 2019). Thus, our findings are more likely to reflect program effects in a longer time horizon than performance in short-term low-stakes tests. Consequently, this paper contributes to the literature by providing credible evidence on the impacts

[4]There is extensive literature on evaluating PD programs in high-income countries. However, most studies have focused on small-scale programs. See Yoon et al. (2007), Walter and Briggs (2012), Gersten et al. (2014) and Fryer Jr (2017) for reviews. Those studies generally find mixed evidence on the effectiveness of teacher PD programs, partly because only a handful use credible research designs such as randomized evaluations or quasi-experimental methods.

[5]Other studies have also found negative impacts of teacher PD programs on student performance. Berlinski and Busso (2017) examines a randomized control trial in Costa Rica where teachers were trained for four weeks to develop instructional approaches incorporating more active learning for high school students in Mathematics. They find that the intervention reduced student achievement, with larger impacts on the best students. Additional studies on the impact of coaching programs in developing countries include Albornoz et al. (2019), Cilliers and Taylor (2017), Harvey (1999), Piper and Zuilkowski (2015), Sailors et al. (2014), and Yoshikawa et al. (2015).

of a multi-year national teacher PD program on student achievement in a high-stakes math exam.

As a secondary contribution, we advance the nascent literature examining the features of PD programs that make them effective at improving student achievement. Some studies argue that teacher PD programs are effective when they possess certain qualities, such as being content-focused, intensive, of sustained duration, incorporating active learning, and follow-up efforts (Cohen and Hill, 1998; Darling-Hammond et al., 2009; Desimone and Garet, 2015; Hill, 2007; Supovitz, 2001; Walter and Briggs, 2012; Yoon et al., 2007). However, these recommendations on what constitutes an effective teacher PD program are mainly theoretical, with limited empirical evidence to support them (Ball et al., 2008; Elmore, 2002; Scher and O'Reilly, 2009). A notable exception is Popova et al. (2022), which finds that teacher PD programs are more effective if they have a single-subject focus, include face-to-face training sessions, and are linked to teacher career incentives. Our paper contributes to this literature by examining the effectiveness of teacher PD programs in a context where student performance on standardized exams is important in determining school reputation. Schools in such systems have an incentive to exempt less prepared students from exit exams to maintain their standings in national school rankings (Marshall, 2013; Gilligan et al., 2022).

2. Background and the Operation Turn Around program

Students in Jamaica are admitted into secondary schools based on their performance in the standardized primary school exit examination, the Grade Six Achievement Test (GSAT). The secondary school system is divided into the junior (lower) and senior (upper) secondary phases. The junior phase consists of grades 7-9, after which students proceed to grades 10-11 in the senior phase. In the junior secondary phase, students take various subjects, including Mathematics, General Science, and Language Arts. There are no external examinations at the end of the junior secondary phase. Instead, students are given the opportunity to choose the subjects they would like to study at the senior secondary stage.[6]

[6]The subjects a student is approved to study when they become seniors are generally based on their preferences, past academic performance, and capacity constraints.

The senior secondary phase, which is the focus of our study, is designed to prepare students for the CSEC high school exit exam. The CSEC exam is a regionally standardized exam administered by the Caribbean Examination Council (CXC) to mark high school completion. All colleges consider students' performance on the CSEC exam during their admission process. Out of the 33 CSEC subjects offered by the CXC, senior secondary students typically study about 8 subjects (English and Mathematics included).[7] The CXC uses a six-point grading scheme for CSEC exams (grades 1, 2, 3, 4, 5, and 6), with grade 1 being the highest quality performance and grade 6 the lowest. However, for all CSEC subjects, regional universities and the CXC defines grades between 1 and 3 as passing scores.

It is important to note that not all grade 11 students studying a particular subject ultimately take the CSEC exam. The subject teacher, in consultation with the school principal, determines which grade 11 students will be recommended to sit the CSEC exam for each subject (Evans, 2001).[8] The student recommendation system is based on students' performance in mock tests and other internal assessments. Some students may also refuse to sign up for the CSEC exam because they cannot pay the required registration fees. For example, based on available data, the percentages of eligible grade 11 students who took the CSEC exams in 2015, 2017, and 2018 were 88 percent, 91 percent, and 92 percent, respectively (Ministry of Education, Youth, and Information, 2019).

Jamaica's Ministry of Education, Youth, and Information (MOE) has expressed a longstanding concern regarding the low level of math proficiency among high school graduates. Based on statistics reported by the MOE, less than 45 percent of students who took the CSEC math exams between 2000 and 2013 passed each year, with an average pass rate of about 36 percent over the period (Bourne, 2019). The perennial poor performance in math has gained national attention, making it a key policy and political concern. One general requirement for acceptance into tertiary education institutions is passing a minimum of 5 CSEC subjects, including Mathematics and English (Wright, 2021). Some employers apply a similar requirement in the labor market for high

[7]The subject choices are generally based on the student's interest and the career path they wish to pursue.

[8]For example, consider the case of John, a student who attended classes for three subjects during senior secondary school (Mathematics, Chemistry, and Biology). However, John might only be allowed to take the math exam because the instructors in Chemistry and Biology chose not to recommend him for those exams.

school graduates in Jamaica.

The government of Jamaica has argued that the small and declining number of qualified math teachers in secondary schools is one of the critical reasons for poor math performance in the CSEC exam.[9] As a policy response, the government introduced the OTA coaching program in 2014 to improve teacher quality and student performance. The stated goal of the OTA program was to improve student performance in math at the senior secondary school level, thereby increasing the number of high school graduates attaining the prerequisite grades for tertiary education. The OTA program deployed math specialists (coaches) to selected public high schools to train the existing teachers that were preparing students for the CSEC math exam. The intervention focused on building teacher quality through professional development and instructional coaching. Schools were assigned one coach who visited the school once per week (or twice for larger schools) to train teachers. The training involved several activities, including observing teachers' lessons and providing feedback, providing demonstration lessons, and organizing various professional development sessions to build teacher capacity. High school principals and heads of math departments also benefited from a special training program to develop their ability to lead the teaching of math.

The OTA program is an all-year program, and the selected schools were unchanged over the three-year post-intervention period for which we have data. The MOE identified the public high school to receive the OTA intervention based on several factors, chiefly their historical math performance. While a school's likelihood of treatment is increasing in the number of students attaining severe failing grades (grades 5 and 6) in the CSEC math exam, the MOE does not disclose the exact threshold. Of 251 public high schools, 58 participated in the OTA program with no reports of non-compliance.

[9]For instance, Senator Ruel Reid is quoted in the *Jamaica Observer* stating that "Based on our 2013/2014 Census, only 207 of the 1,784 mathematics teachers deployed in the secondary-education system are fully qualified to teach mathematics to grade 11. This means only 207 have at least a bachelor's degree in mathematics teaching" (Jamaica Observer, 2016). Moreover, the MOE reports that about 490 Mathematics and Science teachers left the secondary school education system in 2014 and 2015.

3. Data

We use administrative data from Jamaica's Ministry of Education, Youth, and Information on the universe of students taking exams at public and private high schools during the 2009-2016 CSEC exam years. The data contains grade information for all subjects taken by each student, basic demographic information such as gender, and the number of subjects taken on the CSEC examination. In addition to student performance data, we also have data on various school-level characteristics, such as the number of students taking the CSEC exam and an indicator for treated schools. The additional school-level information includes school rank, indicators of geographical location (district identifiers and a urban/rural indicator), the capacity of the school, number of teachers, the student-teacher ratio, the school's technical/non-technical status, whether the school runs a shift or whole day system, and whether the school is a same-sex or co-educational institution.

We measure student performance on the standardized CSEC math exam in various ways. For our main analysis, we create three dependent variables indicating whether the student obtained Grades 1 or 2 (high pass), Grade 1, 2, or 3 (pass), and Grades 5 or 6 (severe fail). In additional analyses, we also examine the indicators of whether the student obtained the highest pass (Grade 1) and whether the student failed the exam (Grades 4, 5, or 6). Note that the indicators for *pass* and *severe fail* correspond to how tertiary institutions interpret high school grades for admission purposes.

To investigate how the OTA program affected access to the CSEC exam, we created two measures of access: the number of students taking the exam at each school and the number of students absent on the test day. Finally, we restrict the primary analysis sample to public high schools. The final sample contains a panel data set of 251 public schools and 1,684 school-year observations.

4. Methods

This section presents our empirical design and discusses the estimation methods. We use the differences-in-differences design to estimate the causal effect of the OTA intervention on our outcomes of interest. To motivate our empirical strategy, consider the standard two-way fixed effects (TWFE) regression model typically used in the DID literature:

$$Y_{st} = \beta_0 + \beta_1 D_s + \tau_{TWFE}(D_s * T_t) + \beta_2 X_s + \gamma_t + \varepsilon_{st}, \tag{1}$$

where Y_{st} is the outcome of interest for school s in year t; $D_s = 1$ if school s received the OTA intervention and $D_s = 0$ otherwise; T_t is an indicator for the post-treatment period; γ_t contains examination year fixed effects to account for general time trends in academic performance; and the vector X_s contains pre-treatment time-invariant school-level characteristics.[10] The treatment effect of interest is τ_{TWFE}, which captures the average change in the grades of treated students due to the OTA program under the parallel trends (PT) assumption. In this paper, the PT assumption implies that the outcomes of interest in the OTA and non-OTA public schools would have continued following parallel paths without the intervention. There is no direct way to test whether the PT assumption holds, but we provide evidence of no differential pre-treatment trends for all outcomes we study.[11]

The TWFE model in equation (1) delivers the average treated effect on the treated (ATT) schools. However, it is well known that the above model specification imposes additional restrictions beyond the parallel trends assumption when we include covariates in the DID design (Lechner, 2011). Under the standard PT assumption, the above model specification assumes that the covariates are balanced between treatment and comparison groups and are unrelated to the evolution of the outcome. In addition, the TWFE model restricts treatment effect heterogeneity based on the covariates, although we may relax this restriction with a more saturated specification. When

[10] The post-treatment dummy, T_t, is omitted because it is co-linear with the time fixed effects.

[11] We do so by showing that the pre-treatment estimates from an event-study model are close to zero and jointly statistically insignificant.

these restrictions do not hold, the ordinary least squares estimator of τ_{TWFE} in equation (1) does not generally produce the ATT.

Therefore, in this paper, we assume the *conditional* parallel trends assumption and pursue more flexible methods of estimation that overcome the limitations of the TWFE model. The conditional parallel trends assumption modifies the standard assumption to hold conditional on covariates (X_s). The conditional PT assumption states that in the absence of the treatment, the average outcomes of treated and comparison group units would have followed parallel paths conditional on covariates. The conditional PT assumption permits the identification of the ATT even when observed differences between the treated and comparison groups lead to non-parallel trends in the outcomes of interest (Abadie, 2005). Several methods exist to estimate the ATT under the conditional PT assumption, but we use the recently developed doubly robust DID (DR-DID) approach in Sant'Anna and Zhao (2020) as our preferred estimation method. We briefly discuss this doubly robust DID method and refer the reader to Sant'Anna and Zhao (2020) for a detailed discussion of technical identification and estimation issues.

The fundamental feature of the DR-DID estimator is that it combines the strengths of two estimation approaches for the ATT—the outcome regression (OR) and inverse probability weighting methods (IPW). The OR method models the evolution of the outcomes over time, while the IPW method uses the propensity score to compute a weighted average of differences in outcomes over time to get the ATT. Both methods produce consistent estimates of the ATT if the conditional mean outcome function (in the OR method) and the propensity score model (in the IPW method) are correctly specified. The DR-DID method combines the outcome and propensity score models, so the resulting estimand possesses the so-called double robustness property. Here, double robustness means that the DR-DID estimator is consistent for the ATT if at least one (but not necessarily both) of the propensity score model and the outcome model for the comparison group is correctly specified. Thus, the DR-DID method is superior to separately using the OR method or methods based on inverse probability weighting.

For estimation, we can use any estimators of the outcome regression and propensity score models in the components of the DR-DID estimand in Sant'Anna and Zhao (2020). We implement two versions of their doubly-robust DID method. The first version is the *generic* DR-DID, which uses maximum likelihood to estimate the propensity score model and ordinary least squares for the outcome regression model. The second version is the *improved* DR-DID, which estimates the propensity score model using inverse probability tilting and uses weighted least squares for the outcome regression model. Our preferred specification is the improved DR-DID method because it is not only doubly robust for estimating the ATT but also doubly robust for conducting inference. That is, the asymptotic variance of the improved DR-DID method is independent of a correctly specified model for the outcome and propensity score models. Sant'Anna and Zhao (2020) show that both versions of the DR-DID method are consistent for the ATT, locally semiparametrically efficient, and asymptotically normal.

We report heteroskedasticity-robust standard errors clustered at the school level in all estimations. In addition, while our main results use non-OTA public high schools as our baseline control group, we demonstrate the robustness of our findings using an expanded sample that combines untreated private and public schools to form the control group. We discuss those additional results in Section 5.4.

5. Results

In this section, we present our empirical findings on the effect of the OTA program on students' academic performance. Using administrative data from 2009 to 2016, our main analysis uses a school-level panel to examine the impact of the intervention on various measures of access and performance in the math exit examination. This section also presents estimates on the spillover and heterogeneous effects of the program and the results for several robustness checks of our main empirical models.

5.1. Main Impacts of the OTA Program

Table 1 shows the main impact of the OTA program on students' access to and performance in the CSEC mathematics exam. These effects are estimated using the improved doubly robust DID estimation approach. We begin by investigating the effect of the OTA program on the number of students sitting the math exam to document any changes in access to the CSEC exam. Since each teacher must recommend the students who are allowed to take the exit exam for the subject they teach, access to the math exam will increase if the intervention encouraged teachers to increase the number of students they nominate.

The point estimate in column 1 of Panel A shows that the OTA intervention increased the number of students taking the exam at treated schools by about 14 students. This estimate represents a 12.7% increase relative to the pre-treatment average number of math test takers at untreated schools. In addition, the estimate in column 2 show that the OTA program increased the number of students who were absent on exam day by about 0.73 or 12.6% of the corresponding pre-treatment mean. Our findings suggest that while the OTA intervention improved access to the CSEC exam in treated schools, it also increased the level of absenteeism on the day of the exam. However, the magnitude of the increase in absenteeism is practically small relative to the increased number of test takers. Since the coaching program was designed to improve student performance, the unintended effect of the program on access to the CSEC math exam is interesting and suggests that policymakers should consider incorporating such considerations in designing teacher coaching interventions.

In Panel B, we present our findings on the impact of the OTA program on three measures of math performance. In particular, in columns 3-5 we examine the share of students that obtained a high passing score (Grades 1 or 2), a passing score (Grades 1, 2, or 3), and a severe failing score (Grades 5 or 6). The results suggest that the OTA intervention significantly reduced math performance in the upper portions of the grade distribution and increased the share of test takers who obtained a score in the lower tail of the distribution. Specifically, we find that the OTA program reduced the share of students obtaining a high pass by 3.3 percentage points and the share obtaining a passing score by 3.6 percentage points. Relative to the baseline performance in non-OTA schools, these estimates translate to reductions in performance of 22% and 10.6% for High Pass and Pass, respectively. In addition, we find that the program increased the share of test takers who received a severe failing score by 2.9 percentage points or 7% of the pre-treatment mean. While the magnitude of this effect is quite large, it is not statistically significant at conventional levels.

The above measures of student performance combine multiple exam grade levels to reflect how tertiary institutions interpret and use the CSEC exams for admissions decisions. For instance, for any given subject, tertiary institutions typically define a passing score as obtaining Grades 1, 2, or 3, which is our Pass measure. While these outcomes provide a good overview of the aggregated impact of the program, such composite measures of performance mask potential differences in the treatment effects across the overall grade distribution. To shed more light on the effect on each exam grade, Figure 1 displays the estimated impact of the program on the shares of students obtaining each possible grade.[12] The results show that the program significantly reduced the share of students obtaining Grade 1 by 2 percentage points, the share obtaining Grade 2 by 1.3 percentage points, but the program had no impact on the share of students obtaining Grade 3. In contrast, we find that the program increased the share of students obtaining Grades 4 and 5, but had no impact on those obtaining Grade 6. Together, these results suggest that the OTA program likely harmed the higher-ability treated students who under-performed relative to their untreated counterparts. In addition, the program appears to negatively affect the lower-ability students who were marginally induced into taking the exam and likely received a failing grade.

5.2. Spillover Effects on Non-Math Subjects

We now present our results for the impacts of the intervention on students' performance in non-math subjects. While the OTA intervention utilized math specialists to train math teachers, the program's effects might spill over to other subjects. For example, the instructors of other quantitative subjects may feel more comfortable recommending students for these non-math exams. In addition, the

[12] Each point on the graph is estimated from a different model where the outcome variable is the share of students obtaining the corresponding grade on the horizontal axis. Note that these shares reflect the net impact of the program from two main sources: (1) the impact on the students who would have been recommended to sit the exam regardless of the OTA program and (ii) the impact on students who are recommended to sit the exam due to the intervention.

students who are induced into sitting the math exam may reallocate their study times and habits in ways that may impact their performance in other subjects. We examine these spillover effects by estimating the impact of the program on students' performance in other subjects using our preferred improved doubly robust DID method. To avoid analyzing all subjects as one homogeneous group or performing subject-specific analysis, we conduct the spillover analysis at the school-subject level. Specifically, we construct the dependent variable for each school as the fraction of students obtaining specific grades separately for all non-Mathematics subjects taken each year.

Table 2 shows the spillover effects of the OTA program. The estimates indicate that the OTA program had no impact on the number of students sitting all other subjects (math excluded) or their average performance in those exams. Moreover, the null effects of the intervention on student performance persist when we classify all other subjects into STEM and non-STEM groups.[13] The fact that we fail to find any meaningful spillover effects of the OTA program on non-Mathematics subjects is reassuring because it suggests that the program was implemented as planned by the Ministry of Education, with any resulting treatment effects being limited to the targeted subject (Mathematics). It also suggests that our results are not detecting unobserved time trends in student performance since performance in STEM courses likely behaves similarly to performance in math. We evaluate the parallel trends assumption in greater detail in Section 5.4 below.

5.3. Heterogeneous Effects of the OTA Program

We also investigate whether the OTA intervention had a differential effect across several pre-treatment measures of school quality and resources. These pre-treatment characteristics include the average pass rate for math, the average pass rate across all subjects, a proxy for the average class size (i.e., the number of students per teacher), the number of students sitting the math exam, and the number of test takers for all subjects. We generally view the average pass rates for math and all subjects as proxies for teacher and school quality while the other school characteristics capture school resourcefulness. For each school attribute, we group the schools based on whether they have

[13] We find a small spillover effect on absenteeism for STEM subjects at the 10 percent level. This effect is not practically meaningful, especially when viewed in the broader context of all the estimated spillover effects.

values below or above the median and estimate our preferred doubly robust specification on each subgroup.

Table 3 presents the heterogeneous effects of the program across the five school-level pre-treatment characteristics. We find that the OTA program significantly increased the number of students sitting the math exam at schools with low and high math performance (columns 1 and 2), lower performance across all subjects (column 3), and at under-resourced schools (column 6). In terms of student performance, the results in column 1 of Panel B suggest that the program reduced the share of students passing the exam and increased the share of students receiving severe failing grades at institutions with poorer quality teachers. In contrast, column 2 shows that the program reduced the performance of higher-achieving students at institutions with better-quality teachers. Columns 3 and 4 show a similar pattern where the program harmed higher achieving students at better quality institutions and had a larger negative effect on lower achieving students at lower quality schools.

Therefore, the impact of the program on the share of students that passed the math exam is concentrated at the lowest quality schools and for worse teachers. These institutions are likely trying to maximize the math pass rate since they do not have many students with higher quality passes.[14] As such, the program effects are most concentrated on the pass rate and severe failure rate at these institutions. These results suggest that the impact of the intervention interacts with school and teacher quality, with the burden of the intervention falling on higher achieving students at higher quality institutions and lower achieving students at lower quality institutions.

Next, we examine the impact of the intervention based on the level of school resourcefulness. The results suggest that the negative impact of the intervention on student performance is concentrated at schools with historically smaller class sizes (column 5), a larger number of math test-takers (column 8), and a larger number of test-takers for all subjects (column 10). As such, the schools that were historically more likely to recommend more students to sit the CSEC exam faced the largest burden of the intervention.

[14] Recall that the OTA program targeted public schools with historically poor math performance. The grade 1 rate is 4.3% and 24% for treated and untreated schools at the 95^{th} percentile of the Grade 1 distribution.

The above heterogeneous effects suggest two broad findings. First, the negative effects of the OTA program appear to be only partially driven by the increased number of students recommended to sit the math exam (i.e., our access outcome variable). This is because in most subgroups where the OTA program increased the number of test takers in Table 3, we do not find statistically significant changes in student performance. Second, the OTA program harmed higher achieving students at higher quality institutions and lower achieving students at lower quality institutions. One likely explanation is that the OTA program coaches guided teachers to abandon their proven teaching methods that usually worked for their cohort of students. However, if teachers in the higher quality treated schools had figured out a way to help their students perform relatively well, then any drastic deviations from their usual teaching approaches could potentially harm student learning and performance in such schools. For example, if teachers in the higher quality treated schools were choosing winners/losers and devoting more time to the promising students, then abandoning this strategic approach because of the intervention would likely affect the students who could achieve scores at the top of the distribution. Loyalka et al. (2019) reports similar results for a national teacher PD program in China, where the program led to negative effects on academic performance among students taught by more qualified teachers.

5.4. Robustness and Alternative Estimation Approaches

In this section, we examine the sensitivity of our main findings by performing a series of additional checks. These exercises include (i) using the randomization inference procedure, (ii) replacing the outcome variables with count data, (iii) exploring alternative estimation methods, (iv) using an alternative comparison group, (v) re-estimating our model at the student level, and (vi) examining the parallel trends assumption.

5.4.1. Randomization Inference

In this section, we present the results of a randomization inference test as an alternative approach for conducting inference based on a re-sampling procedure. The traditional approach to performing inference in our main specifications accounts for sampling variation in the estimated effects based on the assumption that the sample is a random draw from a large population. Given that we have access to the universe of public high schools, randomization inference helps to account for design-based uncertainty (i.e., the variability in the estimates due to how treatments are assigned) (Athey and Imbens, 2017; Abadie et al., 2020).

We implement the randomization inference procedure by randomly assigning public high schools to a placebo treatment and re-estimating the doubly robust treatment effects. This exercise is repeated 500 times to generate an empirical distribution of placebo treatment effects. In all repetitions, we set the number of schools in the placebo treatment group equal to the actual number of treated schools in the data. If our main findings are due to chance, then we will expect to find placebo treatment effects that are meaningfully large and different from zero.

In Figure 2, we present the placebo treatment effects for each outcome of interest from the randomized inference procedure. For comparison, the solid line reproduces the doubly robust estimates presented in Table 1, and the dashed lines indicate the 5^{th} and 95^{th} percentiles of the distribution. As expected, Figure 2 shows that the placebo estimates are small and centered around zero. Moreover, with the exception of one outcome (the number of students absent on the exam day), our estimated treatment effects are outside of the interval defined by the 5^{th} and 95^{th} percentiles.

Similarly, Table A.1 in the online appendix summarizes the estimates from the randomized inference method. Rows 1 and 2 present the baseline results, row 3 shows the mean of the 500 placebo estimates, and row 4 shows the randomized inference p-value. The randomized inference p-value is given by c/S, where c is the number of times the placebo treatment effect was larger than the estimated treatment effect in absolute terms and S is the number of placebo simulations. The results in row 3 confirm that the average value of our placebo estimates is practically zero and the randomized inference p-value in row 4 shows that except for absent, all our estimated point estimates are significantly different from the placebo treatment effects.

5.4.2. Alternative Outcome Variables: Count Performance Data

Our next robustness exercise examines an alternative measure of student performance using count data. Until now, we have examined student performance using the share of students obtaining specific grades in the math exam. In this section, we analyze three alternative performance measures, namely the number of students who received a high passing score, the number passing the exam, and the number of failures.[15]

Table 4 presents the impacts of the OTA intervention on student performance using the count data. The estimate in column 1 shows that the intervention reduced the number of students obtaining a high passing score by about 3 students or 13.5% relative to the pre-treatment average level. This result is consistent with the estimates based on the shares of students reported in Table 1 and further supports our finding that the intervention worsened the performance of high-achievers in the upper portions of the grade distribution. However, the impact of the program on the number of students who obtained a passing score is small and not statistically significant.

In column 3, we find that the intervention increased the number of students who failed the math exam by almost 14 students or 20.8% relative to the pre-treatment average number of failing students. Notably, the magnitude of the increase in the number of students who failed the exam is remarkably similar to the number of new students who were recommended to sit the math exam as shown in Panel A of Table 1. As such, the increased number of students failing the math exam is likely being driven by the marginal students who were induced into sitting the exam due to the OTA intervention.

Our findings on access and student performance can be interpreted in relation to the phenomenon of educational triage, where teachers have incentives to discourage weak students from taking the math exit exam (Gilligan et al., 2022). In particular, the results suggest that the OTA intervention alleviated educational triage by significantly improving access to the CSEC math exam.

[15]Unlike the main set of results which defined a severe fail as obtaining Grades 5 or 6, the failure measure in this section focuses more generally on all failing scores - Grades 4, 5, and 6. Given the documented increase in the number of students sitting the exam and since our variables are now measured as count outcomes, this analysis will help us to assess the impact of the intervention on the students who would have access to the exam regardless of the program and those who are induced into sitting the exam because of the policy.

However, most of the newly recommended students likely failed the exam because they were of lower ability and less prepared on average.

5.4.3. Alternative Model Specification and Estimation

This section investigates the sensitivity of our main findings to alternative DID estimation approaches. The first method is a different version of the baseline improved doubly robust DID method, referred to as the *generic* doubly robust DID method and the second approach is the standard two-way fixed effects method. Both methods attempt to estimate the treatment effect under the parallel trends assumption but differ in terms of how the covariates are used in estimation.

The generic doubly robust method differs from the baseline version in Table 1 in terms of how we estimate the required pieces for the doubly robust procedure (i.e., the outcome regression and propensity score models). In particular, the generic approach uses maximum likelihood to estimate the propensity score model and ordinary least squares for the outcome regression model. According to Sant'Anna and Zhao (2020), the generic approach retains the double robustness property in estimating the ATT but it is not necessarily doubly robust for conducting inference. In other words, the generic doubly robust estimator is also consistent but its asymptotic variance depends on a correctly specified model for the outcome and propensity score models. The two-way fixed effects method estimates a linear regression model that includes the covariates as control variables, but does not explicitly model the assignment into treatment.

Table 5 presents the results from the above alternative methods. The results are qualitatively similar to our baseline findings. In terms of access, Panel A shows that the OTA intervention significantly increased the number of students sitting the CSEC exam across both methods. These estimates indicate that the intervention increased the average number of test takers at treated schools between 14 (12%) and 19 (15.8%) students. Similarly, the alternative estimation approaches suggest that the intervention slightly increased absenteeism on exam day. However, while the estimates in Table 4 column 2 are both larger in magnitude than the corresponding baseline coefficient, they are not consistently statistically significant.

Next, Panel B shows that the OTA program significantly reduced student performance regardless of the estimation method. For High Pass, we find that the intervention reduced student performance by about 5 percentage points (33.6%) and 5.4 percentage points (36.5%) for the generic doubly robust and TWFE methods, respectively. While both alternative methods show a sizable reduction in the share of students receiving a Pass, the generic doubly robust estimates are imprecisely estimated, while the TWFE estimates remain statistically significant. Consistent with the baseline results, we do not find a statistically significant impact on the share of students who severely fail the math exam. Overall, it is reassuring that the qualitative findings are robust to alternative estimation methods. In fact, relative to our baseline results, the magnitudes from the alternative estimation approaches suggest that the OTA intervention had a larger impact on access and performance.

5.4.4. Alternative Control Group: A Sample of Public and Private Schools

To further assess the sensitivity of our main findings, we re-estimate the impacts of the OTA program using an alternative control group that is composed of both public and private high schools. Given that the OTA program was implemented in the public school system, we focused our baseline results on *public* high schools. In this section, however, we augment the public schools in our control group with untreated private schools since no private schools were selected for treatment. On the one hand, the bigger control group should improve the precision of our estimates. On the other hand, we need to be certain that this expanded control group satisfies the conditional parallel trends assumption that is required for causal identification. However, there are no obvious differences between public and private schools or policy differences that would induce differential trends in potential outcomes between treated and control schools. We examine the parallel trends assumption for both samples in greater detail in *section 3.4.6* below.

Table 6 shows the impact of the OTA intervention on the main outcomes of interest using the expanded control group of public and private high schools. While this exercise increases our sample size by 29%, the results are qualitatively the same as the baseline results.[16] We continue to find that

[16]The school-level sample increased by 29%, and the number of test-takers in the sample increased by 41%.

the OTA program increased access to the CSEC math exam and had a negative impact on student performance in this larger sample. In Panel A, we find that the intervention increased the number of students taking the exam by about 15 students (11.88%), but it did not have a statistically significant effect on absenteeism. We also find that the intervention significantly reduced the share of students obtaining a High Pass by 2.5 percentage points (21%). However, the program had no statistically significant effects on Pass or severe failure. While the coefficient estimates for student performance in this expanded sample are smaller than those using only the public schools, the results similarly suggest that the impacts on performance are concentrated on the best students in the upper end of the grade distribution. Thus, the finding that the OTA math intervention program worsened student performance is largely robust to the choice of the control group.

5.4.5. Student-level Analyses

Our data contains a repeated cross-section of all students who completed the CSEC exam across all secondary schools in Jamaica. For our main analysis, we aggregated the data to the high school level due to efficiency reasons related to the doubly robust methods we utilize in this paper. For instance, Sant'Anna and Zhao (2020) shows that the doubly robust methods for panel data attain the semiparametric efficiency bound, while those based on repeated cross-sectional data generally do not. However, the student-level analysis allows us to estimate the impacts of the OTA intervention on outcomes that are measured at a more granular level and with a larger sample. For this reason, in this section, we replicate our previous findings using the student-level data.

We restrict the student-level analysis to measures of student performance. All student performance outcomes are defined similarly to the school-level analysis but measured as binary indicators of whether the student received the specified exam grades in question. We present all student-level results in Online Appendix B. Similar to the main analysis, our preferred estimates are based on the doubly robust DID method for the public school sample. After presenting those results, we then conduct similar robustness checks, including randomization inference, examining the impact of using alternative estimation approaches, and using the extended sample of both public and private schools.

Table B.1 presents the doubly robust estimates showing the impact of the OTA program on student performance. Overall, the results corroborate the school-level findings, showing that the intervention significantly reduced student math performance. Specifically, the estimates suggest that the OTA program reduced the likelihood of obtaining a high passing score by 3.6 percentage points or 18.6% percent relative to the pre-treatment average in the non-OTA high schools. The program also reduced the probability of obtaining a passing score by 4.1 percentage points, which translates to 10% of the pre-treatment average pass rate. Again, we find that the intervention had a slightly positive but insignificant impact on the likelihood of obtaining a severe failing score.

We also show that our student-level findings are not sensitive to the robustness exercises we conducted for the school-level analysis. First, Figure B.1 presents the randomization inference results for the three student-level performance outcomes. For High Pass and Pass–outcomes that were statistically significant – the results show that the estimated placebo treatment effects are small and centered around zero. Moreover, the estimated effects denoted by the solid vertical lines are outside of the interval defined by the 5$^{\text{th}}$ and 95$^{\text{th}}$ percentiles of the placebo effects distribution.

Second, we re-estimate the student-level effects of the intervention using the two alternative methods, namely the generic doubly robust DID and TWFE methods. Table B.2 shows that the alternative estimation methods yield similar results to the preferred doubly robust specification. Finally, Table B.3 presents student-level estimates using the expanded control group that includes both private and public school students. These results are also very consistent with our general findings. Consequently, the specification checks suggest that our findings are robust to the use of student or school-level data, the alternative composition of the control group, and alternative estimation strategies.

5.4.6. Parallel Trends Assumption

In this section, we discuss the dynamic effects of the OTA intervention to provide supporting evidence for the parallel trends assumption. These effects are estimated for our main outcome variables using the improved doubly robust DID method. Table 7 and Figure A.1 show the school-level event study estimates for the public school sample.[17] In general, the point estimates in the pre-intervention period are statistically insignificant and the magnitudes are generally not economically meaningful. In addition, the test of joint statistical significance fails to reject the null hypothesis that the pre-treatment estimates are jointly zero at the 5% level of significance. Consequently, these results suggest that the pre-treatment trends in student performance were similar across OTA and non-OTA schools. This is strong evidence in favor of the parallel trends assumption being satisfied in this context.

6. Conclusion

This paper examines a teacher coaching intervention, dubbed the Operation Turn Around program, which deployed specialized Mathematics specialists to train high school teachers in Jamaica. We use administrative data on all high schools from 2009 to 2016 and the recently developed doubly robust difference-in-differences methods to estimate the causal effect of this intervention on student academic performance and access to the math exam.

Although the intervention was mainly designed to improve student math performance, we find positive unintended effects on access to the CSEC exit exam. The results show that the coaching program increased the number of students taking the exam at treated schools by about 14 students. Thus, our findings suggest that the OTA intervention mitigated the common phenomenon of educational triage in Jamaica, whereby teachers typically recommended only those students they believed were prepared and likely to pass the exit exam.

However, we find that the coaching intervention significantly reduced student performance in math on the standardized high school exit exam. Specifically, the OTA program reduced the probability of passing math by 3.6 percentage points or 10.6 percent of the pre-intervention average

[17] The corresponding school-level estimates for the public and private school sample are presented in Table A.2 and Figure A.2 of online Appendix A. In addition, we present the student-level dynamic effects in Online Appendix B, Table B.4 and Figure B.2. The results of these analyses are consistent with the school-level estimates using the public school sample.

pass rate. Additional analyses indicate that the program's negative effects are driven by declines in the highest-quality grades in the upper portions of the grade distribution. Interestingly, we also find that the OTA program had a larger negative effect at higher-quality treated schools with better teachers and greater resources. However, these adverse effects of the intervention do not spill over to non-math subjects.

In addition, in the lower portions of the grade distribution, we find that the program increased the number of failures (i.e., students obtaining the bottom three grades). This finding is consistent with the intervention encouraging teachers to recommend marginal students (who would otherwise not sit the exam) to take the math exam and the evidence suggests that these marginally induced students likely failed the math exam. Nonetheless, the evidence suggests that the increase in test takers does not entirely explain the negative effects on student performance. For example, our heterogeneous effects analysis shows that the intervention did not change student performance in several subgroups where we simultaneously find increases in the number of test takers.

Our results complement other studies in the literature which suggested that teacher coaching programs are not one-size-fits-all prescriptions to improve student achievement. In fact, while the program being studied costs the public approximately 55 million Jamaican Dollars or about 30 USD per treated student, we find that it mostly harmed students' academic performance. Therefore, in a context where high school teachers are largely untrained, as is the case in many developing countries, policymakers should be very cautious about using similar PD programs to improve the academic achievement of high school students.

References

Abadie, A. (2005). Semiparametric difference-in-differences estimators. *The Review of Economic Studies*, 72(1):1–19.

Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. M. (2020). Sampling-based versus design-based uncertainty in regression analysis. *Econometrica*, 88(1):265–296.

Albornoz, F., Anauati, M. V., Furman, M., Luzuriaga, M., Podesta, M. E., and Taylor, I. (2019). Training to teach science: experimental evidence from argentina. *World Bank Economic Review.*, 0(0):1–25.

Athey, S. and Imbens, G. W. (2017). The econometrics of randomized experiments. In *Handbook of economic field experiments*, volume 1, pages 73–140. Elsevier.

Ball, D., Simons, J., Wu, H., Simon, R., Whitehurst, G., and Yun, J. (2008). Chapter 5: Report of the task group on teachers and teacher education. Technical report, United States Department of Education, National Mathematics Advisory Panel, Washington, DC.

Berlinski, S. and Busso, M. (2017). Challenges in educational reform: An experiment on active learning in mathematics. *Economics Letters*, 156:172–175.

Bold, T., Filmer, D., Martin, G., Molina, E., Rockmore, C., Stacy, B., Svensson, J., and Wane, W. (2017). What do teachers know and do? does it matter? evidence from primary schools in africa. *Does it Matter*.

Bourne, P. A. (2019). Mathematics performance in jamaica. *International Journal of History and Scientific Studies*, 1(4):8–31.

Bruns, B., Filmer, D., and Patrinos, H. A. (2011). Making schools work through accountability reforms.

Chaudhury, N., Hammer, J., Kremer, M., Muralidharan, K., and Rogers, F. H. (2006). Missing in action: teacher and health worker absence in developing countries. *Journal of Economic perspectives*, 20(1):91–116.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., and Yagan, D. (2011). How does your kindergarten classroom affect your earnings? evidence from project star. *The Quarterly journal of economics*, 126(4):1593–1660.

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American economic review*, 104(9):2633–79.

Cilliers, J. and Taylor, S. (2017). How to re-skill a workforce? experimental evidence of in-service teacher training and coaching. Technical report, Working Paper.

Cohen, D. K. and Hill, H. C. (1998). Instructional policy and classroom performance: The mathematics reform in california. Technical report.

Darling-Hammond, L., Wei, R. C., Andree, A., Richardson, N., and Orphanos, S. (2009). Professional learning in the learning profession: A status report on teacher development in the united states and abroad. Technical report, Stanford University National Staff Development Council and The School Redesign Network, Palo Alto, CA.

Desimone, L. M. and Garet, M. S. (2015). Best practices in teacher's professional development in the united states. *Psychology, Society and Education*, 7(3):252–263.

Elmore, R. F. (2002). *Bridging the gap between standards and achievement: The imperative for professional development in education*. Albert Shanker Institute, Washington, DC.

Evans, H. L. (2001). *Inside Jamaican schools*. University of West Indies Press.

Filmer, D. and Rogers, H. (2018). Learning to realize education's promise. *World Development Report. The World Bank*.

Fryer Jr, R. G. (2017). The production of human capital in developed countries: Evidence from 196 randomized field experiments. In *Handbook of economic field experiments*, volume 2, pages 95–322. Elsevier.

Gersten, R., Taylor, M. J., Keys, T. D., Rolfhus, E., and Newman-Gonchar, R. (2014). Summary of research on the effectiveness of math professional development approaches. rel 2014-010. *Regional Educational Laboratory Southeast*. https://files.eric.ed.gov/fulltext/ED544681.pdf.

Gilligan, D. O., Karachiwalla, N., Kasirye, I., Lucas, A. M., and Neal, D. (2022). Educator incentives and educational triage in rural primary schools. *Journal of Human Resources*, 57(1):79–111.

Glewwe, P. and Muralidharan, K. (2016). Improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications. In *Handbook of the Economics of Education*, volume 5, pages 653–743. Elsevier.

Glewwe, P. W., Hanushek, E. A., Humpage, S. D., and Ravina, R. (2011). School resources and educational outcomes in developing countries: A review of the literature from 1990 to 2010.

Hanushek, E. A. and Rivkin, S. G. (2007). Pay, working conditions, and teacher quality. *The future of children*, pages 69–86.

Hanushek, E. A. and Rivkin, S. G. (2010). Generalizations about using value-added measures of teacher quality. *American economic review*, 100(2):267–71.

Harvey, S. (1999). The impact of coaching in south african primary science inset. *International Journal of Educational Development*, 19(3):191–205.

Hill, H. C. (2007). Learning in the teaching workforce. *The future of children*, 17(1):111–127.

Jamaica Observer (2016). Education ministry to address decline in csec math passes. *Jamaica Observer*.

Kerwin, J. T., Thornton, R., et al. (2015). Making the grade: Understanding what works for teaching literacy in rural uganda. Technical report. https://www.psc.isr.umich.edu/pubs/pdf/rr15-842.pdf.

Lechner, M. (2011). The estimation of causal effects by difference-in-difference methods. *Foundations and Trends in Econometrics*, 4(3):165–224.

Loyalka, P., Popova, A., Li, G., and Shi, Z. (2019). Does teacher training actually work? evidence from a large-scale randomized evaluation of a national teacher training program. *American Economic Journal: Applied Economics*, 11(3):128–54.

Marshall, P. (2013). *The Tail: How England's schools fail one child in five-and what can be done*. Profile Books.

Ministry of Education, Youth, and Information (2019). Education statistics 2018-2019. Technical report, Ministry of Education, Youth, and Information.

OECD (2021). Education at a glance 2021: Oecd indicators. Technical report, Organisation for Economic Co-operation and Development (OECD) Publishing.

Piper, B. and Korda, M. (2011). Egra plus: Liberia. program evaluation report. Technical report. https://files.eric.ed.gov/fulltext/ED516080.pdf.

Piper, B. and Zuilkowski, S. S. (2015). Teacher coaching in kenya: Examining instructional support in public and nonformal schools. *Teaching and Teacher Education*, 47:173–183.

Popova, A. (2021). *Measuring the Quality of Teaching and Teacher Training in Low-Income Settings*. Stanford University.

Popova, A., Evans, D. K., Breeding, M. E., and Arancibia, V. (2022). Teacher professional development around the world: The gap between evidence and practice. *The World Bank Research Observer*, 37(1):107–136.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American economic review*, 94(2):247–252.

Sailors, M., Hoffman, J. V., David Pearson, P., McClung, N., Shin, J., Phiri, L. M., and Saka, T. (2014). Supporting change in literacy instruction in malawi. *Reading Research Quarterly*, 49(2):209–231.

Sant'Anna, P. H. and Zhao, J. (2020). Doubly robust difference-in-differences estimators. *Journal of Econometrics*, 219(1):101–122.

Scher, L. and O'Reilly, F. (2009). Professional development for k–12 math and science teachers: What do we really know? *Journal of Research on Educational Effectiveness*, 2(3):209–249.

Supovitz, J. A. (2001). *Translating teaching practice into improved student achievement*, volume 2, pages 81–98. Chicago: University of Chicago Press.

Taylor, N., Deacon, R., and Robinson, N. (2019). Secondary level teacher education in sub-saharan africateacher preparation and support. Technical report, JET Education Services.

Walter, C. and Briggs, J. (2012). What professional development makes the most difference to teachers. Technical report, University of Oxford Department of Education, Oxford. https://www.oupjapan.co.jp/sites/default/files/contents/events/od2018/media/od18_Walter_reference.pdf.

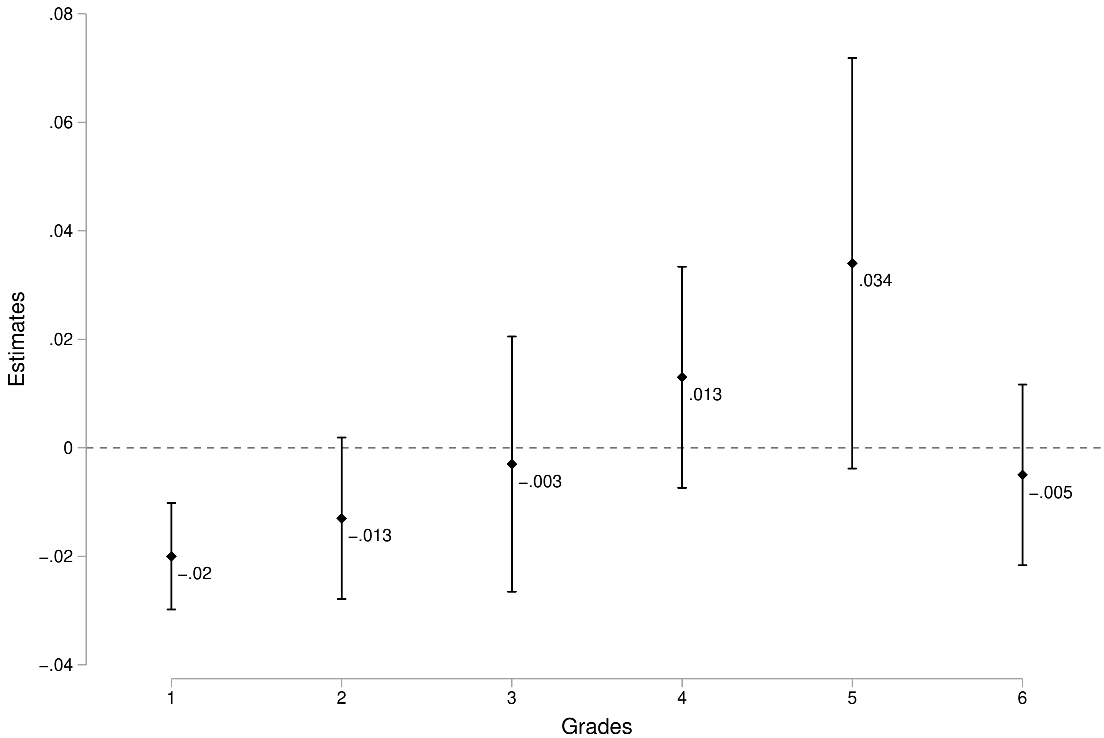
World Bank (2018). World Development Report 2018: Learning to realize education's promise. Technical report, World Bank, Washington, DC: World Bank.

Wright, N. A. (2021). Need-based financing policies, college decision-making, and labor market behavior: Evidence from jamaica. *Journal of Development Economics*, 150:102617.

Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., and Shapley, K. L. (2007). Reviewing the evidence on how teacher professional development affects student achievement. (Issues & Answers Report, REL 2007-No. 033). *Regional Educational Laboratory Southwest (NJ1)*. https://files.eric.ed.gov/fulltext/ED498548.pdf.

Yoshikawa, H., Leyva, D., Snow, C. E., Treviño, E., Barata, M., Weiland, C., Gomez, C. J., Moreno, L., Rolla, A., D'Sa, N., et al. (2015). Experimental impacts of a teacher professional development program in chile on preschool classroom quality and child outcomes. *Developmental psychology*, 51(3).

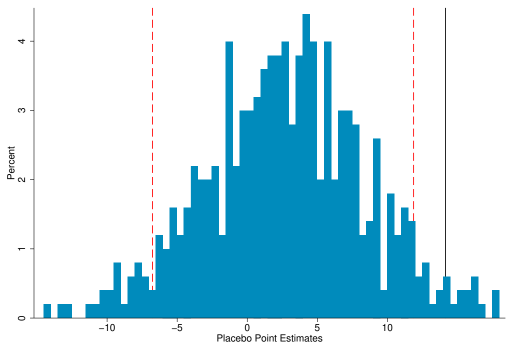
Figure 1: Impact of the OTA Program on Performance



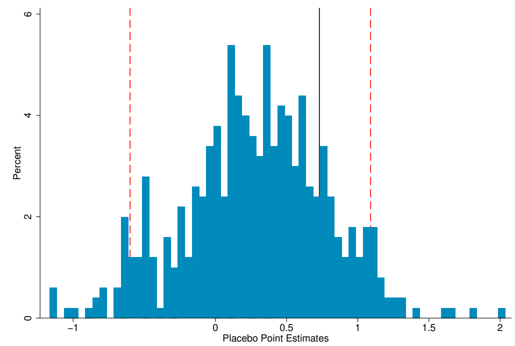
Notes. This figure shows the improved doubly robust differences-in-differences estimates (with 95% confidence intervals) of the impact of the Operation Turn Around program on school-level performance in Mathematics (Sant'Anna and Zhao, 2020). The sample is based on administrative data on all public high schools in Jamaica from 2009 through 2016, including performance on the Caribbean Secondary Education Certificate examination (CSEC). The estimates come from separate regressions for each grade (i.e., Grades 1 through 6) available on the CSEC exam.

Figure 2: Distribution of Placebo Treatment Effects from Randomization Inference

Panel A: Access

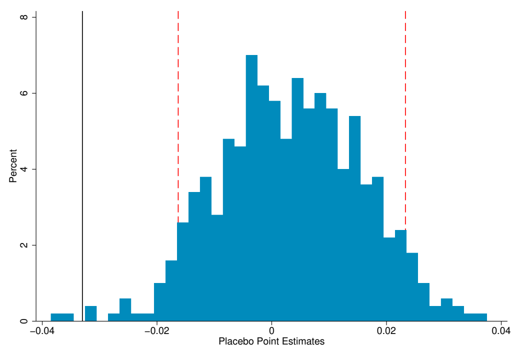


(a) Number of test takers

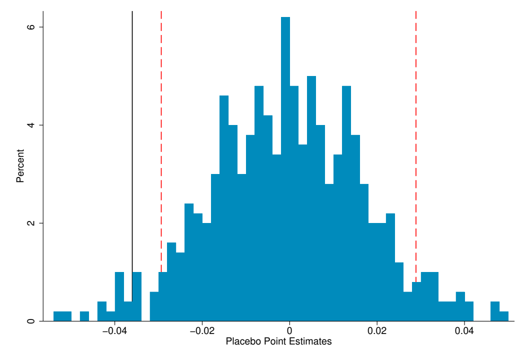


(b) Absent on exam day

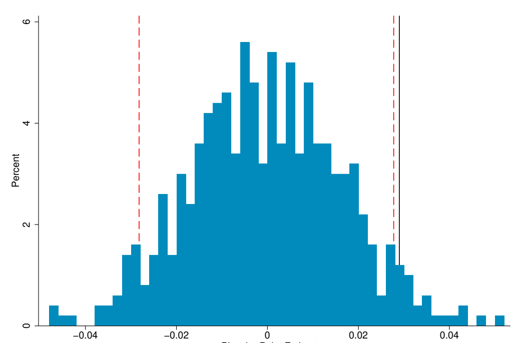
Panel B: Student Performance



(c) High Pass



(d) Pass



(e) Severe Fail

Notes. This figure plots histograms of the distribution of placebo effects from randomization inference using the improved doubly robust difference-in-differences estimator (Sant'Anna and Zhao, 2020). The solid (black) line denotes our preferred estimates reported in Table 1. The dashed (red) lines are the 5th and 95th percentiles of the placebo effect distribution. The sample is based on administrative data on all public high schools in Jamaica from 2009 through 2016, including performance on the Caribbean Secondary Education Certificate examination.

Table 1: Main Results - Impact on Access and Performance

Method	Panel A: Access		Panel B: Performance		
	Number of test takers	**Absent on exam day**	**High Pass**	**Pass**	**Severe Fail**
	(1)	(2)	(3)	(4)	(5)
Improved DR-DID	14.14**	0.73*	-0.033***	-0.036**	0.029
	(6.80)	(0.39)	(0.011)	(0.018)	(0.021)
Mean	111.71	5.80	0.15	0.34	0.40
Std. Deviation	(81.66)	(7.97)	(0.20)	(0.23)	(0.20)
Observations	1,684	1,684	1,684	1,684	1,684

Notes. This table reports the improved doubly robust differences-in-differences estimates of the impact of the Operation Turn Around program on school-level access and performance in Mathematics (Sant'Anna and Zhao, 2020). The sample is based on administrative data on all public high schools in Jamaica from 2009 through 2016, including performance on the Caribbean Secondary Education Certificate examination. Panel A presents measures of access to the CSEC exam. In Panel B, "High Pass" denotes obtaining Grades 1 or 2, "Pass" denotes obtaining Grades 1, 2 or 3, and "Severe Fail" denotes Grades 5 or 6. For each outcome, the pre-intervention mean and standard deviation for schools in the control group are presented below each estimate. Robust standard errors are clustered at the school level in parenthesis.

Table 2: Spillover Effects - Impact on Non-Mathematics Subjects

Dependent variable	**All Subjects**	**Subject Classification**	
		STEM	**Non-STEM**
	(1)	(2)	(3)
Panel A: Access			
Number of test takers	1.45	1.57	1.41
	(1.05)	(1.14)	(1.08)
Absent on exam day	0.16	0.26*	0.09
	(0.11)	(0.15)	(0.10)
Panel B: Performance			
High Pass	0.004	-0.006	0.006
	(0.011)	(0.013)	(0.012)
Pass	0.005	0.005	0.001
	(0.012)	(0.016)	(0.013)
Severe Fail	-0.001	-0.007	0.004
	(0.006)	(0.007)	(0.007)
Observations	32,483	12,464	20,019

Notes. This table reports the improved doubly robust differences-in-differences estimates of the impact of the Operation Turn Around program on school-level non-Mathematics performance (Sant'Anna and Zhao, 2020). The data analysis is done using data at the school-subject level. The sample is based on administrative data on all public high schools in Jamaica from 2009 through 2016, including performance on the Caribbean Secondary Education Certificate examination. Column (1) reports results for all non-Mathematics subjects while Columns (2) and (3) present results grouped by STEM classification. In Panel B, "High Pass" denotes obtaining Grades 1 or 2, "Pass" denotes obtaining Grades 1, 2 or 3, and "Severe Fail" denotes Grades 5 or 6. Robust standard errors are clustered at the school level in parenthesis.

Table 3: Heterogeneous Effects of the OTA Program

	Average Pass Rate (Math)		**Average Pass Rate (All Subjects)**		**Number of students per teacher**		**Number of test takers (Math)**		**Number of exam takers (All Subjects)**	
	Below (1)	Above (2)	Below (3)	Above (4)	Below (5)	Above (6)	Below (7)	Above (8)	Below (9)	Above (10)
Panel A: Access										
Number of test takers	23.57***	17.36***	30.91**	7.17	5.31	23.85**	17.11**	17.23	6.89	15.16*
	(7.44)	(7.29)	(12.35)	(9.41)	(7.12)	(12.17)	(6.68)	(15.67)	(6.33)	(8.37)
Absent on exam day	2.05***	0.21	1.25	0.42	0.41	0.95	0.52	1.45	-0.42	1.09*
	(0.61)	(0.73)	(1.12)	(0.63)	(0.59)	(0.83)	(0.57)	(1.11)	(0.61)	(0.58)
Panel B: Performance										
Highest Pass	-0.002	-0.014*	-0.007	-0.014**	-0.021***	-0.003	-0.008*	-0.020**	-0.004	-0.016**
	(0.003)	(-0.008)	(0.005)	(0.006)	(0.006)	(0.006)	(0.005)	(0.009)	(0.007)	(0.007)
High Pass	-0.01	-0.025	-0.008	-0.027	-0.041***	0.007	-0.017	0.033**	0.003	-0.037***
	(0.010)	(0.028)	(0.012)	(0.019)	(0.014)	(0.021)	(0.015)	(0.015)	(0.018)	(0.014)
Pass	-0.045**	0.001	-0.031	-0.003	-0.047**	0.008	-0.024	-0.019	0.008	-0.039*
	(0.019)	(0.058)	(0.024)	(0.033)	(0.022)	(0.040)	(0.025)	(0.021)	(0.033)	(0.021)
Severe Fail	0.053**	-0.005	0.043	-0.007	0.041	0.004	0.026	(0.011)	(0.018)	0.032
	(0.025)	(0.054)	(0.031)	(0.029)	(0.030)	(0.027)	(0.026)	(0.024)	(0.041)	(0.024)
Number of treated schools	47	11	28	30	34	24	40	18	28	30

Notes. This table reports the improved doubly robust differences-in-differences estimates of the impact of the Operation Turn Around program on school-level access and performance in Mathematics (Sant'Anna and Zhao, 2020). The sample is based on administrative data on all public high schools in Jamaica from 2009 through 2016, including performance on the Caribbean Secondary Education Certificate examination. In Panel B, "Highest Pass" denotes obtaining Grade 1, "High Pass" denotes obtaining Grades 1 or 2, "Pass" denotes obtaining Grades 1, 2 or 3, and "Severe Fail" denotes Grades 5 or 6. For each school characteristic, "Below" and "Above" corresponds to schools with values that are below and above the pre-treatment median using the distribution of all schools. The last row shows the number of treated schools in each subgroup. Robust standard errors are clustered at the school level in parenthesis.

Table 4: Impact on Count Performance Outcomes

Method	Number of High Passes (1)	Number of Passes (2)	Number of Failures (3)
Improved DR-DID	-3.15**	0.35	13.79**
	(1.58)	(2.41)	(5.38)
Mean	23.27	45.52	66.19
Std. Deviation	(43.58)	(57.00)	(53.84)
Observations	1,684	1,684	1,684

Notes. This table reports the improved doubly robust differences-in-differences estimates of the impact of the Operation Turn Around program on the number of High Passes, Passes, and Failures at the school level (Sant'Anna and Zhao, 2020). The sample is based on administrative data on all public high schools in Jamaica from 2009 through 2016, including performance on the Caribbean Secondary Education Certificate examination. "High Pass" denotes obtaining Grades 1 or 2, "Pass" denotes obtaining Grades 1, 2 or 3, and "Failures" denotes Grades 4, 5 or 6. For each outcome, the pre-intervention mean and standard deviation for schools in the control group are presented below each estimate. Robust standard errors are clustered at the school level in parenthesis.

Table 5: Alternative Empirical Models - Impact on Access and Performance

Method	Panel A: Access		Panel B: Performance		
	Number of test takers	**Absent on exam day**	**High Pass**	**Pass**	**Severe Fail**
	(1)	(2)	(3)	(4)	(5)
Generic DR-DID	13.50**	0.78	-0.0498**	-0.079	0.100
	(6.59)	(0.72)	(0.027)	(0.049)	(0.070)
TWFE	18.67***	1.66***	-0.054***	-0.038**	0.005
	(6.49)	(0.46)	(0.010)	(0.016)	(0.017)
Mean	111.71	5.80	0.15	0.34	0.40
Std. Deviation	(81.66)	(7.97)	(0.20)	(0.23)	(0.20)
Observations	1,684	1,684	1,684	1,684	1,684

Notes. This table reports alternative differences-in-differences estimates of the impact of the Operation Turn Around program on school-level access and performance in Mathematics (Sant'Anna and Zhao, 2020). The alternative methods are the generic doubly robust DID and the standard two-way fixed effects (TWFE) methods. The sample is based on administrative data on all public high schools in Jamaica from 2009 through 2016, including performance on the Caribbean Secondary Education Certificate examination. Panel A presents measures of access to the CSEC exam. In Panel B, "High Pass" denotes obtaining Grades 1 or 2, "Pass" denotes obtaining Grades 1, 2 or 3, and "Severe Fail" denotes Grades 5 or 6. For each outcome, the pre-intervention mean and standard deviation for schools in the control group are presented below each estimate. Robust standard errors are clustered at the school level in parenthesis.

Table 6: Alternative Control Group - Impact on Access and Performance

Method	Panel A: Access		Panel B: Performance		
	Number of test takers	Absent on exam day	High Pass	Pass	Severe Fail
	(1)	(2)	(3)	(4)	(5)
Improved DR-DID	14.66**	0.63	-0.025***	0.021	0.004
	(6.22)	(0.46)	(0.010)	(0.018)	(0.018)
Mean	123.41	10.25	0.12	0.31	0.41
Std. Deviation	(126.38)	(18.59)	(0.17)	(0.21)	(0.18)
Observations	2,380	2,380	2,380	2,380	2,380

Notes. This table reports the improved doubly robust differences-in-differences estimates of the impact of the Operation Turn Around program on school-level access and performance in Mathematics (Sant'Anna and Zhao, 2020). The sample is based on administrative data on all public and private high schools in Jamaica from 2009 through 2016, including performance on the Caribbean Secondary Education Certificate examination. Panel A presents measures of access to the CSEC exam. In Panel B, "High Pass" denotes obtaining Grades 1 or 2, "Pass" denotes obtaining Grades 1, 2 or 3, and "Severe Fail" denotes Grades 5 or 6. For each outcome, the pre-intervention mean and standard deviation for schools in the control group are presented below each estimate. Robust standard errors are clustered at the school level in parenthesis.

Table 7: School-level Event Study Estimates

	Panel A: Access		Panel B: Performance		
Time Period	**Number of test takers**	**Absent on exam day**	**High Pass**	**Pass**	**Severe Fail**
T-4	-8.01	-0.98	-0.010	-0.021	-0.006
	(4.83)	(0.58)	(0.010)	(0.021)	(0.027)
T-3	6.43	-0.39	0.009	-0.014	0.073
	(5.91)	(0.59)	(0.013)	(0.027)	(0.033)
T-2	4.91	-0.21	-0.004	0.027	-0.039
	(6.97)	(0.72)	(0.013)	(0.022)	(0.028)
T-1	-3.34	0.51	0.015	-0.002	-0.024
	(4.74)	(0.61)	(0.011)	(0.017)	(0.025)
0	6.85	0.17	-0.024	-0.016	-0.006
	(6.37)	(0.64)	(0.012)	(0.026)	(0.037)
T+1	13.22	0.92	-0.061	-0.068	0.023
	(6.87)	(0.67)	(0.018)	(0.034)	(0.031)
T+2	18.58	1.39	-0.024	-0.015	-0.003
	(8.05)	(0.65)	(0.010)	(0.023)	(0.032)
F-statistic	5.22	6.31	5.13	3.07	9.13
P-value	0.27	0.18	0.27	0.55	0.06
Observations	1,684	1,684	1,684	1,684	1,684

Notes. This table reports event study estimates of the impact of the Operation Turn Around program on student performance in Mathematics based on the improved doubly robust differences-in-differences method (Sant'Anna and Zhao, 2020). The sample is based on administrative data on all public high schools in Jamaica from 2009 through 2016, including performance on the Caribbean Secondary Education Certificate examination. Panel A presents measures of access to the CSEC exam. In Panel B, "High Pass" denotes obtaining Grades 1 or 2, "Pass" denotes obtaining Grades 1, 2 or 3, and "Severe Fail" denotes Grades 5 or 6. Robust standard errors are clustered at the school level in parenthesis.

Appendix A: Additional Results

Table A.1: Summary of Randomization Inference Results

	Panel A: Access		**Panel B: Performance**		
	Number of test takers	**Absent on exam day**	**High Pass**	**Pass**	**Severe Fail**
DR-DID Estimate	14.14	0.73	-0.033	-0.036	0.029
DR-DID p-value	0.038	0.14	0.003	0.046	0.172
RI Estimate	2.7	0.28	0.003	-0.0006	-0.0003
RI p-value	0.03	0.21	0.01	0.046	0.026

Notes. This tables summarizes the estimates from the randomized inference procedure described in Section 5.4 of the paper. Rows 1 and 2 present the baseline coefficient estimates reported in Table 1, row 3 shows the mean of the 500 placebo estimates, and row 4 shows the randomized inference p-value. The sample is based on administrative data on all public high schools in Jamaica from 2009 through 2016, including performance on the Caribbean Secondary Education Certificate examination. Panel A presents measures of access to the CSEC exam. In Panel B, "High Pass" denotes obtaining Grades 1 or 2, "Pass" denotes obtaining Grades 1, 2 or 3, and "Severe Fail" denotes Grades 5 or 6.

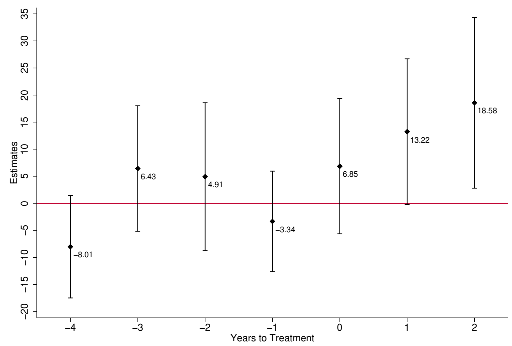
Table A.2: School-level Event Study Estimates (Public and Private Schools)

	Panel A: Access		Panel B: Performance		
Time Period	**Number of test takers**	Absent on exam day	**High Pass**	**Pass**	**Severe Fail**
T-4	-6.70	-0.52	-0.010	-0.013	-0.009
	(4.64)	(0.59)	(0.011)	(0.021)	(0.029)
T-3	1.75	-0.98	0.018	-0.003	0.077
	(5.72)	(0.51)	(0.011)	(0.021)	(0.027)
T-2	7.87	0.14	-0.002	0.012	-0.029
	(6.15)	(0.62)	(0.008)	(0.017)	(0.025)
T-1	-3.01	0.32	0.011	0.001	-0.021
	(4.11)	(0.53)	(0.007)	(0.016)	(0.023)
0	6.70	0.73	-0.032	-0.022	0.009
	(6.04)	(0.60)	(0.013)	(0.024)	(0.029)
T+1	14.54	1.38	-0.063	-0.059	0.022
	(6.51)	(0.64)	(0.016)	(0.028)	(0.028)
T+2	18.29	1.90	-0.041	-0.033	0.013
	(7.58)	(0.58)	(0.013)	(0.024)	(0.029)
F-statistic	3.06	8.80	5.66	1.15	8.01
P-value	0.55	0.07	0.23	0.89	0.09
Observations	2,380	2,380	2,380	2,380	2,380

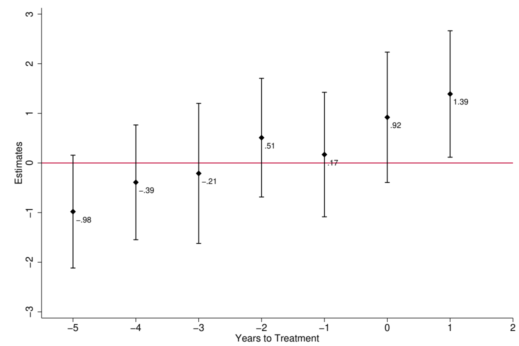
Notes. This table reports event study estimates of the impact of the Operation Turn Around program on student performance in Mathematics based on the the improved doubly robust differences-in-differences method (Sant'Anna and Zhao, 2020). The sample is based on administrative data on all public and private high schools in Jamaica from 2009 through 2016, including performance on the Caribbean Secondary Education Certificate examination. Panel A presents measures of access to the CSEC exam. In Panel B, "High Pass" denotes obtaining Grades 1 or 2, "Pass" denotes obtaining Grades 1, 2 or 3, and "Severe Fail" denotes Grades 5 or 6. Robust standard errors are clustered at the school level in parenthesis.

Figure A.1: School-level Event Study Estimates (Public Schools)

Panel A: Access

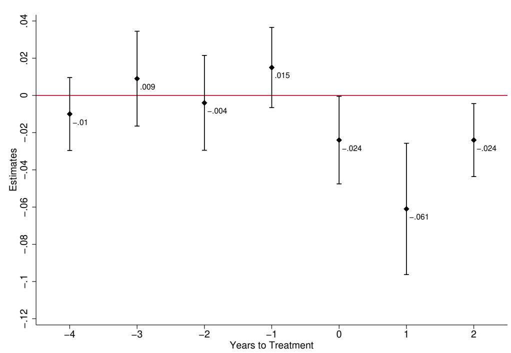


(a) Number of test takers

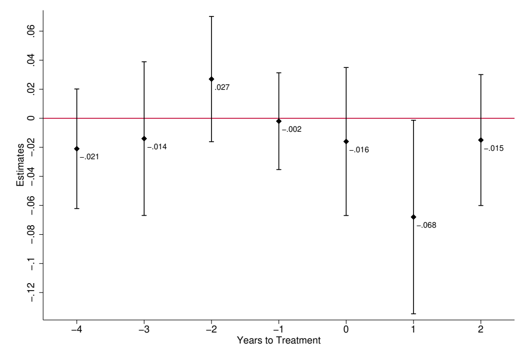


(b) Absent on exam day

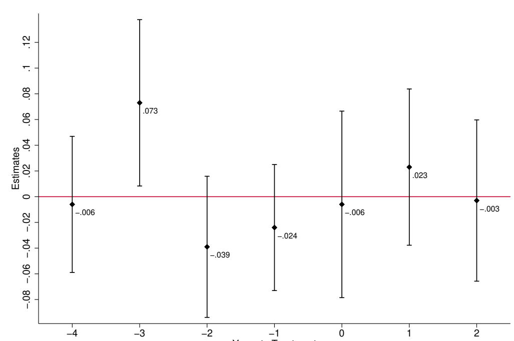
Panel B: Student Performance



(c) High Pass



(d) Pass

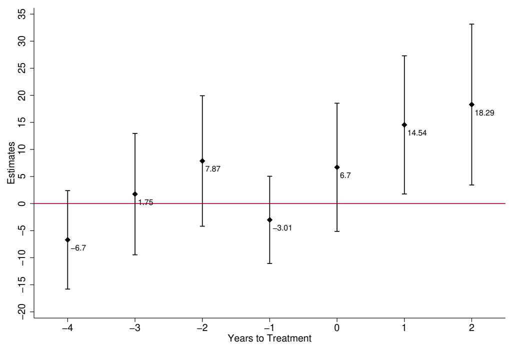


(e) Severe Fail

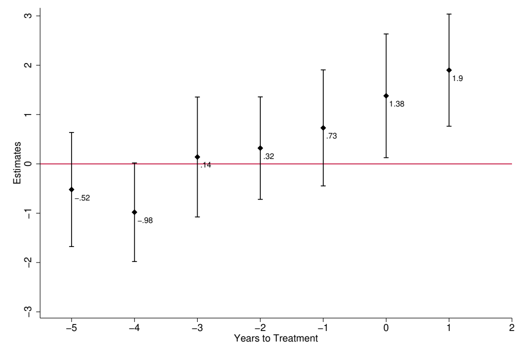
Notes. This figure displays the event study estimates of the impact of the Operation Turn Around program on student performance in Mathematics based on the improved doubly robust differences-in-differences method (Sant'Anna and Zhao, 2020). The estimates are also reported in Table 7 in the paper. The sample is based on administrative data on all public high schools in Jamaica from 2009 through 2016, including performance on the Caribbean Secondary Education Certificate examination. Panel A presents measures of access to the CSEC exam. In Panel B, "High Pass" denotes obtaining Grades 1 or 2, "Pass" denotes obtaining Grades 1, 2 or 3, and "Severe Fail" denotes Grades 5 or 6.

Figure A.2: School-level Event Study Estimates (Public and Private Schools)

Panel A: Access

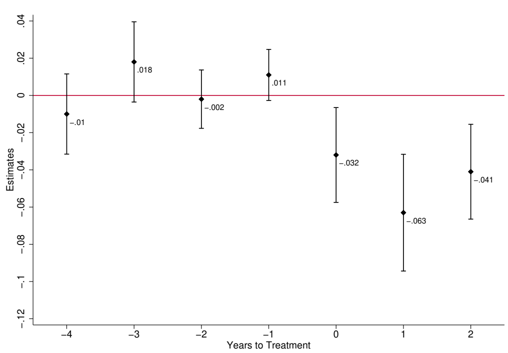


(a) Number of test takers

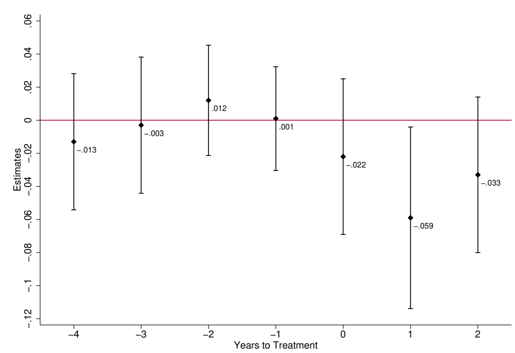


(b) Absent on exam day

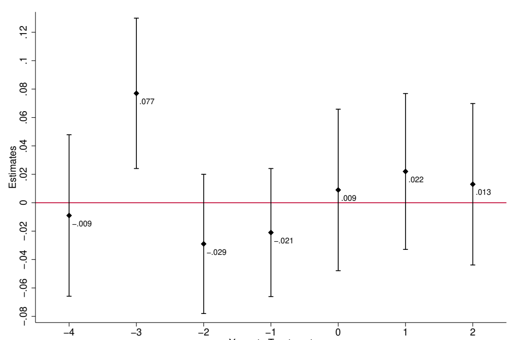
Panel B: Student Performance



(c) High Pass



(d) Pass

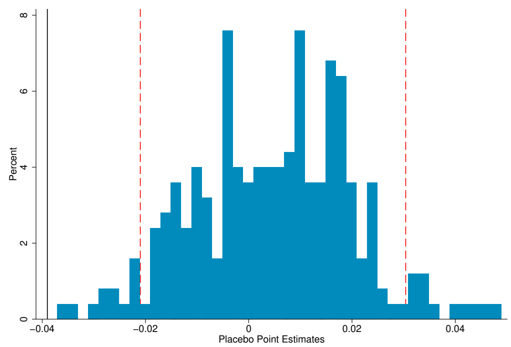


(e) Severe Fail

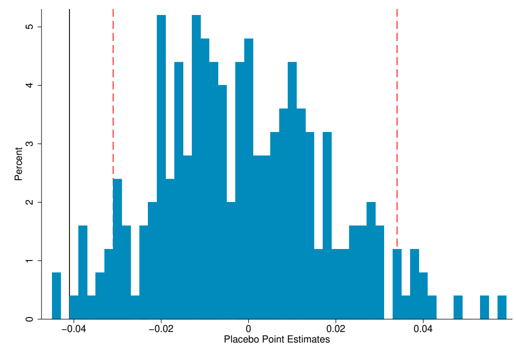
Notes. This figure displays the event study estimates of the impact of the Operation Turn Around program on student performance in Mathematics based on the improved doubly robust differences-in-differences method (Sant'Anna and Zhao, 2020). The estimates are also reported in Appendix A Table A.2. The sample is based on administrative data on all public and private high schools in Jamaica from 2009 through 2016, including performance on the Caribbean Secondary Education Certificate examination. Panel A presents measures of access to the CSEC exam. In Panel B, "High Pass" denotes obtaining Grades 1 or 2, "Pass" denotes obtaining Grades 1, 2 or 3, and "Severe Fail" denotes Grades 5 or 6.

Appendix B: Student-level Analysis

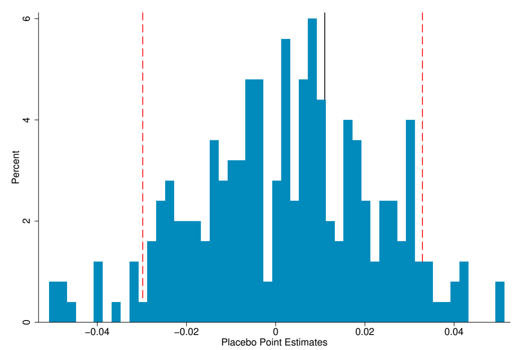
Figure B.1: Distribution of Placebo Treatment Effects from Randomization Inference (Student Performance)



(a) High Pass



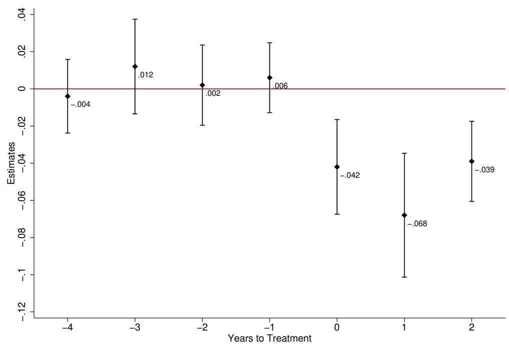
(b) Pass



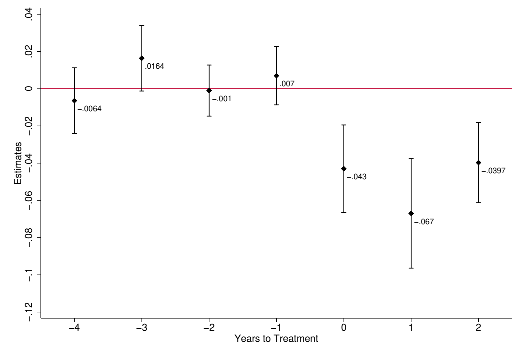
(c) Severe Fail

Notes. This figure plots histograms of the distribution of placebo effects from randomization inference using the improved doubly robust difference-in-differences estimator (Sant'Anna and Zhao, 2020). The solid (black) line denotes our preferred estimates reported in Appendix B Table B.1. The dashed (red) lines are the 5th and 95th percentiles of the placebo effect distribution. The sample is based on administrative data on all public high school students in Jamaica from 2009 through 2016, including performance on the Caribbean Secondary Education Certificate examination.

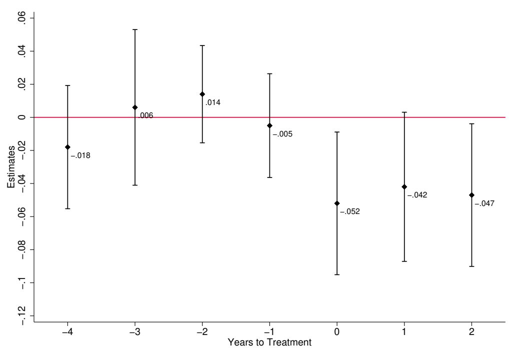
Figure B.2: Student-level Event Study Estimates



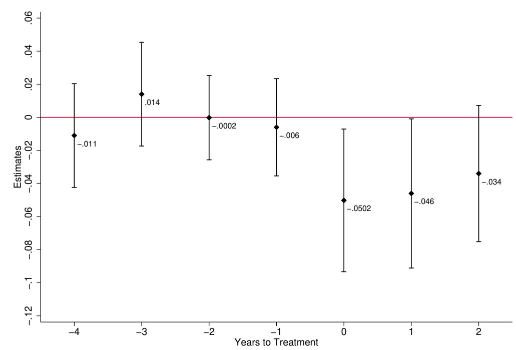
(a) High Pass



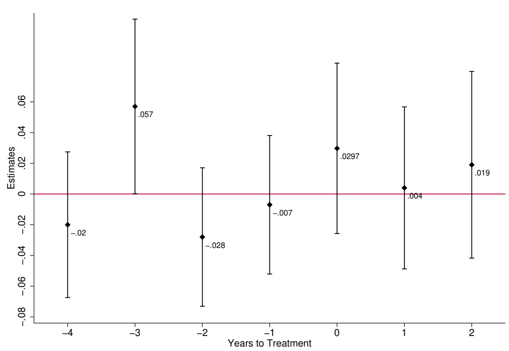
(d) High Pass



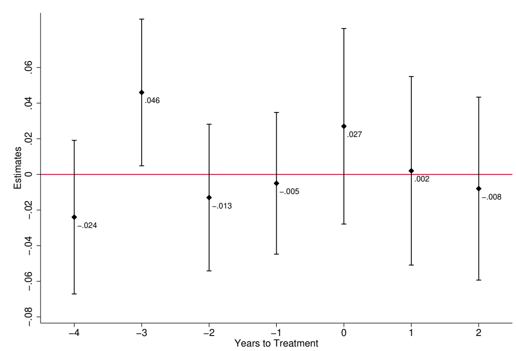
(b) Pass



(e) Pass



(c) Severe Fail



(f) Severe Fail

Notes. This figure displays the student-level event study estimates of the impact of the Operation Turn Around program on student performance in Mathematics based on the improved doubly robust differences-in-differences method (Sant'Anna and Zhao, 2020). The estimates are also reported in Appendix B Table B.4. The sample is based on administrative data on all high school students in Jamaica from 2009 through 2016, including performance on the Caribbean Secondary Education Certificate examination. The left column presents results using the public schools while the right column shows results based on the combined public and private school sample. "High Pass" denotes obtaining Grades 1 or 2, "Pass" denotes obtaining Grades 1, 2 or 3, and "Severe Fail" denotes Grades 5 or 6.

Table B.1: Impact of the OTA Program on Performance

Method	High Pass	Pass	Severe Fail
Improved DR-DID	-0.039***	-0.041**	0.011
	(0.011)	(0.016)	(0.019)
Mean	0.21	0.41	0.35
Std. Deviation	(0.41)	(0.49)	(0.48)
Observations	172,280	172,280	172,280

Notes. This table reports the improved doubly robust differences-in-differences estimates of the impact of the Operation Turn Around program on student-level performance in Mathematics (Sant'Anna and Zhao, 2020). The sample is based on administrative data on all public high school students in Jamaica from 2009 through 2016, including performance on the Caribbean Secondary Education Certificate examination. "High Pass" denotes obtaining Grades 1 or 2, "Pass" denotes obtaining Grades 1, 2 or 3, and "Severe Fail" denotes Grades 5 or 6. For each outcome, the pre-intervention mean and standard deviation for schools in the control group are presented below each estimate. Robust standard errors are clustered at the school level in parenthesis.

Table B.2: Alternative Methods: Impact of the OTA Program on Performance

Method	High Pass	Pass	Severe Fail
Generic DR-DID	-0.042***	-0.043**	0.012
	(0.013)	(0.017)	(0.019)
TWFE	-0.061***	-0.027	-0.022
	(0.013)	(0.017)	(0.022)
Mean	0.21	0.41	0.35
Std. Deviation	(0.41)	(0.49)	(0.48)
Observations	172,280	172,280	172,280

Notes. This table reports alternative differences-in-differences estimates of the impact of the Operation Turn Around program on student-level performance in Mathematics (Sant'Anna and Zhao, 2020). The alternative methods are the generic doubly robust DID and the standard two-way fixed effects (TWFE) methods. The sample is based on administrative data on all public high school students in Jamaica from 2009 through 2016, including performance on the Caribbean Secondary Education Certificate examination. "High Pass" denotes obtaining Grades 1 or 2, "Pass" denotes obtaining Grades 1, 2 or 3, and "Severe Fail" denotes Grades 5 or 6. For each outcome, the pre-intervention mean and standard deviation for schools in the control group are presented below each estimate. Robust standard errors are clustered at the school level in parenthesis.

Table B.3: Alternative Control Group: Impact of the OTA Program on Performance

	High Pass	Pass	Severe Fail
	-0.039***	-0.043**	0.01
	(0.012)	(0.018)	(0.021)
Mean	0.15	0.34	0.38
Std. Deviation	(0.36)	(0.47)	(0.49)
Observations	244,718	244,718	244,718

Notes. This table reports the improved doubly robust differences-in-differences estimates of the impact of the Operation Turn Around program on student-level performance in Mathematics (Sant'Anna and Zhao, 2020). The sample is based on administrative data on all public and private high school students in Jamaica from 2009 through 2016, including performance on the Caribbean Secondary Education Certificate examination. "High Pass" denotes obtaining Grades 1 or 2, "Pass" denotes obtaining Grades 1, 2 or 3, and "Severe Fail" denotes Grades 5 or 6. For each outcome, the pre-intervention mean and standard deviation for schools in the control group are presented below each estimate. Robust standard errors are clustered at the school level in parenthesis.

Table B.4: Student-level Event Study estimates

Time Period	Panel A: Public Schools			Panel B: Public and Private Schools		
	High Pass	**Pass**	**Severe Fail**	**High Pass**	**Pass**	**Severe Fail**
T-4	-0.004	-0.018	-0.02	-0.0064	-0.011	-0.024
	(0.010)	(0.019)	(0.024)	(0.009)	(0.016)	(0.022)
T-3	0.012	0.006	0.057	0.0164	0.014	0.046
	(0.013)	(0.024)	(0.029)	(0.009)	(0.016)	(0.021)
T-2	0.002	0.014	-0.028	-0.001	-0.0002	-0.013
	(0.011)	(0.015)	(0.023)	(0.007)	(0.013)	(0.021)
T-1	0.006	-0.005	-0.007	0.007	-0.006	-0.005
	(0.010)	(0.016)	(0.023)	(0.008)	(0.015)	(0.020)
0	-0.042	-0.052	0.0297	-0.043	-0.0502	0.027
	(0.013)	(0.022)	(0.028)	(0.012)	(0.022)	(0.028)
T+1	-0.068	-0.042	0.004	-0.067	-0.046	0.002
	(0.017)	(0.023)	(0.027)	(0.015)	(0.023)	(0.027)
T+2	-0.039	-0.047	0.019	-0.0397	-0.034	-0.008
	(0.011)	(0.022)	(0.031)	(0.011)	(0.021)	(0.026)
F-statistic	2.78	1.76	4.52	5.67	0.97	4.82
P-value	0.60	0.78	0.34	0.23	0.91	0.31
Observations	172,280	172,280	172,280	244, 718	244, 718	244, 718

Notes. This table reports the student-level event study estimates of the impact of the Operation Turn Around program on student performance in Mathematics based on the the improved doubly robust differences-in-differences method (Sant'Anna and Zhao, 2020). The sample is based on administrative data on all public and private high school students in Jamaica from 2009 through 2016, including performance on the Caribbean Secondary Education Certificate examination. Panel A presents results using public high school students while Panel B shows results using the combined public and private high school students sample. In each panel, "High Pass" denotes obtaining Grades 1 or 2, "Pass" denotes obtaining Grades 1, 2 or 3, and "Severe Fail" denotes Grades 5 or 6. Robust standard errors are clustered at the school level in parenthesis.