Northwestern University

From the SelectedWorks of C. Kirabo Jackson

2019

Can Introducing Single-Sex Education into Low-Performing Schools Improve Academics, Arrests, and Teen Motherhood?

C. Kirabo Jackson

Available at: https://works.bepress.com/c_kirabo_jackson/29/

Can Introducing Single-Sex Education into Low-Performing Schools Improve Academics, Arrests, and Teen Motherhood?

C. KIRABO JACKSON
Northwestern University,
Department of Education and Social Policy, 2120 Campus Drive,
Evanston 60208
email: kirabo-jackson@northwestern.edu

Can Introducing Single-Sex Education into Low-Performing Schools Improve Academics, Arrests, and Teen Motherhood?

By C. KIRABO JACKSON[1]

In 2010, the Ministry of Education in Trinidad and Tobago converted 20 low-performing secondary schools from coeducational to single-sex. I exploit these conversions to identify the policy-relevant causal effect of introducing single-sex education into existing schools (holding other school inputs constant). After accounting for student selection, boys in single-sex cohorts at conversion schools score higher on national exams taken around age 15, both boys and girls take more advanced coursework, and girls perform better on secondary-school completion exams. There are also important non-academic effects; all-boys cohorts have fewer arrests as teens, and all-girls cohorts have lower teen pregnancy rates. Survey evidence suggests that these single-sex conversion effects reflect both direct gender peer effects due to interactions between classmates, and indirect effects generated through changes in teacher behavior. (*JEL* I20, J00)

Policy-makers, researchers, and parents have debated the merits of single-sex education for decades.[2] A recent literature has emerged that relies on quasi-random variation to estimate the causal impact of attending single-sex schools relative to coed schools.[3] These studies generally find that students have better test scores at single-sex schools than at coed schools (Lee et al. 2014; Jackson 2012; Park, Behrman and Choi 2013). This pattern has led some to advocate for introducing single-sex education into existing coed schools. However, the current evidence does not speak to the policy question of whether one should expand single-sex education into existing schools for two important reasons. The first reason is that the cross-school comparisons do not answer the policy question of how expanding single-sex education to existing coed schools affects students holding other school inputs fixed. Because schools do not become single-sex at random, single-sex and coed schools may differ in important ways so that cross-school differences may not

[1] Jackson: Department of Education and Social Policy, Northwestern University, 2120 Campus Drive, Evanston 60208 (kirabo-jackson@northwestern.edu). This project was supported by the Spencer Foundation. I thank Brian Jacob, Julie Cullen, Gordon Dahl, Kitt Carpenter, Heyu Xiong, and Alexey Makarin for useful comments. I thank Carol Singh for invaluable project management and data collection efforts, and Igor Uzilevskiy, Patrick Peters, Diana Balitaan, Kevin Malis, Rodrigo Braga, Hao (Leo) Hu, Mathew Steinberg, Richard Yu, and Ben Henken for excellent research assistance. I am also grateful to Brenda Moore, Harilal Seecharan, and Peter Smith at TTMOE. All errors are mine.

[2] In theory, holding other schooling attributes fixed, single-sex education may improve outcomes because single-sex classrooms may (a) allow for instruction tailored to the specific needs of each sex, (b) allow teachers to focus better on instruction, and (c) remove opposite-sex peers who may affect social dynamics in ways that are deleterious to student outcomes. Conversely, by depriving youth of opposite-sex peers, single-sex education could lead to poor socialization, generate negative social interactions, and reinforce sex stereotypes (Eliot 2016).

[3] There is also an older literature in which researchers were unable to credibly disentangle the effects of single-sex schooling from the characteristics of the students who chose to attend single-sex schools (see Mael et al. 2005).

be due to single-sex education *per se*.[4] To make this clear, suppose single-sex education had no effect on students, but higher-quality teachers sorted into single-sex schools. Single-sex schools would outperform coed schools (due to better teachers shifting from single-sex to coed schools), while a policy of expanding single-sex education into schools would yield zero benefits in the aggregate. The second reason the current evidence does not speak to the policy question is that the causal literature on single-sex schooling has focused on academic outcomes. However, parents often chose single-sex education for their children based on perceived impacts on social development and behaviors rather than test scores, so that understanding the impacts on a broad set of outcomes is important. To shed light on these unresolved issues, this study presents the first analysis of (a) a large-scale policy to introduce single-sex education to existing schools, and (b) the causal effect of single-sex schooling on a wide array of outcomes including short-run test scores, advanced course taking, secondary school completion exams, arrests, and teen motherhood.

I analyze a policy experiment in Trinidad and Tobago. The Ministry of Education (MOE) noticed that students at elite single-sex schools (studied in Jackson 2012) had good outcomes, and thus decided to experiment with single-sex education at low-performing schools. The MOE identified ten pairs (20 in total) of geographically close, similarly sized, low-performing coed public secondary schools. One school in each pair was converted to all-boys, and the other to all-girls. The transition was phased such that the incoming 6^{th}-grade cohorts after 2010 were single-sex while the previously admitted cohorts remained coed. Importantly, (a) selected schools had no control over this decision, and (b) to ensure a clean experiment, the MOE dictated that there be no other changes at these 20 schools. Because this experiment allows one to compare students who attended the same school under both coed and single-sex regimes, one can isolate the effect of adopting a single-sex policy from that of unobserved school-level differences that might exist between coed and single-sex schools in other settings. I link student records before secondary school entry to national exam data three and five years later, arrest records, and birth registry data. These data allow me to analyze the effect of single-sex education on a rich set of outcomes. I also analyze survey data collected during the policy experiment to present evidence on mechanisms.

This paper innovates over the extant literature in a few important ways. Jackson (2012)

[4] In related work, Booth, Cardona-Sosa and Nolen (2013) randomly assign students to all-girls and all-boys discussion sections in a University economics course. The authors find positive effects for females. It is unclear that these results will generalize to the context of secondary school children, or those from disadvantaged backgrounds.

examines the effect of attending elite single-sex schools relative to coed schools in Trinidad and Tobago. Here, I study a policy experiment and compare outcomes in low-performing schools before and after a conversion to single-sex. This allows me to difference out other school characteristics and isolate the policy effect of transitioning. Two studies, Strain (2013) and Dustmann et al. (forthcoming), also examine within-school conversion variation in single-sex education.[5] In both settings, the number of converting schools is small, the choice to convert was potentially endogenous, and other inputs were not constant.[6] Moreover, these two studies come to opposite conclusions. I improve upon these studies by examining a larger sample of converting schools, relying on exogenous conversions outside schools' control, and examining a context in which other inputs were held constant by design. Another key innovation is to move beyond academics and examine impacts on social outcomes. This study is the first to quasi-experimentally identify single-sex schooling effects on outcomes such as crime and teen motherhood. Finally, the use of surveys allows me to shed some much-needed light on underlying mechanisms.

To identify the effect of the transition from coed to single-sex holding other school inputs constant, I compare the outcomes of students who attended the same secondary school under coed versus single-sex regimes. While this transition allows one to hold other school inputs fixed, it does not ensure that students are comparable across cohorts. To address this concern, I follow Jackson (2010, 2012, 2013) and exploit discontinuities in the school assignment rules to isolate exogenous variation in school attendance and to remove bias due to student selection. Among students who apply to a given secondary school, there is a test score cut-off above which students are admitted and below which they are not. This allows for a fuzzy regression discontinuity (RD) estimate of the effect of attending an experimental school in any given year. I combine this exogenous variation in school assignments (in a given year) with exogenous within-school changes in single-sex status (across years) to compare the causal effect of attending an experimental school under the coed regime to the causal effect of attending that same school under the single-sex regime. The resulting estimator is a Difference-in-Regression Discontinuity (DiRD) design. I present several empirical tests to support the validity of the DiRD estimates. While no single test is dispositive, the array of tests presented supports a causal interpretation of the results.

---

[5] Strain (2013) analyzes the introducing single-sex classes into nine schools in North Carolina. Math scores declined after the conversions. Dustmann et al. (forthcoming) examine seven all-boys and four all-girls schools that changed to coed (i.e., *the reverse conversion*). Boys' and girls' scores declined after the reverse conversion.

[6] Dustmann, et. al. (forthcoming) document sizable changes in teacher composition after the reverse transition.

Given the large number of outcomes examined, to avoid mutliple inference issues, I combine the various outcomes into a small set of indexes. I find large positive effects of the transition to single-sex on academic achievement (about 0.2 sd) three years after secondary-school entry. These effects are similar in size to reducing class size by 30 percent (Chetty et al. 2011; Krueger 1999) or increasing teacher quality by two sd (Jackson et al. 2014) but were gained at zero financial cost. These short-run test score impacts were driven largely by boys, but both girls and boys had improvements in the upper tails of the achievement distribution. Five years after secondary-school entry, both boys and girls are more likely to take advanced courses in the single-sex cohorts, and girls had much better performance on secondary-school leaving exams. Looking at social outcomes, boys in single-sex cohorts are roughly six percentage points (60 percent) less likely to be arrested by the age of 18, and the all-girls cohorts are about four percentage points (about 40 percent) less likely to have a live birth by the age of 18. Survey evidence suggests positive direct peer effects in all-girls settings through less peer distraction and more peer learning. For boys, the direction of the direct peer effects is unclear. Consistent with Lee et al. (2014), I find evidence of positive indirect peer effects through changes in teacher behaviors in single-sex settings. That is, there is evidence of efficiency gains for teachers to the more homogeneous single-sex classrooms. The generally positive effects for both sexes echo Duflo, Dupas, and Kremer (2011) who find that both low- and high-achievement students benefit from achievement tracking.[7]

This is the first evaluation of a large-scale policy to intruduce single-sex education into existing coed schools. The results reveal that this policy can be a low-cost way to improve academic and non-academic outcomes for low-achieving students. I present a theoretical framework that reveals that this result may not generalize to all settings. While the findings are consistent with cross-school evidence that single-sex schools tend to outperform coed schools, they underscore the importance of using policy variation to answer policy questions.

The remainder of the paper is as follows: Section I lays out a theoretical framework. Section II describes the policy landscape, the policy experiment, and the data. Section III lays out the empirical strategy. Section IV presents the empirical results. Section V concludes.

[7] The results speak to the effect of having all instruction in single-sex classes, which may be different from that of having a few single-sex classes in a coed school. Also, I present the effect of schools transitioning from coed to single-sex, holding teacher quality and other inputs fixed. This short-run effect may not be the same as the long-run general equilibrium effects of introducing single-sex schooling to an education system.

I Theoretical Framework

To motivate the empirical work, Appendix A presents a model of single-sex schooling. I summarize it here. In the model, the gender composition in a classroom affects student outcomes in multiple ways. First is through direct gender peer effects that operate through peer interactions. Some studies find that more female classmates improves all students' outcomes— arguably because boys are disruptive (Hoxby 2000; Lavy and Schlosser 2011). Others find that more same-gender peers improve student outcomes— arguably because the opposite sex is distracting (Black et al. 2013; Ooserbeek and van Ewijk 2014; Lu and Anderson 2015). Overall, the direct impact of gender composition on the outcomes of each sex is uncertain. Importantly, the gender composition may also *indirectly* affect outcomes through teacher action. Certain teaching practices may benefit one sex more than the other, and teachers may align instruction to each type based on the gender composition. If teachers care about average student outcomes and are rational, they will align all instruction to boys in all-boys settings and will align all instruction to girls in all-girls settings. As such, all single-sex classrooms may enjoy a "boutique effect" if students benefit from similar peers because instruction is aligned to their type (Hoxby and Weingarth 2006; Duflo, Kremer, and Dupas (2011), and a "focus effect" if students benefit from more homogeneous classroom environments because teachers can focus on one type (e.g., Hoxby and Weingarth 2006, Ding and Lehrer 2007). I refer to the combination of these two indirect effects as an *indirect* alignment effect.

The central insight from the model is that the single-sex classroom effect (relative to coed) reflects the differences in both the direct and indirect effects between coed and single-sex settings. Because the potential indirect alignment effect is greatest in single-sex settings, under rational teacher behavior, the indirect effects are non-negative. However, the direct peer effects can lead to negative single-sex effects. As such, single-sex schooling is neither always good nor always bad and depends on the mechanisms at play in the specific context.[8] The model does not predict what one may observe in any particular situation. However, section IV.H employs survey data to shed light on whether these mechanisms operate in the Trinidad and Tobago context.

[8] This is helpful for thinking about how single-sex schools may differ from single-sex classrooms within coed schools. If teachers have a greater incentive to align instruction to one sex in single-sex schools than in single-sex classrooms within coed schools, one may see larger benefits to single-sex schools than single-sex classrooms. In addition, if direct gender peer interactions in the classroom are affected by the gender composition of the school, single-sex classrooms may have different effects from that of single-sex schools.

II. The Trinidad and Tobago Context and the Pilot Program

The Trinidad and Tobago education system evolved from the English system. At the end of primary school, (after grade 5, typically at age 11) students take the Secondary Entrance Assessment (SEA) examinations and are assigned to secondary schools (in part, based on this exam) by the Ministry of Education (MOE). The school assignment algorithm used by the MOE (discussed in Section II.B) generates exogenous variation in secondary school attendance that plays a key role in isolating school effects from selection effects in this study.

Secondary school begins in form 1 (grade 6) and ends at form 5 (grade 10). All the experimental schools (called pilot schools) are on the main island, Trinidad, which is roughly 50 miles long and 37 miles wide. All of the pilot schools are government (traditional public) schools.[9] These schools provide instruction from forms 1 through 5 and teach the national curriculum. Students take two externally graded exams at the secondary level. These are key outcomes in this study. The first is the National Certificate of Secondary Education (NCSE) taken at the end of form 3 (grade 8) by all students (both in public and private schools) in eight subjects.[10] The second exam is the Caribbean Secondary Education Certification (CSEC) taken at the end of form 5 (grade 10) which is equivalent to the British Ordinary levels exam.[11] Students seeking to continue their education take five or more subjects, and all testers take the English language and math exams.

II.A. Description of the Data

This project uses administrative SEA data from 2006 through 2012. These data include scores on the national exam taken at the end of grade 5, the school choices made by the student

---

[9] There are two types of public secondary schools: Government schools and Government Assisted schools (Assisted schools). Government schools are fully funded and operated by the Government while Assisted schools are run by private bodies (usually a religious board) and *at least* half of their expenses are paid for by the Government. Along all other dimensions, Government and Government Assisted schools are identical. Assisted schools are similar to charter schools in the United States. Unlike in many countries where private schools are often of higher perceived quality, private schools in Trinidad and Tobago account for a small share of student enrollment and tend to serve those who "fall through the cracks" in the public system (roughly 2 percent). Students who attend private secondary schools have test scores a third of a standard deviation lower than the average 5th grade student does, and half a standard deviation lower than the average among those students who take the secondary school completion exams.

[10] In some years, the NCSE was not required in all schools. The students' final score is a weighted average of course grades in that subject (submitted by their teachers) and an externally graded exam. The MOE awards certificates to all students who attain a combined final mark of at least 60% for each of the eight core subjects.

[11] There are 31 CSEC subjects covering a range of academic subjects such as Physics and Geography, and vocation subjects such as Technical Drawing and Principles of Business. The CSEC exams are accepted as an entry qualification for higher education in Canada, the UK and the United States. After taking the CSEC, students may continue to take the Caribbean Advanced Proficiency Examinations at the end of grade 12, which is considered tertiary level education but is a prerequisite for admission to the University of the West Indies (the largest University in the Caribbean). The CAPE is the Caribbean equivalent of the English Advanced Levels (A-Levels) examinations.

before sitting the SEA exam, and the administrative secondary school assignment. The data also include age, gender, primary school, and religious affiliation. The final dataset contains information on 124,382 students across the seven cohorts. I link the SEA data to NCSE data for 2009 through 2015 by full name and date of birth.[12] The NCSE data contain scores earned on NCSE exams in eight subjects taken at the end of form 3 (8th grade) when the typical student is 15 years old. To summarize these numerous scores, I compute the average academic score – the average standardized NCSE math, English, science, social studies, and Spanish scores. I also link the SEA data to the administrative exam data for the 2012 through 2016 CSEC (taken five years after secondary school entry) by name and date of birth.[13] Because the first treated cohort entered school in 2010 and took the CSEC in 2015, this allows for an analysis of the CSEC for the first two treated cohorts. The typical student is 17 years old when taking the CSEC and virtually all students take this exam (in both public and private schools). The CSEC data record the subjects that students decide to take, final grades on the exams taken, and whether a student earned a secondary school-leaving credential. With these data, I compute two summary measures. The first is the CSEC index which is the average of passing English, passing math, the number of CSEC courses taken, the number of CSEC courses passed, and obtaining a secondary-school leaving credential (*note*: each variable is standardized before creating the average). The second summary measure is the number of advanced CSEC courses taken (among math, English and science). A key innovation of this paper is to study effects on social outcomes. To look at crime, I link the SEA data by name and date of birth to official arrest records between January 2000 and February 2015. To track teen fertility outcomes, I link the SEA data by name and date of birth to maternal data on all official birth records from the Registrar General's office. These data cover all live births between January 1st 2010 and September 1st 2016.

Table 1 provides summary statistics for the population of interest. Columns 1 and 4 show the means and standard deviations of the main variables for boys and girls separately. Roughly half of the sample is male. On average, 4.9 percent of SEA-takers is a repeat SEA-taker. One notable pattern is that females outperform males on average. Females score about one-third of a standard deviation higher than males on the SEA exams at the end of 5th grade, and this gap is

[12] The match rate of 85 percent is consistent with the national dropout rate of 20 percent by age 17.
[13] The match rate is 80 percent. As expected, this is lower than the match rate for the NCSE (taken two years earlier).

similar for the NCSE exams at the end of 8th grade.[14] Also, by age 17 the dropout rate (i.e., CSEC nonparticipation) is 0.156 for females compared to 0.229 for males.[15] Dropout rates are even larger in pilot schools. While 38.4 percent of females in 5th grade earn the secondary school leaving certificate required for tertiary education entry, only 26.6 percent of males do.

On average, the likelihood that a male is arrested by age 16 and 18 is 3.5 and 7.5 percent, respectively. The comparable figures for females are 0.7 and 1.4 percent. Arrest rates are higher at pilot schools. Roughly 4.5 and 9.4 percent of males at pilot schools before the transition had been arrested by the age of 16 and 18, respectively. These arrest rates are comparable to those in high poverty schools in the U.S. (Stevens et al. 2015). The likelihoods of a girl having a baby by the age of 16 and 18 were 0.9 and 4.3 percent, respectively. Teen motherhood rates are higher among girls at pilot schools where 1.3 and 6.1 percent had a birth by age 16 and 18, respectively. These teen motherhood rates are similar to those in the United States (Bialik 2009; Kearney and Levine 2012). These figures paint a picture of male underperformance at all schools, and underperformance (both academically and socially) at pilot schools for both sexes.

II.B. Student Assignment to Secondary Schools

At the end of primary school, students take the SEA exams. Before the exams, students submit a list of four ordered school secondary school choices. The exams are sent to the MOE where they are externally scored. The test score and the school choices are used by the MOE to assign students to secondary schools using a deferred acceptance algorithm (Gayle and Shapley 1962).[16] This algorithm creates a test score cutoff for each government school above which applicants are admitted, and above which they are not (Jackson 2010). However, there is not full compliance with the cutoffs because the MOE makes administrative assignments in certain circumstances.[17] Importantly, the noncompliance is due to MOE adjustments that are outside the control of parents and students. As such, one can use the cutoffs to estimate a selection-free effect of attending a pilot school using a fuzzy-regression-discontinuity-type design.

The MOE does not report the cutoffs for schools. However, because I have access to the

---

[14] Females outperform males on average in all subjects, but the relative male underperformance is greater in language and humanities subjects.

[15] These high school completion numbers are comparable to those in the United States in 2012.

[16] For full details on how students are assigned to schools see Appendix C.

[17] For example, students who score below the cutoffs for all the schools in their choices receive an MOE assignment that may not comply with the cutoffs. Also, when schools are forced to restrict capacity for unforeseen reasons (e.g. one school was closed down for a year due to fire and another could not open due to flooding), students receive administrative assignments to accommodate these circumstances.

administrative assignment (which is outside the control of the students or their parents), there *are* real cutoffs that were used to assign students, and any non-compliance with the cutoffs are orthogonal to students' attributes, one can uncover the cutoffs empirically. For each of the 20 pilot schools, I regress whether an applicant is assigned to the school by the MOE on an indicator variable denoting scoring above each possible test score.[18] I select the cutoff that yields the largest *F*-statistic to be the cutoff for that school. If one used the actual schools attended, one might worry that this empirical approach would result in endogenous cutoffs. However, this approach uses the MOE assignment and not the actual school attended to infer the location of the cutoffs.

To show that this procedure approximates the actual cutoffs, the top panel of Figure 1 shows the likelihood of receiving the administrative assignment to pilot school A as a function of applicants' incoming SEA scores in 2009 and 2010 (before and after the transition). The dashed vertical line indicates the estimated cutoff. In this school, the administrative assignments follow the cutoff rules virtually all the time so that it is clear where the cutoff was for this school in both years. However, the MOE did not assign *all* students to schools using the cutoff rules in all years. The lower panel shows the likelihood of attending pilot school B in 2008 and 2011. Even though compliance is imperfect for this school in these specific years, within a narrow range, one can easily infer the location of the cutoff from the data.[19]

II.C. The Pilot Program

The MOE noticed that students who attended elite single-sex schools (studied in Jackson 2012) had good outcomes, and thus decided to experiment with single-sex education at low-performing schools. In March of 2010, the MOE identified 20 low-performing coed government schools to transition to single-sex starting in September 2010 (there were 90 government schools in the country at the time). The MOE selected these 20 experimental schools (called pilot schools) in pairs. One school in each pair converted to all-girls and the other to all-boys. The MOE made all these decisions without the approval or consultation of the schools.[20] These 20 pilot schools were to be converted on a phased basis (with each successive incoming cohort) to single-sex over a period of five years. Commencing with the 2010 SEA placement process, all incoming 6th-grade

[18] Note: An applicant to a school listed that school in their choices and was not already assigned to a preferred school.
[19] While some schools' assignments are better approximated by the cutoff rule than others, all results are robust to only using cutoffs that explain more than seventy percent of the variation in the administrative assignments.
[20] There has harsh been criticism of this initiative, particularly from the Trinidad and Tobago Unified Teachers Association. They argued that there was not prior consultation and there was insufficient time for proper preparation.

students at pilot schools were of the same sex. To avoid disrupting the incumbent students, students who were previously admitted to coed cohorts remained in coed cohorts, while those admitted after the transition were admitted to single-sex cohorts. The phased cohort-level transition was to continue until all pilot schools were single-sex in all grades.[21]

The pilot schools were chosen based on three criteria. Each pilot school had to be (1) non-selective, (2) a traditional government school, and (3) close to another potential pilot school of similar selectivity and size. This last condition was necessary because each school in a pair of pilot schools would need to take half of the students who would have attended the other school in the pair. To demonstrate that the selection of pilot schools followed the stated MOE criteria, Figure 2 plots the likelihood of being a pilot school for different groups of schools by the distance to the nearest government school.[22] Only non-selective traditional government schools were chosen for the pilot. Among these, the likelihood of being a pilot is strongly associated with being close to another government school.[23] In sum, pilot school status was involuntary, and the MOE selected pilot schools based on known criteria. As such, one can be confident that pilot schools were not chosen based on a trajectory of improving or worsening outcomes. Indeed, I show that the effect of attending pilot schools did not exhibit any differential pre-trending in Section IV.

[21] In 2010, the 6th graders were in single-sex classes, while grades 7 through 10 were in coed classes. In 2011, the new incoming 6th graders and the 7th graders were single-sex, while earlier cohorts (now in grades 8 through 10) were in coed classes. However, in 2013 there was a change in government, there was growing frustration from the teachers' union that they were never consulted on the transition, and single-sex pilot was abandoned. In 2013, the Ministry of Education announced that the incoming 6th graders in 2014 would be coed. Appendix B presents the status of each grade in pilot school by year. Note that because I study outcomes several years after secondary school entry, the end of the pilot program occurs after secondary school entry of youngest cohort examined in this study.

[22] I obtained address data for each secondary school in Trinidad and Tobago, geocoded each address, and computed the shortest distance (as the crow flies) between each secondary school and the closest government school. Addresses were obtained from a variety of sources: contacting individual schools, school websites, and official databases with school information. Addresses were geocoded using the reported longitude and latitudes from google maps. In a few cases street addresses could not be geocoded so that area centroids and street mid points were used.

[23] Among non-selective traditional government schools, over 40 percent of those that were within one kilometer of another government school were chosen as pilot sites, and none more than 2.5 kilometers from a government school was. In a regression, school type, school selectivity, and distance "explains" over half of the variation in pilot school status. One may wonder why these variables do not "explain" all of the variation. One reason is that the distance calculations are based on as the crow flies and my geocoding was not 100 percent accurate. These factors introduce measurement error and reduce the explained variation. A second reason is that some schools that were close to a government school and non-selective were a different size from the closest school – making the transition less feasible. Third, some schools are clustered such that at times three schools are closer to each other but only two schools (one pair) could be chosen for the pilot. Finally, because the MOE wished to have pilot schools in different areas where several schools were clustered closer to each other only two (one pair) would be chosen for the pilot.

III. Empirical Framework

To isolate the effect of transitioning to single-sex from that of attributes of schools that happen to be single-sex, I identify the causal impact of attending a pilot school both before and after the transition to single-sex, and then to compare the two. As shown in Figure 1, among pilot school applicants in a given year, the likelihood of being assigned to a pilot school increases in a discontinuous manner as one's score goes from below to above the cutoff for that pilot school. If the locations of the cutoffs are orthogonal to student characteristics, and the effect of incoming test scores on outcomes is smooth through the cutoffs, one can attribute any sudden jumps in outcomes through the cutoffs to the suddenly increased likelihood of being assigned to, and therefore attending, a pilot school *in a given year*. Using the sudden jump in the likelihood of attending a pilot school through the cutoffs, one can remove student selection bias and identify the causal impact of attending a pilot school (relative to comparison schools) in any given year using a fuzzy regression discontinuity (RD) design.

The experiment caused pilot schools to go from coed for the pre-period cohorts to single-sex for the post-period cohorts. Comparison schools remained coed in both periods. I model outcome Y_{ijc} of student i at school j in SEA cohort c as [1].

[1] $$Y_{ijc} = \delta T_{ij} + \sigma\left(T_{ij} \times POST_{ic}\right) + \pi_c + v_{ijc}$$

In [1], T_{ij} is an indicator that is equal to 1 for student i who attended pilot school j and 0 otherwise. $Post_{ic}$ is equal to 1 for student i in cohort c who took the SEA after 2010 (the post transition period) and 0 otherwise. π_c are cohort fixed effects, and v_{ijc} captures all other determinants of the outcome. The parameter δ is the differential effect of attending a pilot school when they were coed (this reflects underlying school quality differences between pilot and comparison schools when both groups of schools were coed). The parameter σ is the effect of single-sex education holding these aforementioned underlying school quality differences (i.e., those differences due to teacher quality, input quality, etc.) fixed. This policy-relevant parameter captures the *change* in school effectiveness caused by the adoption of single-sex instruction into an existing school.[24]

[24] During the post period $Y_{ijc} = T_{ij}(\sigma + \delta) + v_{ijc}$. During the pre-period $Y_{ijc} = T_{ij}(\delta) + v_{ijc}$. From [1], during the post period, the causal impact of attending a pilot school (relative to a coed school) reflects the single-sex schooling effect (σ) plus the impacts of any underlying school level differences between pilot and non-pilot schools (δ). Only in the unlikely event that single-sex and coed schools are the same in expectation in all other dimensions does this cross-sectional estimate uncover the policy parameter. As such, *all* cross sectional studies (e.g. Jackson 2012; Park, Behrman and Choi 2013, and others) may confound single-sex effects with underlying school differecnes. Because pilot schools were coed during the pre-period, the causal impact of attending a pilot school during the pre-period

To isolate the effect of transitioning from coed to single-sex from other school attributes, I (a) use the cutoffs in pre-transition years to obtain an RD estimate of δ — the effect of attending a pilot school before the transition, (b) use the cutoffs in post-transition years to obtain an RD estimate of $\sigma + \delta$ — the effect of attending a pilot school *after* the transition, and then (c) use these two RD estimates to compute σ — the *change* in relative effectiveness of pilot schools after the transition to single-sex. The proposed estimator of the single-sex schooling effect is a difference-in-regression-discontinuity (DiRD) design. The validity of this approach requires that the RD identifying assumptions be satisfied (i.e., smoothness of potential outcomes through cutoffs), and that the DiD identifying assumptions be satisfied (i.e., common trends between pilot and other schools, and no other changes at pilot schools). The DiRD design also requires that, in expectation, the pilot school Local Average Treatment Effect (LATE) is the same over time – that is, those who attended pilot schools both before and after the transition benefit the same from attending a pilot school.[25] In section IV.G, I present evidence that each of these sets of assumption is likely satisfied.

Because students apply to four schools, students can be used to identify multiple cutoffs. For example, a student may be above the cutoff for their third choice school, but below the cutoff for their second and first choice schools. To exploit the fact that each student can be used to identify the effect of multiple cutoffs, following Jackson (2010, 2012), each student is in the dataset once for each cutoff for which they are marginal. Students are marginal for a school's cutoff if they applied to that school and were not already admitted to a more preferred school.[26] While the data contain 124,382 students, the analytic dataset contains 366,536 student-by-cutoff observations.

This DiRD estimate is obtained by the two-stage-least-squares (2SLS) model described below where equation (2) is the second-stage equation.

$$[2] \qquad Y_{ijc} = \left(\widehat{T_{ij} \times Post_{ic}}\right)\sigma_1 + \sigma_2 \hat{T}_j + f_c(SEA_{ic}) + \pi_{choice} + C_{ijc} + \pi_c + \varepsilon_{ijc}$$

All variables are defined as in [1]. Y_{ijc} is the outcome of person i, applying to school j in SEA

reflects only the impacts of any school-level differences between pilot and non-pilot schools (δ). If the school quality differences are constant over time, and the causal impact of attending pilot schools after the transition minus that from before will difference out the underlying school quality differences and isolate the policy parameter.

[25] If the LATE is not the same before and after the transition, then then one could not simply "difference out" the pilot school effect between the two RD estimates and the DiRD estimates would be biased. I discuss this in Section IV.G.

[26] That is, students assigned to their top choice school are in the dataset once for the cutoff of their top choice school. Students assigned to their second choice school are in the dataset twice (once for the top choice and once for the second choice). Students assigned to their third choice school are in the dataset three times (once for each of the top, second, and third choice). Finally, students who were either assigned to their fourth choice school or received an administrative assignment are in the data four times (once for each choice).

cohort *c*. $f_c(SEA_{ic})$ is a fourth-order polynomial in the incoming SEA score interacted with the SEA cohort and gender[27], and $Post_{ic}$ is an indicator equal to 1 if student *i* in cohort *c* took the SEA in 2010 or thereafter. π_c are a cohort fixed effects. To make comparisons among students of the same gender who apply to the same school in the same year, the model includes cutoff-by-gender fixed effects, C_{ijc}. I also include the student choice fixed effects, π_{choice}, to serve as additional controls. The choice fixed effect denotes the unique combination and order of the four school choices during the pre and post period. As such, all comparisons are made among students who chose the same set of schools in the exact same order during the same regime (i.e. pre or post transition).[28] This accounts for differences in preferences and other attributes that predict parental decisions of where to send their children, but are typically not observed by researchers.

The endogenous regressors are attending a pilot school, T_{ij} , and attending a pilot school after the transition, $T_{ij} \times Post_{ic}$. To remove selection bias, I instrument for these two endogenous regressors with 7 indicator variables connoting scoring above the cutoff for a pilot school in each year between 2006 and 2012. The pilot school cutoffs for all seven years identify the impact of attending a pilot school across all years, and the pilot school cutoffs for years 2010 through 2012 identify the impact of attending a pilot school after the transition. The fitted values $\hat{T}_{ij}$ and $T_{\iota j} \widehat{\times Post}_{\iota c}$ are predicted values of the endogenous regressors estimated in the first-stage regressions. Because the model includes the estimated effect of attending a pilot school across all years, the coefficient on $(T_{\iota j} \widehat{\times Post}_{\iota c})$ identifies the parameter of interest σ – the causal effect of the *change* in the effectiveness of pilot schools after the transition to single-sex. Standard errors are clustered at the assigned school level (which is a larger group than the individual student). I test all of the identifying assumption of this DiRD model in Section IV.

IV. Main Results

IV.A. Illustrating the DiRD Variation Visually

The first outcome I examine is average academic scores on the NCSE exams. While I exmaine impacts on various subjects in Section IV.B, I first focus on this summary measure to

---

[27] I also present the main preferred model with different smooth functions of the running variable in Appendix H.

[28] For example students who chose schools A,B,C, and D are in a different group from those who list B,A,C, and D. Moreover, students who chose who chose schools A,B,C, and D before 2010 are in a different group than those who chose schools A,B,C, and D after 2010.

avoid problems of mutilple hypthesis testing. Before presenting the regression results, I provide a visual illustration of the identifying variating. The DiRD estimate is the difference between fuzzy-RD estimates obtained before and after the transition. The top panel of Figure 3 presents the first stage for all pilot schools combined.[29] It shows the likelihood of *attending* a pilot school during the pre-transition years on the left and the post-transition years on the right, as a function of applicants' scores relative to the cutoff. The figure shows data for test scores within 40 points (about 1.2 standard deviations) of the cutoff divided into equally spaced bins. The figure includes the 90% confidence interval for the mean in each test score bin and the smooth fit based on a fourth-order polynomial of the relative score. For both the pre and the post years, there is a shift in the likelihood of attending a pilot school through the cutoff (a relative score of 0). The increase in the likelihood of attending a pilot school is about 40 and 60 percentage points in the pre and post-transition periods, respectively.[30]

To illustrate the mechanics of the DiRD estimate, the bottom panel of Figure 3 shows the change in 8^{th}-grade average academic scores (net of the cutoff and choice effects) through the cutoffs before and after the transition. Scores are about 0.09sd lower through the cutoff for the pre-transition period (left) and about 0.02sd higher through the cutoff in the post period (right).[31] Using this discontinuity variation, the Wald estimate of the pilot school effect is -0.09÷0.4= -0.225sd in the pre-period and -0.02÷0.6=-0.033sd in the post period. The DiRD estimate is the difference between the RD effect in the post period (which includes the single-sex effect and school effects) and that for the pre-period (which includes only school effects) which is -0.033+0.225=0.19sd. In words, the causal effect of attending a pilot school relative to a comparison school on math scores increased by about 0.19sd after the transition to single-sex compared to before. This is similar to the 2SLS regression estimate. Figure 3 presents a visual representation, of the DiRD variation used in the 2SLS model outlined above. Figure 3 is to illustrate the nature of the variation. Analogous figures for each of the other main outcomes examined in this paper are in Figure 4.

IV.B. Regression Results

---

[29] Following Jackson (2010,2012,2013), I show multiple cutoffs in a single graph by taking the applicants for each pilot school in each year (140 cutoffs; 20 schools over 7 years), re-centering their SEA scores around the cutoff for that school in that year and then stacking the data for each cutoff.

[30] To assuage concerns that the different first stages drive the results. I present the main results omitting data for the non-compliers within 10 and 20 points below the cutoff and the results are very similar. See appendix H.

[31] The fact that the pre-transition RD effects were negative suggests that the parents of low-achieving students may have been uninformed about the relative impacts of pilot schools. Indeed, Abdulkadiroğlu et. al. (2017) find that well-meaning parents can chose schools that reduce child outcomes when they are poorly informed about school quality.

Table 2 presents the main regression results for academic scores.[32] Each column represents a different econometric specification, and each model represents a point estimates from a separate regression. The coefficient on "Boys-Pilot*POST" is the estimated *change* in the effectiveness of pilot schools after the transition to all-boys, the coefficient on "Girl-Pilot*POST" is the estimated *change* in the effectiveness of pilot schools after the transition to all-girls, and the coefficient on "Pilot*POST" is the estimated *change* in the effectiveness of pilot schools after the transition to either all-boys or all-girls. I discuss the effect on the boys and girls in turn.

Model 1 presents the raw pre versus post comparisons with no controls for both boys and girls combined, while that for boys only and girls only are in models 7 and 13, respectively. Overall, test scores were largely unchanged after the transition to single-sex at pilot schools. However, average scores of boys were 0.1sd lower (p-value<0.1), while those for girls was 0.12sd higher (p-value<0.01). Models 2, 9, and 14 control for students' incoming test scores. Conditional on test scores, average academic scores (both boys and girls) at pilot schools increased by 0.099sd (p-value<0.01). This increased point estimate is driven by the fact that boys had lower incoming test scores after the transition than before. Once this is accounted for, the estimated transition effect is positive. Models 3, 9 and 15 presents results from specifications that condition on both incoming scores and choices. Conditional on test scores and choices, average academic scores at pilot schools increased by 0.08sd (p-value<0.05). This increase is similar for boys (0.149sd) and girls (0.113sd) and is statistically significantly different from zero for each group.

The models thus far include powerful controls. However, to address bias due to selection in *unobserved* dimensions, models 4, 10, and 16 presents the 2SLS models that isolate exogenous variation in school attendance due only to a students' score relative to that of the test score cutoffs of pilot schools. The first stage F-statistics for all endogenous regressors are well above 100 so that there is no weak identification problem. This 2SLS model yields a single-sex transition effect of 0.22 (p-value<0.01). This positive effect appears to be driven entirely by boys who have an all-boys transition effect of 0.236 (p-value<0.01) rather than girls who have an all-girls transition effect of only 0.044 (p-value>0.1). Because choices are an important control in the models, models 5, 11, and 17 presents the 2SLS models that include fixed effects for each cutoff in addition to

[32] Because not all students who took the SEA are linked to NCSE data, the main test score outcome is the average academic score conditional on taking the NCSE exams. In Section IV.D. I present results that account for the NCSE participation margin explicitly. The main conclusions are unchanged in such analyses.

fixed effects for the student choices. In such models, the point estimates are similar. In these preferred models, the all-boys transition effect is 0.231 (p-value<0.05), the all-girls transition effect is -0.0275 (p-value>0.1), and the single-sex transition effect overall is 0.181 (p-value<0.05). The overall point estimate is very similar to that implied by the simple Wald estimate discussed in above. In sum, in the most credible of models, the transition to single-sex improved students' academic scores, and this increase was driven almost entirely by benefits for boys.

One key innovation of this paper is to account for school-level characteristics that are correlated with single-sex schools in cross-sectional analyses. To show the importance of this, models 6, 12, and 18 shows the estimated single-sex schooling effect for the post-transition cohorts only (i.e., using only the cross-school variation). The estimated single-sex school effect is small and is not statistically significantly different from zero for all groups. This demonstrates how relying on cross-school variation (even if it is clean) confounds single-sex status with other school differences, and may be very misleading about the causal effect of single-sex education *per se*.

IV.C. Effects on Test Scores for Individual Subjects

Students take several subjects for NCSE certification. This allows for an exploration of the effects of single-sex schooling on a broad set of subjects. Table 3 presents the estimated 2SLS single-sex effect in academic subjects (Math, English, Spanish, Science, Social Studies) and non-academic subjects (Arts, Physical Education, Technology). Looking at the combined single-sex transition effect (top panel), the point estimate for each of the academic subjects is positive and economically meaningful. However, the effects are only statistically significant in Math (0.213sd), and Social Studies (0.289sd). This suggests that the improvement in academic subjects was largely consistent across the individual subjects. In contrast, the estimated impacts on non-academic subjects are small, statistically insignificant, and differ in sign across subjects.

The theory suggests that because teachers may align instruction to boys in some subjects and girls in others, there could be some differences in the transition effects across subjects by sex. To explore this possibility, the middle panel (models 9 through 16) presents the effects for boys, and the lower panel (models 17 through 24) presents the effects for girls. For boys, the transition effect is positive in all subjects except Technical, and statistically significant in Math, Spanish, Social Studies, and Arts. The results for boys are consistent with an across the board improvement in test scores of about 0.2sd. Indeed, for none of the subjects can one reject that the real effect is 0.2 at traditional levels of statistical significance. Contrary to some hypotheses regarding the effect of single-sex schooling, there is no tendency for the all-boys effect to be any larger or smaller in more male-dominated subjects, or in subjects in which males tend to underperform historically. The results for females in the lower panel tell a different story than those for boys. The point estimates are positive in math and negative in the others, and none is statistically significant. This is consistent with the small average effects for girls in Table 2. The results suggest positive impacts on the average test scores of males (across all subjects) and no average effect for females.

IV.D. Distributional Effects

I now examine whether the average effects mask important impacts at other points of the test-score distribution. To examine this, I put each student's NCSE scores into a test score quintile (across all years) and I estimate the full 2SLS models where the outcome is an indicator variable denoting taking the NCSE and scoring at each quintile.[33] I present the coefficient estimates for boys and girls in Table 4. The top panel presents the marginal effects for math scores.

Because the average impacts were particularly pronounced for math (as is common for educational interventions) I first explore impacts on the math score distribution, and then examine academic and non-academic scores. The outcome in the first column is an indicator for taking the NCSE math exam. In the early years of the NCSE (which straddle the sample period) some schools did not participate in the NCSE. As such, NCSE non-participation is not necessarily an indication of dropout. However, any changes on this margin need to be taken into account. The point estimate indicates that after the transition boys were about 7.4 percentage points less likely to have taken the NCSE math exam (p-value<0.01). One may wonder if students who would have been low scoring were more likely to be missing, hence driving up the average scores. The data are inconsistent with this explanation. After the transition, boys were 12.4 percentage points less likely to be in the bottom quintile (p-value<0.05). If even the full reduction in NCSE taking came from this bottom quartile, there remains a five percentage point reduction in scoring at the bottom that cannot be accounted for with reduced test taking. Indeed, much of this reduction in the proportion of low-scoring boys is accounted for by a 4.9 percentage-point increase in the proportion scoring in the top 80 percent of the distribution (p-value<0.01). While the point estimates are both positive, there are no statistically significant changes in the proportion of boys scoring in the second or third or fourth quintile. However, boys after the transition are 6.27 percentage points more likely to be

[33] Because the NCSE is designed to test the same skills every year and are scored the same way every year, I define the test score quartiles pooling all the years of data.

in the top quintile (p-value<0.01). This indicates that the large improvements in average math scores for males are driven in large part by having fewer students with low scores (in the bottom 20%) and more students scoring at the top of the test score distribution (in the top 20%). This also suggests that much of the test score gains for boys are among the most able students.

For girls, the estimated effect on having missing NCSE data is a small 0.0135 and is not statistically significant – an indication that compositional change does not drive the results. While the positive impact of the transition on girls' math scores was not statistically significant, the quantile analysis suggests that there were some real improvements. The patterns for girls suggest that the small average gains obscured some real improvements at the top of the achievement distribution with little effect at the bottom. That is there is a 6.7 percentage point increase in the likelihood of scoring in the top 20 percent (p-value<0.01) in math. These patterns suggest a shifting of girls towards the top of the distribution. Single-sex education improved math scores throughout the achievement distribution for boys and led to some improvements among the ablest girls.

The middle panel presents the same analysis for academic subjects overall. These patterns largely echo those for math scores. Both girls and boys are more likely to score in the top 20 percent of the academic score distribution after the transition than before. Also, both girls and boys are less likely to score between the 41st and 60th percentiles. There are no statistically significant effects at other quintiles, suggesting that after the transition, both boys and girls who would have been in the middle of the achievement distribution were more likely to have high average performance in academic subjects than before. However, while there are fewer boys in the bottom 40 percent of the distribution, this is not true for girls so that the overall impact on average scores is more positive for boys than for girls. Finally, the bottom panel presents results for all subjects combined. Here, there is no impact for girls, but there are clear increases in the proportion of boys scoring in the top 20 percent of the average of all subjects (both academic and non-academic).

In sum, boys are more likely to score in the top of the achievement distribution in both academic and non-academic subjects. In contrast, girls are more likely to score in the top of the achievement distribution only in academic subjects. However, these gains in the top of the distribution for girls are somewhat offset by fewer students in the middle—leaving average test scores largely unchanged. Overall, the results reveal a robust positive test score effect for boys across most subjects. For girls, the results suggest modestly improved outcomes in academic subjects and possibly worse outcomes in non-academic subjects. While speculative, this pattern of results suggests that girls may have had a more academic orientation after the transition to all-girls, while boys experienced a general increase in scholarly engagement. The results in Section IV.E on advanced course taking and high school completion (which is achieved by those at the top of the academic achievement distribution) are consistent with this interpretation.

IV.E. Longer-Run Education Outcomes

To provide evidence on longer-run effects, Table 5 presents estimated effects on taking the school-leaving exam (the CSEC taken in 10th grade), and performance on that exam. These models include all students who took the SEA and not just those who take the exams. Note that these outcomes are only available for the first two post-transition cohorts. The top panel present results for boys and girls combined, the middle panel presents results for boys, and the lower panel presents results for girls. As before, each point estimate is obtained from a separate regression.

To avoid problems of multiple hypothesis testing, I focus on the two summary outcomes of CSEC performance outlined in Section II; the CSEC index, and the number of advanced courses taken. For the CSEC Index (Models 1,11, and 21), in the pooled sample of both boys and girls, the transition effect is 0.0791 (p-value<0.1). This modest and marginally statistically significant effect masks heterogeneity by gender. For boys, the transition effect on the CSEC index is 0.0002 (p-value<0.1), while that for girls is a sizable 0.2 (p-value<0.01). To better understand how female outcomes are improved, I examine effects on the individual components of the CSEC index for girls. In model 23, there is no appreciable change in the likelihood of dropping out of school by 10th grade, so that this is not driving the effect. A key academic outcome is earning a secondary school completion credential. Passing five subjects including math and English will earn one a CSEC certificate. Earning this certificate is a prerequisite to continuing in the traditional academic education system. After the transition, girls were 13.6 percentage points more likely to pass the English exam (p-value<0.01), 8.8 percentage points more likely to pass the math exam (p-value<0.01), and passed 0.64 more subjects (p-value<0.01). As such, girls were about 8.25 percentage points more likely to earn a secondary school completion certificate (p-value<0.01). Only 38 percent of students who enter secondary school leave with the prerequisites to enter tertiary education so that this represents a sizable 20 percent increase. Consistent with the overall CSEC index results, there are no significant impacts on the individual outcomes for boys.

The second summary measure is the number of advanced CSEC courses taken (English, math, science) in upper secondary school. While all students taking these exams will take English, math, and science, some students choose to take these subjects at the advanced level. Taking advanced level courses is a strong indicator of likely tertiary enrollment and signals motivation. This signaling value is similar to that of taking AP courses in the United States (Jackson 2010). The estimated transition effects in pooled model 2, reveals that students took about 0.184 more advanced courses afer the trasition to single sex (p-value<0.01). The effects are somewhat larger for males (0.180) than for females (0.121), but the two impacts cannot be distinguished from eachother. Looking at the individual components of the summary measure in the pooled sample, Models 4, 5, and 6 reveal that students were 6.3 percentage points more likely to take advanced English (*p*-value<0.01), 3.11 percentage points more likely to take advanced math (*p*-value<0.05), and 8.93 percentage points more likely to take advanced science (*p*-value<0.01) after the transition to single-sex. There is little effect on the participation margin so that these increases represent an increase in advanced course-taking conditional on staying in school. The positive effects on these higher-level academic outcomes are consistent with the large effects on test scores at the top of the achievement distribution for both boys and girls in academic subjects (Table 4).

The results suggest largely positive effects of single-sex education on the short-run academic outcomes of both boys and girls (with stronger short-run impacts for boys). The short-run impacts are likely driven by improvements at the upper end of the achievement distribution. Consistent with this, both boys and girls are more likely to take advanced courses in upper secondary school. However, examination performance is similar for boys after the transition, but increases markedly for girls. Overall, girls are about 20 percent more likely to complete secondary school after the transition. Even though boys are no more likely to complete secondary school, they do graduate from school with more advanced coursework. As such, the overall long-run academic impacts are likely positive for both boys and girls (albeit larger for girls).

IV.F. Social Outcomes

While the objective of all schools is to improve educational outcomes, parents often choose single-sex schools because they perceive that single-sex education can enhance social outcomes. For example, parents may send their daughters to all-girls schools in an attempt to reduce the likelihood of teen pregnancy and may send their sons to all-boys schools with the hope that such schools may be better able to control the boys and decrease the chances of crime and other poor behaviors. Conversely, opponents of single-sex schooling worry that single-sex schooling may lead to poor socialization with the opposite sex and reinforce negative gender views.[34] While parent and advocates have strong priors on the direction of these effects, there is little causal evidence on the topic. To speak to this issue, I analyze the effects of transitioning from coed to single-sex on arrests as a teenager and the likelihood of being a teen mother (for girls only). Table 6 presents the results for boys and girls in the top and bottom panels, respectively.

Model 1 shows that the likelihood of being arrested by the age of 17 is lower at pilot schools after the transition than before. Students at pilot schools were 2.86 percentage points less likely to have been arrested by age 17 (p-value<0.05). Looking at being arrested by age 18, the estimated effect increases in magnitude to a 3.12 percentage-point reduction (p-value<0.05). These reductions in arrests are driven by boys. Given that boys are much more likely to be arrested than girls, the marginal effects on girls are small and not statistically significant. In contrast, boys after the transition were 2.86 percentage points less likely to have been arrested by age 17 (p-value>0.1), and 5.96 percentage points less likely to have been arrested by age 18 (p-value<0.01). Since roughly 10 percent of boys at pilot schools before the transition had been arrested by age 18, this all-boys effect represents about a 60 percent reduction in arrest rates. This is similar in magnitude to the effect of attending a better school on crime (Deming 2011).

The last outcome studied is teen pregnancy. The independent variables are having a baby by the ages of 17 and 18. Because most teen births occur after the age of 17 (see Table 1), one might expect larger effects on having a baby at older ages. Indeed, this is the case. While the reduction on having a baby by age 17 is not significant in Model 7, Model 8 shows that the transition of pilot schools to all-girls reduced the likelihood of having a baby by the age of 18 by 4.54 percentage points (p-value<0.05). Relative to the baseline rates of teen pregnancy at pilot schools, this represents a reduction of about 80 percent. To put these estimates in perspective, these estimated reductions in teen pregnancy are as large as those from comprehensive teenage-pregnancy prevention programs (Rosenthal et al. 2009). Because pregnancy is likely more disruptive for females than for males, the reduced teen pregnancy provides a potential explanation for the improved CSEC performance of females but not for males.

The results paint a picture of improved outcomes across the board after the transition to single-sex. While I have not explored an exhaustive list of outcomes, and not all outcomes

[34] There is also a growing literature showing that test score impacts may not be related to important effects on longer run outcomes (Jackson 2018; Heckman 1999; Deming 2011).

improved for all subpopulations, the consistent pattern of generally positive results (across academic and social outcomes) suggests that there were some real improvements in human capital. I now examine the extent to which these results can be taken as causal.

IV.G. Specification Checks and Tests of Validity

As discussed in Section III, the DiRD model requires that the RD identifying assumptions are valid, the DiD identifying assumptions are valid, and the DiRD identifying assumption is valid. I summarize tests of these assumptions here (see Appendix D for a fuller discussion). All of the specification checks described below indicate that the DiRD estimates can be interpreted causally.

Testing the Regression Discontinuity Identifying Assumptions

For the RD design to be valid, there should be no gaming of the cutoffs, and outcomes must vary smoothly through the cutoffs. One diagnostic for gaming is to test for non-smoothness in density through the cutoff. Such tests reveal no change in density through the cutoffs (appendix Figure D1). To test for smoothness in outcomes through cutoffs, one can test for smoothness of latent outcomes (as predicted by covariates) through cutoffs. Predicted outcomes (i.e. fitted values from a regression of academic scores on repeater status in 5th grade, religion, primary school fixed effects, selectivity of each secondary school choice, age at sitting the SEA, and month of birth) vary smoothly through the cutoffs (see column 1 in Table 7). Estimated impacts on each individual covariate are presented in Appendix Table D1. Of the 41 covariates tested, 4 are significant at the 10 percent level – this is consistent with random chance.

Testing the Difference-in-Difference Identifying Assumptions

For the DiRD estimates to represent the causal effect of single-sex education requires that there were no changes in the pilot schools that coincided with the pilot program. Confounding changes within schools is unlikely because (a) schools had no control over pilot status, (b) the MOE stipulated that no other changes take place in pilot schools, and (c) schools had little time to react to the pilot school announcement before the start of the school year. Even so, one may worry that pilot schools were already improving before transition. To test for this, Figure 5 shows the RD estimates for attending a pilot school for each year three years before and after the transition on academic scores, the number advanced courses, the CSEC index, having a baby by 18 and being arrested by 18. Each data point is an RD point estimate. The 90 percent confidence interval is presented for each estimate. For all the outcomes for which there are significant transition effects, there is little evidence of pre-trending and the change in outcomes occurs at the transition of soon thereafter – compelling evidence that the observed changes in school effectiveness are real.

Even though the MOE stipulated that there be no changes, as a final check on the DiD assumptions of no other changes, I supplement the testing data with survey data on a sample of teachers at all schools available for the years 2009 and 2013.[35] In a difference-in-difference regression model (Table 5), I find no evidence of systematic changes in teacher gender, teacher age, teacher education, or school cohort sizes after the transition to single-sex at pilot schools. These results support the MOE's claim that there were no other changes at pilot schools.

Testing the Additional DiRD Identifying Assumption

Because the DiRD strategy is a comparison of two Local Average Treatment Effects (LATE), for it to uncover the transition effect requires that the LATEs identified in both periods be the same. There are two ways that a violation of this condition could look like a positive transition effect. The first is if the marginal applicants to pilot schools (i.e. the compliers) after the transition benefited more from attending pilot schools than those before. If so, due to treatment heterogeneity it would appear that the transition increased the effectiveness of pilot schools.[36] As documented in Appendix E, using an approach similar to that outlined in Abadie (2003) and Angrist and Fernandez-Val (2014) the compliers were very similar before and after the transition so that the LATE should also be similar.[37] To assess whether any small differences pose a problem, I interact attending a pilot school (in any year) with the rich set of observed student characteristics.[38] If heterogeneous pilot-school treatment effects (along the numerous measured student dimensions) explain the transition effect, then the transition effect will go away in models with these interactions. Figure 6 presents the single-sex transition effect in models that account for pilot-school treatment heterogeneity in each of these dimensions. In all models, the single-sex

[35] Ideally, one would have administrative data on individual teacher that can be linked over time to track turnover and classroom placements more closely. Unfortunately, such data are not available.

[36] Figure 5 provides evidence that this does not drive the results. The RD estimate is similar for the 2010 and 2012 cohorts. While the 2012 cohort made their school choices with full knowledge of the transition, students in the 2010 cohort who were assigned to a pilot school were automatically assigned to the sister/brother school in each pair that matched their sex. As such, the applicants to the pilot schools in 2009 and 2010 should have been more similar than those between 2009 and 2012. Despite this, the treatment effect is very similar for 2010 and 2012.

[37] Specifically, both boys and girls were less likely to list a pilot school as one of their four secondary school choices, and while female applicants had similar incoming test scores before and after the transition, the average male applicant had lower incoming scores after the transition.

[38] These interacted characteristics include incoming test scores, whether the top choice school was single-sex, the number of single-sex schools chosen (a measure of preferences for single-sex education), the selectivity of schools in the student choices, the selectivity of their primary school, whether the primary school was single-sex, and predicted academic scores (based on repeater status in 5th grade, religion, primary school fixed effects, selectivity of each secondary school choice, age at sitting the SEA, and month of birth).

effect is similar to the main model. This suggests that the transition effect cannot be explained by different pilot school LATEs with respect to the rich set of student characteristics so that the DiRD identifying assumption is likely satisfied.

A second way the pilot school LATE could change is if the schools that students list just below the pilot school (i.e., the counterfactual school) tended to be worse after the transition than before. I assess this directly by testing for whether the ranking (in average SEA scores) of the school listed just below the pilot school (among pilot school applicants) changed after the transition. Reassuringly, among the compliers, the selectivity of the next ranked school is unchanged after the transition.[39] In sum, while no individual test is dispositive, the set of empirical tests as a whole suggest that heterogeneity in the pilot school LATE does not drive the main results.

Testing that These Effects Are Unlikely to Have Arisen by Random Chance

One may worry that the documented treatment effects could have happened by random chance by designating any 20 schools as pilot schools. To assess this, I estimated each school's likelihood of being a pilot school based on achievement level, school type, and proximity to other schools and identified 18 non-pilot schools with similar likelihoods as the actual pilot schools. These schools could have been chosen for the pilot and form a natural placebo test. I estimate the change in the effect of attending a placebo school before and after 2010 (see Appendix Table D2). For academic scores, advanced courses, arrests, and teen birth, the placebo estimates are small and not statistically significant. The placebo transition effect is statistically significant for the CSEC index, but it has the opposite sign as the pilot transition effect. This suggests that the estimated positive effects are real and are unlikely to have arisen by the choice of pilot schools.

IV.H. Evidence on Mechanisms

Single-sex education effects likely reflect some combination of the three mechanisms outlined in the model: (a) direct gender peer interaction effects, and the two indirect peer effects; (b) the boutique effect, and (c) the focus effect. Even though it is impossible to disentangle these different mechanisms, I can test for these mechanisms using survey data. In 2013, 2014 and 2015, I administered anonymous surveys to students in pilot schools and 20 comparison coed schools.[40] I designed the survey questions to be sensitive to the mechanisms outlined in the model.

---

[39] Using the approach laid out in Angrist and Fernandez-Val (2014), I find that the selectivity of the next ranked school of the compliers was unchanged after the transition.

[40] The 20 comparison schools were chosen based on selectivity, school type and location (as were the pilot schools).

To identify the effect of the single-sex transition on survey responses, I rely on the assumption that students in the same school in the same grade across cohorts will be exposed to essentially the same school inputs. Also, I only analyze answers to questions about classroom peers and teachers and not about the students themselves so that any changes in student selection are less of a concern. I compare survey responses of students in the same school and grade across cohorts. By looking within school grade cells, one holds school-grade level inputs constant, and one also holds the level of student maturity constant. I estimate the following model by OLS.

[3] $$Q_{ijg} = (All\_Boys_{jg})\delta_1 + (All\_Girls_{jg})\delta_2 + \pi_{jg} + \varepsilon_{ijg}.$$

Q_{ijg} is the response of student i in school j in grade g to the survey question, π_{gj} is a fixed effect for the school and grade, and ε_{ijg} is the error term. *All_Boys* and *All_Girls* are indicator variables equal to *1* if the school grade cell is all boys or all girls respectively in that year and 0 otherwise. The parameters of interest are δ_1 and δ_2 — the effect of being in an all-boys or all-girls cohort (holding other school attributes fixed). Standard errors are clustered at the school-grade level.

Before presenting results, I first establish that the survey results are likely valid. The overall survey response rate was 67 percent. Appendix F shows that there is no differential response by treatment status. To provide additional checks on the validity of the survey responses, I included three validation questions. The first question is whether most of the student's friends are the same gender as them, the second is whether their parents think education is important, and the third is the reported gender of the teacher. If the survey results are valid (i.e. students respond as expected, there is no differential selection of students, and there were no personnel changes at treatment schools), one should see significant effects on students reporting that most of their friends are the same gender as them, we should see no effect on whether an individual's parents think education is important, and we should see no effect on teacher gender. These are the patterns observed.

Are the positive effects driven by direct peer effects?

Lavy and Schlosser (2012) and Hoxby (2001) find that classrooms with larger shares of boys (on the margin) tend to be more disruptive and that more boys hurts all students. However, for single-sex schooling to generate positive single-sex effects for both boys and girls, either the indirect peer effects have to be large enough to offset any negative direct peer effect for boys, or direct peer effects would have to be asymmetric such that boys have better outcomes when exposed to more boys, while girls have better outcomes when exposed to more girls (as found in Whitmore 2005, Black et al 2013; Ooserbeek and van Ewijk 2014; and Lu and Anderson 2015). To test for this, I created four indexes designed to capture the direct peer effect mechanisms presented in the gender peer effects literature. The four measures describe whether (1) peers are disruptive, (2) peers distract students from doing schoolwork, (3) students learn from their peers, and (4) peers make students anxious.[41] The estimates are presented in Table 8.

Girls in the all-girls cohorts are less likely to report that their peers are disruptive and that their peers are distracting, and more likely to report that they learn from their classmates. These effects are large and statistically significant. Consistent with the "boys are bad" hypothesis, boys in the all-boys cohorts are more likely to report that their peers are disruptive and that their peers are distracting, and less likely to report that they learn from their peers. However, the effects for boys are much smaller in magnitude than for girls, and only the effect on peer learning is statistically significantly different from zero. Also, while boys learn less from their peers, there is some evidence of *positive* direct gender peer effects for males in other dimensions. Specifically, boys are less likely to report that their peers make them anxious in all-boys cohorts. That is, boys were less likely to worry about impressing classmates, less afraid to raise their hands in class, and were less nervous when they were compared to their classmates. There is no such effect for girls. Taken together, while the results indicate positive direct peer effects for girls, the overall direct peer effects for boys is ambiguous. If the benefits of being less anxious around boys are larger (or smaller) than the benefits of learning more from one's classmates in coed settings, the direct peer effect for boys would be positive (or negative).[42] However, the positive effects for boys could also be due to large positive indirect peer effect. I explore this possibility below.

Are the effects driven by the boutique or the focus effect?

The other hypothesized mechanisms behind a single-sex schooling effect are indirect peer effects driven by changes in teacher instructional practices that may lead to efficiency gains (the focus effect) or alignment of instruction to the particular needs of each sex (the boutique effect). While I am unable to observe the extent to which instruction is aligned toward boys or girls, I can test for changes in student reports of teachers' instructional practices. Specifically, if there are different changes in teacher practices between the all-girls and all-boys environments, it would be evidence of the boutique effect. Conversely, if there are similar changes in teacher practices in the

[41] See Appendix G for detailed discussion of the individual survey questions used to create these measures.

[42] Note that boys do not report much higher levels of disruption in the all-boys cohorts so that the negative direct gender peer effects through increase disruption for boys may be small. That is, even if having more boys is disruptive, boys do not perceive it as such and may not be harmed by it.

all-girls and all-boys settings, it would be evidence of the focus effect.

Because there are several survey questions, I created six indexes that measure teacher practices: The first is the extent to which teachers give students individual attention. The second is whether teachers use examples in their instruction. The third is the degree to which teachers monitor and track student understanding during class. The fourth is whether teachers are warm toward students. The fifth is whether teachers involve students in instruction. The sixth is whether the teacher is strict.[43] Each measure is standardized to be mean zero with unit variance.

Table 8 indicates that both boys and girls report higher levels of individual teacher attention in the single-sex cohorts. The fact that there is more reported teacher attention in both all-boys and all-girls cohorts is indicative of an indirect peer effect and is suggestive of the focus effect. That is, there may be some efficiency gains associated with having a more homogenous student population that allows teachers to provide more individualized attention. This is generally supported by the fact that both boys and girls report higher levels of teacher warmth in the single-sex cohorts (only the effect for boys is statistically significant). The evidence of the boutique effect (i.e., using different instructional practices in the all-girls and all-boys environments) is weak. There are no statistically significant effects on teachers using examples in class, on teachers tracking student understanding, or on whether teachers are strict. The one practice for which there is a change is the extent to which teachers involve students. Teachers appear to be marginally more likely to involve students in the all-girls cohorts and slightly *less* likely in the all-boys cohorts.

In sum, the results suggest that single-sex environments provide positive direct peer interaction effect for girls, while the sign of the direct peer interactions for boys is ambiguous. Looking at indirect peer effects through teacher behaviors, the results are consistent with efficiency gains to being in a single-sex classroom that allows teachers to give students more individualized instruction and exhibit greater warmth (i.e., a focus effect). However, I find little evidence of greater alignment of instruction to the particular needs of each sex (i.e., a boutique effect).

V. Discussion and Conclusions

The merits of single-sex education have been debated for decades. A few recent studies have relied on quasi-random variation to provide credible evidence that attending a single-sex school tends to confer benefits over attending coed schools (Lee et al. 2014; Jackson 2012; Park,

[43] See Appendix F for details on the individual survey questions used to create these measures.

Behrman and Choi 2013). However, because coed and single-sex schools may differ in unobserved ways, these cross-school comparisons may not isolate the impact of single-sex schooling *per se,* and do not speak to the policy question of whether expanding single-sex education into existing coed schools is good for students. To answer the policy question, this paper examines a policy in which twenty low-performing secondary schools in Trinidad and Tobago were chosen by the Ministry of Education to transition from coed to single-sex (with no other changes in personnel, inputs, or school policy). By exploiting exogenous variation in school assignments, I isolate the effect of schools from that of students. By exploiting changes in single-sex status within schools over time, I isolate the effect of single-sex status from that of the fixed characteristics of the schools that happen to provide single-sex instruction.

Three years after attending a single-sex secondary school, boys have higher scores on standardized tests (relative to students at the same school when it was coed). Five years later, both boys and girls are more likely to take and pass advanced courses, and girls are more likely to have completed secondary school. One limitation of existing studies on single-sex education is the lack of evidence on nonacademic outcomes. This paper presents some new results in this regard. Boys are less likely to have been arrested, and girls are less likely to be teen mothers in the single-sex environments. The theoretical framework highlights that the single-sex effect will depend on the nature of direct gender peer effects and also indirect peer effects that operate through the behaviors of teachers in both single-sex and coed environments. This is supported by survey results indicating that both direct and indirect peer effect mechanisms may have played an important role.

It is worth noting that the benefits to single-sex education were achieved at zero additional financial cost. To achieve comparable results through increases in school spending, reductions in class size, tutoring, or other interventions would require a sizable financial outlay. The results of this study illustrate the potential cost-effectiveness of leveraging peer effects (both direct and indirect) to improve student outcomes (both academic and otherwise). The evidence demonstrates that single-sex education *can be* an effective, low-cost way to improve outcomes for low-achieving students. However, further work is needed to identify better the contexts in which single-sex instruction is likely to improve the outcomes for both girls and boys. The theoretical framework and the evidence on mechanisms presented herein may guide future research in these areas.

Works Cited

Abadie, A. (2003), "Semiparametric Instrumental Variable Estimation of Treatment Response Models." *Journal of Econometrics*, 113(2), 231–63.

Abdulkadiroğlu, Atila, Parag A. Pathak, and Christopher R. Walters. 2018. "Free to Choose: Can School Choice Reduce Student Achievement?" *American Economic Journal: Applied Economics*, 10 (1): 175-206.

Angrist Joshua, and Ivan Fernandez-Val, ExtrapoLATE-ing: External Validity and Overidentification in the LATE Framework, published in Advances in Economics and Econometrics Tenth World Congress (2014).

Booth, Alison L., Lina Cardona-Sosa, and Patrick Nolen. 2013. "Do Single-Sex Classes Affect Exam Scores? An Experiment in a Coed University." *IZA Discussion Paper Number 7207.*

Black, Sandra E. & Paul J. Devereux & Kjell G. Salvanes, 2013. "Under Pressure? The Effect of Peers on Outcomes of Young Adults," Journal of Labor Economics, University of Chicago Press, vol. 31(1), pages 119 - 153.

Brame, Robert; Shawn Bushway, Raymond Paternoster and Michael Turner (2014) "Demographic Patterns of Cumulative Arrest Prevalence by Ages 18 and 23." Crime and Delinquency 60:3:471-486

Chetty, Raj, John N Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. 2011. "How does your kindergarten classroom affect your earnings? Evidence from Project STAR." The Quarterly Journal of Economics, 126(4): 1593–1660.

Chade, H., Smith, L., 2006. Simultaneous search. Econometrica 74 (5), 1293–1307.

Chavous, Kevin P. 2013. "Single-Sex Education: A Viable Option for Today's Students." *The Huffington Post.* 5 4. Accessed 3 5, 2015. http://www.huffingtonpost.com/kevin-p-chavous/singlesex-education-a-via_b_3015145.html.

Deming DJ. Better Schools, Less Crime?. Quarterly Journal of Economics. 2011;126 (4) :2063-2115.

Ding, Weili, Lehrer, Steven F., 2007. Do peers affect student achievement in China's secondary schools? The Review of Economics and Statistics 89 (2), 300–312.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2011. "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." *American Economic Review* (101) 5: 1739-74.

Dustmann, Ku, and Kwak (2017) "Why Are Single-Sex Schools Successful?," CESifo Working Paper Series 6535, CESifo Group Munich.

Eliot, Lise "FORGET WHAT YOU THINK YOU KNOW ABOUT THE BENEFITS OF SINGLE-SEX SCHOOLING" Newsweek, December, 31, 2016.

Gale, David, and Shapley, Lloyd, 1962. College admissions and the stability of marriage. Am. Math. Mon. 69 (1), 9–15.

Heckman, James. 1999. "Policies to Foster Human Capital." *NBER Working Paper No 7288.*

Hoxby, Caroline M. 2000. "Peer effects in the classroom: Learning from gender and race variation." *NBER Working Paper no. 7867.*

Hoxby, Caroline and Gretchen Weingarth. 2006. "Taking Race Out of the Equation: School Reassignment and the Structure of Peer Effects." Presented at the 2006 American Economic Association Annual Meetings.

Jackson, C. Kirabo. (2010) "Do Students Benefit From Attending Better Schools?: Evidence From Rule-based Student Assignments in Trinidad and Tobago" *The Economic Journal*, 120(549): 1399-1429.

Jackson, C. Kirabo. 2013. "Can Higher-Achieving Peers Explain the Benefits to Attending Selective Schools?: Evidence from Trinidad and Tobago." *Journal of Public Economics.*

Jackson, C. Kirabo, (2018), What Do Test Scores Miss? The Importance of Teacher Effects on Non–Test Score Outcomes, *Journal of Political Economy*, **126**, issue 5, p. 2072 - 2107.

Jackson, C. Kirabo. 2012. "Single-sex schools, student achievement, and course selection: Evidence from rule-based student assignments in Trinidad and Tobago." *Journal of Public Economics* 96 (2012): 173-187.

Jackson, C, K, J. E. Rockoff, and D. O. Staiger. Teacher Effects and Teacher-Related Policies. Annual Review of Economics, 6(1):801–825, 2014.

Kearney, Melissa S. and Phillip B. Levine. 2012. "Why Is the Teen Birth Rate in the United States So High and Why Does It Matter?" *Journal of Economic Perspectives*, 26(2): 141-63.

Krueger, Alan B. 1999. "Experimental estimates of education production functions." The quarterly journal of economics, 114(2): 497-532.

Lavy, Victor, and Analia Schlosser. 2012. "Mechanisms and Impacts of Gender Peer Effects at School." *American Economic Journal: Applied Economics* 3 (2): 1-33.

Lee, Soohyung, Lesley J. Turner, Seokjin Woo, and Kyunghee Lim. 2014. "All or Nothing? The Impact of School and Classroom Gender Composition on Effort and Academic Achievement." *NBER Working Paper No 20722.*

Lu, Fangwen, and Michael L. Anderson. 2015. "Peer Effects in Microenvironments: The Benefits of Homogeneous Classroom Groups". *Journal of Labor Economics* 33 (1): 91–122.

McCrary, Justin, (2008), Manipulation of the running variable in the regression discontinuity design: A density test, *Journal of Econometrics*, 142, issue 2, p. 698-714.

Mael F., A. Alonso, D. Gibson, K. Rogers, and M. Smith. 2005. Single-sex Versus Coeducational Schooling: A Systematic Review. Washington, DC: US Department of Education.

Oosterbeek, Hessel., Reyn van Ewijk "Gender peer effects in university: Evidence from a randomized experiment" Economics of Education Review Volume 38, February 2014, Pages 51–63.

Park, Hyunjoon, Jere R. Behrman, and Jaesung Choi. 2013. "Causal Effects of Single-Sex Schools on College Entrance Exams and College Attendance: Random Assignment in Seoul High Schools." *Demography* 50 (2): 447-469.

Rosenthal, Marjorie S., Joseph S. Ross, RoseAnne Bilodeau, Rosemary S. Richter, Jane E. Palley, and Elizabeth H. Bradley, "Economic Evaluation of a Comprehensive Teenage Pregnancy Prevention Program: Pilot Program" American Journal of Preventative Medicine. 2009 Dec; 37(6 Suppl 1): S280–S287.

Stevens, W. David, Lauren Sartain, Elaine M. Allensworth, and Rachel Levenstein; with Shannon Guiltinan, Nick Mader, Michelle Hanh Huynh, and Shanette Porter. "Discipline Practices in Chicago Schools: Trends in the Use of Suspensions and Arrests" University of Chicago Report 2015.

Strain, Michael (2013) "Single-sex classes & student outcomes: Evidence from North Carolina" Economics of Education Review Volume 36, October 2013, Pages 73–87

U.S. Department of Education. 2014. *Digest of Education Statistics.* National Center for Educational Statistics.

Whitmore, Diane, "Resource and Peer Impacts on Girls' Academic Achievement: Evidence from a Randomized Experiment," American Economic Review: Papers and Proceedings, 2005, 95 (2), 199-203.

Table 1:
Summary Statistics

	1	2	3	4	5	6
	All Boys Taking the SEA	Boys at All-Boy's Pilot Schools Pre-Transition	Boys at All-Boy's Pilot Schools Post-Transition	All Girls Taking the SEA	Girls at All-Girl's Pilot Schools Pre-Transition	Girls at All-Girl's Pilot Schools Post-Transition
Assigned to Pilot School	0.184	0.878	0.925	0.175	0.915	0.946
	(0.388)	(0.328)	(0.263)	(0.380)	(0.279)	(0.227)
Times Repeated SEA	0.050	0.045	0.068	0.047	0.055	0.046
	(0.218)	(0.208)	(0.251)	(0.212)	(0.229)	(0.210)
Std. Total SEA Score	-0.172	-0.332	-0.636	0.177	-0.503	-0.364
	(1.047)	(0.746)	(0.792)	(0.916)	(0.741)	(0.715)
Std. Math NCSE Score	-0.175	-0.566	-0.702	0.183	-0.270	-0.117
	(0.987)	(0.825)	(0.848)	(0.960)	(0.840)	(0.784)
Std. English NCSE Score	-0.251	-0.644	-0.721	0.254	-0.265	-0.101
	(0.985)	(0.866)	(0.940)	(0.933)	(0.922)	(0.796)
Academic NCSE	-0.216	-0.602	-0.686	0.220	-0.261	-0.116
	(0.891)	(0.703)	(0.789)	(0.864)	(0.759)	(0.703)
CSEC Index	0.053	-0.227	-0.245	0.338	-0.188	0.067
	(0.926)	(0.711)	(0.661)	(0.968)	(0.750)	(0.792)
Total Advanced Courses	0.518	0.319	0.310	0.670	0.264	0.372
	(0.826)	(0.536)	(0.578)	(0.889)	(0.539)	(0.611)
Take the NCSE	0.835	0.916	0.847	0.872	0.832	0.852
	(0.372)	(0.278)	(0.360)	(0.335)	(0.374)	(0.355)
Take the CSEC	0.771	0.75	0.729	0.844	0.814	0.794
	(0.420)	(0.433)	(0.444)	(0.363)	(0.389)	(0.404)
Earn Secondary Certificate	0.266	0.123	0.106	0.384	0.132	0.214
	(0.442)	(0.329)	(0.308)	(0.486)	(0.339)	(0.410)
Arrested by age 16	0.034	0.044	0.032	0.007	0.014	0.007
	(0.182)	(0.205)	(0.176)	(0.081)	(0.116)	(0.082)
Arrested by age 18	0.075	0.094	0.059	0.014	0.025	0.011
	(0.263)	(0.292)	(0.235)	(0.116)	(0.156)	(0.106)
Baby by age 16				0.009	0.012	0.014
				(0.093)	(0.109)	(0.117)
Baby by age 18				0.043	0.076	0.046
				(0.203)	(0.266)	(0.209)
Students	62953	3259	4627	61429	3311	4016

Notes: This dataset is based on the population of SEA takers during years 2006 through 2012. All SEA and NCSE scores are standardized to be mean zero and unit variance in each cohort. Earning a certificate means passing five subjects in the CSEC exams (10th grade) including math and English. This is the prerequisite to entering tertiary education. Summary statistics are provided for all boys and girls taking the SEA and also the subsample of boys and girls who attended the pilot school pre and post transition to single-sex. Standard errors are presented in parentheses.

Table 2:

Single-Sex Transition Effects on Academic Scores in 8th Grade

	All Years					Post Only [a.]
	OLS	OLS	OLS	2SLS	2SLS	2SLS
	Both Boys and Girls (306,445 observations)					
	1	2	3	4	5	6
Pilot*POST	0.0249	0.0999**	0.0860*	0.222**	0.181*	-0.0168
	[0.0729]	[0.0367]	[0.0402]	[0.0553]	[0.0825]	[0.0926]
	Boys Only (150,296 observations)					
	7	8	9	10	11	12
Boys-Pilot*POST	-0.102+	0.104+	0.149**	0.236**	0.231*	-0.0292
	[0.0565]	[0.0546]	[0.0508]	[0.0619]	[0.111]	[0.141]
	Girls Only (156,149 observations)					
	13	14	15	16	17	18
Girls-Pilot*POST	0.129**	0.0595+	0.113*	0.0444	-0.0275	0.00109
	[0.0398]	[0.0351]	[0.0451]	[0.0786]	[0.109]	[0.0889]
Test Scores	N	Y	Y	Y	Y	Y
Choices	N	N	Y	N	Y	Y
Applicant Group	N	N	N	Y	Y	Y

Robust standard errors in brackets adjusted for clustering at the assigned school level.

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

Notes: Each column represents a separate regression. All models include the effect of attending a pilot school so that the coefficients on Boys-Pilot*POST and Girl-Pilot*POST represent the post-transition change in the effect of attending an all-boys pilot school and all-girls pilot school. The coefficient on Pilot*POST represents the difference between the effect of attending any pilot school after transition and attending any pilot school prior to transition. Test scores refers to fourth order polynomial of incoming SEA test scores interacted with SEA cohort and gender.

a. The post only models exclude the SEA cohorts prior to 2010. Accordingly, the number of observations for this model is less than that for the full model. There are 62,011, 64,310, and 126,321 male, female, and total observations in the post only models, respectively.

Table 3:
Effects on Scores by Subject

	Math	English	Spanish	Science	Social Studies	Arts	Technical	Physical Education
	Boys and Girls Combined: 300,412 observations							
	1	2	3	4	5	6	7	8
Pilot*POST	0.213*	0.0707	0.126	0.182	0.289**	0.0557	-0.111	0.0639
	[0.0898]	[0.0814]	[0.0993]	[0.116]	[0.107]	[0.168]	[0.204]	[0.123]
	Boys Only: 147,468 observations							
	9	10	11	12	13	14	15	16
Boys-Pilot*POST	0.271**	0.0515	0.265*	0.205	0.331**	0.590*	-0.15	0.0759
	[0.102]	[0.128]	[0.107]	[0.155]	[0.123]	[0.252]	[0.197]	[0.207]
	Girls Only: 153,244 observations							
	17	18	19	20	21	22	23	24
Girls-Pilot*POST	0.0784	-0.0982	-0.0583	-0.0392	-0.0295	-0.318	-0.103	-0.022
	[0.0948]	[0.109]	[0.167]	[0.138]	[0.108]	[0.188]	[0.199]	[0.135]
Choice Effects	Y	Y	Y	Y	Y	Y	Y	Y
Cutoff Effects	Y	Y	Y	Y	Y	Y	Y	Y
SEA Scores	Y	Y	Y	Y	Y	Y	Y	Y

Robust standard errors in brackets are adjusted for clustering at the assigned school level.

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

Notes: Each column in each panel (top and bottom) represents a separate regression and is indicated with a specification number (1 through 20). The dependent variable is indicated in the top row. All models include the effect of attending a pilot school so that the coefficients on Boys-Pilot*POST and Girls-Pilot*POST represent the post-transition change in the effect of attending an all-boys pilot school and all-girls pilot school. Similarly, the coefficient on Pilot*POST represents the difference between the effect of attending any pilot school after transition and attending any pilot school prior to transition. All models include fourth order polynomial of incoming SEA test scores interacted with SEA cohort and gender, cohort fixed effects, indicators for gender, cutoff fixed effects, and choice group indicator variables.

Table 4:

Nonlinear Effect on Test Scores on Eighth Grade Test Scores

	Math					
	Missing	Bottom 20%	21st to 40th Percentile	41st to 60th Percentile	61st to 80th Percentile	Top 20%
Pilot*POST	0.0557**	-0.0740**	-0.0275	-0.00915	-0.0215	0.0765***
	[0.0262]	[0.0345]	[0.0265]	[0.0274]	[0.0193]	[0.0194]
Boys-Pilot*POST	0.0747**	-0.124***	-0.00589	0.0187	-0.0261	0.0627***
	[0.0355]	[0.0434]	[0.0359]	[0.0350]	[0.0268]	[0.0221]
Girls-Pilot*POST	0.0135	-0.0396	0.0121	-0.045	-0.00842	0.0674***
	[0.0325]	[0.0471]	[0.0383]	[0.0395]	[0.0295]	[0.0253]
	Academic Subjects					
	Missing	Bottom 20%	21st to 40th Percentile	41st to 60th Percentile	61st to 80th Percentile	Top 20%
Pilot*POST	0.0540**	-0.0282	-0.0207	-0.0673***	-0.0123	0.0744***
	[0.0258]	[0.0332]	[0.0250]	[0.0226]	[0.0327]	[0.0183]
Boys-Pilot*POST	0.0720**	-0.0534	-0.00656	-0.0790**	0.0117	0.0552***
	[0.0349]	[0.0455]	[0.0345]	[0.0394]	[0.0245]	[0.0175]
Girls-Pilot*POST	0.0129	0.0187	0.0141	-0.0600*	-0.0265	0.0408*
	[0.0325]	[0.0539]	[0.0424]	[0.0310]	[0.0630]	[0.0210]
	All Subjects (Both Academic and Non-Academic Subjects)					
	Missing	Bottom 20%	21st to 40th Percentile	41st to 60th Percentile	61st to 80th Percentile	Top 20%
Pilot*POST	0.0540**	-0.0587*	0.00435	-0.0413	-0.00222	0.0438**
	[0.0258]	[0.0300]	[0.0256]	[0.0290]	[0.0227]	[0.0207]
Boys-Pilot*POST	0.0720**	-0.0735*	0.0263	-0.0574	-0.0137	0.0463**
	[0.0349]	[0.0432]	[0.0453]	[0.0408]	[0.0278]	[0.0190]
Girls-Pilot*POST	0.0129	-0.00244	0.0422	-0.0231	-0.0183	-0.0113
	[0.0325]	[0.0498]	[0.0386]	[0.0400]	[0.0316]	[0.0220]

Robust standard errors in brackets adjusted for clustering at the assigned school level.

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

Each column represents a separate regression. All models include the effect of attending a pilot school so that the coefficients on Boys-Pilot*POST and Girls-Pilot*POST represent the post-transition change in the effect of attending an all-boys pilot school and all-girls pilot school. All models include fourth order polynomial of incoming SEA test scores interacted with SEA cohort and gender, cohort fixed effects, indicators for gender, cutoff fixed effects, and choice group indicator variables. The number of observations is larger than in the test score regressions because these models include observations that are missing NCSE test score data. The quintile dummies connote both taking the NCE exam and scoring at a particular quintile of the distribution. The male sample includes 180,932 observations and the female sample has 178,859 observations.

Table 5:

Effects on Educational Attainment Outcomes (5 Years later)

	Summary Outcomes		Individual Outcomes							
	CSEC Index	Advanced Courses	Dropout Before 10th grade	Take Advanced English	Take Advanced Math	Take Full Science	Pass Math	Pass English	Subjects Passed	Earn Certificate
	Boys and Girls Combined: 307,129 observations									
	1	2	3	4	5	6	7	8	9	10
Pilot*Post	0.0791+	0.184**	-0.000258	0.0634**	0.0311*	0.0893**	0.0244	0.0564*	0.310+	0.0219
	[0.0471]	[0.0429]	[0.0278]	[0.0228]	[0.0145]	[0.0309]	[0.0251]	[0.0283]	[0.165]	[0.0231]
	Boys Only: 154,616 observations									
	11	12	13	14	15	16	17	18	19	20
Boys-Pilot*Post	0.00215	0.180**	-0.00807	0.0730*	0.0559**	0.0515	-0.0377	0.0132	0.164	-0.0027
	[0.0633]	[0.0395]	[0.0378]	[0.0325]	[0.0178]	[0.0313]	[0.0364]	[0.0341]	[0.212]	[0.0233]
	Girls Only: 152,513 observations									
	21	22	23	24	25	26	27	28	29	30
Girls-Pilot*Post	0.201**	0.121+	0.0225	0.0665	0.00696	0.0473	0.0881**	0.136**	0.638**	0.0825**
	[0.0539]	[0.0709]	[0.0498]	[0.0443]	[0.0183]	[0.0682]	[0.0292]	[0.0437]	[0.179]	[0.0282]
Cutoff Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Choice Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
SEA score	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Robust standard errors in brackets adjusted for clustering at the assigned school level.

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

Notes: These outcomes are measured using the CSEC data from 2012-2016. As such, the regression results are only included for the SEA cohorts before 2012. Each column represents a separate regression. The dependent variable is indicated in the top row. All models include the effect of attending a pilot school so that the coefficients on Boys-Pilot*POST and Girls-Pilot*POST represent the post-transition change in the effect of attending an all-boys pilot school and all-girls pilot school. Similarly, the coefficient on Pilot*POST represents the difference between the effect of attending any pilot school after transition and attending any pilot school prior to transition. All models include fourth order polynomial of incoming SEA test scores, cohort fixed effects, indicators for gender, cutoff fixed effects, and choice group indicator variables.

Table 6:

Effects on Arrests and Teen Births

	Arrest by 17	Arrest by 18	Baby by 17	Baby by 18
	Boys and Girls: 359,791 Observations			
	1	2		
Pilot * Post	-0.0286*	-0.0312*	-	-
	[0.0128]	[0.0158]	-	-
	Boys Only: 180,932 Observations			
	3	4		
Boys-Pilot * Post	-0.0269	-0.0596**	-	-
	[0.0186]	[0.0212]	-	-
	Girls Only: 178,859 Observations			
	5	6	7	8
Girls-Pilot * Post	0.0124	0.0116	-0.0211	-0.0454*
	[0.00882]	[0.00984]	[0.0176]	[0.0219]
Choice Group	Y	Y	Y	Y
Application Group	Y	Y	Y	Y
SEA Scores	Y	Y	Y	Y

Robust standard errors in brackets adjusted for clustering at the assigned school level.

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

Each column represents a separate regression. The dependent variable is indicated in the top row. All models include the effect of attending a pilot school so that the coefficients on Boys-Pilot*POST and Girls-Pilot*POST represent the post-transition change in the effect of attending an all-boys pilot school and all-girls pilot school. All models include fourth order polynomial of incoming SEA test scores interacted with SEA cohort and gender, cohort fixed effects, indicators for gender, cutoff fixed effects, and choice group indicator variables.

Table 7:
Evidence of no other Changes

	1	2	3	4	5
	2SLS	DiD	DiD	DiD	DiD
		Teacher Variables at School-Year Level			
	Predicted Average Academic Scores [b.]	Female	BA Degree	Year of Birth	Cohort Size
Girls-Pilot*Post	0.00127	0.0243	0.108	-0.469	5.534
	[0.0163]	[0.0473]	[0.131]	[1.325]	[10.75]
Boys-Pilot*Post	-0.0113	0.0287	-0.0254	1.413+	-12.36
	[0.0132]	[0.0279]	[0.177]	[0.788]	[14.53]
Years	2006-2012		2008 & 2013		2006-2012
Observations	Males: 150,296 Females: 156,149	240	240	240	1,060

Robust standard errors in brackets

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

Models 2 through 5 are simple difference-in-difference regression models that include school fixed effects, year fixed effects, and an indicator for being a pilot school post-transition Girls-Pilot*Post and Boys-Pilot*Post. Models 1,2 and 3 are based on a survey on teachers administered in 2008 and 2013. Column 4 is based on the administrative SEA data for years 2006 through 2012. Cohort Size is the admitted cohorts size.

b. Predicted scores are fitted values from a regression of math scores on the number of SEA attempts (repeater status in 5th grade), the student's religion, indicators for the primary school, selectivity of the student's first second third and fourth secondary school choices, month of birth (to measure quarter of birth effects), and age at SEA. Results using the predicted average of all academic subjects are similar.

This 2SLS model is as in equation 2. However, student choices are not included as a covariate.

Table 8:

Student Survey Results

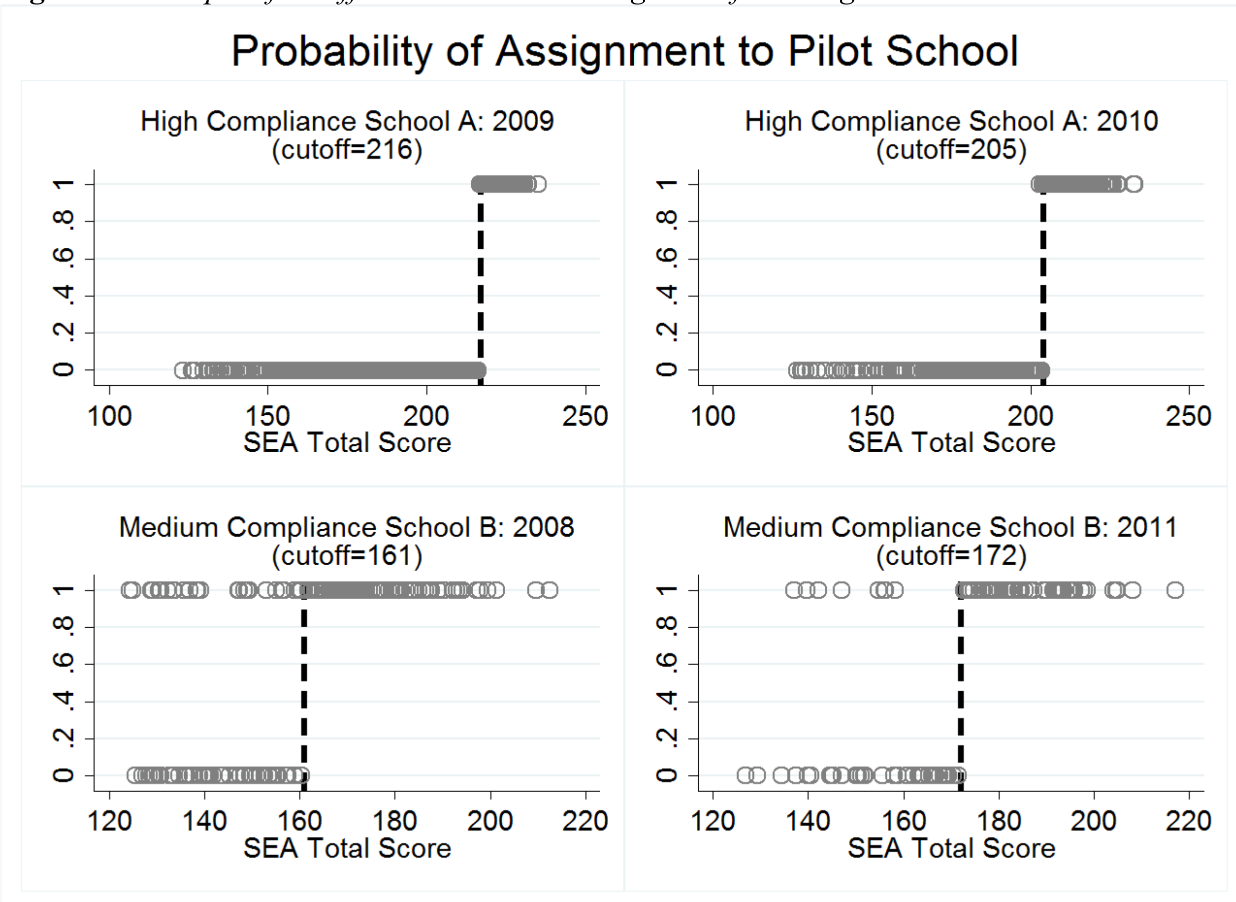
	1	2	3	4	5	6	7	8	9	10
	Peers are disruptive	Peers are distracting	Students learn from peers	Peers make students anxious	Teachers give individual attention	Teachers use examples	Teachers track student understanding	Teacher is warm toward students	Teachers involve students	Teacher is strict
Boys-Pilot*Post	0.0357	0.0433	-0.116*	-0.107*	0.114*	0.0297	0.0115	0.128*	0.0327	0.0325
	[0.0625]	[0.0547]	[0.0504]	[0.0440]	[0.0512]	[0.0438]	[0.0432]	[0.0526]	[0.0516]	[0.0445]
Girls-Pilot*Post	-0.194**	-0.130*	0.168**	0.0272	0.0989*	0.00381	0.0288	0.0494	0.0646+	0.0264
	[0.0593]	[0.0527]	[0.0464]	[0.0392]	[0.0462]	[0.0389]	[0.0427]	[0.0434]	[0.0386]	[0.0484]
Observations	25,250	25,250	27,948	26,596	27,845	26,538	27,239	26,378	27,554	27,991

Robust standard errors in brackets are adjusted for clustering at the school-grade level.

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

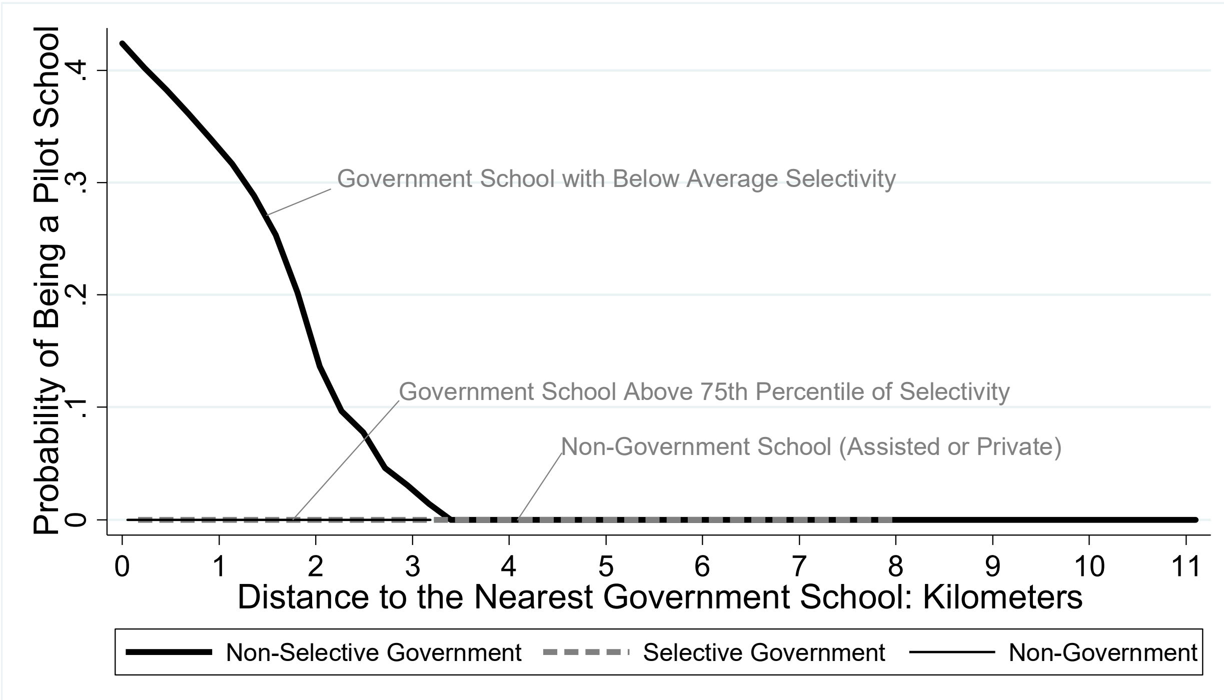
Notes: Each column represents a separate regression. The second row indicates the dependent variable. The sample is all students who attended a pilot school or one of the comparisons schools between 2012 and 2015 and also completed a survey. Because all models include school fixed effects, the coefficients on Boys-Pilot*Post and Girls-Pilot*Post represent the post-transition change in the effect of attending an all-boys pilot school and all-girls pilot school. All models include survey-year fixed effects and school-gender-form fixed effects. As such, all comparisons are made among student of the same gender at the same school (with the same teachers) in the same form but in different SEA cohorts.

Figure 1: *Example of cutoff based on school assignment for a single school in 2009 and 2010*



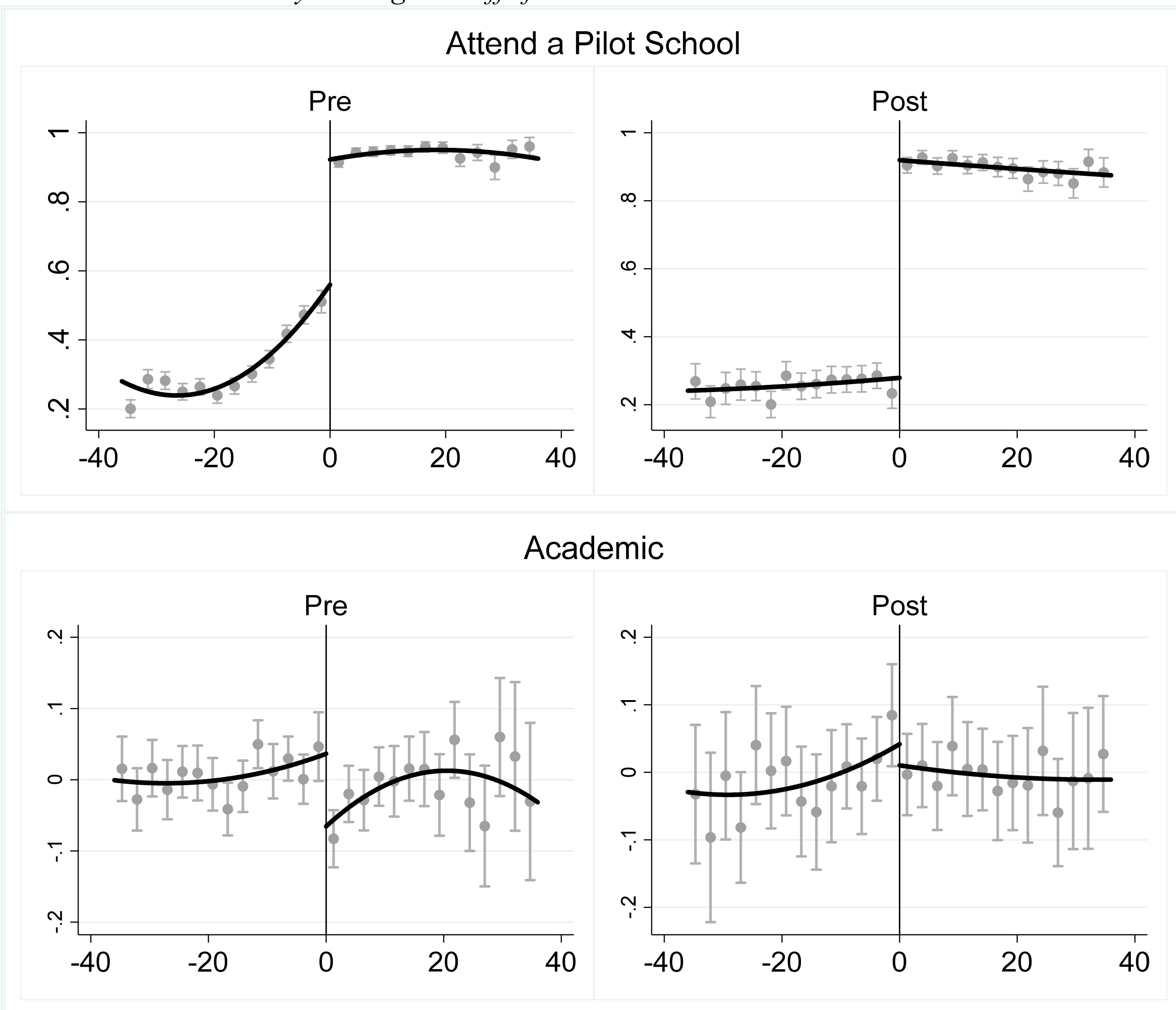
This figure depicts the probability that a student applicant to pilot school X is assigned by the MOE to that school as a function of their incoming SEA score relative to the estimated cutoff for pilot school X. The probabilities are shown for raw SEA test scores. The dashed line is the estimated cutoff for school X. The top two panels present data for a school with high compliance with the cutoffs (in a pre and post transition year) and the lower two panels present data for a low compliance school (in a pre and post transition year).

Figure 2: *Predictors of Pilot School Status*



This figure depicts the probability that each school was chosen as a pilot school as a function of the distance (as the crow flies) between that school and the nearest government school. Schools with below average selectivity have incoming SEA scores below the mean for all schools. Schools with above 75th percentile of selectivity have incoming SEA scores above the 75th percentile for all schools.

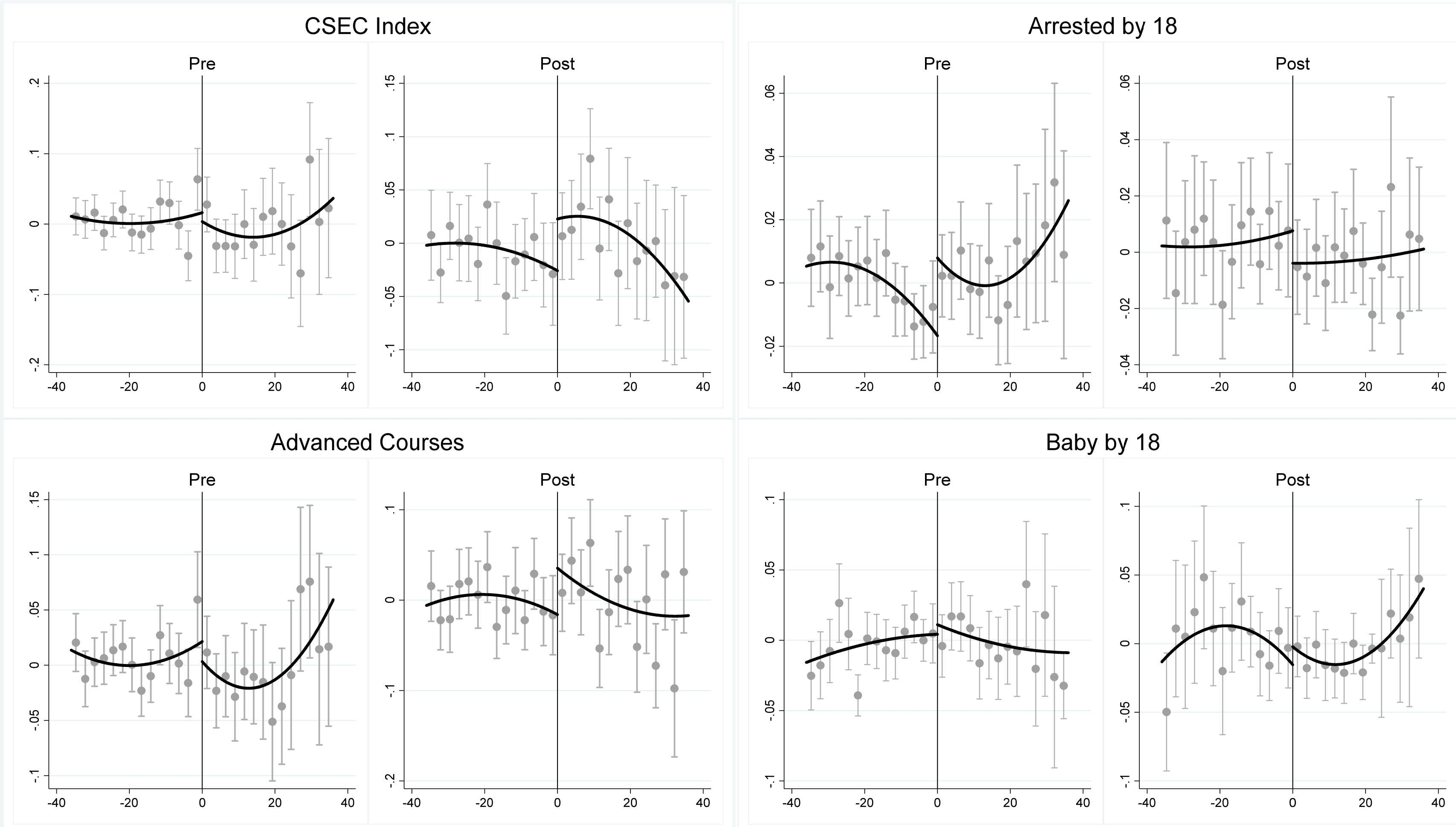
Figure 3:
Discontinuity Through Cutoffs for Pilot Schools: Pre and Post Transition



This top figure depicts the probability that a student applicant to any pilot school attends a pilot school as a function of their incoming SEA score relative to the estimated cutoff for the pilot school to which they are an applicant. The standard deviation of SEA scores is 29. The data for the cutoffs for each pilot school in each year are stacked and the applicant's tests scores are centered around the respective cutoff. Relative scores below 0 are below the cutoff and those above 0 are above the cutoff. The figures show the estimated outcome for each test score bin. The 90 percent confidence interval for the outcome in each bin is presented with the grey error bars. The solid black line on each side of the cutoff depicts a fourth order polynomial in the relative test scores (on either side of the cutoff). The lower panel provides analogous figures where the outcome is the average academic NCSE score. That is, the lower panel presents students average academic NCSE scores as a function of their incoming SEA scores relative to the estimated cutoff for the pilot school to which they are applicants.

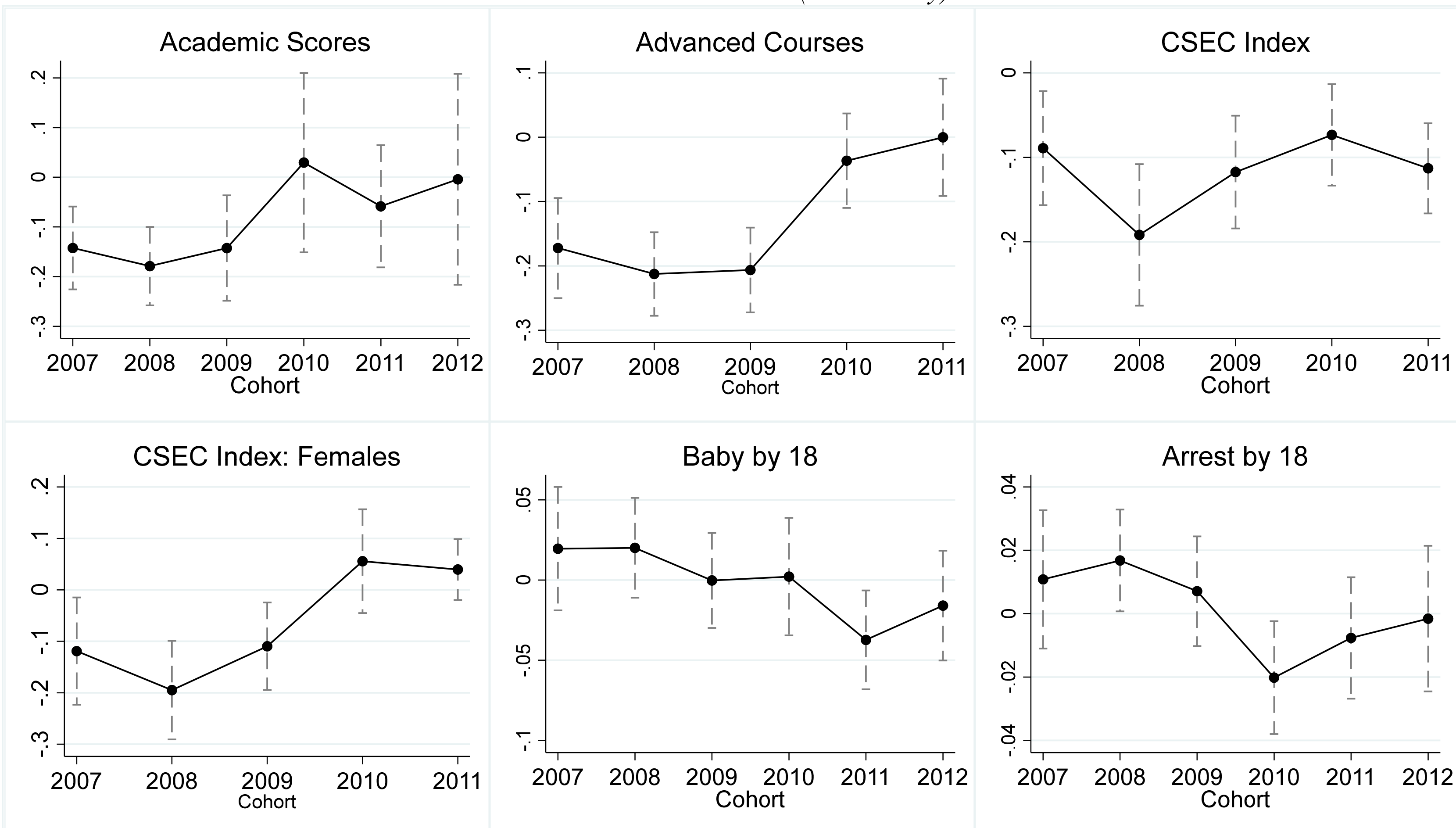
Figure 4:

Discontinuity Through Cutoffs for Pilot Schools: Pre and Post Transition



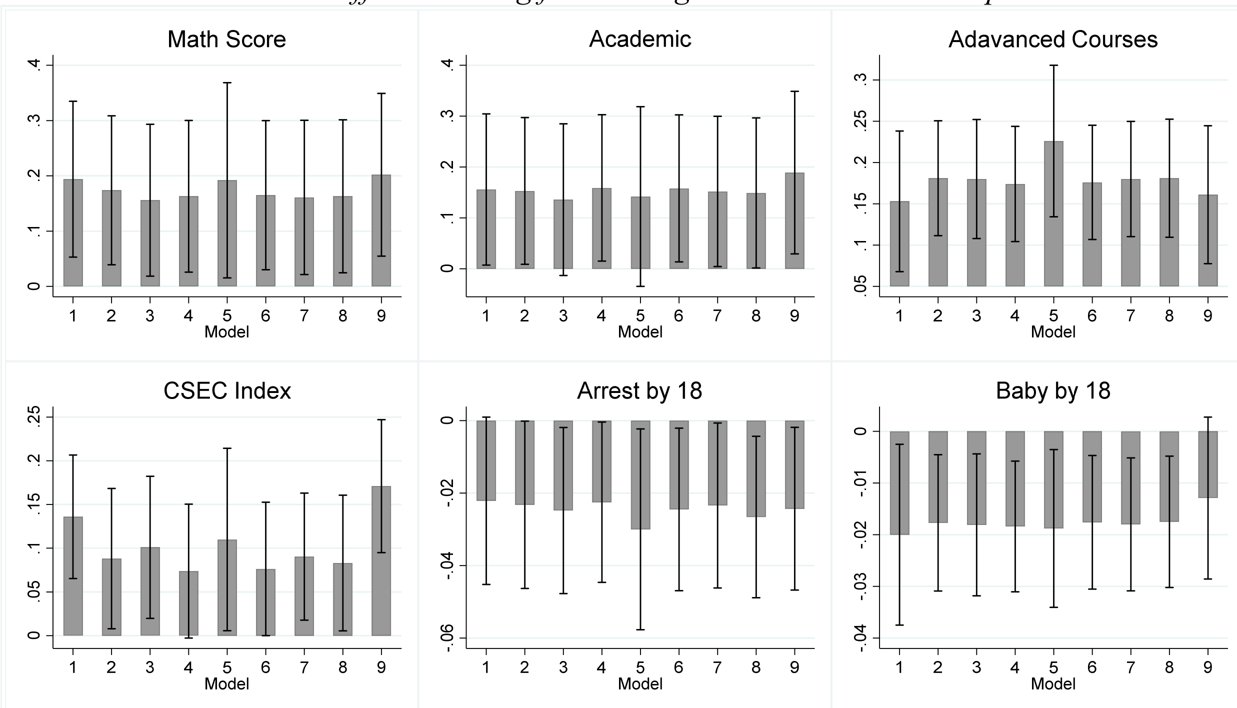
This figures present average outcomes for student applicant to pilot schools as a function of their incoming SEA score relative to the estimated cutoff for the pilot school to which they are an applicant. The standard deviation of SEA scores is 29. The data for the cutoffs for each pilot school in each year are stacked and the applicant's tests scores are centered around the respective cutoff. Relative scores below 0 are below the cutoff and those above 0 are above the cutoff. The figures show the estimated outcome for each test score bin. The 90 percent confidence interval for the outcome in each bin is presented with the grey error bars. The solid black line on each side of the cutoff depicts a fourth order polynomial in the relative test scores (on either side of the cutoff).

Figure 5:
RD Estimates Over Time (Event Study)



Notes: Each data point represents the 2SLS regression estimate of attending a pilot school three years before and after the transition. The endogenous variables are attending a pilot school in 2007, 2008, 2009, 2010, 2011, and 2012. The excluded instruments are indicator variables denoting scoring above the threshold for an all-male or an all-female pilot school in 2007, 2008, 2009, 2010, 2011, and 2012. All models include fourth order polynomial of incoming SEA test scores interacted with SEA cohort and gender, cohort fixed effects, indicators for gender, cutoff fixed effects, and choice group indicator variables. Standard errors are adjusted for clustering at the assigned school level.

Figure 6:
Transition Effect Allowing for Heterogeneous Pilot School Impacts



Notes: Each data point represents the 2SLS regression estimate of the attending a pilot school after the transition relative to pre-transition. In each model, the pilot school indicator is interacted with a pre-treatment covariate to account for possible pilot-school treatment heterogeneity. The endogenous variables are attending a pilot school pre and post transition and the interaction between pre-treatment covariates and attending a pilot school. The excluded instruments are indicator variables denoting whether a student scored above the threshold for a pilot school, whether a student scored above the threshold for a pilot school in 2010 or after, and interactions between these instruments and pre-treatment student characteristics. All models include fourth order polynomial of incoming SEA test scores interacted with SEA cohort and gender, cohort fixed effects, indicators for gender, cutoff fixed effects, and choice group indicator variables. Each model 1-9 includes interactions with pilot boy and girl school attendance and observed student characteristics. Model 1 uses incoming SEA scores. Models 2 uses the average SEA scores of the top choice, and Model 3 uses the average SEA scores of all the choices. Models 4 and 5 use whether the student's top choice school was single-sex, and the number of single-sex schools in the student's choice set, respectively. Model 6 includes interactions between attending a pilot school and the average SEA scores at the student's primary school. Model 7 uses the fraction of own gender peers at the student's primary school. Model 8 includes interactions with the average selectivity of the students next ranked school (i.e. the counterfactual school), and Model 9 includes interactions with the predicted academic scores.

FOR ONLINE PUBLICATION

Appendix A:

A Simple Model of Single-Sex Education

Much of the debate on single-sex education is focused on whether single-sex education works or does not work, and results from the empirical literature are mixed. However, there are several mechanisms through which single-sex education may affect student outcomes such that "work" versus "not work" dichotomy may be too simplistic. To help make sense of the disparate findings in the literature and to motivate the empirical work, I present a model that nests common explanations for single-sex schooling effects and allows for three separate (not mutually exclusive) mechanisms through which single-sex education can affect student outcomes. The model shows that single-sex education (relative to coed) may have different effects depending on context, and it highlights the conditions under which single-sex education effects may emerge.

Student Outcomes:

Student outcomes are given by (1) where Y_{ijg} is the outcome of student i with teacher j of gender $g \in \{girl, boy\}$, G_i is an indicator equal to 1 if student i is a girl and 0 otherwise. $\bar{G}_j$ is the share of girls in class with teacher j, the proportion of time that teacher j aligns instruction to girls is $p_j \in [0,1]$, and u_{ij} is a random mean zero error term.

(1) $$Y_{ijg} = \alpha G_i + f_g(\bar{G}_j) + h_g p_j + \psi\|2p_j - 1\| + u_{ij}.$$

There are three channels through which gender peer composition affects student outcomes. There is a direct gender peer effect and two indirect effects; the boutique effect and the focus effect. In (1), $f_g(\bar{G}_j)$ captures the **direct gender peer effect** that operates through peer interactions. Some studies find female classmates improve all students' outcomes— arguably because boys are disruptive (Hoxby 2000; Lavy and Schlosser 2011). However, others find that students have better outcomes when exposed to same gender peers— arguably because the opposite sex is distracting (Black et al 2013; Ooserbeek and van Ewijk 2014; Lu and Anderson 2015). To allow for differential responses to the proportion of female classmates, I include the subscript g on f_g.

The term, $h_g p_j$ captures the idea that that certain teaching practices may benefit girls more than boys and *vice versa.* Where p_j is the proportion of time the teacher spends "aligning" classroom practices to the needs of girls, $h_{girls} \geq 0$ and $h_{boys} \leq 0$. The idea that students may benefit from similar peers because instruction can be aligned to their type undergirds the "Boutique" model of peer effects (Hoxby and Weingarth 2006), and is posited in Duflo, Kremer and Dupas (2011) model of ability tracking. This "**Boutique effect**" captures an indirect peer effect because the teacher's choice of p_j may depend on the gender composition in the classroom.[44]

The term $\psi\|2p_j - 1\|$ captures the idea that there may be some efficiency gains for teachers associated with focusing their instruction to one group (either boys or girls). Hoxby and Weingarth (2006) call this the "focus" mode of peer interactions. Importantly, unlike the boutique effect, the focus effect is the same for all students. This focus term is motivated by the finding that students tend to have better outcomes in more homogeneous classroom environments (e.g. Hoxby

[44] This gender alignment effect could be driven by the kinds of teaching examples used, the pedagogical practices employed, the discipline practices used, or even the ability level to which the class is pitched. As an example of gendered content, sports-based examples might be more engaging for boys than for girls.

and Weingarth 2006, Ding and Lehrer 2007). The efficiency gains may be due to teachers spending less time planning lessons for two separate audiences, or teachers spending less time managing the disengagement of those students whose type is not being catered to at any given point in time. This "**focus effect**" term captures another *indirect* peer effect because the teacher's choice of alignment (p_j) may depend on the gender composition in the classroom.

Teacher's Choice of Alignment (p_j):

To understand how the *indirect* peer effects (i.e. the focus and boutique effects) operate across coed and single-sex settings, one must model how teachers chose gender alignment (p_j) as a function of the gender composition ($\bar{G}_j$). A teacher's payoff is an increasing function of the outcomes of her class $W(Y_{i\in j})$ so that $\partial W/\partial Y_i > 0\ \forall i \in j$. Teachers chose how much time to spend aligning instruction to girls (p_j) in order to maximize their individual payoff. As shown formally below, teachers will employ entirely male-aligned instruction (p_j=0) in all-male classrooms and entirely female-aligned instruction (p_j=1) in all-female classrooms. Intuitively, *if teachers prefer it when their students have better outcomes*, they will align all their instruction to girls in all-girls classroom and to boys in all-boys classrooms. The proof below also shows that a teacher's choice of alignment in mixed-gender classrooms is unclear *ex-ante*, and depends on the parameters of the production function and the incentives faced by teachers. This result implies that the behaviors of teachers in single-sex relative to coed settings are context specific.

Expected Benefits of Single-Sex Instruction:

Taking expectations of equation (1) for students in single-sex environments minus that for coed environments will yield the single-sex treatment effect β, the expected difference in outcomes in single-sex environments (*s=1*) relative to coed environments (*s=0*). Because coed classrooms are roughly half female, I assume $(\bar{G}_j|s=0)=\frac{1}{2}$. The single-sex treatment effect for girls and boys is given by (2) and (3) below, respectively.

(2) $$\beta_{boys} = f_{boy}(0) - f_{boy}\left(\frac{1}{2}\right) + h_{boy}(0) - h_{boy}(p_j|s=0) + \psi - \psi\|2(p_j|s=0) - 1\|.$$

(3) $$\beta_{girls} = f_{girl}(1) - f_{girl}\left(\frac{1}{2}\right) + h_{girl}(1) - h_{girl}(p_j|s=0) + \psi - \psi\|2(p_j|s=0) - 1\|.$$

The single-sex schooling effect for each sex depends on three factors. The first factor is the "*direct peer interactions effect*" from $f_{boy}(0) - f_{boy}\left(\frac{1}{2}\right)$ and $f_{girl}(1) - f_{girl}\left(\frac{1}{2}\right)$. If all students benefit from more female classmates then, all else equal, girls will be better off in single-sex settings and boys will be worse off. However, if both boys and girls benefit from more same-sex classmates, then both boys and girls in single-sex settings will enjoy positive direct peer effects. Importantly, the direct peer interaction effect can be positive or negative for either boys or girls.

The next two factors reflect indirect peer effects through teacher action. The second factor is the "*boutique effect*" (i.e. the effect of having greater alignment to one's own type) from $h_{boy}(0) - h_{boy}(p_j|s=0)$ and $h_{girl}(1) - h_{girl}(p_j|s=0)$. The boutique effect is nonnegative, but the magnitude depends on alignment in coed settings. That is, the benefits of all-boys classes to boys is larger if coed classes are more girl aligned and *vice versa*. This is important, because the choice of alignment in coed settings depends on teachers' incentives, which may vary from context to context. If teachers split their attention in coed settings (i.e. $0 < (p_j|s=0) < 1$), both boys and girls in single-sex settings may benefit from the boutique effect. The third factor is the "*focus effect*" (i.e. the positive effect of having a teacher focus her instruction to only one type) and is summarized with $\psi - \psi\|2(p_j|s=0)-1\|$. The more teachers split their time between aligning instruction to both girls and boys in coed settings, the greater is the benefit to single-sex schooling for *both* boys and girls. However, if teachers are already fully aligned to any one group in coed settings, then there would be no additional focus effect in the single-sex settings for that group.

A key implication of the model is that under rational behavior, the indirect effects of single-sex schooling are non-negative. However, the direct peer effects can lead to negative single-sex schooling effects. As such, single-sex schooling is neither always good nor always bad and depends on the mechanisms at play in the specific context. This is important for thinking about how single-sex schools may differ from single-sex classrooms within coed schools. If teachers have a greater incentive to align instruction to one sex in single-sex schools than in single-sex classrooms within coed schools, one may see larger benefits to single-sex schools than single-sex classrooms. In addition, if the direct gender peer interactions in the classroom are affected by the gender composition of the school, single-sex classrooms may have different effects from that of single-sex schools. The model does not predict what one may observe in any one situation. However, it does organize thinking around the effects, and shows that the single-sex effect depends on three key factors; (a) the size and direction of the direct gender peer effects (b) the change in alignment of instruction between coed and single-sex settings, and (c) the size of the teacher "focus effect". After presenting the effects of single-sex schooling in the Trinidad and Tobago context, Section IV.G employs survey data to present suggestive evidence on mechanisms.

Proofs for Claims

Student outcomes are given by (1) where Y_{ijg} is the outcome of student i with teacher j of gender $g \in \{girl, boy\}$, G_i is an indicator equal to 1 if student i is a girl and 0 otherwise. $\bar{G}_j$ is the share of girls in class with teacher j, the proportion of time that teacher j aligns instruction to girls is $p_j \in [0,1]$, and u_{ij} is a random mean zero error term.

(A1) $$Y_{ijg} = \alpha G_i + f_g(\bar{G}_j) + h_g p_j + \psi\|2p_j - 1\| + u_{ij}.$$

The term $f_g(\bar{G}_j)$ captures the direct gender peer effect that operates through peer interactions. To allow for differential responses to the proportion of female classmates by males and females, I include the subscript g on f_g. The term, $h_g p_j$ captures the idea that that certain teaching practices may benefit girls more than boys and *vice versa.* Where p_j is the proportion of time the teacher spends "aligning" classroom practices to the needs of girls, $h_{girls} \geq 0$ and $h_{boys} \leq 0$. The term $\psi\|2p_j - 1\|$ captures the idea that there may be some efficiency gains for teachers associated with focusing their instruction to one group (either boys or girls).

A teacher's payoff is an increasing function of the outcomes of her class $W(Y_{i\in j})$ so that $\partial W/\partial Y_i > 0 \ \forall i \in j$. Teachers chose how much time to spend aligning instruction to girls (p_j) in order to maximize their individual payoff.

<u>Proposition 1</u>: *Teachers will employ entirely male-aligned instruction in all-male classrooms and entirely female-aligned instruction in all-female classrooms*

<u>Proof:</u> If the classroom is all-boys, then the for all $i \in j$, $h_g p_j < 0$, so that $(Y_{ij}|p_j = 0) > (Y_{ij}|p_j \neq 0)$. By assumption, $\partial W/\partial(Y_i) > 0 \ for\ all\ i \in j$, so that $W(Y_{i\in j}|p_j = 0) >$

$W(Y_{i\in j}|p_j \neq 0)$. As such, teachers in all-boys classes will chose to align instruction to boys. Similarly, if the classroom is all-girls, then the for all $i \in j$, $h_g p_j > 0$, so that $(Y_{ij}|p_j = 0) < (Y_{ij}|p_j \neq 0)$. By assumption, $\partial W/\partial(Y_i) > 0\ for\ all\ i \in j$, so that $W(Y_{i\in j}|p_j = 0) < W(Y_{i\in j}|p_j \neq 0)$. As such, teachers in all-girls classes will chose to align instruction to girls.

<u>Proposition 2</u>: *In mixed-gender classrooms, some teachers may align instruction to girls all the time, others may align instruction to boys all the time, and some teachers may align instruction to boys and girls some of the time.*

The choice of how to align instruction will depend on the specific parameters of the production function and the incentives faced by teachers and cannot be determined *ex-ante*. I outline three illustrative cases. In all cases, I make the assumption that teacher utility is a weighted average of the utility for each student in her class, so that $W(Y_{i\in j}) = \sum_{i\in j} l(y_i)$.

Case 1a: If the costs of adopting a mixed strategy are sufficiently large (i.e. ψ is very large), and teachers payoffs are convex in individual student test scores, then teachers will choose to align instruction only to girls. Intuitively, with very large costs to adopting a mixed strategy, teachers will align instruction to only girls or only boys even in mixed-gender classrooms. If teacher payoffs are convex in individual student scores, then the marginal increase in utility will be higher for increasing tests scores at the top of the distribution than at the bottom. Because girls are more highly represented at the top of the distribution, teachers will choose only girl aligned instruction in coed classrooms.

Case 1b: If the costs of adopting a mixed strategy are sufficiently large (i.e. ψ is very large), and teacher payoffs are concave in individual student test scores, then teachers will choose to align instruction only to boys. Intuitively, with very large costs to adopting a mixed strategy, teachers will align instruction to only girls or only boys even in mixed-gender classrooms. If teacher payoffs are concave in individual student scores, then the marginal increase in utility will be higher for increasing tests scores at the bottom of the distribution. Because boys are more highly represented at the bottom of the distribution, teachers will choose only boy aligned instruction in coed classrooms.

<u>Proof:</u> Assume that the distribution of female incoming achievement has the same shape as that of males, but is a right shift of that for males by some positive constant *a* -- this is consistent with my data. If the teacher aligns instruction to males/females, then the male/female latent outcome distribution is shifted to the right by some constant δ (i.e. $h_g p_j$=$-(h_b p_j) = \delta$). For each percentile, p, of the male distribution with latent outcome $y_{p,male}^{latent}$, there is a female with latent outcome $y_{p,male}^{latent} + a$. If the teacher aligns instruction to males, then the increase in payoff for the teacher is $W\left(y_{p,male}^{latent} + \delta,\ \cdot\right) - W(y_{p,male}^{latent},\ \cdot)$ for that student. If the teacher aligns instruction to females, then the increase in payoff for the teacher is $W\left(y_{p,male}^{latent} + a +\ \delta,\ \cdot\right) - W(y_{p,male}^{latent} + a,\ \cdot)$ for the analogous female student. If teacher payoffs are convex, then $W\left(y_{p,male}^{latent} + a +\ \delta,\ \cdot\right) - W\left(y_{p,male}^{latent} + a,\ \cdot\right) > W\left(y_{p,male}^{latent} + \delta,\ \cdot\right) - W(y_{p,male}^{latent},\ \cdot)$ for all percentiles of the male distribution so that the teacher's payoff is higher if she aligns instruction to females. Conversely,

if teacher payoffs are concave, then this inequality is reversed for all percentiles of the male distribution so that the teacher's payoff is higher if she aligns instruction to males.

Case 3: If ψ is small, teachers may adopt some mixed approach. In the extreme case where there is no cost to adopting a mixed strategy, teacher payoffs are linear in the average for the classroom, there are equal number of boys and girls in the classroom, and boys and girls are equally responsive to alignment, teachers will be indifferent between adopting a mixed strategy or aligning instruction to only one sex. In such cases, a mixed strategy may be adopted in coeducation classrooms.

<u>Proof:</u> With linear payoffs, equal numbers of boys and girls, and equal responsiveness to alignment for boys and girls, the teachers expected payoff function under female alignment minus her payoff under male alignment is $\tau[h_g(p) - h_b(p)]$, where τ is some scalar. If the average marginal effect of alignment is the same for both male and female students, then $h_g(p) = h_b(p)\ \forall\ p$, so that $\tau\left[h_g(1) - h_b(1)\right] = 0$. In such a scenario, teachers are indifferent between aligning instruction to boys only, girls only, or adopting some mixed strategy. It is easy to come up with other scenarios in which teacher will chose some mixed strategy.

Appendix B:

Single-sex Status by Year and Grade in Pilot Schools

Appendix Table B1:

Status of Grades by Academic Year and Grade

	Calendar Year Admitted Form 1 (6^{th} grade) class						
	2009	2010	2011	2012	2013	2014	2015
Grade 6	Coed	**Single-Sex**	**Single-Sex**	**Single-Sex**	**Single-Sex**	Coed	Coed
Grade 7	Coed	Coed	**Single-Sex**	**Single-Sex**	**Single-Sex**	**Single-Sex**	Coed
Grade 8	Coed	Coed	Coed	**Single-Sex**	**Single-Sex**	**Single-Sex**	**Single-Sex**
Grade 9	Coed	Coed	Coed	Coed	**Single-Sex**	**Single-Sex**	**Single-Sex**
Grade 10	Coed	Coed	Coed	Coed	Coed	**Single-Sex**	**Single-Sex**

Note that the single-sex pilot program was started for the 2010 SEA cohort and was abandoned for the 2014 SEA cohort.

Appendix C:
The School Assignment Algorithm

School slots are assigned in rounds such that the most highly subscribed/ranked school fills its spots in the first round, then the next highly subscribed school fills its slots in the second round, and so on until all school slots are filled. This is done as follows: (1) the number of school slots at each school n_j is predetermined based on capacity constraints. (2) Each student is tentatively placed in the applicant pool for her first choice school and is ranked by SEA score. (3) The school at which the n_j^{th} ranked student has the highest SEA score is determined to be the most highly subscribed/ranked school and the top n_{j1} students in the applicant pool for top-ranked school j_1 are admitted to school j_1. The SEA score of the n_{j1}-th student is the cutoff score for school j_1. (4) The top-ranked school slots and the admitted students are removed from the process, and the second choice becomes the new "first choice" for students who had the top-ranked school as their first choice but did not gain admission. (5) This process is repeated in round two to assign students to the second highest ranked school j_2 and determine the cutoff score for the second-ranked school, and this is repeated in subsequent rounds until all slots are filled. This assignment mechanism is a deferred acceptance algorithm (Gale and Shapley, 1962) in which students have incentives to truthfully reveal their rankings among chosen schools.

While the optimal set of school choices is difficult to solve, Chade, and Smith (2006) demonstrate that the choice set should include the school with the largest expected payoff (utility conditional on attendance times the likelihood of admission), students should rank selected schools in order of actual preferences, and should include a “reach” school for which admission is unlikely but the utility conditional on attendance is high.

This process is used to assign over 90% of all students. As such, as a practical matter, one can consider this applying to all students. However, there are a few exceptions to this rule. First, Government Assisted schools (not analyzed in this study) are allowed to admit 20% of their incoming class at the principal’s discretion. None of the pilot school is Government assisted so that there is no problem of principals hand picking students at the pilot schools. However, there are also assignments that do not follow this rule because students who do not score high enough to be assigned to a school on their choice list receive an administrative assignment from the Ministry of Education (these assignments are made to balance space considerations). Finally, due to unforeseen circumstances some schools may have less capacity than expected or may close (this may happen due to flooding etc.). In such rare cases, the Ministry will assign students to schools based on open slots in nearby schools, open slots in other schools in the choice list, and proximity.

I aim to use this assignment rule to isolate exogenous variation to the 20 pilot schools. A key feature of this assignment rule is that each school has a test score cutoff above which applicants are very likely to be assigned and below which applicants are very unlikely to be assigned. Even though the cutoffs are not known to the public and not all the administrative assignments follow the cutoff rule (due to a few exceptions *made by the MOE*), because I have access to the administrative assignment (which is outside the control of the students or their parents), the cutoffs can be recovered empirically for the 20 pilot schools.

Appendix D:
Robustness Checks and Test of Validity

Validity of the RD Variation

The exogenous variation used in this paper is driven by the test-score cutoffs. Even though there is no way to prove for certain that the cutoff variation is valid, here I present evidence that this identification strategy is valid. One key diagnostic is to test for excess density above the cutoff and less than expected density below the cutoff (McCrary 2008). The first *prima facie* evidence of no change in density through the cutoff is simply the histogram of relative scores. As one can see in Figure D1, there is no uncharacteristic spike in density above the cutoff or dip in density just below the cutoff. If one computes the density of observations at each relative score and regresses this on scoring above the cutoff along with smooth functions of the relative score, there is no statistically significant relationship between scoring above the cutoff and the density. The point estimate is -0.0006 (p-value=0.53) – negative and not statistically significant. Taken together, the patterns suggest that the variation due to the test score cutoffs is likely valid. The other common test is for smoothness of latent outcomes (as proxied by covariates) through the cutoffs. In presenting the results I present effects on both actual outcomes and predicted outcomes (based on covariates) that show that the cutoff variation in pilot school attendance is not associated with any changes in predicted outcomes – consistent with the cutoff variation being valid.

Figure D1 shows that there is little evidence of gaming around the cutoffs regarding the density of observation at each test score. However, the validity of the design also requires that there be no sorting of students around the cutoff (i.e. that latent outcomes are smooth through the cutoff). Given that students are unaware of the location of the cutoffs and are forced to make school choices before they take the SEA examinations, it is very unlikely that there is any sorting around the test score cutoffs. However, to provide further evidence that the variation employed (due to the cutoffs) is valid, I create a predicted academic score variable and test for whether the 2SLS model (using the cutoff variation) predicts any change in predicted math scores. Specifically, I regress the NCES academic on the number of SEA attempts (repeater status in 5th grade), the student's religion, indicators for the primary school, selectivity of the student's first second third and fourth secondary school choices, month of birth (to measure quarter of birth effects), and age at SEA. These variables are very strong predictors of math scores such that they yield an R-squared of 0.42. I then take the fitted value from this regression as my predicted academic score. If there is some gaming of the cutoff, one would likely see that attending a pilot school as a result of scoring above the cutoff should be associated with better "predicted" scores. However, with no gaming there should be no relationship between scoring above the cutoff and one's predicted score. Consistent with no gaming, there is no relationship between scoring above the cutoff and one's predicted academic score (note that this model does not include control for choices). Column 1 of Table 7 shows the estimated effect on predicted scores for the all-boys pilot schools and the all-girls pilot schools. The coefficients on the variables of interest are both small and statistically insignificant – indicating no gaming of the cutoff to the all-boys pilot schools. While the impact on predicted outcomes is small, it is helpful to see the estimated impact on the individual covariates. This is presented in Table D1. Of the 41 covariates tested, 4 are significant at the 10 percent level – this is consistent with random chance.

To assuage any lingering concerns that the estimated impacts are driven by the choice of bandwidth, Appendix Figure D2 shows the estimated impacts for the main outcomes in the preferred model for different bandwidth windows around the cutoff. With the sole exception of having a baby (which is unstable at very narrow bandwidths), the estimated impacts are generally invariant to the bandwidth.

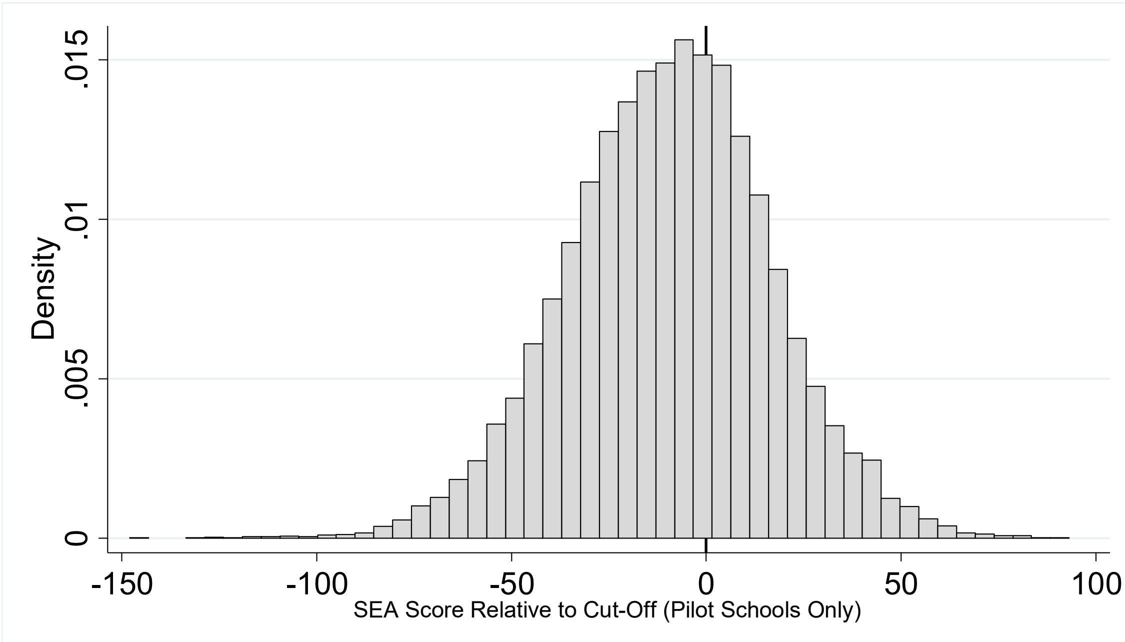
Another possible concern is that the documented treatment effects could have happened by random chance by designating any 20 schools as pilot schools. To assess this, I estimated each school's likelihood of being a pilot school based on achievement level, school type, and proximity to other schools and identified 18 non-pilot schools with similar likelihoods as the actual pilot schools. This were all schools that had an estimated propensity score of greater than 0.34 (as did the actual pilot schools). These schools could have been chosen for the pilot and form a natural placebo test. I estimate the change in the effect of attending a placebo school before and after 2010 (see Appendix Table D2). For academic scores, advanced courses, arrests and teen birth, the placebo estimates are small and not statistically significant. The placebo transition effect is statistically significant for the CSEC index, but it has the opposite sign as the pilot transition effect. This suggests that the estimated positive effects are real and are unlikely to have arisen by the choice of pilot schools.

Validity of DiD Variation

For the DiRD estimates to represent the causal effect of single-sex education requires that there were no changes in the pilot schools over time that coincided with the pilot program. Confounding changes within schools is unlikely because, (a) the individual schools had no control over when they would become pilot schools, (b) the government stipulated that no other changes take place in these schools, and (c) schools were not made aware of the changes until the summer preceding the change so that schools had no time to react before the start of the school year. Even so, one may worry that the pilot schools were already on an upward trajectory prior to transition. To show evidence that this does not drive the results, Figure 5 presents the RD estimates for attending a pilot school for each year three years before and three years after the transition on the main outcomes. Each data point is an RD point estimate, and the 90 percent confidence interval is presented for each RD estimate. For all outcomes there is no indication of any differential pre-trending and all the improvements occur after the transition. Figure 5 presents a clear visual representation of the DiD variation employed and provides compelling evidence that the DiD identifying common trends assumption is valid.

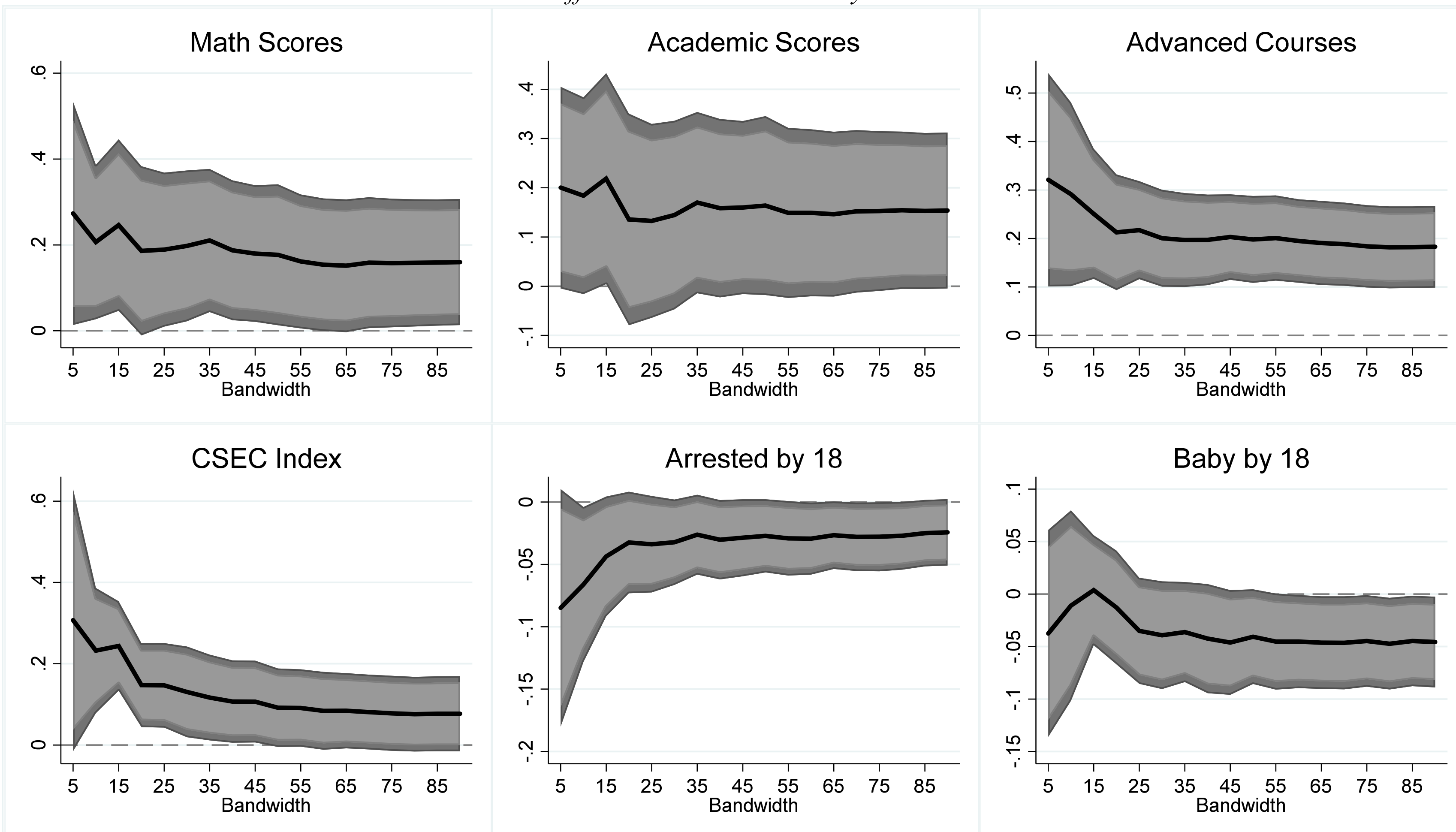
Even though the MOE clearly stipulated that there be no changes, as a final check on the DiD assumptions of no other changes, I used survey data collected by the MOE on a sample of teachers at all schools obtained in 2009 and 2013 to see if there were any systematic personnel changes that occured during the transition. In a simple difference in difference regression model predicting school level teacher characteristics with school fixed effects and year fixed effects (Table 5), between 2009 and 2013 the transition schools saw no differential change in the percentage of female teachers, or the percentage of teachers with a Bachelor's degree. There is a marginally statistically significant effect on year-of-birth at the all-male pilot schools. This is consistent with sampling variability and is in the opposite direction of what would be required to generate a positive transition effect for the all-male pilot schools. I also explore if differences in class size (proxied by cohort size) can explain the results. In a simple model predicting *initial* cohort size (i.e. the size of the assigned cohort) with school fixed effects and year fixed effects, the post-transition pilots have no statistically significant differences in cohort size (and the effect for the all-male and all-female pilot schools are in opposite directions). The results corroborate the notion that there were no other changes at pilot schools.

Appendix Figure D1:
Distribution of Incoming SEA Scores around the cutoff



For each relative SEA score for each cutoff, the density of observations is computed. This figure represents the density of incoming SEA test scores relative to the test score of the applicant student. The cutoff for the school to which the student has applied is 0.

Appendix Figure D2:
DiRD Effect on Main Outcomes by Bandwidth



This graph depicts the estimated 2SLS regression estimated transition effect on various key outcomes in a model with cutoff fixed effects, cohort fixed effect, choice fides effects, and a fourth order polynomial in test scores. The estimates include both boys and girls (with the exception of having a baby) The 2SLS models are estimated among observations within a particular bandwidth of the cutoffs. The estimated 2SLS single-sex transition effects are reported for each bandwidth between 15 (0.5sd) and 90 (3sd). The 90 percent confidence interval for the estimate is presented in light grey, and the 95% confidence interval is in dark gray.

Appendix Table D1:

Transition Impacts on Covariates

Covariate	Pilot*Post	SE
Religion 1	0.000368	[0.000379]
Religion 2	-0.0239	[0.0191]
Religion 3	0.0126	[0.0178]
Religion 4	0.00883	[0.0263]
Religion 5	-0.00144	[0.00314]
Religion 6	0.023	[0.0150]
Religion 7	-0.00102	[0.00133]
Religion 8	0.0227	[0.0209]
Religion 9	0.00924	[0.00625]
Religion 10	0.00479	[0.00887]
Religion 11	-0.0116*	[0.00594]
Religion 12	-0.0217	[0.0172]
Religion 13	0.000425	[0.00127]
Religion 14	0.00211	[0.00135]
Religion 15	-0.0144*	[0.00786]
SEA at age 10	-0.000259	[0.000237]
SEA at age 11	0.0102+	[0.00571]
SEA at age 12	0.0073	[0.0193]
SEA at age 13	-0.00692	[0.0200]
SEA at age 14	-0.00759	[0.0237]
SEA at age 15	-0.00376	[0.0145]

Covariate	Pilot*Post	SE
Born January	0.00528	[0.0142]
Born February	0.00429	[0.0103]
Born March	-0.0196*	[0.0110]
Born April	-0.0048	[0.0128]
Born May	-0.00025	[0.0127]
Born June	0.00201	[0.00955]
Born July	0.0192	[0.0137]
Born August	0.00186	[0.0131]
Born September	-0.0158	[0.0139]
Born October	-0.0031	[0.0149]
Born November	-0.0133	[0.0146]
Born December	0.0242+	[0.0146]
Times Taken SEA	-0.00417	[0.0151]
Mean Sea Choice 1	-0.395	[0.860]
Mean Sea Choice 2	1.081	[0.712]
Mean Sea Choice 3	-0.575	[0.845]
Mean Sea Choice 4	1.367	[1.459]
Mean SEA Primary School	0.876	[0.690]
Fraction Male at Primary School	0.00252	[0.00876]

Robust standard errors in brackets

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

Notes: Each point estimate is from a separate regression. The reported coefficients are from the 2SLS model outlined in [2] but excluding the choice fixed effects. All models include fourth order polynomial of incoming SEA test scores interacted with SEA cohort and gender, cohort fixed effects, indicators for gender, and cutoff fixed effects.

Appendix Table D2

Transition Effect with Placebo Cutoffs

	Math Scores	Academic Scores	Advanced Courses	CSEC Index	Arrest by 18	Baby by 18
	Boys and Girls Combined					
	1	2	3	4	5	
Pilot*POST	0.158*	0.181*	0.184**	0.0791+	-0.0312*	
	[0.0741]	[0.0825]	[0.0429]	[0.0471]	[0.0158]	
Placebo*POST	-0.0784	0.0444	-0.0319	-0.187**	-0.0222	
	[0.101]	[0.0823]	[0.0436]	[0.0533]	[0.0205]	
	Boys Only					
	6	7	8	9	10	
Boys-Pilot*POST	0.271**	0.231*	0.180**	0.00214	-0.0596**	
	[0.102]	[0.111]	[0.0395]	[0.0633]	[0.0212]	
Placebo*POST	-0.0772	0.0587	-0.082	-0.228**	-0.0256	
	[0.112]	[0.0819]	[0.0512]	[0.0782]	[0.0376]	
	Girls Only					
	11	12	13	14	15	16
Girls-Pilot*POST	0.0785	-0.0275	0.121	0.201**	0.0116	-0.0454*
	[0.0947]	[0.109]	[0.0739]	[0.0539]	[0.00984]	[0.0219]
Placebo*POST	-0.047	0.0747	0.0319	-0.132*	-0.0178	0.00537
	[0.138]	[0.132]	[0.0664]	[0.0594]	[0.0171]	[0.0266]
Choice fixed effects	Y	Y	Y	Y	Y	Y
App Group Fixed Effects	Y	Y	Y	Y	Y	Y
SEA Scores	Y	Y	Y	Y	Y	Y

Robust standard errors in brackets adjusted for clustering at the assigned school level.

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

Notes: The dependent variable is indicated in the top row. Each row represents a separate regression. All models include the effect of attending a pilot school so that the coefficients on Boys-Pilot*POST and Girls-Pilot*POST and Pilot*Post represent the post-transition change in the effect of attending an all-boys pilot school, an all-girls pilot school, and any Pilot School, respectively. All models include fourth order polynomial of incoming SEA test scores interacted with SEA cohort and gender, cohort fixed effects, indicators for gender, cutoff fixed effects, and choice group indicator variables. The placebo schools are the 18 schools with very similar likelihoods of having been a pilot school based on school type, school achievement level, an proximity to another school

Appendix E:

Evidence of Changes in Applicant Pool and Compliers

After the transition from coed to single-sex, the applicant pool to the pilot schools changed. To gain a sense of this, I regressed listing a pilot school as a first choice, second choice, third choice, and fourth choice on a simple Post indicator denoting whether the observation is for an SEA cohort in 2010 or after. The constant term and the coefficient on Post are reported in Table E1. The constant term represents the proportion of the sample that lists a pilot school in their choices during the pre-transition period, and the coefficient on Post is the change in proportion after the transition.

The constant terms in Models 1 through 4 show that the pilot boy schools were typically listed as students third of fourth choices. About 8.5 percent have it as a third choice and 8.5 percent list an all-boys pilot as the fourth choice, while the comparable numbers for the first and second choice are about 2.7 and 5.5 percent, respectively. After the transition, boys were 0.4 percentage points more likely to list an all-boys pilot as their second choice, 2.54 percentage points more likely to list an all-boys pilot as their third choice, and to 6.36 percentage points more likely to list an all-boys pilot as their fourth choice. Some of this increased demand would have come from boys who would have listed the all-girls pilot switching over to the all-boys pilot, so that this may not be indicative of increased demand for all-boys schooling. To assess this, the dependent variable in models 5 through 8 is listing any pilot school as a first, second third or fourth choice. If students were indifferent to the change and simply listed the all-boys school in each pair if they would have chosen the all-girls pilot school, there should be no effect on choosing any pilot school. The coefficient on the post indicator is negative in all models, suggesting that, overall boys were less likely to list pilot schools after the transition than before. The results for girls in models 9 through 16 tell a similar story. Girls are more likely to list an all-girls pilot school after the transition than before, but are less likely to list a pilot school in general after the transition. Also, the magnitudes of the changes are very similar to those for boys.

Given that there was a change in the likelihood that students would select a pilot school, it is helpful to determine the extent to which the marginal applicant differed before versus after the transition. That is, because the estimate pilot school impacts are based on the impact for the compliers, it is helpful to determine whether the characteristics of compliers were markedly different before versus after the transition. As pointed out in Abadie (2003) and later Angrist and Fernandez-Val (2014), so long as there are no defiers, for any characteristic (X=x), the relative likelihood that a complier has a characteristic X=x is given by the ratio of the first stage for those with X=x to the overall first stage. Using the population means, one can therefore uncover the proportion of the compliers with characteristic X=x. Based on this insight, I compute complier means for some key characteristics before and after the transition. For continuous variables, I put each observation into one of 10 deciles of the distribution for that variable. For each decile group, I compute the average of the continuous variable. I then use the approach from above to compute the fraction of the compliers that falls into each decile of the continuous variable. To recover an estimate of the average, I then multiply the fraction of compliers in each decile by the average for that decile and sum across all ten deciles. The complier means are reported in Appendix Table E2.

Using this approach, the complier means are quite similar before versus after the transition. The average scores of compliers were -0.189 and -0.236 standard deviations below the mean before and after the transition, respectively. This decline of 0.047 standard deviations represents a modest move from the 41st percentile to the 39st percentile of the SEA distribution. To summarize student incoming ability, I also compute the complier mean for predicted academic scores. Before versus after the transition, these are essentially unchanged (going from -0.41 to -0.40 before versus after the transition). I also compute changes in the complier means of the selectivity of the top choice school, the selectivity of all the chosen schools, the average SEA scores at the primary school, the proportion of males at the primary school, and the selectivity of the next ranked school. The only variable that changes in any appreciable way is the selectivity of the next ranked school. Before the transition, the selectivity of the net ranked school was -0.8 and after it increased to -0.72. Though this was not the same, it goes in the opposite direction of what would generate a spurious positive transition effect. In sum, a comparison of complier means suggests that any changed in the LATE before versus after the transition are likely to be small.

Table E1:
Changes in Likelihood of Choosing a Pilot School

	Boys: 60,133 observations							
	Pilot Boys School				Any Pilot School			
	1	2	3	4	5	6	7	8
	First	Second	Third	Fourth	First	Second	Third	Fourth
Post	-0.00107	0.00441*	0.0254**	0.0635**	-0.0144**	-0.0349**	-0.0411**	-0.0397**
	[0.00132]	[0.00191]	[0.00244]	[0.00269]	[0.00145]	[0.00216]	[0.00273]	[0.00307]
Constant	0.0269**	0.0544**	0.0824**	0.0851**	0.0403**	0.0941**	0.149**	0.189**
	[0.000868]	[0.00122]	[0.00147]	[0.00150]	[0.00105]	[0.00156]	[0.00191]	[0.00210]

	Girls: 59,114 observations							
	Pilot Girls School				Any Pilot School			
	9	10	11	12	13	14	15	16
	First	Second	Third	Fourth	First	Second	Third	Fourth
Post	0.00370**	0.00675**	0.0203**	0.0420**	-0.0129**	-0.0350**	-0.0468**	-0.0381**
	[0.000789]	[0.00141]	[0.00203]	[0.00259]	[0.00105]	[0.00177]	[0.00240]	[0.00291]
Constant	0.00693**	0.0257**	0.0520**	0.0844**	0.0237**	0.0677**	0.120**	0.165**
	[0.000449]	[0.000857]	[0.00120]	[0.00150]	[0.000823]	[0.00136]	[0.00175]	[0.00201]

Robust standard errors in brackets
** $p<0.01$, * $p<0.05$, + $p<0.1$
Notes: This table reports the coefficients on the constant terms and a Post 2010 indicator dummy in a model predicting the likelihood that a student lists an all-boys pilot school (models 1 through 4) and all girls pilot school (models 9 through 12) or any pilot school (models 5 through 8 and models 13 through 16) as their first choice (models 1, 5, 9, and 13), second choice (models 2, 6, 10, and 14), third choice (models 3, 7, 11, and 15) or fourth choice (models 4, 8, 12, and 16).

Table E2:

Estimated Complier Means Before vs. After the Transition

Variable	Pre Mean	Post Mean
Total SEA Score	-0.1891	-0.2364
Selectivity of Top Choice	1.0563	1.0806
Average SEA of All Choices	0.6061	0.5793
Selectivity of Next Choice	-0.8088	-0.7212
Average SEA of Primary School	-0.0056	0.0341
Average Proportion Male	0.4885	0.5092
Predicted Academic Score	-0.4136	-0.4057
Number Single Sex Choices	1.7330	2.6469
Top Choice is Single Sex	0.4683	0.5705

Appendix F:

Student Survey Specification Checks

Appendix Table F1:

Student Survey Specification Checks

	1	2	3	4
	Survey Participation Rate (Cohort Level)	My Parents think Education is Important	Most of my friends are the same gender as me	Teacher is Female
Boys-Pilot*Post	-0.0317	0.00476	0.299**	-0.0202
	[0.0356]	[0.0496]	[0.0527]	[0.018]
Girls-Pilot*Post	-0.00231	0.0117	0.243**	0.0058
	[0.0423]	[0.0363]	[0.0587]	[0.0175]
Survey Year Effects	Y	Y	Y	Y
School-Gender-Form Effects	N/A	Y	Y	Y
Observations	609	27,477	27,514	25,886

Robust standard errors in brackets

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

Each column represents a separate regression. The sample is all students who attended a pilot school or one of the comparisons schools between 2012 and 2015 and also completed a survey. Because all models include school fixed effects, the coefficients on Boys-Pilot*Post and Girls-Pilot*Post represent the post-transition change in the effect of attending an all-boys pilot school and an all-girls pilot school, respectively. All models include survey year fixed effects and school-gender-form fixed effects. As such, all comparisons are made among student of the same gender at the same school (with the same teachers) in the same form but in different SEA cohorts.

Appendix G:
Student Questions used to Construct Indexes

Appendix Table G1:
Student Questions About Peers

	Peers			
	Disruptive peers	Distracting peers	Learn from Peers	Nervous around peers
Survey Question				
My classroom is orderly	0.4749	-0.0619		
I feel safe in the classroom	0.4083	-0.0541		
My classmates distract me from my schoolwork	-0.0697	0.3345		
We do not waste time in my classes	0.3506	-0.1319		
My classmates sometimes encourage me to misbehave	-0.1112	0.3252		
I learn from my classmates			1	
I worry about impressing classmates while in class				0.4548
I am afraid to raise my hand in class				0.4748
I get nervous when I am compared with classmates				0.3498

Notes: This table presents the factor loadings for each survey question used to construct each factor. The factor is listed at the top of each column. The individual survey items are listed in the rows.

Appendix Table G2:

Student Questions about Teachers

	Teacher Related Questions					
	Spend one-on-one time	Use examples to relate topics	Check for understanding	Warmth toward students	Involves students	Teacher is strict
Survey Question						
My teachers pay attention to me in	0.4839					
My teachers spend one-on-one time	0.4839					
Teachers use many examples that help		0.5344				
Teachers ask us for several ways to		0.6048				
The teachers require us to relate		0.521				
Teachers check whether we know the			0.6148			
Teachers give assignments that help			0.621			
Teachers hold discussions that help us			0.5791			
Teachers ask us to explain our			0.4202			
My teachers like me				0.5572		
My teachers care about me				0.5777		
My teachers often make me feel bad				-0.2335		
Teachers praise my efforts				0.5149		
My teachers listen to my ideas					0.4996	
My teachers involve students in					0.4996	
My teachers are strict						1

Notes: This table presents the factor loadings for each survey question used to construct each factor. The factor is listed at the top of each column. The individual survey items are listed in the rows.

Appendix Table H1:

Main 2SLS Estimate for Different Polynomial Order of the SEA Score

Polynomial Order	4th Order	3rd Order	Quadratic	Linear
	Math Score (8th Grade)			
Pilot*Post	0.158*	0.113*	0.157*	0.150*
	[0.0741]	[0.0564]	[0.0716]	[0.0646]
	Academic Scores (8th Grade)			
Pilot*Post	0.181*	0.159*	0.156*	0.164*
	[0.0825]	[0.0795]	[0.0777]	[0.0720]
	Advanced Courses			
Pilot*Post	0.184**	0.172**	0.233**	0.125**
	[0.0429]	[0.0446]	[0.0420]	[0.0476]
	CSEC Index			
Pilot*Post	0.0791*	0.0887*	0.0916*	0.0337
	[0.0471]	[0.0440]	[0.0461]	[0.0510]
	Baby by 18 (Girls Only)			
Pilot*Post	-0.0454*	-0.0396*	-0.0463*	-0.0589*
	[0.0219]	[0.0179]	[0.0229]	[0.0244]
	Arrested by 18 (Boys Only)			
Pilot*Post	-0.0596*	-0.0330*	-0.0623**	-0.0581**
	[0.0212]	[0.0158]	[0.0216]	[0.0203]

Robust standard errors in brackets adjusted for clustering at the assigned school level.

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

Notes: Each column represents a separate regression. All models include a polynomial (of differing order) in the SEA score interacted with cohort and gender, choice fixed effects, and cutoff fixed effects. All models include the effect of attending a pilot school so that the coefficients on Pilot*POST represents the difference between the effect of attending any pilot school after transition and attending any pilot school prior to transition.

Appendix Table H2:
Donut RD Estimates

	Academic Scores	Advanced Courses	CSEC Index	Baby by 18 (Girls Only)	Arrest by 18
	Omitting scores between 0 and -10				
	1	2	3	4	5
Pilot*POST	0.169**	0.173***	0.04	-0.0603**	-0.0221*
	[0.0842]	[0.0476]	[0.0519]	[0.0283]	[0.0129]
Observations	234,284	243,248	243,248	137,918	285,400
	Omitting scores between 0 and -20				
	6	7	8	9	10
Pilot*POST	0.121*	0.203***	0.0476	-0.0654***	-0.0231*
	[0.0719]	[0.0444]	[0.0505]	[0.0249]	[0.0126]
Observations	174,994	184,989	184,989	102,154	217,167

Robust standard errors in brackets adjusted for clustering at the assigned school level.

*** p<0.01, ** p<0.05, * p<0.1

Notes: Each column represents a separate regression. All models include a fourth order polynomial in the SEA score interacted with cohort and gender, choice fixed effects, and cutoff fixed effects. All models include the effect of attending a pilot school so that the coefficients on Pilot*POST represents the difference between the effect of attending any pilot school after transition and attending any pilot school prior to transition.