Appendix For:

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

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.[1]

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

[1] 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.

Proposition 2: *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.

Proof: 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.

Proof: 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 (6th 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.

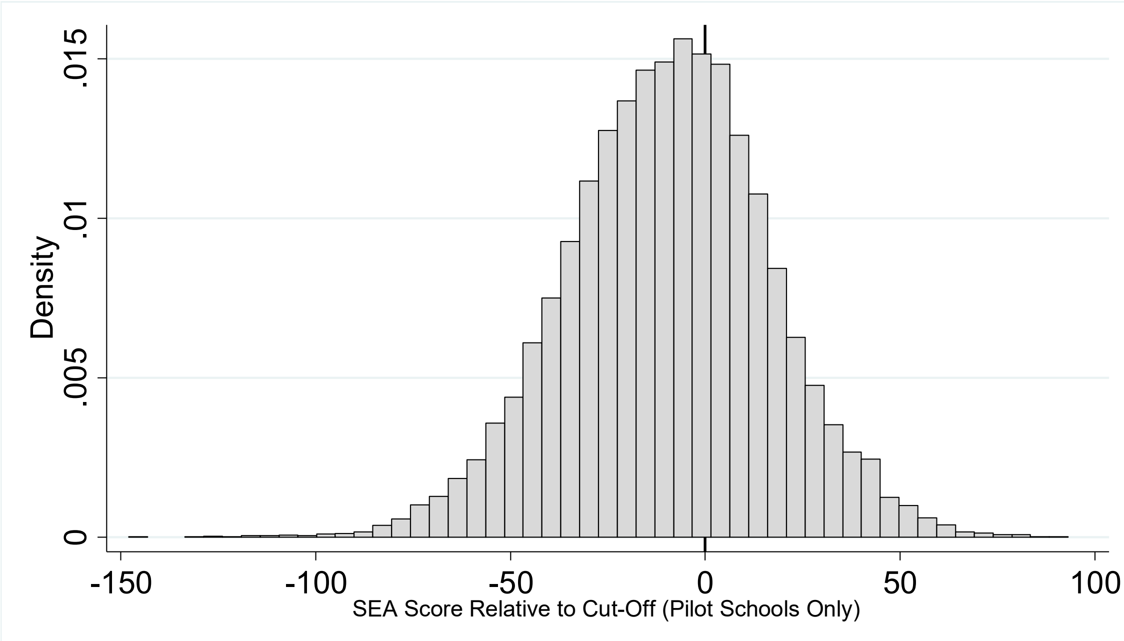
Another possible concern is that the documented treatment effects could have happened by random chance by designating any 20 schools as pilot schools. To assess this, I estimated each school's likelihood of being a pilot school based on achievement level, school type, and proximity to other schools and 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.

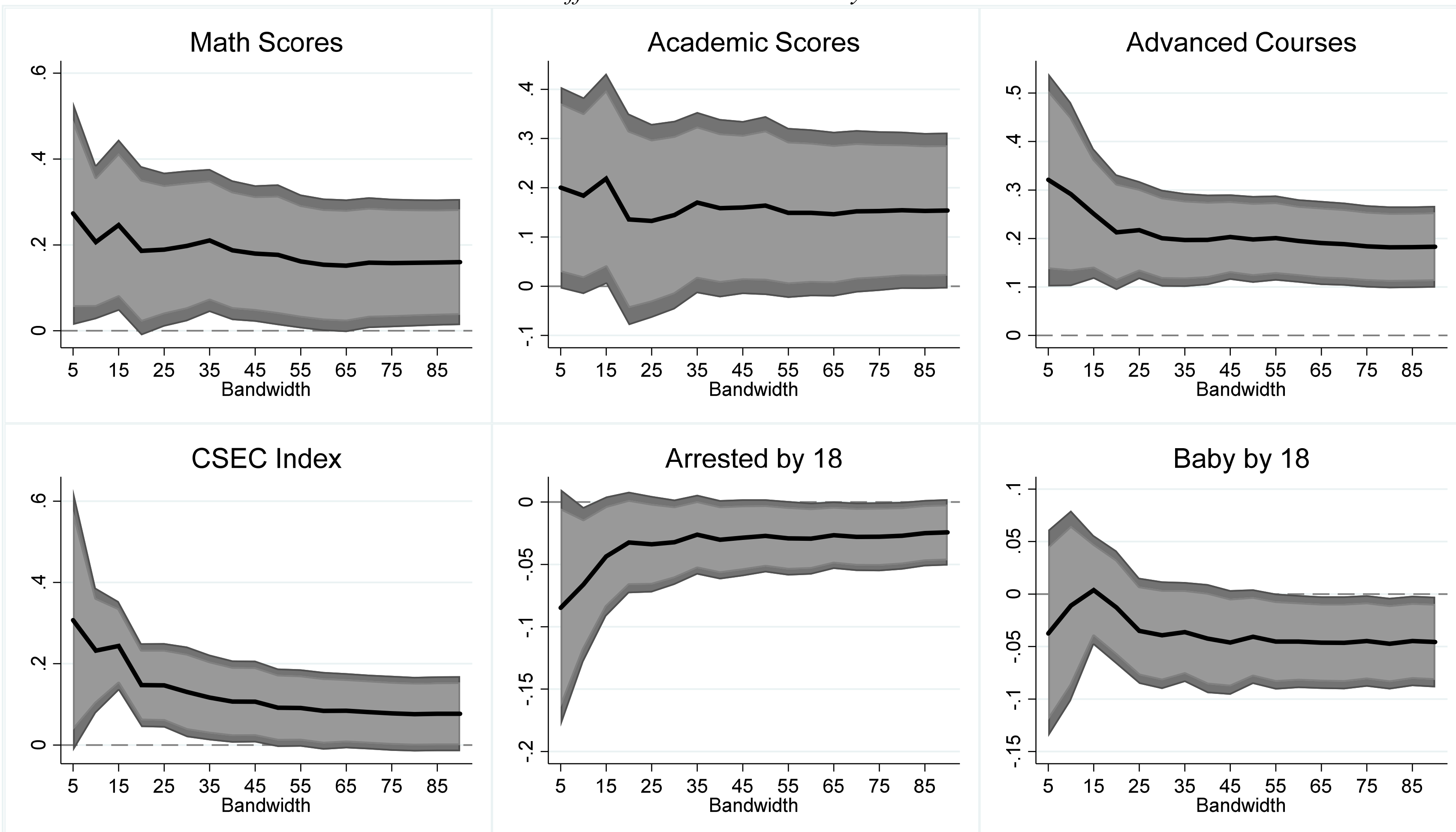
Even though the MOE clearly stipulated that there be no changes, as a final check on the DiD assumptions of no other changes, I used survey data collected by the MOE on a sample of teachers at all schools obtained in 2009 and 2013 to see if there were any systematic personnel changes that occured during 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.