Appendix A – Implementation Summary

Program Model – MSQI, ExpandED, and Tutoring

In April 2013, roughly 150 low-performing, high-poverty New York City Department of Education (NYC DOE) middle schools were invited to apply for a chance to participate in a district literacy intervention through the Middle School Quality Initiative (MSQI). Of the invited schools, 129 schools responded that they were interested in participating. To control costs and maximize the probability of success for the program, experimental schools were selected for the randomization based upon school size and subject to a minimum school environment grade on the NYC DOE school survey. The sixty smallest interested schools with school environment grades of D or higher formed the experimental group. From this group of sixty, twenty schools were randomly chosen to implement the full MS ExTRA program, with the MSQI, ExpandED, and guided reading tutoring providing a school-wide literacy focus and additional extracurricular opportunities. From the remaining forty experimental schools, twenty schools were randomly chosen to implement the MSQI only, and twenty schools were selected as control schools, receiving no special intervention as a result of the experiment.[1] The table below indicates the services provided within each set of schools in the experimental group:

Group	Schools	MSQI	ExpandED	Reading Tutoring
MS ExTRA	20	X	X	X
MSQI Only	20	X		
Control	20			

Middle School Quality Initiative
NYC DOE began implementing the Middle School Quality Initiative in 2011. The group of treatment and comparison schools participating in MSQI beginning in fall of 2013 comprise the third cohort of MSQI schools.

MSQI schools work over several years to build a schoolwide literacy culture by providing professional development for teachers and administrators, including ongoing literacy coaching, access to literacy curricula and software, and programming support in implementing a daily period of differentiated literacy support (a strategic reading period or "SRP"). In the first year of MSQI implementation, few schools choose to implement the full MSQI model, which requires modifying school schedules to accommodate a differentiated SRP for targeted grades, but some start by introducing specific curricula and software,[2] and most schools send teachers and administrators to professional development focused on integrating literacy across content areas and best practices in literacy instruction. By the third year of implementation, all participating MSQI schools are expected to implement an SRP where students participate in a range of differentiated, smaller-group interventions (usually 10-15 students, grouped homogenously by reading ability), and to implement

[1] Throughout the paper, we refer to the MS ExTRA schools as the "treatment" schools, the MSQI-only schools as the "main control", and the control schools as the "supplemental control."
[2] Two common examples of literacy curricula provided by the MSQI and used in MSQI schools are Word Generation and Wilson Just Words.

Word Generation, a vocabulary program that challenges students to use specific words across content areas each week.[3]

In MS ExTRA and MSQI-only schools, each principal nominated a teacher leader or junior administrator as the school's MSQI coordinator each year. MSQI coordinators served as the main point of communication between MSQI district administration and MSQI schools and facilitated implementation of the various components of the MSQI program at the school.

After-School Programming by ExpandED Schools

ExpandED Schools created and manages a model of after-school programming that was implemented in the MS ExTRA schools. ExpandED Schools partners with community-based organizations (CBOs) to extend the length of the school day and expand the types of learning experiences available to students. Typically, ExpandED programs add 2.5 hours of programming, offering students a mixture of academic and non-academic activities as well as an additional meal. In contrast to typical after-school programs, ExpandED schools attempt to create a seamless transition between the traditional school day and after-school hours, where students perceive their after-school classes to be as important as core content classes. The types of programming offered varied by site, depending upon a school's particular CBO partner, the skills of the CBO employees, and the involvement of teachers and school administrators in ExpandED. Examples of typical activities include dance, sports, science labs, homework help, robotics, photography, and debate club.

In MS ExTRA schools, each ExpandED program was led by an MS ExTRA site director, employed by the CBO, who was responsible for overseeing program operations, staffing, budgeting, and logistics. Program Managers worked closely with MS ExTRA site directors to ensure that ExpandED operations and scheduling met quality expectations and were compatible with requirements for high-dosage tutoring.

Guided Reading Tutoring

MS ExTRA tutoring provided participating students with daily small-group literacy instruction centered around building student reading comprehension while reading high-interest fiction or non-fiction chapter books. Guided reading tutorials were modeled after the guided reading framework described in Fountas and Pinnell (2001). In guided reading tutorials, students read leveled high-interest chapter books in groups of four or fewer with a single tutor one period per day[4], five days per week. During the 2013-14 school year (Year 1), only 6th grade students were eligible to participate in tutoring. During the 2014-15 school year (Year 2), 6th and 7th grade students were eligible to participate in tutoring. During the 2015-16 school year (Year 3), 6th, 7th, and 8th grade students were eligible to participate in tutoring.

In Year 1, each MS ExTRA site was assigned a Regional Tutoring Coordinator (RTC) who managed tutor staffing, coordinated tutoring logistics, and provided instructional coaching for tutors at 2 or 3 MS ExTRA schools. In Year 2, each MS ExTRA site was assigned a Program Manager (PM), who managed tutor staffing, coordinated tutoring logistics, provided instructional coaching for tutors, and oversaw ExpandED programming at 2 or 3 MS ExTRA schools. RTCs (PMs) worked closely with EdLabs to ensure fidelity of implementation to the program's research design and instructional model.

[3] See http://wordgen.serpmedia.org/ for more information about the program.

[4] ExpandED periods are approximately 50-70 minutes long; SRP periods are approximately 45-60 minutes long.

Project Timeline

In April 2013, roughly 150 low-performing, high-poverty NYC DOE middle schools were invited to apply for a chance to participate in a district literacy intervention through the Middle School Quality Initiative. In May 2013, the randomization was conducted and schools were notified as to whether they were selected. Two weeks later, MS ExTRA school leaders were invited to a program launch, hosted by the NYC DOE Chancellor of Schools. Tutor recruitment and selection began in June 2013. Year 1 tutor training was held from August 27 to August 30, 2013. Year 1 Diagnostic DRP assessments were administered between September 12 and September 26, 2013. Year 1 tutoring began at all sites between September 24 and October 2, 2013 and continued through at least June 13, 2014, with some sites continuing through the last day of school, June 26, 2014. The number of Year 1 active programming days varied by school from 128 to 158.

Year 2 tutor training was held from August 26 to August 29, 2014. Year 2 Diagnostic DRP assessments were administered between September 5 and September 23, 2014. Year 2 tutoring began at all sites between September 17 and October 6, 2014 and continued through at least May 29, 2015, with some sites continuing through the last day of school, June 26, 2015. The number of Year 2 active programming days varied by school from 107 to 151.

Year 3 tutor training was held from August 25 to August 27, 2015. Year 3 Diagnostic DRP assessments were administered between September 11 and October 2, 2015. Year 3 tutoring began at all sites between September 9 and October 5, 2015 and continued through at least June 2, 2016, with some sites continuing through June 27, 2016. The number of Year 3 active programming days varied by school from 88 to 173.

Guided Reading Tutoring Model

Curriculum development

EdLabs partnered with teachers and middle school literacy experts around the country to refine the guided reading lesson structure, to identify a library of high-quality, high-interest chapter books, and to develop a structured curriculum to support guided reading tutoring instruction for each of these books. This panel chose the books that form the foundation of the guided reading curriculum libraries. The 6th grade curriculum library included 36 fiction and 16 non-fiction chapter books spanning a wide range of genres and reading levels.[5] The 7th grade curriculum library included a similarly diverse selection of 53 fiction and 27 non-fiction chapter books.[6,7] The 8th grade curriculum library included 72 fiction and 40 non-fiction chapter books.[8] There was a significant amount of overlap between the lists to allow for 7th and 8th graders access to lower-level texts as appropriate.

[5] The 6th grade library included texts with Lexile levels ranging from 360L to 1190L, and Fountas and Pinnell reading levels P through Z+. The median book was F&P level W and Lexile level 785L.

[6] The 7th grade library included texts with Lexile levels ranging from 310L to 1310L, and Fountas and Pinnell reading levels O through Z+. The median book was F&P level V and Lexile level 770L.

[7] It may be noted that the median book in the 7th grade library was of a slightly lower reading level than that of the 6th grade library. Responding to feedback from program managers, tutors, and students, we expanded the breadth of offerings at lower reading levels when constructing the 7th grade library for Year 2. This expansion accounts for the unexpected direction of the change.

[8] The 8th grade library included texts with a similar range as the 7th grade library, but extended the offerings to allow additional variety as students entered the third year of tutoring.

To plan their lessons, tutors were provided with a detailed guided reading curriculum for each book. For each chapter of each book (or each 8-20 page section of text for books with short chapters), a master curriculum document provided questions to activate students' prior knowledge, thinking tasks to assign a purpose to reading, vocabulary definitions, and questions to facilitate discussion. For each page of text, the curriculum provided detailed comprehension questions with exemplary student responses, follow-up questions or instructions to correct common misconceptions, the common-core standard to which the question is aligned, and a reference location in the text (page and paragraph). Curriculum documents were either printed for tutors or accessed via Google Drive on WiFi-enabled tablets (iPads and Android tablets).

Student book selection

The library of books available to a given student was determined by that student's grade level and most recent score on the Degrees of Reading Power (DRP) assessment. The DRP is a formative reading assessment that measures student reading comprehension. Books in the 6th grade library were assigned to one or more of Tiers 1, 2, 3, and 4, corresponding to DRP ranges 40-45, 46-50, 51-55, and 56-60. Books in the 7th and 8th grade library were assigned to one or more of Tiers 1, 2, 3, 4, and 5, corresponding to DRP ranges 40-45, 46-50, 51-55, 56-60, and 61-67. Students were allowed to choose among books within 1 tier of their group's assigned tier, so a student in a Tier 2 group had access to books that were labeled Tier 1, 2, or 3.

To choose a book, tutors selected 3-5 titles from this subset of leveled texts to present to each group. Tutors shared brief summaries or samples of each book with students, who voted on which book to read. Tutors who taught more than one group at a given grade level and tier could hold individual votes in each group or pool votes across groups in a given grade level to determine which book(s) to read next.

Program Structure

A typical sixty-minute lesson began with a 3-5 minute do-now activity centered around an "activating prior knowledge" question to help remind students of important plot details or foundational knowledge that was important to that day's reading. Tutors then spent 3-5 minutes reviewing 2-3 important vocabulary terms and assigning a thinking task to set a purpose for reading. Once this introduction was complete, tutors transitioned students to 30-40 minutes of independent reading, assigning students a section of text to read and providing extension activities for students who finished early. Extension activities included comprehension quizzes, writing prompts, and extension readings that were provided for each section of text. During independent reading, students were pulled one at a time for 5-10 minute individual conferences with their tutor. During conferences, tutors listened as students read passages aloud and asked them open-book comprehension questions about the text they had just read. After each student had at least one chance to conference with their tutor, the tutor would convene the group for a 5-20 minute closing discussion, analyzing material from that day's section of text.

Scheduling

During Year 1, 6th grade students attended tutorials during 9th or 10th period, as a part of regularly scheduled ExpandED programming. During Year 2, some schools scheduled reading tutoring during a strategic reading period in the traditional school day, some schools provided tutoring during

ExpandED after school programming, and most schools provided tutoring during both an SRP and ExpandED. SRP students typically attended tutorials in 7th or 8th period, during the traditional school day, while ExpandED students attended tutorials during 9th or 10th period. While schedules varied by school, 7th and 8th grade tutorials were more likely to be scheduled during an SRP in the traditional school day, while 6th grade tutorials were more likely to be scheduled during after-school programming.

Tutors

Selection Process

In Year 1, tutors were recruited and hired by a third party, ReServe Inc., as hourly employees. ReServe, Inc. – a subsidiary of Fedcap – is a non-profit organization that helps retired professionals and older individuals find part-time work in the non-profit and public sectors.[9] In Years 2 and 3, tutors were recruited and hired either by ReServe or by the school's CBO, depending on the school. All tutors were required to hold a Bachelor's degree, pass a high school reading assessment, and pass a background check administered by the NYC DOE. School leaders had the option to interview tutors themselves or to delegate the responsibility to school personnel, CBO staff, or their school's RTC (PM in years 2 and 3).

In Year 1, tutors worked five days per week, and a tutor's typical workday was 2pm to 5pm, including two one-hour tutoring sessions and a one-hour planning period, but exact schedules varied by site. In Years 2 and 3, tutors worked five days per week, and a tutor's typical workday was 1pm to 5pm, including three one-hour tutoring sessions and a one-hour planning period, but exact schedules varied by site. In Year 1, at full capacity, 110 tutors served students across the 20 full treatment schools. In Years 2 and 3, at full capacity, 145 tutors served students across 19 of the 20 full treatment schools.[10]

Tutor Characteristics

Because ReServe specifically seeks to match retired individuals to service positions, at the start of year 1, 59% of tutors who worked in MS ExTRA schools were over 55 years old, and 41% were under 55. In Years 2 and 3, there was more diversity in tutor age as more tutors were recruited and employed by CBOs. On their resumes, 19% of tutors listed some K-12 education experience in a traditional school setting as a teacher, substitute teacher, or paraprofessional. Similarly, 33% of tutors listed some experience tutoring children in grades K-12 prior to joining MS ExTRA. All tutors held Bachelor's degrees from accredited universities, and 29% held advanced graduate or professional degrees (e.g., MA, MSc, JD, MD, or PhD).

In Years 2 and 3, there were several tutors who returned from the previous school year. At the beginning of Year 2, 56 tutors were returning from Year 1. At the beginning of Year 3, 68 tutors were returning from Year 2. In all, 26 tutors worked in MS ExTRA schools for all three years of the project.

Training

[9] Traditionally, ReServe limits participation to individuals over 55 years of age, but for this project, ReServe accepted participants, who they call ReServists, of all ages.

[10] One school opted to discontinue ExpandED programming and reading tutoring after Year 1.

Tutors who were hired during the summer before Year 1 or Year 2 participated in a twenty-hour summer training held over four days before joining their respective school teams each year; the Year 3 training was held over the course of three days but was also twenty hours. EdLabs hired an adolescent literacy expert to train tutors in instructional pedagogy and a classroom management specialist to develop tutors' classroom management skills. The summer training focused on the guided reading instructional model, lesson planning using the curriculum, managing student behavior, building relationships with students, and establishing systems and procedures to run effective tutorials. The training was differentiated for the returning and the new tutors in Years 2 and 3. Tutors hired after the start of each school year were trained by their RTC (PM) one-on-one or in small groups.

Supervision
In Year 1, management for the reading tutoring and the after-school programming was split between two groups. Seven RTCs oversaw tutoring at 2 or 3 sites each, conducting annual formal observations of each tutor and ongoing informal observations and providing coaching to the tutors. MS ExTRA site directors reported to two PMs, who oversaw the budget process and after-school programming at 10 sites each. RTCs were required to visit each of their sites during reading tutoring at least one day every week during the school year.

In Years 2 and 3, the RTC role was eliminated and nine PMs oversaw tutoring and ExpandED programming at 2 or 3 sites each, conducting 2-3 formal observations of each tutor in addition to ongoing informal observations and coaching. The consolidation of RTC and PM responsibilities in the PM role was designed to streamline communication between ExpandED Schools personnel, school leaders, and site directors and to foster a sense of shared ownership of the tutoring program among these stakeholders. MS ExTRA site directors reported to PMs, who also oversaw the budgeting process and ExpandED programming at the same 2 or 3 sites. PMs were required to visit each of their schools during guided reading tutoring at least one day every week during the school year.

Coaching and Development
RTCs (Year 1) and PMs (Years 2 and 3) served as instructional leaders for tutors, providing regular informal feedback, instructional coaching, and material resources. Tutors also attended professional development seminars during the school year that focused on the guided reading instructional model, lesson planning using the curriculum, managing student behavior, building relationships with students, and establishing systems and procedures to run effective tutorials. Half-day professional development seminars were offered 3-5 times annually, typically during school holidays. Seminars were led by DOE staff, RTCs/PMs, and EdLabs staff. Tutors were also encouraged to attend CBO staff meetings and professional development sessions when possible.

Student Assignment

Eligibility for tutoring
Students were assigned to tutoring based upon their diagnostic score on the DRP assessment, a multiple choice reading comprehension assessment, and were grouped homogenously by their DRP

performance.[11] In Year 1 and Year 2, 6th grade students who scored greater than or equal to 40 and less than or equal to 60 on their beginning of year DRP assessment were eligible to participate in guided reading tutoring. In Year 2, 7th grade students who were eligible in Year 1, scored below 64 on their Year 2 beginning of year DRP assessment, and remained enrolled in a treatment school remained eligible to participate in guided reading tutoring. The same process was in place for the 2015-16 school year, with the upper cut-off increasing to 67 on the DRP for 8th grade students. Students whose initial Year 1 DRP scores qualified them for guided reading tutoring services could "graduate out" of the program for one school year by scoring above 64 at the start of Year 2 or a 67 at the start of Year 3.

Students who qualified for tutoring in Year 1 as 6th graders comprise the "Experimental Cohort" for whom the program was designed to provide three years of reading tutoring from 6th through 8th grade. In negotiations with schools, CBOs, and partner organizations, a non-negotiable criteria of the partnership was the fidelity of the Experimental Cohort's experience to the original program design in terms of student-to-tutor ratio (4:1), tutoring period length (at least 40 minutes each day), tutoring frequency (5 days per week), and student eligibility (Year 1 initial DRP score $\geq$40 and $\leq$60). Students who entered the program as 6th graders in Year 2 or Year 3 comprise the "Non-Experimental Cohorts" for whom schools were allowed more flexibility in program design, enrollment, and scheduling. In Year 2, students in the Non-Experimental Cohort were more likely to attend tutoring 4 days per week, and a small number of students (<10) who scored slightly below the eligible range (36-39) were allowed to participate in tutoring in selected schools. In Years 2 and 3, tutoring for the Non-Experimental Cohort was scheduled for less than 40 minutes per day in some schools.

Assignment to Tutoring

The eligible range was established so that 6th grade students who demonstrated at least basic middle-elementary level fluency, but remained below grade level in their comprehension skills were eligible to participate in the program. Assignment was done so that the lowest performing eligible students received priority when there were more eligible students than available seats at a school.[12] Eligible students who were not assigned to tutoring were assigned to a waitlist, ordered by diagnostic DRP score (from low to high), to fill open seats as the year went on. If an assigned student stopped attending after-school programming for an extended period of time (typically one month), and their parent was unreachable or confirmed that they were removing their student from the program, the empty seat was filled by an eligible student from the waitlist. Assigned students who returned after a prolonged absence were moved to the top of the waitlist.

During Year 1, approximately 45% of 6th grade students in MS ExTRA schools were offered the opportunity to participate in guided reading tutoring during the initial assignment. In Year 2, approximately 35% of 6th grade students and 35% of 7th grade students in MS ExTRA schools were

[11] The DRP was administered three times during each school year, a diagnostic assessment in September, a mid-year assessment in January, and an end-of-year assessment in June. Some groups were shifted based upon mid-year results, and some groups were shifted based upon personality conflicts among students or between students and tutors.

[12] Due to an administrative error, students in three MS ExTRA schools were assigned to tutoring starting with the median-scoring eligible student (on their diagnostic DRP assessment) and moving outward, rather than starting with eligible students who scored the lowest on the beginning of year DRP assessment and moving up the list of eligible students. Students who should have been assigned to tutoring initially were moved to the top of the waitlist. In Year 2, all tutoring assignments gave priority to eligible students who scored the lowest on their beginning of year DRP assessment.

offered the opportunity to participate in guided reading tutoring during the initial assignment. In Year 3, approximately 28% of 6th grade students, 25% of 7th grade students, and 29% of 8th grade students were offered spots in the reading tutoring program.

In Year 1, the number of tutoring seats at each school was determined by taking 45% of 2012-13 student enrollment and rounding up or down to the nearest multiple of eight (since each additional tutor could serve two periods of four students). In Year 1, 886 students were initially assigned to tutoring out of 1,251 eligible students and 2,104 total 6th graders in MS ExTRA schools. Accounting for student turnover and use of the waitlist, 980 different students participated in guided reading tutoring for at least one day during Year 1.

In Year 2, the number of 7th grade tutoring seats at each school was determined by rounding to the nearest multiple of four the number of students who were initially assigned to tutoring in Year 1 and remained (1) eligible for tutoring in Year 2 and (2) enrolled at a treatment school. In Year 2, 719 7th grade students and 753 6th grade students were initially assigned to tutoring.[13] In Year 3, 576 8th grade students, 480 7th grade students, and 534 6th grade students were initially assigned to tutoring.

Group assignment

Students were assigned to groups of four by DRP score in five point bands. Students scoring 40-45 (tier 1), 46-50 (tier 2), 51-55 (tier 3), 56-60 (tier 4), or 61-64 (tier 5) were grouped together. Adjustments were made to initial group assignments to avoid personality conflicts. To accommodate these adjustments, students could be moved to groups one tier above or below their assigned tier as long as the range of the most recent DRP assessment scores in a given group did not exceed 10 points.

Student Attendance

The observed attendance rate for after-school programs varied widely by school. After-school attendance could not technically be mandatory, and students could not be penalized for not attending. However, in schools where the leaders purposefully messaged to parents that the after-school programming and reading tutoring were important parts of providing their child a well-rounded education and encouraged parents to see the program as a continuation of the traditional school day, there were markedly higher rates of program attendance. Within after-school reading tutoring, Year 1 daily attendance rates varied by school from 30% to 79%**,** with an overall average attendance rate of 47%. During Year 1, the number of active programming days varied by school from 128 to 158**,** with 1,905 of 2,260 enrolled 6th grade students participating in after-school programming for at least one day.

During Year 2, tutoring attendance rates varied by school from 25% to 63%, with an overall average attendance rate of 40%. During Year 3, tutoring attendance rates varied by school from 14% to 71%, with an overall average attendance rate of 43%.

[13] In Year 2, the number of 6th grade students assigned to tutoring was influenced by the school's scheduling model. The number of tutors hired for each school was determined assuming that approximately 45% of 6th grade students would be able to be served, however this projection assumed that the school would implement an SRP, allowing tutors to teach three periods per day and serve 12 students per day. At schools that did not implement an SRP, tutors could only serve 8 students per day and, as a result, fewer 6th grade students were assigned to tutoring.

Data Collection

Site Visits

Fidelity of implementation was monitored by a combination of site-based management, remote management, summer training, and ongoing professional development. RTCs (PMs) conducted weekly informal observations and annual formal observations of tutors. In these observations, RTCs (PMs) evaluated tutors on specific aspects of classroom management and instructional pedagogy, provided in-person feedback and coaching, and used a detailed rubric to record observation data.

In addition to RTC (PM) supervision, a team of EdLabs staff conducted quarterly site visits to each MS ExTRA school in Years 1 and 2, and one site visit in Year 3. EdLabs site visits included 5- to 10-minute observations of each tutor. Observations were recorded using a detailed rubric, and school-level aggregated data was shared with RTCs (PMs) after each visit. EdLabs recorded and reported information about tutor professionalism, classroom management, instructional fidelity to the guided reading model, student attendance, and student engagement at each site during each visit.

Technology

At the end of Year 1, EdLabs and ExpandED Schools partnered with LightSail Education and Baker and Taylor to conduct a pilot to test the viability of transitioning the majority of the 7th grade library to eBooks rather than physical books. Book transportation and storage proved to be one of the largest logistics challenges in Year 1, and allowing students to read eBooks on tablets was piloted to solve problems associated with these challenges. Two sites piloted the LightSail eReader application during the last month of programming in Year 1 and continued to use LightSail during summer school programming at their school.

At the beginning of Year 2, 7th grade students in all MS ExTRA schools were provided electronic access to a subset of the 7th grade library using tablet eReaders rather than traditional physical books. Of the 81 books in the 7th grade library, 60 were made available to students in guided reading tutoring exclusively as eBooks. All but two sites (17 of 19) used Apple iPad Mini 2 tablets to access Lightsail and read eBooks in the Lightsail environment. Because of budget constraints, the two largest sites used Android tablets to read eBooks in the Axis Reader environment. Axis Reader is a simple, free eReader that allows students to check out and read books from the digital lending library. Lightsail is a paid eReader app that embeds cloze vocabulary questions into the text of chapter books, adds comprehension questions at key points of the book, allows teachers (or tutors) and students to share written messages about the text, allows teachers to track student progress in real time, and displays student Lexile growth based upon their performance answering cloze vocabulary questions. In some circumstances, students can earn badges within the app for meeting certain benchmarks of Lexile growth or pages read.

References

Fountas, Irene C., and Gay Su Pinnell. (2001). *Guiding Readers and Writers, Grades 3-6: Teaching Comprehension, Genre, and Content Literacy*. Connecticut: Heinemann.

Appendix B – Variable Construction

A. Administrative District Data

Economically Disadvantaged
Students were tagged as economically disadvantaged if they qualify for free or reduced lunch or attend a universal feeding school – i.e. if the free lunch flag in the administrative data was a "1."

English Language Learners
Students were tagged as English Language Learners if the ELL flag in the administrative data was a "Y" and not otherwise.

Special Education Services
The administrative indicates whether students have an Individualized Education Program. Students are designated as receiving special education services if the IEP special education flag is a "Y" and not otherwise.

Home Language & Country of Birth
NYC DOE provides codes for the language spoken in a student's home and his or her country of birth. Students are marked as being born in the US if they were born in any of the 50 US states and as not born in the US if they are born in any other country.

All demographic variables listed above vary over time. We use the value of the variable in the year before treatment for all students present in the district in 2012-13. For students new to the district in 2013-14, we use the value of the variable from October of 2013-14.

Race/Ethnicity
We code the race variables such that the five categories – white, black, Hispanic, Asian and other – are complete and mutually exclusive. Hispanic ethnicity is an absorbing state. Hence "white" implies non-Hispanic white, "black" non-Hispanic black, and so on.

Race and gender variables are assigned using the most recently available data and are filled in from past years of data if unavailable in the most recent data.

DRP Scores
The Degrees of Reading Power assessment was used to determine students' reading level and to assign students of the targeted ability range to attend tutoring. If a student was present in the DRP file but had one observation missing a DRP score and another non-missing, we used the non-missing score. If a student was recorded with multiple DRP scores, we used the earlier of the two tests as a student's DRP score.

Treatment
Students were assigned to the last school they were in before October 31st in 2013-14. Students were first assigned to the school they took the DRP in. If a student did not take the DRP, she was assigned to the school that she was enrolled in at the time of the October

snapshot. If a student was missing from the October snapshot but was present in the end-of-year file with an admit date before October 31st, she was assigned to that school.

The tutoring cohort in each school was determined by DRP scores; the range of DRP scores that qualified a student for tutoring was school-specific due to size constraints. Students were assigned to the experimental tutoring cohort *for this analysis* if they took the DRP and scored between the range of scores that tutored students enrolled in their assigned school (as described above) had scored.

B. Administrative Project Data

Student Attendance at Tutoring
Student attendance at tutoring was recorded by RTCs (PMs in years two and three) each day of the tutoring program. Students were recorded as present each day if they were marked as present and zero otherwise; this includes days in which tutoring was cancelled at a given school. Attendance rates are calculated as the days present divided by the total number of school days that were eligible for tutoring; this excludes all weekends and official holidays. Days in which a school or tutor cancelled tutoring were counted in the denominator.

Books Read at Tutoring
Tutors recorded the books that they read with each tutoring group and the dates during which those books were read. These were matched to students' assignments to groups and the days that students were present at tutoring to calculate the number of books that students read. Each book came from the designated curriculum described in detail in Appendix A; for each book we record the number of pages and words in the book to calculate the number of pages and words that students read in tutoring. Whether books were of particular "black interest" or "Hispanic interest" was determined by the race of the protagonist(s), the race of the author, the image on the cover, or the historical relevance of the plot.

Tutor Characteristics
Tutor characteristics were culled from resumes submitted at the time of their application. This includes whether or not tutors had a bachelor's degree in either English or education, whether or not tutors had K-12 teaching experience, whether or not tutors had previous tutoring experience, and whether or not tutors had a graduate degree. Tutors were also observed during four site visits in years one and two and one site visit in year three. Observers recorded an overall tutor quality score, which is also used in analysis.

C. Neighborhood Characteristics

Each school lies in a census tract; median household income and the percent of single parent households for each tract were gathered from the ACS and linked to each school. All Stop and Frisk interactions are recorded with longitude and latitude coordinates; these interactions were linked to census tracts using ArcGIS and collapsed at the census tract level to be linked to schools. In cases where one interaction lied on a border between two census tracts, one census tract was randomly assigned.

Appendix Table 1: Number of Tutors by School

School	2013-14	2014-15	2015-16
Renaissance School of the Arts	2	4	4
Thurgood Marshall Academy for Learning and Social Change	3	3	4
J.H.S. 143 Eleanor Roosevelt	8	8	8
M.S. 223 The Laboratory School of Finance and Technology	5	8	9
South Bronx Academy for Applied Media	7	8	9
J.H.S. 123 James M Keiran	9	17	11
Eagle Academy for Young Men	5	6	8
The Highbridge Green School	7	10	12
Frederick Douglass Academy V. Middle School	5	6	8
Juan Morel Campos Secondary School	7	0	0
I.S. 340	4	5	2
East Flatbush Community Research School	5	7	6
I.S. 30 Mary White Ovington	6	10	11
P.S. 109	5	5	8
Andries Hudde	12	16	13
P.S. 043	6	6	8
Waterside School for Leadership	3	4	4
Village Academy	6	7	8
PS/IS 116 William C. Hughley	3	3	5
Queens United Middle School	5	5	7

Appendix Table 2: Attrition

	Pooled	Black	Hispanic	*p-value*
	(1)	(2)	(3)	(2)=(3)
Missing Math	-0.010	-0.020	-0.005	0.326
	(0.012)	(0.016)	(0.015)	
	5,090	2,227	2,299	
Missing ELA	-0.008	-0.012	-0.009	0.735
	(0.007)	(0.008)	(0.011)	
	5,090	2,227	2,299	
Missing Attendance	-0.002	-0.008	0.002	0.194
	(0.004)	(0.007)	(0.006)	
	5,090	2,227	2,299	

Notes: This table reports ITT results of the tutoring program. The sample is students in treatment and main control schools with DRP scores within the range that would qualify them to receive tutoring services. Students who qualify for tutoring or after-school programs in years 2014-15 and 2015-16 who are not in the experimental cohort are not included in any sample in those years. Students are assigned to the school that they are enrolled in by October 31st in the first year of treatment. Testing variables are drawn from district test score files and are standardized to have a mean of zero and a standard deviation of one within each year and grade among students with valid test scores. All specifications control for matched-pair fixed effects and the student-level demographics summarized in Table 2 plus three years of baseline reading and math scores and their squares. When the outcome is student attendance, controls also include student attendance in the year prior to treatment. All controls are interacted with indicators for whether a student is Hispanic, black, or other race. Standard errors, reported in parentheses, are clustered at the school-year level. Significance at the 1%, 5%, and 10% levels indicated by ***, **, and *, respectively.

Appendix Table 3: Effects on Student State Test Scores and Attendance (Lee Bounds)

	2013-14	2014-15	2015-16	Pooled
	(1)	(2)	(3)	(4)
Panel A: Main Results				
Math	-0.085	-0.043	0.078	-0.020
	(0.054)	(0.049)	(0.114)	(0.047)
	1,707	1,647	1,447	4,802
ELA	0.067*	0.004	0.033	0.034
	(0.039)	(0.038)	(0.048)	(0.027)
	1,714	1,654	1,571	4,940
Attendance	0.012***	0.010***	0.013***	0.011***
	(0.004)	(0.003)	(0.003)	(0.002)
	1,758	1,680	1,601	5,041
Panel B: Black Students				
Math	0.014	-0.034	0.284**	0.071
	(0.078)	(0.070)	(0.128)	(0.065)
	749	713	607	2,068
ELA	0.107**	0.055	0.064	0.074**
	(0.042)	(0.037)	(0.051)	(0.029)
	755	721	677	2,153
Attendance	0.016***	0.020***	0.021***	0.019***
	(0.005)	(0.003)	(0.005)	(0.003)
	772	740	686	2,199

Notes: This table reports Lee Bounds on the ITT results of the tutoring program. As described in Lee (2009), we calculate lower bounds by dropping the highest achieving treatment students, or lowest achieving control students, until attrition is equal between treatment and control. This process occurs independently for each outcome. The sample is students in treatment and main control schools with DRP scores within the range that would qualify them to receive tutoring services. Students who qualify for tutoring or after-school programs in years 2014-15 and 2015-16 who are not in the experimental cohort are not included in any sample in those years. Students are assigned to the school that they are enrolled in by October 31st in the first year of treatment. Testing variables are drawn from district test score files and are standardized to have a mean of zero and a standard deviation of one within each year and grade among students with valid test scores. All specifications control for matched-pair fixed effects and the student-level demographics summarized in Table 2 plus three years of baseline reading and math scores and their squares. All controls are interacted with indicators for whether a student is Hispanic, black, or other race. Standard errors, reported in parentheses, are clustered at the school-year level. Significance at the 1%, 5%, and 10% levels indicated by ***, **, and *, respectively.

Appendix Table 4: School-Level Regressions

	2013-14	2014-15	2015-16	Pooled
	(1)	(2)	(3)	(4)
Panel A: Full Tutoring Sample				
Math	-0.118	-0.066	0.058	-0.044
	(0.084)	(0.068)	(0.114)	(0.073)
	36	36	36	36
ELA	0.000	-0.035	0.016	-0.007
	(0.079)	(0.059)	(0.069)	(0.061)
	36	36	36	36
Attendance	0.012*	0.010*	0.014**	0.012**
	(0.007)	(0.006)	(0.006)	(0.005)
	36	36	36	36
Panel B: Black Students				
Math	-0.022	-0.026	0.274*	0.070
	(0.102)	(0.088)	(0.141)	(0.082)
	34	34	34	34
ELA	0.061	0.030	0.025	0.038
	(0.078)	(0.082)	(0.089)	(0.070)
	34	34	34	34
Attendance	0.023**	0.029***	0.033***	0.028***
	(0.008)	(0.008)	(0.008)	(0.007)
	34	34	34	34
Panel C: Hispanic Students				
Math	-0.086	-0.014	-0.092	-0.060
	(0.081)	(0.090)	(0.118)	(0.086)
	36	36	36	36
ELA	0.026	-0.035	-0.007	-0.006
	(0.085)	(0.064)	(0.075)	(0.064)
	36	36	36	36
Attendance	0.010	0.005	0.006	0.007
	(0.009)	(0.009)	(0.008)	(0.007)
	36	36	36	36

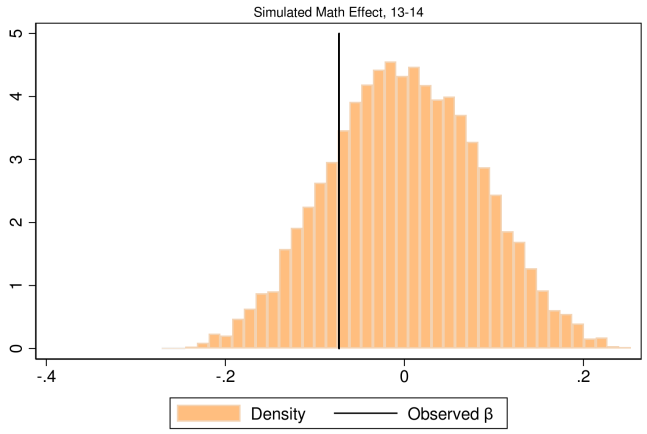
Notes: This table reports school-level results of the tutoring program. School-level means are calculated over the sample of students in treatment and main control schools with DRP scores within the range that would qualify them to receive tutoring services. In Panel A, that includes all students eligible for tutoring. In Panel B, it includes only black students and in Panel C, it includes only Hispanic students. Students are assigned to the school that they are enrolled in by October 31st in the first year of treatment. Testing variables are drawn from district test score files and are standardized to have a mean of zero and a standard deviation of one within each year and grade among students with valid test scores. All specifications control for matched-pair fixed effects. All regressions are weighted by the number of students with valid outcome test scores in each year. Standard errors, reported in parentheses, are robust to heteroskedasticity. Significance at the 1%, 5%, and 10% levels indicated by ***, **, and *, respectively.

Appendix Table 5: Main Effects - Predicted DRP for Missing Matchpairs (ITT)

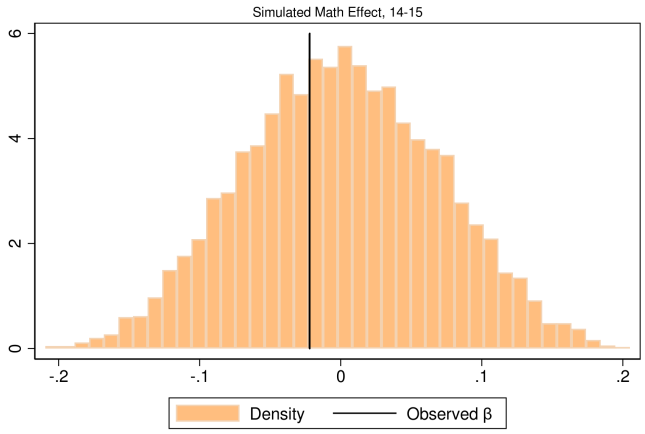
	2013-14	2014-15	2015-16	Pooled
	(1)	(2)	(3)	(4)
Panel A: Effects on State Test Scores				
Math	-0.111*	-0.057	0.053	-0.049
	(0.056)	(0.047)	(0.117)	(0.049)
	2,087	1,996	1,664	5,747
ELA	0.048	0.003	-0.003	0.015
	(0.039)	(0.037)	(0.052)	(0.027)
	2,083	2,017	1,915	6,015
Reading Subscore	0.009	-0.005	-0.019	-0.006
	(0.034)	(0.030)	(0.043)	(0.021)
	2,083	2,017	1,915	6,015
Writing Subscore	0.125*	-0.010	0.021	0.045
	(0.065)	(0.077)	(0.078)	(0.048)
	2,083	2,017	1,915	6,015
Panel B: Effects on Student Attendance at School				
Attendance	0.011***	0.008**	0.014***	0.011***
	(0.003)	(0.003)	(0.003)	(0.002)
	2,140	2,046	1,947	6,133
Control Mean	0.923	0.922	0.916	0.920

Notes: This table reports ITT results of the tutoring program. The sample is students in treatment and main control schools with DRP scores within the range that would qualify them to receive tutoring services. Students who qualify for tutoring or after-school programs in years 2014-15 and 2015-16 who are not in the experimental cohort are not included in any sample in those years. Students are assigned to the school that they are enrolled in by October 31st in the first year of treatment. Testing variables are drawn from district test score files and are standardized to have a mean of zero and a standard deviation of one within each year and grade among students with valid test scores. All specifications control for matched-pair fixed effects and the student-level demographics summarized in Table 2 plus three years of baseline reading and math scores and their squares. All controls are interacted with indicators for whether a student is Hispanic, black, or other race. Standard errors, reported in parentheses, are clustered at the school-year level. Significance at the 1%, 5%, and 10% levels indicated by ***, **, and *, respectively.

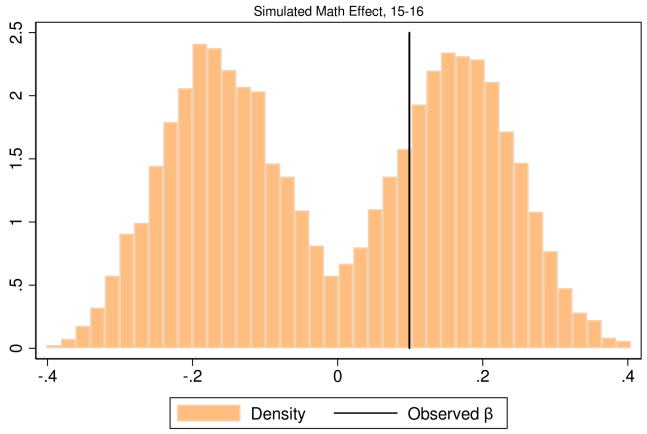
Appendix Figure 1: Permutation Tests



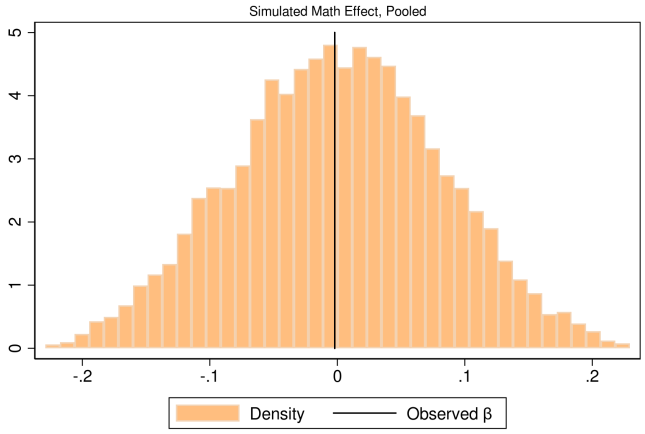
N = 10000 Replications. **P-value** (two-tailed) = 0.402



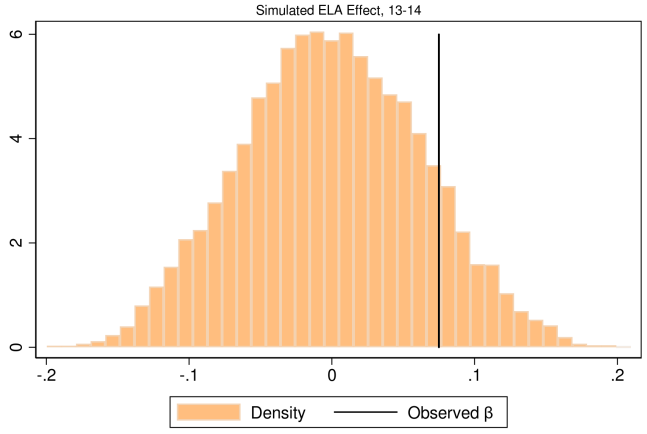
N = 10000 Replications. **P-value** (two-tailed) = 0.760



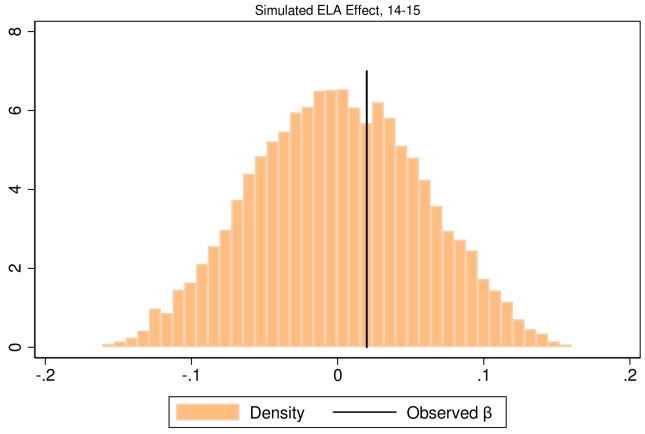
N = 10000 Replications. **P-value** (two-tailed) = 0.787



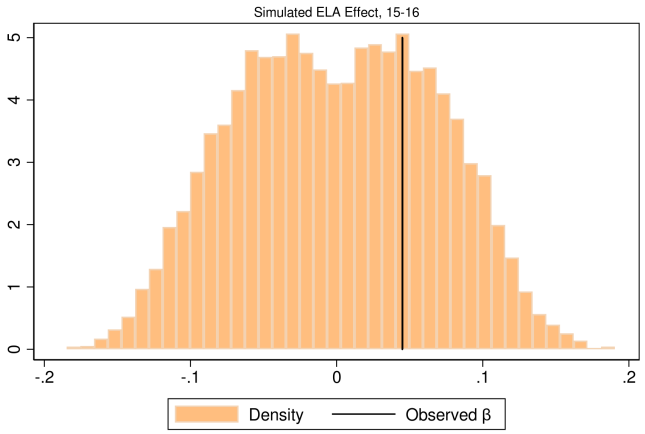
N = 10000 Replications. **P-value** (two-tailed) = 0.982



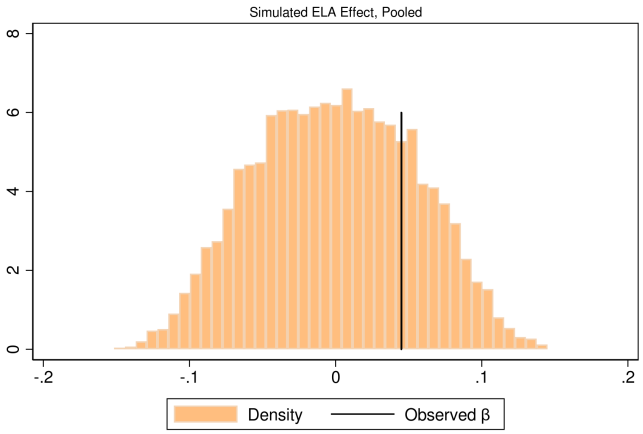
N = 10000 Replications. **P-value** (two-tailed) = 0.245



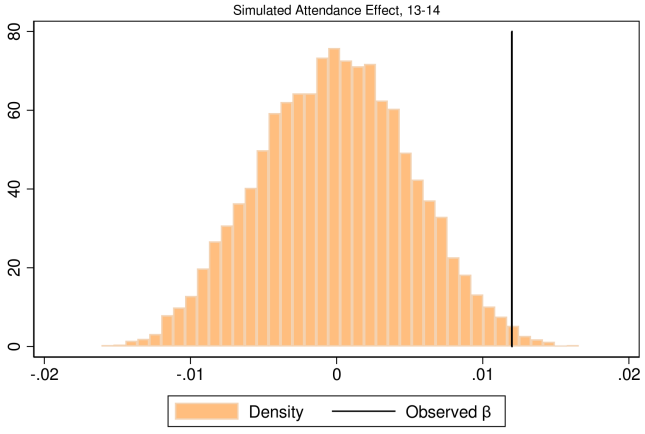
N = 10000 Replications. **P-value** (two-tailed) = 0.749



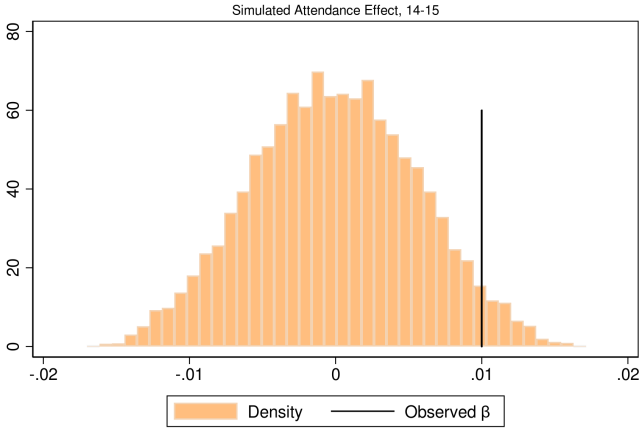
N = 10000 Replications. **P-value** (two-tailed) = 0.577



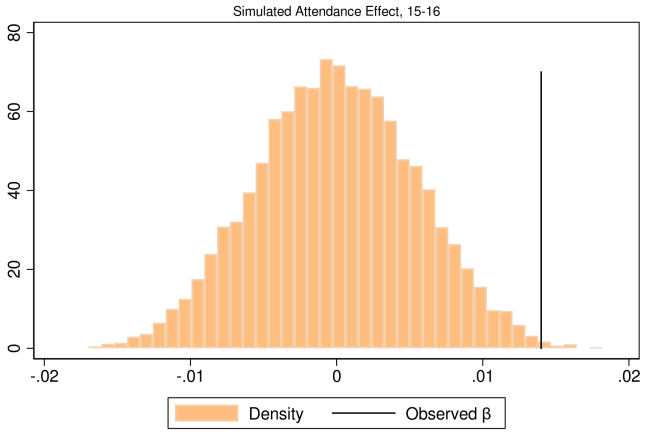
N = 10000 Replications. **P-value** (two-tailed) = 0.455



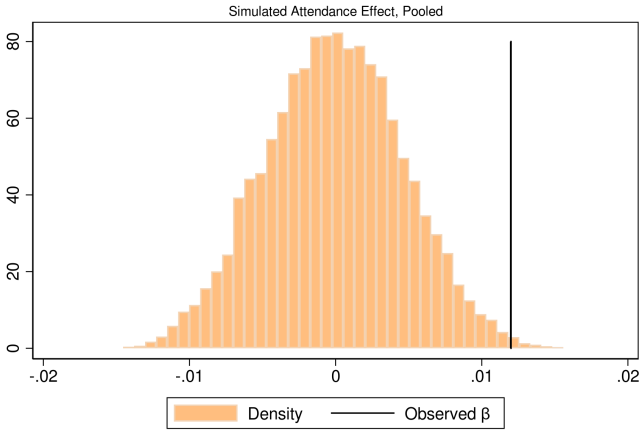
N = 10000 Replications. **P-value** (two-tailed) = 0.014



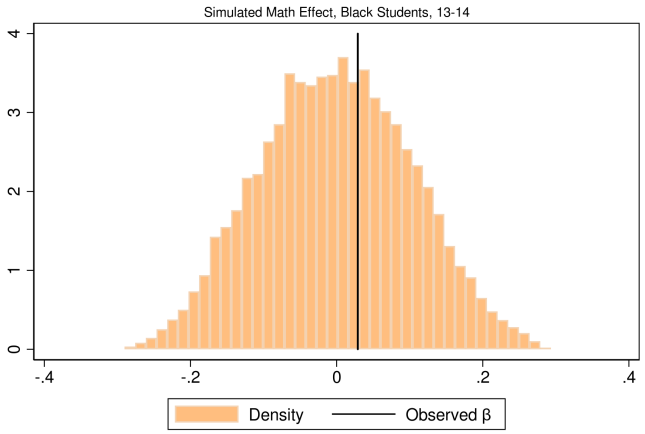
N = 10000 Replications. **P-value** (two-tailed) = 0.078



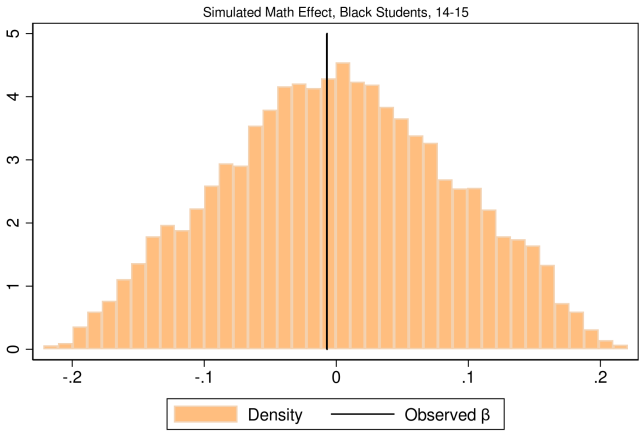
N = 10000 Replications. **P-value** (two-tailed) = 0.006



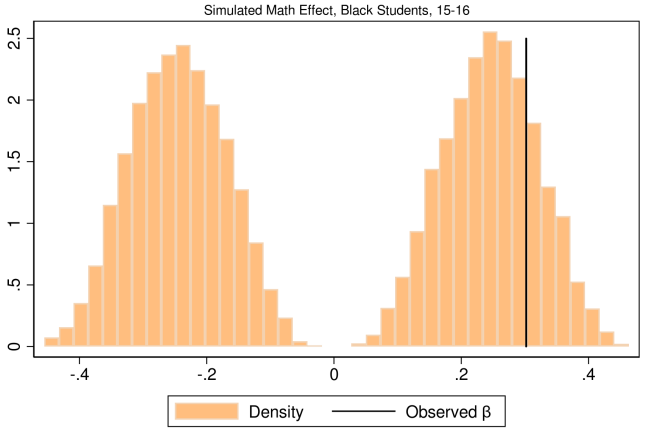
N = 10000 Replications. **P-value** (two-tailed) = 0.007



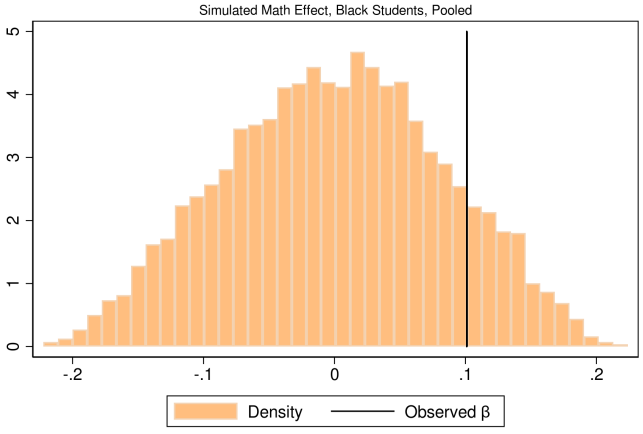
N = 10000 Replications. **P-value** (two-tailed) = 0.795



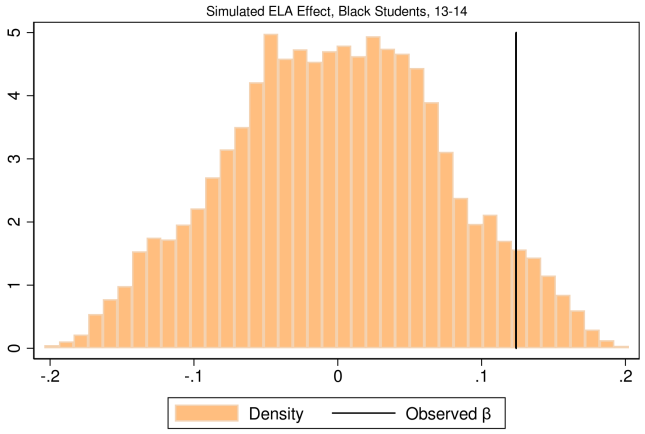
N = 10000 Replications. **P-value** (two-tailed) = 0.938



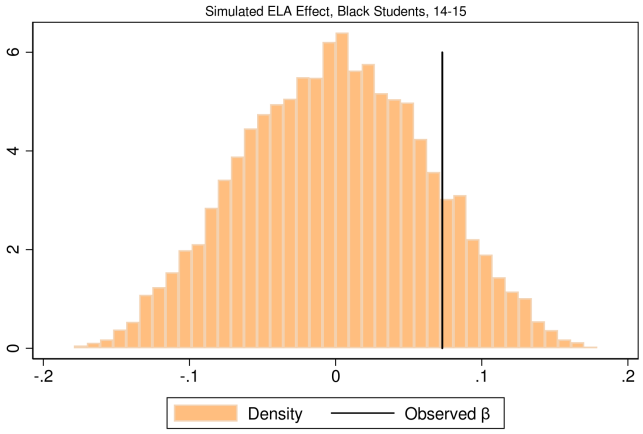
N = 10000 Replications. **P-value** (two-tailed) = 0.240



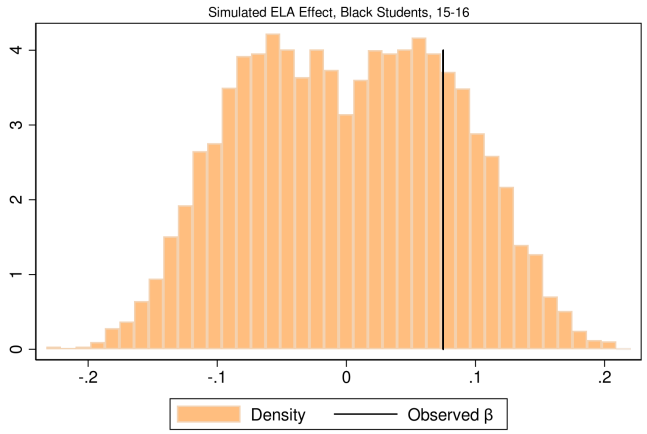
N = 10000 Replications. **P-value** (two-tailed) = 0.254



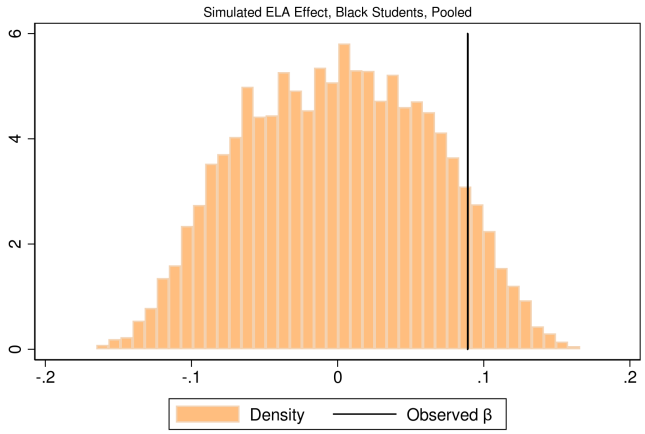
N = 10000 Replications. **P-value** (two-tailed) = 0.115



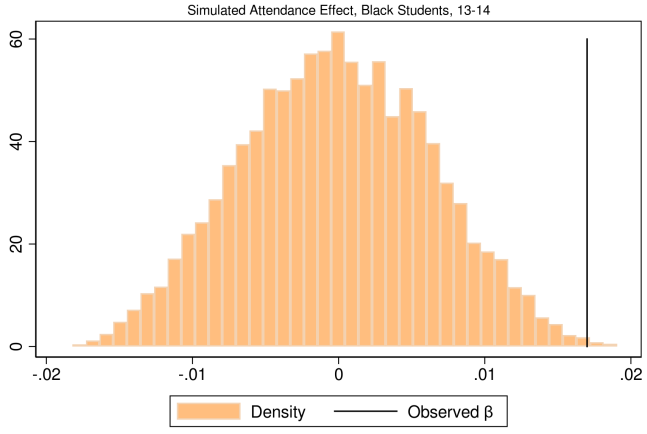
N = 10000 Replications. **P-value** (two-tailed) = 0.265



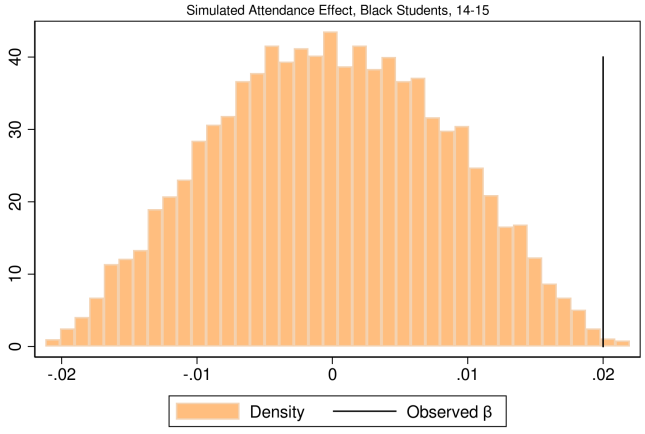
N = 10000 Replications. **P-value** (two-tailed) = 0.419



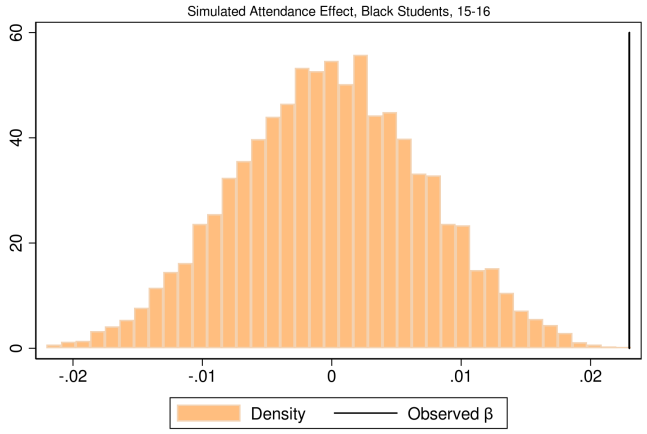
N = 10000 Replications. **P-value** (two-tailed) = 0.173



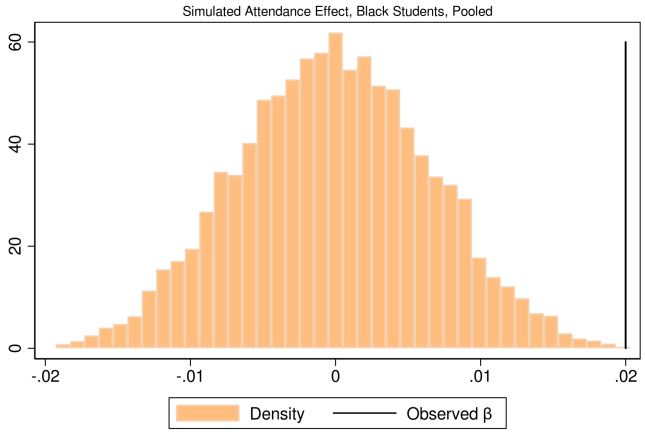
N = 10000 Replications. **P-value** (two-tailed) = 0.002



N = 10000 Replications. **P-value** (two-tailed) = 0.003



N = 10000 Replications. **P-value** (two-tailed) = 0.000



N = 10000 Replications. **P-value** (two-tailed) = 0.000

Notes: These figures plot the distribution of treatment coefficients obtained by conducting OLS regressions on 10000 re-randomized samples. Re-randomization is done by randomly assigning one school to treatment and one school to control within each matched pair 10000 times. The main specifications are re-run using this simulated treatment assigment and the simulated betas are stored. The exact two-sided p-value is the number of simulated betas that are greater than the observed beta in absolute value.