Vertical versus Horizontal Incentives in Education:

Evidence from Randomized Trials

Online Appendix (Not for Publication)

Contents

1 Appendix A: Technical Appendix

Below, we extend our price theory model to explain the empirical results of effort substitution by ability. We turn off the channel for multiple agents and restrict ourselves to a student's decision problem (instead of household) to focus on the key result that effort substitution is higher for lower math ability students.

Imagine a student who is assigned two tasks $i = \{1, 2\}$. The student chooses effort levels e_1 and e_2. Each effort level takes values in $\mathbb{R}_+$ and generates performance measures $\alpha_1 = f(e_1, \theta_1)$ and $\alpha_2 = g(e_2, \theta_2)$. We will refer to θ_i as the "type" of the student on task i. Note that $\theta_i \in \mathbb{R}_+$ and θ_1 is independent of θ_2 i.e. student's type in task 1 is independent of agent's type in task 2.

We assume that the production function of performance measure is twice differentiable, increasing and concave in both arguments.

We further assume that the student has preferences that can be represented by a utility function that is additive in performance measures:

$$\begin{aligned} u(e_1, e_2) &= \alpha_1(e_1, \theta_1) + \alpha_2(e_2, \theta_2) \\ &= f(e_1, \theta_1) + g(e_2, \theta_2) \end{aligned}$$

Finally we assume that choosing effort levels is costly and the student has a fixed amount of "effort income" he can expend[1]:

$$c_1 e_1 + c_2 e_2 \leq B$$

where c_1 and c_2 are marginal costs of increasing effort e_1 and e_2 and B is effort income.[2]

1.1 Analysis

1.1.1 Feature 1: Effort Substitution

Solving the constrained optimization problem for optimal e_1^* and e_2^*, we get

$$\begin{aligned} \frac{f_e(e_1^*, \theta_1)}{g_e(e_2^*, \theta_2)} &= \frac{c_1}{c_2} \\ \implies f_e(e_1^*, \theta_1) &= \frac{c_1}{c_2} g_e\left(\frac{B - c_1 e_1^*}{c_2}, \theta_2\right) \end{aligned} \tag{1}$$

where f_e and g_e are first derivatives of f and g with respect to effort. To see how optimal effort levels change with costs, we differentiate e_1^* and e_2^* with respect to c_1. Taking derivatives we have

$$\begin{aligned} \frac{\partial e_1^*}{\partial c_1} &= \frac{\left[\frac{1}{c_2}\right] g_e(e_2^*, \theta_2) - \frac{e_1^*}{c_2}\frac{c_1}{c_2} g_{ee}(e_2^*, \theta_2)}{f_{ee}(e_1^*, \theta_1) + \left[\frac{c_1}{c_2}\right]^2 g_{ee}(e_2^*, \theta_2)} \\ &< 0 \end{aligned}$$

[1] "Effort income" corresponds to a time constraint. One can think of it as the total effective hours that a student has in a day.

[2] This model can easily accommodate dynamic complementarities if we extend the model to two periods and assume that ability in the second period is a function of ability and effort in the first period. This can explain our post treatment empirical results in Houston.

and

$$\frac{\partial e_2^*}{\partial c_1} = -\left[\frac{1}{c_2}e_1^* + \frac{c_1}{c_2}\frac{\partial e_1^*}{\partial c_1}\right] \tag{2}$$

This is greater than 0 (i.e. optimal effort in task 2 increases with cost of effort in task 1) as long as $\frac{\partial e_1^*}{e_1^*} > -\frac{\partial c_1}{c_1}$ or the percentage change in e_1^* is greater than the percentage change in c_1. This leads us to the following result.

Proposition 1 *An increase in incentives or decrease in cost c_i on task i always leads to an increase in agent effort on task i. It leads to a decrease in agent effort on the other task j as long as $\frac{\partial e_1^*}{e_1^*} > -\frac{\partial c_1}{c_1}$ or the percentage change in e_1^* is greater than the percentage change in c_1.*

We are also interested in how this *effort substitution problem* differs by agent type. Taking second derivatives of equation (2) with respect to ability, we have

$$\frac{\partial^2 e_2^*}{\partial c_1 \partial \theta_1} = -\frac{1}{c_2}\left[\underbrace{\frac{\partial e_1^*}{\partial \theta_1}}_{\text{(a)}} + c_1 \underbrace{\frac{\partial^2 e_1^*}{\partial c_1 \partial \theta_1}}_{\text{(b)}}\right]$$

To sign $\frac{\partial^2 e_2^*}{\partial c_1 \partial \theta_1}$ we need to sign terms (a) and (b) from the equation above.

- To sign (a), we use equation (1) and differentiate with respect to θ_1

$$\frac{\partial e_1^*}{\partial \theta_1} = \frac{-f_{e\theta}(e_1^*, \theta_1)}{f_{ee}(e_1^*, \theta_1) + \frac{c_1^2}{c_2^2} g_{ee}(e_2^*, \theta_2)}$$

  Assuming $f_{e\theta}(e_1^*, \theta_1) > 0$, we get that (a) > 0.

- To sign (b), we differentiate equation (1) twice with respect to c_1 and θ_1

$$\frac{\partial^2 e_1^*}{\partial c_1 \partial \theta_1} = \frac{-\frac{2}{c_2}\frac{c_1}{c_2}\frac{\partial e_1^*}{\partial \theta_1} g_{ee}(e_2^*, \theta_2) - \frac{\partial e_1^*}{\partial c_1} f_{ee\theta}(e_1^*, \theta_1)}{f_{ee}(e_1^*, \theta_1) + \frac{c_1^2}{c_2^2} g_{ee}(e_2^*, \theta_2)}$$

  where we have assumed that third derivatives approach 0 ($f_{eee}(e_1^*, \theta_1) \to 0$, $g_{eee}(e_2^*, \theta_2) \to 0$).

This leads us to the following proposition.

Proposition 2 *Assume $f_{ee\theta}(e_1^*, \theta_1) = 0$. If $\frac{\partial e_1^*}{\partial \theta_1} > c_1 \frac{\partial^2 e_1^*}{\partial c_1 \partial \theta_1}$, then $\frac{\partial^2 e_2^*}{\partial c_1 \partial \theta_1} < 0$.*

We have shown that if $\frac{\partial e_1^*}{\partial \theta_1} > c_1 \frac{\partial^1 e_1^*}{\partial c_1 \partial \theta_1}$, we observe effort substitution by ability when $f_{ee\theta}(e_1^*, \theta_1) = 0$. Of course this is not needed. If $f_{ee\theta}(e_1^*, \theta_1)$ is negative and sufficiently big, then the same holds.

2 Appendix B: Implementation Manual

2.1 Washington DC

Schools

On August 8, 2008, DCPS Chancellor Michelle Rhee and Roland Fryer conducted an introductory meeting with all principals of schools with students in the sixth, seventh or eighth grades in Washington DC. Newly minted Chancellor Rhee made the Capital Gains program one of her signature initiatives. As such, schools and students were expected to participate unless they had a compelling reason not to do so.

After hearing the premise of the program, 28 principals asked for their schools to be included in the randomization process. Fourteen schools were selected as treatment schools, but one declined to participate. The remaining thirteen treatment schools that were selected were provided with school-specific training to help set up the program. After the initial randomization, five more schools that had not originally attended the introductory meeting were also added to the pool. Three schools were selected for treatment: two of these schools chose to participate in the program (the other did not respond to EdLabs within the required 24 hours). In total, there were 34 schools: 17 selected into treatment (two of which did not participate) and 17 selected into control. All analyses are done using the full 17 school treatment sample.

Each principal of a treatment school received sample student consent forms, brochures, and general overviews to share with their staff. Each treatment school was asked to identify a school coordinator to manage the on-site operations of the program. In year 1, members of the Capital Gains team conducted meetings at each of the treatment schools to explain the program to the school's staff during the first two weeks of school and to help them select the school-specific metrics that would be used to assess and reward their students. In year 2, members of the Capital Gains team held meetings with treatment school principals in March and April of 2009 and again during the summer. All metrics were decided upon and in place before school started in August.

Students

In September 2008, students in treatment schools were given Capital Gains "parent packets" to take home with them. These packets included: A letter from Chancellor Rhee with details about the program, a letter from the Capital Gains team with details about the partnership between the program and SunTrust banks, a parental consent/opt-out form, a list of frequently asked questions about the program, an overview of school-specific metrics, and a program calendar with details about pay periods and payment dates. For the second year of the program, students were once again given Capital Gains "parent packets" that included the following, in addition to those items listed above: A letter from SunTrust banks about student accounts and proper behavior inside bank branches, and a customer information sheet from SunTrust banks, needed for students to establish savings accounts.

The school district determined that students in grades 6-8 in each year of treatment at selected treatment schools would be automatically enrolled in the program unless a parent consent form was returned indicating that the parent did not want their student to participate. In year 1, 9 students out of 3,269 were opted out by their parents. In year 2, 8 students out of 3,186 were opted out by their parents

Performance Metrics and Incentive Structure

In year 1, each school selected 3 metrics, along with attendance and behavior, which were used to evaluate students. The most popular metrics included homework completion, grades on tests, and wearing a proper uniform. Students could earn up to ten points for each of the five metrics, and each point was worth $2. Teachers kept track of the students' performance for a two week period and rewards were distributed in the week following the close of the previous period. There were a total of 15 two week periods in the first year.

In year 2, school-specific metrics became more standardized, and a third mandatory metric was added: short-cycle assessment. Once during every two-week pay period, teachers assessed students in a rotation of subjects, typically by pulling ten questions from the DC Benchmark Assessment System database. In year 2, the point system was also deemed an unnecessary step and was discarded. For the first four pay periods of year 2, each metric was worth a maximum of $20; for each subsequent pay period, academic metrics such as short-cycle assessment and grades are worth $35 while non-academic metrics such as attendance, behavior, and uniform are worth $10. The metrics used by each school in each year are listed in Appendix Table 1.

Payment Process

Preparation and set-up: Student rewards were distributed via direct deposit into savings accounts or by check. Deposits were heavily promoted by schools as the safest distribution method and as a means of encouraging fiscal responsibility and increasing familiarity with banking. In order to set up and deposit funds, a partnership was formed with SunTrust to create and manage student savings accounts that were interest-earning and child-owned (child is sole custodian).

SunTrust organized "Bank Days" at each of the participating schools at the start of the program. Representatives from the bank visited the schools and signed up students for accounts during their lunch and free periods. All students were required to have a social security number and picture ID before setting up an account. Social security numbers were verified by the Capital Gains project managers, who also attended Bank Days. After establishing their accounts, students signed forms authorizing EdLabs to make direct deposits over the course of the year.

Students and families who could not (no social security number) or would not (unwilling to provide personal identification) open saving accounts were paid by check. EdLabs contracted with Netchex, a check processing vendor, to process check payments.

Payment logistics: In the first year, teachers were responsible for filling out hard copy spreadsheets every two weeks. The sheets allowed teachers to record individual student performance on each of the metrics for the two-week reward period. The spreadsheets were shipped to a scanning company which scanned to spreadsheets and sent the images to a data entry company. The data entry company entered all student performance data into electronic spreadsheets that EdLabs project managers accessed via a secure (File Transfer Protocol) site. Once the sheets were downloaded by EdLabs, payment amounts were calculated and and audited for accuracy.

In the second year, the DCPS Office of the Chief Technology Officer created an interface that allowed teachers to enter student performance data directly into a database, which was accessible by EdLabs project managers for download, payment calculation, and audit.

Once student payments were calculated and audited, a "pay list" was sent to a payroll vendor. The vendor then accessed a Harvard-owned bank account set up specifically for processing student payment transactions to initiate direct deposits (for those students who signed up for a savings account) and create checks for the remaining students. Those checks were delivered to DCPS project management staff for distribution to school coordinators, who then handed them out to

students. In year 1, spreadsheets were collected from teachers on Friday at the end of a two week pay period and checks were delivered the following Thursday. In year 2, teachers were required to enter information into the database by Saturday evening and payments were delivered the following Wednesday.

Program Support

Throughout the program, targeted strategies were employed to increase participation and awareness and to ensure smooth implementation in all schools.

Student Support:

- Certificates: Certificates were sent to each participating student displaying the amount of money earned based on their performance on each of his or her school's metrics. Certificates both described the student's behavior (e.g. "You were late to class 6 times this pay period") as well as reported the amount earned for each metric.

- Knowledge Quizzes: To gauge students' understanding of the basic elements of the Capital Gains program, a short quiz was administered to participating students in the fall and spring of the first year of treatment. In the second year of treatment, students were given quizzes during mandatory financial literacy sessions throughout the school year. A final quiz was given to students during the spring of 2010.

- Check Cashing Letters: Letters were provided to all students with instructions of free check-cashing options.

- Student Survey: At the end of each year of treatment, students were surveyed about their attitude, effort, and motivation in school. The questions were not specific to the programmatic structure of Capital Gains. Student responses to the surveys were included in analysis, detailed descriptions of the variables used from the survey can be found in Appendix C.

School Support:

- Parents' Nights: During the first year of Capital Gains, community forums (or "parents' nights") were held to inform parents of the details of the program, but turnout was low. In the second year, the program manager held information sessions during Back-to-School Night at selected schools.

- Assemblies: Schools held school assemblies and/or pep rallies to further introduce the program. School administrators and coordinators used these forums to generate excitement about the program, to go over details about earning money and getting paid, and answer any questions students might have

- Materials: Each school also hung posters throughout the building to promote the program and to explain the school-specific performance metrics.

- School Communication: Capital Gains project managers contacted all coordinators regularly to confirm that rewards were being distributed in a timely manner, and contacted the principal via e-mail or phone to provide updates on program operations or to address potential concerns.

- Coordinator Reports and Graphs: For each pay period, EdLabs sent the school coordinator an overall report that presented data on each of their student's performance (i.e. scores on each metric, consent status, bank account status, and reward history). Coordinators also received lists of the top ten earners in each grade for a given pay period as well as a list of the top ten students with the largest increase in rewards from the last period. Additionally in year 1, schools were provided with graphs that showed how each grade level scored on each of the metrics so they could compare performance across grade levels. Some schools requested that these graphs compare classrooms instead of grades. Halfway through the program and at the end of the year, schools were provided with graphs that showed their performance across periods on each metric so they could see how student performance was changing over time.

- Dashboards: In the second year of treatment, EdLabs project managers created a dashboard to help schools monitor their students' progress. Dashboards were sent to coordinators for distribution to principals and teachers at the end of each pay period. They reported school- and grade-level averages as well as the top earning and most improved earners at each school.

- School Stipends: Each school received a stipend to help offset the additional work the program created for its staff. The stipend amounts were based on the number of students participating in the program, with small schools receiving around $1000-$3000 and the largest school receiving around $20,000. The principal decided whether the funds were to be given to the coordinator or split among the coordinator and other staff members.

- Implementation Reviews: In January of year 1, Capital Gains project managers invited all coordinators, principals, and other staff members to complete an online survey as part of an effort to further understand the effects of the program. The survey results contain valuable insights and feedback from schools on program implementation and impact. Project Managers also visited each of the schools at the end of the first year to discuss possible improvements for the second year.

2.2 Houston

Schools

We identified 71 low-performing elementary schools in the district (based upon the average 5th grade scores on the Texas Assessment of Knowledge and Skills (TAKS)) that could benefit from inclusion in the Math Stars incentive program. On Thursday, September 2, 2010, HISD leadership held an introductory meeting with principals and math teachers from these low-performing elementary schools. After presenting an overview of the research design we invited them to commit to participate by signing a pledge to implement the Math Stars program with fidelity to the research design.

Schools had five days to consider their commitment to the program (within a day, however, over two-thirds of the schools invited had already indicated their commitment and interest by signing a School Commitment Letter.) By Tuesday, September 7, 60 schools had elected to participate in the random selection process. We selected the 50 schools with the smallest enrollment to participate in the experiment in order to minimize costs.

Students

HISD decided that students and parents at selected schools would be automatically enrolled in the program. Parents could choose not to participate and return a signed opt-out form at any point during the school year. HISD further decided that students and parents were required to participate jointly: students could not participate without their parents and vice versa.

Software and Incentive Structure

The Accelerated Math (AM) platform creates math assignments tailored to each student's ability level, enabling students to take brief online assessments to gauge achievement in mathematics. For 5th grade, math objectives fall into the following subject areas: Number Sense and Operations; Algebra; Geometry and Measurement; and Data Analysis, Statistics, and Probability.

Students began the program year by taking an initial diagnostic assessment to measure mastery of math concepts, after which AM created customized practice assignments that focused specifically on areas of weakness. Teachers assigned these custom assignments and students were then able to print the assignments and take them home to work on (with or without their parents). Each assignment had six questions, and students needed to answer at least five questions correctly to receive credit. Students scanned their completed assignments into AM, and the assignments were graded electronically. Teachers then administered an AM test that served as the basis for potential rewards: students were given credit for official mastery by answering at least four out of five questions correctly.

Students: Students earned $2 for every objective mastered. Students who reached the 200 objectives threshold were declared "Math Stars" and received a $100 completion bonus and special certificate. Additional monetary incentives were introduced during the program: during the sixth pay period (mid-February to mid-March) students received $4 for every objective mastered; during the final week of the eighth pay period (the first week of May), students received $6 for every objective mastered.

Parents: Parents of children at treatment schools earned up to $160 for attending eight parent-teacher review sessions ($20/session) in which teachers presented student progress using Accelerated Math Progress Monitoring dashboards. Parents and teachers were both required to sign the student progress dashboards and submit them to their school's Math Stars coordinator in order to receive credit. Additionally, parents earned $2 for their child's mastery of each AM curriculum objective, as long as they attended at least one conference with their child's teacher (these were regularly scheduled conferences as per previous years and were also held as usual in control schools). This requirement also applied retroactively: if a parent first attended a conference during the final pay period, the parent would receive a lump sum of $2 for each objective mastered by their child to date. Parents were not instructed on how to help their children complete math worksheets.

Teachers: Fifth grade math teachers at treatment schools received $6 for each academic conference held with a parent in addition to being eligible for monetary bonuses through the HISD ASPIRE program, which rewards teachers and principals for improved student achievement. Each treatment school also appointed a Math Stars coordinator responsible for collecting parent/teacher conference verification forms and printing and distributing student reward certificates, among other duties. Each coordinator received a stipend of $500, but this amount was not tied to performance.

Principals: Principals at treatment schools were eligible for monetary bonuses through the HISD ASPIRE program, which rewards teachers and principals for improved student achievement.

Training and Program Launch

Once schools were selected, the Accelerated Math program was ordered for treatment and control schools, as well as computers and scanners for each school (depending on the number of students and classrooms). AM was installed in treatment schools on September 10 and control schools on September 20. HISD also hired a district-based program manager who was trained in using AM as well as a technology support staff member.

On September 10, a welcome packet in English and Spanish was sent home with students. The packet included a detailed description of the program, a program calendar, answers to frequently asked questions, and an opt-out form. Parents who decided they did not want their student(s) to participate in the incentive component of the Math Stars program were able to return a signed opt-out form at any point during the school year; however, students were not able to opt out of using the Accelerated Math platform.

Meanwhile, treatment schools identified in-school coordinators within one day of being randomly selected; coordinators' primary duties included collecting parent-teacher conference sheets and distributing checks and reward certificates to students on pay day. To effectively train participating schools' staff to use the Accelerated Math program, Renaissance Learning staff conducted teacher and coordinator training in treatment schools the week beginning September 13 (teachers in control schools were trained from September 28-29.)

Teacher training consisted of coaching teachers in how to use the Accelerated Math platform to provide practice and assessment opportunities for students at different skill levels. To ensure differentiated instruction, students were able to test within multiple grade levels of objectives. Therefore, a library or bank of Accelerated Math objectives, practice questions, and assessments – spanning second through seventh grades – were available from which teachers could pull assignments that students could master. However, starting in February – four full months after the beginning of the program – teachers were restricted from drawing objective assignments from libraries below fourth grade equivalency.

After brief site visits to ensure that experimental schools' technological infrastructures were properly in place, teachers were re-trained in how to use Star Math (a companion program to the Accelerated Math platform that was already in place in the HISD schools), which allows classroom teachers to administer a customized diagnostic test to students to assess skill levels within certain grade-level objectives. Therefore, to determine the grade level at which each student should begin their mastery of objectives, teachers began administering student diagnostic assessments the week beginning Monday, September 20. Within two days, 92 percent of students in treatment schools had taken the diagnostic assessment.

Payment Process

Preparation and Set-up: At the conclusion of each pay period, the district-based program manager would begin processing student and parent payments along two fronts: first, extracting student performance data from the Accelerated Math platform, removing students who opted out, and calculating student rewards ($2/objective mastered); second, collecting parent-teacher conference dashboards from school coordinators and inputing parent attendance figures. These two data points were consolidated in a pay file and organized by school.

After all parent conference data was collected and inputed, the pay file was sent to EdLabs to complete the payment algorithm and conduct a few basic audits. The pay file was then sent back to the district program manager, who reformatted and finalized the file for the HISD finance office, who uploaded payment information to JP Morgan Chase. Checks were printed, bundled by school,

and delivered to each school.

EdLabs also used the pay file to create reward certificates for every student receiving a payment. The certificate detailed how many math objectives the student mastered during the last period, the cumulative total, and the current financial earnings. When students passed the 200 objective threshold, they received a special certificate in addition to their $100 bonus.

Payment Logistics: School coordinators received student and parent checks and student certificates one day prior to pay day. Each school planned pay day differently, but there was striking uniformity: typically a small assembly was held in the cafeteria during which checks and certificates were distributed and students were recognized for their achievements. Parents were often in attendance as well to acknowledge their children and receive their checks.

Bonus Rounds

The first several pay periods of Math Stars yielded high rates of participation among both students (i.e. percentage of students mastering at least one objective and receiving payment) and parents (i.e. percentage of parents attending a conference with their student's teacher). As a result of smooth implementation and general enthusiasm about the program among students and staff members, HISD and EdLabs introduced two bonus rounds: during the entire sixth pay period, (February 14 through March 11), students received $4 (rather than the usual $2) for each objective mastered. During the final week of the eighth pay period (May 2 through May 5), students received $6 for each objective mastered. These changes were communicated to students primarily through posters hung throughout the school and flyers sent home in weekly folders.

There were two primary objectives in introducing these bonus rounds: first, the additional incentive was meant to strengthen students' preparation for end-of-year testing. The first ($4) bonus round took place just prior to the Texas Assessment of Knowledge and Skills (TAKS), while the second ($6) bonus round took place prior to the Stanford 10. Second, a sub-experiment was being conducted to estimate a demand curve for math objectives; i.e. asking whether a student will devote more effort to mastering math objectives relative to the increase in the reward.

Site Visits

In an effort to gather extensive qualitative data on the implementation of HISD's Math Stars program, EdLabs conducted brief site visits to all 25 treatment schools.

EdLabs observed classrooms, interviewed students, teachers, and school leaders, and developed, with extensive help from HISD program personnel, a site visit rubric. In addition to providing a comprehensive collection of qualitative school-level data to use in the evaluation of the Math Stars program (i.e. correlating school-level performance with observed implementation indicators), the site visits also supplied the district-based program manager with additional best practices to share with other schools during the last few pay periods of the program.

3 Appendix C: Variable Construction

3.1 Washington DC

Demographic Variables
Demographic variables that should not vary over time (rage, gender) were pulled from the following files (in order of precedence): 2008-09 DCPS enrollment file, 2008-09 DCCAS file, 2007-08 DCCPS enrollment file, 2006-07 DCPS enrollment file. Demographic variables that may vary from year to year (free lunch eligibility, LEP status, and special education indicators) were only pulled from the 2008-09 DCPS enrollment file and the 2008-09 DCCAS file.

- *Race/Ethnicity*: We code the race variables such that the five categories – white, black, Hispanic, Asian and other – are collectively exhaustive and mutually exclusive. Hispanic ethnicity is an absorbing state. Hence "white" implies non-Hispanic white, "black" non-Hispanic black, and so on.
- *Gender*: Gender was coded as male, female, or missing.
- *Free Lunch*: A student was considered free lunch eligible if he was coded as "Free" or "Reduced" in the DC enrollment file, and considered non-free lunch eligible if he was coded as "Pay All" in the enrollment file. All blanks were coded as missing.
- *Limited English Proficient*: A student was considered LEP if he had a status of "English Language Learner (ELL)", "ELL Level 1" through "ELL Level 4," or "ELLm (Return to ESL)" in the DC enrollment file or a "Y" in the ELL variable in the 2008-09 DCCAS file. All blanks in the enrollment file were coded as non-LEP since those students had a value of "N" in the 2008-09 DCCAS file.
- *Special Education*: A student was considered enrolled in special education if he had a status of "Special Education" or "Referred" in the DC enrollment file or a "Y" in the special education variable in the 2008-09 DCCAS file.
- *School-Level variables*: School level variables were constructed for each school based on the population of students assigned to that school by the following rules: students enrolled in school before October 1st, 2008 were assigned to the school they attended first. Students enrolled after October 1st were assigned to the school that they attended for longest. The first school attended was determined by looking at the pattern of attendance across schools. School-level demographic variables included in the analysis included percent of the student population that is black, Hispanic, and eligible for free lunch as defined above.

State Test Scores
The state test is the DC Comprehensive Assessment System (DCCAS), which is administered to third through eighth grade and tenth grade students every spring. Baseline test scores were pulled from the DCCAS 2006-07 and 2007-08 files. Outcome test scores were pulled from the DCCAS 2008-09 and 2009-10 files. Outcome test scores were standardized in each year to have a mean of zero and standard deviation of one in each grade by subject over the DC school district. Proficiency

levels were also taken directly from these files - each student is marked as achieving at the "below basic," "basic," "proficient," or "advanced" level in each subject.

Grades
Grades were pulled from files containing the transcripts for all students in DC public schools from 2008-09. Letter grades were converted to a standard 4.0 scale. Each student's grades from each semester were averaged to yield a GPA for the year.

Attendance Rates
Each student's attendance rate in each year 2007-08 through 2009-10 was calculated as the total number of days present in any DC public school divided by the total number of days enrolled in any DC public school, according to the DCPS attendance file.

Behavioral Offenses
The number of behavioral incidents for each student was pulled from an administrative file listing all behavioral incidents in each year. Students not listed in this file were assumed to have zero behavioral incidents. Students were flagged as committing a behavioral offense if they showed up in the file. School-level behavior variables were created by summing the total number of behavioral incidents per school, and was assumed to be zero if a school had no students who attended that school showing up in the behavioral incident file.

Survey Responses
Survey variables were constructed from a file containing paper survey responses that were manually entered into a computer. If a student had more than one response to a particular question and those responses conflicted, those responses were dropped. All question responses were converted to a numerical scale so that higher numbers indicated more effort. Individual variables were then converted to binary indicators so that they had a value one if a student's response was in the top half of of all responses and zero otherwise.

- *Completing Homework*: Students were asked "About how much of your assigned homework to you usually complete, either during school hours or outside of school?" and could respond "All," "Three quarters," "Half," "One quarter," or "Almost none." We code "All" or "Three quarters" as one and the rest as zero.

- *Arrive On Time*: Students were asked how much they agree with the following statement: "I don't really care whether I arrive on time for class." They could respond "Totally untrue," "Mostly untrue," "Somewhat true," "Mostly true," or "Totally true." We code "Totally untrue" as one and the rest as zero.

- *Behavior Not a Problem*: Students were asked how much they agree with the following statement: "My behavior is a problem for the teachers in my classes." They could respond "Totally untrue," "Mostly untrue," "Somewhat true," "Mostly true," or "Totally true." We code "Totally untrue" as one and the rest as zero.

- *Work Hard in School*: Students were prompted "I work very hard on my schoolwork." and could respond "Not at all true," "Not very true," "Sort of true," or "Very true." We code "Sort of true" or "Very true" as one and the rest as zero.

- *Push Self in School*: Students were asked how much they agree with the following statement: "I have pushed myself hard to completely understand my lessons in school." They could respond "Totally untrue," "Mostly untrue," "Somewhat true," "Mostly true," or "Totally true." We code "Totally true" or "Mostly true" as one and the rest as zero.

- *Intrinsic Motivation Inventory*: We disseminated part of the Intrinsic Motivation Inventory, developed by Ryan (1982), to students in our experimental group. The instrument contains many modules, but we limited our questions to those in the interest/enjoyment subscale in our surveys as it is considered the self-reported measure of intrinsic motivation. The interest/enjoyment subscale consists of seven statements on the survey: (1) I enjoy doing schoolwork very much; (2) doing schoolwork is fun; (3) I thought this was a boring activity; (4) doing schoolwork does not hold my attention at all; (5) I would describe doing schoolwork as very interesting; (6) I think doing schoolwork is quite enjoyable; and (7) while I am doing schoolwork, I think about how much I enjoyed it. Respondents are asked how much they agree with each of the above statements on a seven-point Likert scale ranging from "not at all true" to "very true." To get an overall intrinsic motivation score, one adds up the values on each statement (reversing the sign on statements (3) and (4)). Only students with valid responses on each statement are included in our analysis of the overall score, as non-response may be confused with low intrinsic-motivation. When reporting results, we report effects on scores normalized to have a mean of zero and a standard deviation of one.

Summary Index Measures All indices were calculated only for students who had non-missing values of *all* of the components of the index.

- *Incentivized Outcomes Index*: An administrative measure of GPA was coded to be an indicator variable that was one if the student's GPA was above median and zero otherwise. Survey measures of whether a student completes his or her homework, is on time to class, and is *not* a behavioral problem for teachers were similarly constructed. The administrative measure of behavior was a one if the student *did not* committ any behavioral offense and zero otherwise. The index is the sum of those five measures standardized over the sample to have a mean of zero and a standard deviation of one.

- *Academic Achievement Index*: State test scores in math and reading were standardized in each grade over all students in DC. The index is the unweighted average of those two standardized measures.

- *Behavior Index*: An administrative measure of attendance was coded as an indicator variable that was one if the student attended more days of school than the median student and zero otherwise. The survey measures of working hard in school, pushing oneself in school, and intrinsic motivation were coded as indicator variables that were one if the student scored above the median and zero otherwise. The index is the sum of those four measures standardized over the sample to have a mean of zero and a standard deviation of one.

Assignment to Treatment
The DC 2008-09 attendance file was used to determine the first school that each student attended. A student was assigned to the control group of the student's first school was a control school, if the

student was in 6th, 7th, or 8th grade, and if the student was present in that school before October 1, 2008. The treatment group was defined similarly.

3.2 Houston

Demographic Variables
Demographic variables that should not vary over time (race, gender) were pulled from the 2010-11 HISD enrollment file and were filled in with values from the 2010-11 attendance file if missing in the enrollment file. They were filled in with previous years' enrollment files (through 2007-08) if missing, with the most recent data given precedence. Demographic variables that may vary from year to year (free lunch eligibility, economically disadvantaged status, LEP status, a special education indicator and a gifted and talented indicator) were only pulled from the 2010-11 enrollment and attendance files (with precedence to the enrollment file).

- *Race/Ethnicity*: We code the race variables such that the five categories – white, black, Hispanic, Asian and other – are collectively exhaustive and mutually exclusive. Hispanic ethnicity is an absorbing state. Hence "white" implies non-Hispanic white, "black" non-Hispanic black, and so on.
- *Gender*: Gender was coded as male, female, or missing.
- *Free Lunch and Economically Disadvantaged*: A student is considered free lunch eligible if he was coded as "Free" or "Reduced" in the HISD administrative data, and considered non-free lunch eligible if he is coded as "Economically Disadvantaged but not FRL" or "Ineligible." A student is considered economically disadvantaged if he is eligible for free or reduced price lunch or is flagged as economically disadvantaged without free and reduced lunch.
- *Limited English Proficient and Special Education*: These statuses are determined by HISD Special Education Services and the HISD Language Proficiency Assessment Committee, respectively; they enter into our regressions as dummy variables. We do not consider students who have recently transitioned out of LEP status to be of limited English proficiency.
- *Gifted and Talented*: HISD offers two Gifted and Talented initiatives: Vanguard Magnet, which allows advanced students to attend schools with peers of similar ability, and Vanguard Neighborhood, which provides programming for gifted students in their local school. We consider a student gifted if he or she is involved in either of these programs.

State Test Scores The state test is the Texas Assessment of Knowledge and Skills (TAKS) in years previous to 2011-12 and the State of Texas Assessment of Academic Readiness (STAAR) in all following years. The Stanford 10 is a low-stakes test administered in every year. Baseline test scores were pulled from the TAKS 2008-09 and 2009-10 files. Outcome test scores were pulled from the TAKS 2010-11 and STAAR 2012-13 files. For ease of interpretation, we normalize raw scores to have a mean of zero and a standard deviation of one within grades, subjects, and years. Proficiency levels were also taken directly from these files – each student is marked as meeting either the minimum proficiency level determined by the state of Texas or as achieving a commended performance.

Attendance Rates

When calculating the school-level attendance rate, we consider all the presences and absences for students when they are enrolled at each school. Individual attendance rates account for all presences and absences for each particular student, regardless of which school in HISD the student was enrolled in when the absence occurred.

Behavioral Offenses
The number of behavioral incidents for each student was pulled from an administrative file listing all behavioral incidents in each year. Students not listed in this file were assumed to have zero behavioral incidents. Students were flagged as committing a behavioral offense if they showed up in the file.

Survey Responses

- *Student Survey*: First, students were asked "Did your parents check your homework this year more than last year?" We code responses of "more this year" as one and responses of either "more last year" or "about the same" as zero. Second, students were asked "What subject do you like better, math or reading?" We code responses of "math" as one and "reading" as zero.

- *Parent Survey*: Parents were asked "Do you ask your 5th grade student more often about how he/she is doing in math class or reading class?" We code responses of "math class" as one and responses of either "reading class" or "no difference" as zero.

- *Intrinsic Motivation Inventory*: We disseminated part of the Intrinsic Motivation Inventory, developed by Ryan (1982), to students in our experimental group. The instrument contains many modules, but we limited our questions to those in the interest/enjoyment subscale in our surveys as it is considered the self-reported measure of intrinsic motivation. The interest/enjoyment subscale consists of seven statements on the survey: (1) I enjoy doing schoolwork very much; (2) doing schoolwork is fun; (3) I thought this was a boring activity; (4) doing schoolwork does not hold my attention at all; (5) I would describe doing schoolwork as very interesting; (6) I think doing schoolwork is quite enjoyable; and (7) while I am doing schoolwork, I think about how much I enjoyed it. Respondents are asked how much they agree with each of the above statements on a seven-point Likert scale ranging from "not at all true" to "very true." To get an overall intrinsic motivation score, one adds up the values on each statement (reversing the sign on statements (3) and (4)). Only students with valid responses on each statement are included in our analysis of the overall score, as non-response may be confused with low intrinsic-motivation. When reporting results, we report effects on scores normalized to have a mean of zero and a standard deviation of one.

Summary Index Measures All indices were calculated only for students who had non-missing values of *all* of the components of the index.

- *Incentivized Outcomes Index*: Individual measures of the number of parent conferences attended and the number of objectives mastered were standardized over the entire sample to have mean zero and standard deviation one. The index is the unweighted average of those two standardized measures.

- *Incentivized Achievement Index*: TAKS/STAAR and Stanford 10 test scores in math were standardized over all 5th graders in HISD. The index is the unweighted average of those two standardized measures.

- *Non-Incentivized Achievement Index*: TAKS/STAAR and Stanford 10 test scores in reading and Stanford 10 test scores in science and social studies were standardized over all 5th graders in HISD. The index is the unweighted average of those four standardized measures.

- *Survey Outcomes Index*: Individual measures of whether or not parents check homework more, whether a student prefers math to reading, and whether a parent asks about math more than reading were all coded as binary variables as described above. The index is the sum of those three measures standardized over the sample to have a mean of zero and a standard deviation of one.

- *Behavior Index*: An administrative measure of attendance was coded as an indicator variable that was one if the student attended more days of school than the median student and zero otherwise. The measure of behavioral offenses was a one if a student *did not* commit any behavioral offense and zero otherwise. The survey measure of intrinsic motivation was coded as an indicator variable that was one if the student scored above the median and zero otherwise. The index is the sum of those three measures standardized over the sample to have a mean of zero and a standard deviation of one.

Assignment to Treatment

Due to a limitation in the attendance data provided by HISD, we are unable to determine the dates on which students enrolled in their current schools. AM registration files provide a "snapshot" file that records each students' enrolled school as of October 8. We assign students in one of the 25 treatment schools on October 8, 2010 to our treatment group (the control group is defined similarly). Our results are not sensitive to changing the treatment assignment based on the first school attended during the 2010-11 school year.

Teacher Value-Added

HISD officials provided us with 2009-10 value-added data for 3,883 middle and elementary school teachers. In Table 2, we present calculations based on the district-calculated Cumulative Gain Indices. We normalize these indices such that the average teacher in each subject has a score of zero and the sample standard deviation is one. These scores are then averaged within each school.

4 Appendix D: Cost-Benefit Analysis

We calculate back-of-the-envelope Internal Rates of Return (IRRs) based on the expected income benefits associated with increased student achievement. We calculate these for DC and Houston as well as 14 other education interventions to place the returns of our experiments within the context of other interventions. The interventions covered include experiments on charter schools, teacher incentives, class size reductions, and curriculum changes.

We follow Krueger (2003) to calculate the IRRs. Let E_t denote an individual's real annual earnings at time t and β denote the percentage increase in earnings resulting from a one standard deviation increase in test scores. The IRR is the discount rate r^* that sets costs equal to the discounted stream of future benefits:

$$C_0 = \sum_{t=T_0}^{T_N} E_t * \beta(\tau_m + \tau_r) * (\tfrac{1+g}{1+r})^t$$

where T_0 is the time period in which the individual turns 18 and enters the labor market, T_N is the time period in which the individual turns 65 and retires, τ_m and τ_r denote the treatment effects for math and reading, respectively, and g is the annual rate of real wage growth.

According to the literature on the relationship between test score gains and lifetime earnings, β lies somewhere between 8 percent and 12 percent (Krueger, 2003). Krueger also notes that real earnings and productivity have historically grown at rates between 1 percent and 2 percent, which are plausible rates for g. For the purpose of this cost benefit analysis, we set $\beta = 0.12$ and g equal to 0.02, and approximate E_t using the Current Population Survey. For each intervention, we calculate cost per student per year for both treatment and control, the age at which intervention starts, treatment effects in math and reading (where applicable) and the year the individual enters the labor market. Below we describe these calculations in greater detail, and present the resulting IRRs in Appendix Table 7.

Financial Incentives

DC: For our experiment in DC, we distributed a total of $4.0 million in incentives across two years of treatment to approximately 3,580 students, and spent about $231,000 on other administrative costs such as program manager salaries. Approximately $40,000 was spent on incentives to collect surveys from treatment and control schools, so this cost has been removed from the overall cost of the program. Because the average number of years spent in intervention was 1.2 years, the incremental cost per student per year of treatment is $981.84. Using an initial age of 13 at the beginning of the intervention, we get an IRR of 31.77%.

Houston: For our experiment in Houston, we distributed a total of $875,382 in incentives across 1,554 treatment students, and spent $367,000 on other administrative costs such as purchasing the AM software. Roughly $46,000 of the administrative costs were spent on incentives to collect surveys in both treatment and control schools, so this cost has been removed to more accurately reflect the true cost of the program. This brings the cost per student per year to $666.53 for treatment students and $99.45 for control students, in 2011 dollars. Using 11 as the age of students at the time of intervention with one year spent in treatment, we get an IRR of 14.68%.

Coschocton Incentive Program: Bettinger (2012) evaluated a pay-for-performance program for students in grades three through six in Coschocton, Ohio from 2004-2007. Eligible students received cash payments for improving achievement in standardized tests for five core subjects: math, reading, writing, science and social studies. He reports a 0.133σ (0.0485) increase in math scores, a 0.01σ (0.0454) increase in reading scores, a 0.23σ (0.041) increase in social studies scores and a 0.048σ (0.039) decrease in science scores. Writing scores are excluded from the analysis because in any given year, different grades took different writing tests that were not comparable. We include the social studies and science effects in the calculation for the Coschocton Incentive Program's IRR because the experiment provided incentives for improving test scores across all subjects. Pooling these effects therefore gives a more comprehensive view of treatment effects. In this experiment, randomization was done at the grade-school level each year of the experiment, yielding a total of 1,615 students in the experimental sample, with 801 students being eligible for treatment overall (source: personal communication). As the program cost $52,000 in incentives and administrative costs across three years, the cost per student per year is approximately $65. Using an average initial age during intervention as 11, we get an IRR of 51.08%.

NYC, Dallas, and Chicago: Fryer (2011) summarizes the results of financial incentives on student achievements in New York, Chicago, and Dallas. In Dallas, second grade students were paid to read books. In New York, students were rewarded for performance on interim assessments. In Chicago, students were paid for classroom grades. The incremental cost per student per year in Dallas was $62.21 which included $13.81 paid on average in incentives and $86,000 in administrative costs. In New York, the incremental cost per student per year was $377.04 for 7th graders and $339.25 for 4th graders, including average incentives paid and administrative costs, but excluding $500 spent per school to collect surveys. The incremental cost per student per year in Chicago was $373.76 which included incentive payments and $85,000 in administrative costs. The weighted average cost for an extra student in the three experiments was $323.47. The estimated treatment effect on math and reading scores is zero when pooled across all three cities. As a result, we are unable calculate an IRR, because the discount rate would have to be a very large negative number to bring the net present value of costs equal to zero.

Teacher Certification

Teach for America: Teach for America is a non-profit organization that recruits recent college graduates to teach for two years in low-income communities. Glazerman et al. (2006) report findings from a national randomized evaluation of the impact of TFA on student outcomes. The experiment involved approximately 100 elementary classrooms, grades 1 through 5, from 17 schools across Baltimore, Chicago, Compton, Houston, New Orleans, and the Mississippi Delta. Students were stratified by grade and school and randomly assigned to either a TFA or non-TFA teacher. Glazerman et al. (2006) report that students assigned to a TFA teacher score about 0.15σ (0.04) higher in math and 0.03σ (0.04) higher in reading than students assigned to non-TFA teachers. In an interview, the national spokesperson for TFA, Takirra Winfield, claims that TFA spent around $16,400 to recruit and select each new teacher, $7,000 to train them, and $14,000 per year on stipends for the two years of the program. Thus, we get a total cost of $51,400 per TFA recruit per year. This study had a total of 44 TFA

teachers teaching 785 students, giving a per student cost of $2,881. Using an average initial age during intervention of 9, we get an IRR of 11.82%.

Early Childhood Intervention

Head Start Impact Study: Head Start is a preschool program funded by federal grants, and is designed to serve 3- to 5-year-old children living at or below the federal poverty line. Puma et al. (2010) evaluate Head Start by studying randomized admission into the program. They investigate the impact on two different cohorts, a 3-year-old cohort, which is exposed to the program for two years, and a 4-year-old cohort which is exposed to the program for just one year. Puma et al. (2010) report that winning a lottery to attend Head Start resulted in an increase of 0.135σ (0.071) in math test scores and 0.188σ (0.064) in reading test scores. According to an National Institute for Early Education Research report, the average spending per child in Head Start was $9,198 in 2010. However, this is not necessarily the marginal cost of Head Start because as Puma et al. note, approximately 60 percent of the control group children in their study participated in child care or other early education programs. Based on the same report, average spending per child on other pre-K programs was $4,831. Using these cost calculations and an average initial age of 4, we get an IRR of 9.26%.

Class Size

Tennessee STAR experiment: Project STAR was an experiment carried out in 79 Tennessee schools from 1985 to 1989 where 11,600 students in kindergarten to third grade were randomly assigned to small classes (13-17 students), regular classes (22- 25 students), or regular classes with a full-time aide. Krueger (1999) estimates the impact of reduced class size on test scores using a student's initial assignment to one of the three groups. He reports that students in smaller classes had a 0.133σ (0.033) increase in reading test scores and a 0.107σ (0.033) increase in math test scores, compared to students assigned to a regular class without an aide. In conducting a cost benefit analysis, Krueger (2003) assumes that since, class size reduced from about 22 to about 15 students, funds are allocated to create $7/15 = 47\%$ more classes. Accordingly, the marginal cost per student for each year a student is in a small class is $3,501, or 47% of the nationwide total expenditure per student in 1997-1998. The average number of years spent in a small class was 2.3 years. Using this and an average initial age of 7, we get an IRR of 8.55%.

Charter Schools

Harlem Children's Zone: The Harlem Children's Zone (HCZ) is a 97-block area in central Harlem, New York that combines reform-minded charter schools with a web of community services designed to ensure that the social environment outside of school is positive and supportive for children from birth to college graduation. Dobbie and Fryer (2009) estimate the causal impact of attending the Promise Academy in the HCZ by exploiting the fact that HCZ charter schools are required to select students by lottery when the number of applicants exceeds the number of available slots for admission. In this scenario, the treatment group is composed of students who are lottery winners and the control group consists of students who are lottery losers. The two-stage-least-squares (2SLS) estimates for attending these charter schools during middle school are 0.229σ (0.037) in math scores and a 0.047σ (0.033)

in reading scores. Similarly, the 2SLS estimates for elementary school imply that attending Promise Academy charter schools for one year increases reading scores by 0.114σ (0.095) and math scores by 0.191σ (0.116) relative to the control group. Dobbie and Fryer (2009) state that the New York Department of Education provided every charter school, including the Promise Academy, $12,443 per student in 2008-2009. HCZ estimates add another $4,657 per student for in-school costs and approximately $2,172 per pupil for after-school and "wrap-around" programs. This implies that HCZ spends $19,272 per student per year. Using this number and adjusting for average number of years spent in treatment (1.24 years for middle school and 0.834 years for elementary school), we get an IRR of 10.84% and 11.92% for middle and elementary school respectively.

Apollo: Fryer (2014) examines the impact on student achievement of implementing a bundle of best practices from high-performing charter schools into low-performing, traditional public schools in Houston, Texas. Fryer uses school-level randomized field experiments and quasi-experimental comparisons. Treatment schools implemented the following five practices: increased instructional time; replacement of principals and teachers who failed to adequately increase student achievement; implementation of daily high-dosage mathematics tutoring for fourth graders; use of data-driven curricula; and fostering a culture of high expectations. The intervention was done in 8 elementary schools and 9 middle and high schools. Fryer reported a yearly increase of 0.072σ (0.039) in reading test scores and an increase of 0.184σ (0.06) in math test scores for elementary school students over an average of 1.34 years spent in treatment. For middle and high schools, Fryer reports a yearly decrease of 0.012σ (0.022) in reading scores and an increase of 0.146σ (0.031) in math test scores, over an average of 1.31 years in treatment. The reported costs per student per year were $355 for elementary school students and $1,837 for secondary school students. Using an average initial age of 10 for elementary school and 14 for secondary school, we have an IRR of 36.10% and 19.28% for elementary and secondary schools respectively.

SEED: SEED schools are five-day-a-week urban boarding schools that have an extended school day, provide extensive after-school tutoring, utilize data-driven curricula, and maintain a culture of high expectations. Curto and Fryer (2014) utilize the fact that when a SEED school is oversubscribed, it determines admission via a random lottery. Thus, the treatment group is composed of lottery winners and the control group consists of lottery losers. Curto and Fryer (2014) report that winning the lottery increases math achievement by 0.218σ (0.082) and reading achievement by 0.201σ (0.086). Using data from District of Columbia Public Schools (DCPS) and SEED schools' financial reports, Curto and Fryer report that SEED's cost per student per year in 2008-09 were $39,275. According to the National Center for Education statistics, the total expenditure per student in DCPS was $20,523 for the same year, giving us an incremental cost of attending a SEED school of around $18,752 per student per year. Using an average initial age of 13 and an average 2.33 years of being enrolled in SEED, we get an IRR of 8.66%.

Managed Professional Development

Success for All: Success for All is a school-level elementary school intervention that focuses on improving literacy outcomes for all students in order to improve overall student achievement.

In 2007, it was used in 1,200 schools across the country (Borman et al., 2007). The program is designed to identify and address deficiencies in reading skills at a young age using a variety of instruction strategies, ranging from cooperative learning to data-driven instruction. Borman et al. (2007) use a cluster randomized trial design to evaluate the impacts of the Success for All model on student achievement. Thirty-five schools from eleven states volunteered and were randomly assigned to either the treatment or control group for a 3-year longitudinal study. Control schools implemented Success for All in grades 3-5, while treatment schools implemented Success for All in grades K-2. Comparisons were then made between the treated K-2 students and the untreated K-2 students. Borman et al. report a 0.09σ (0.06) increase in reading test scores. Implementing Success for All would cost schools $75,000 the first year, $35,000 the second year, and $25,000 the third year, for a total of $135,000. For the purpose of this evaluation, all participating schools received Success for All but in different grades. However, for a more realistic cost of implementing this program, we only consider the incremental cost for the treatment schools, which is roughly $746 per student per year, using 18 treatment schools and 1,085 treatment students across the 3 years. Using an initial age of 6, we get an IRR of 13.97%.

Curriculum

Enhanced Reading Opportunities: The US Department of Education initiated the Enhanced Reading Opportunities (ERO) study to evaluate supplemental literacy programs targeted at 9th graders whose reading levels were between two and five years below grade level. As part of the study, two cohorts of ninth grade students from 34 high schools and 10 school districts implemented one of two reading interventions: Reading Apprenticeship Academic Literacy (RAAL) and Xtreme Reading. Students were selected based on being two to five years below grade level on reading comprehension test scores, and were randomly assigned to enroll in an ERO class or not. Experienced English and social studies teachers volunteered to teach the ERO class for two years, and were provided training and technical assistance by the program's developers (Somers et al., 2010). Somers et al. (2010) find an increase of 0.11σ (0.037) in reading test scores and a 0.07σ (0.035) increase in math test scores as a result of the program. The average annual cost per student of implementing the programs was $1,931. Using an initial age of 15, we get an IRR of 22.26%.

Teacher Incentives

Talent Transfer Incentives: Glazerman et al. (2013) use a randomized experiment in 10 districts across the nation to investigate the impact of filling vacancies with high-achieving teachers through the Talent Transfer Initiative (TTI). In each district, the TTI offered teachers with consistently high value-added (ranking in the top 20 percent within their subject and grade) $20,000, paid over two years, to teach at low-achieving schools randomly assigned to treatment. Across the 10 districts included in the study, 165 teacher teams from 114 schools were randomly assigned to treatment or control spanning grades 3 through 8. The initiative began in 2009 with 7 districts (cohort 1) and 3 additional districts were added in 2010 (cohort 2). Each team consisted of focal teachers, who were the teachers that filled the vacancies, and non-focal teachers who constituted the rest of the team. Glazerman et al. report positive impacts on test scores for elementary school students as a result of the TTI. The cumulative

effect of focal teachers in elementary school on cohort 1 is a 0.22σ (0.06) increase in math scores and a 0.25σ (0.05) increase in reading scores, which we divide by two to get the yearly effect. The sample of treatment students in cohort 1 is roughly 2,451, which is half of the sample size reported for grades 3 through 8 with unique student-focal teacher combinations. Glazerman et al. estimate the cost of implementing TTI was $36,382 per team, the majority of which included transfer stipends and retention stipends over the two years. Half of this cost multiplied by 87 teams gives a per student per year cost of $645.70. Using an average initial age of 10, we get an IRR of 28.66%.

High Dosage Tutoring

Experience Corps: This program trains older adults, aged 55 and above, to tutor and mentor elementary school children who are at risk of academic failure. Volunteers receive training focused on literacy and relationship building, as well as a stipend based on number of hours worked. Volunteers work with students one-on-one for about 15 hours a week. Morrow-Howell et al. (2009) use a randomized experiment across 23 schools in Boston, New York City, and Port Arthur, Texas to evaluate the effectiveness of this program. At the beginning of the school year in 2006, all students in need of reading assistance were referred to the Experience Corps program. All referred students were then randomly assigned to the treatment or control group. The EC program tutored 430 students in total, with 451 students in the control group. Morrow-Howell et al. report an average increase of 0.075σ (0.067) on reading test scores. To calculate cost per student per year, we first calculated average cost per tutor. Based on its IRS 990 form, Experience Corps had a total cost of $1,343,936 in 2009, when the program had 2,000 tutors (Morrow-Howell et al., 2009). This gives us a per tutor cost of $671. With 505 tutors in the evaluation and 430 students tutored, we have a per student per year cost of $788. Using an average initial age of 8, we get an IRR of 13.81%.

Appendix Tables and Figures

Appendix Table 1: School Metrics in DC

School	Metric 1	Metric 2	Metric 3	Metric 4	Metric 5
Year 1					
Brightwood Education Campus			Homework	Assessments	Classwork
Browne Education Campus			Homework	Grades	Uniform
Burroughs Education Campus			Homework	Assessments	Writing
Eliot-Hine Middle School			Homework	Grades	Tests
Emery Education Campus			Homework	Tests	Uniform
Hart Middle School			Uniform	Homework	Grades
Jefferson Middle School			Homework	Grades	Out of Class Time
Kelly Miller Middle School			Uniform	Grades	Tests/Quizzes
Langdon Education Campus	**Attendance**	**Behavior**	Uniform	Grades	Book Quizzes
Lincoln Middle School			Homework	Daily Objective	Service
Raymond Elementary School			Homework/Projects	Assessments	Uniform
Shaw Middle School			Math Grades	English Grades	Science/Social Studies Grades
Stuart-Hobson Middle School			Uniform	Homework	Grades
Takoma Education Campus			Uniform	Homework	Tests
Whittier Education Campus			Uniform	Homework	Academics
Year 2					
Brightwood Education Campus				Homework	4 Subject Assessment
Browne Education Campus				Grades in 4 Subjects	Uniform
Burroughs Education Campus				Math and ELA Homework	Science and Social Studies Homework
Columbia Heights Education Campus*				Homework	Grades in 4 Subjects
Eliot-Hine Middle School				Grades in 4 Subjects	Homework
Emery Education Campus				Grades in 4 Subjects	Uniform
Hart Middle School				Grades in 4 Subjects	Uniform
Jefferson Middle School				Grades in Math/ELA	Homework
Kelly Miller Middle School	**Attendance**	**Behavior**	**Short Cycle Assessment**	Grades in 4 Subjects	Uniform
Langdon Education Campus				Grades in Math/Science	Uniform
Raymond Elementary School				Homework	Uniform
Shaw Middle School			(Reading only)	Short Cycle Assessment, Math	Short Cycle Assessment, Sci/Soc
Stuart-Hobson Middle School				Grades in 4 Subjects	Homework
Takoma Education Campus				Homework	Uniform
Whittier Education Campus				Class Preparation	Grades in Math/ELA

*formerly Lincoln Middle School

Note: bolded metrics were determined by EdLabs. Non-bolded metrics were determined by each school.

Appendix Table 2A: Mean Effect Sizes (Two Stage Least Squares Estimates) in DC

		Payment Periods		Attendance	
	ITT	Ever Treated	Periods Treated	Ever Treated	Days Treated
	(1)	(2)	(3)	(4)	(5)
A. Incentivized Outcomes					
Behavioral Offense	-0.032***	-0.037***	-0.029***	-0.036***	-0.029***
	(0.008)	(0.010)	(0.008)	(0.009)	(0.008)
GPA	0.118***	0.131***	0.137***	0.127***	0.140***
	(0.026)	(0.029)	(0.030)	(0.028)	(0.030)
Complete Homework	0.087***	0.098***	0.075***	0.096***	0.077***
	(0.016)	(0.017)	(0.013)	(0.017)	(0.014)
Arrive on Time	0.058***	0.062***	0.064***	0.061***	0.065***
	(0.021)	(0.022)	(0.023)	(0.022)	(0.023)
Behavior Not a Problem	0.048**	0.052**	0.054**	0.051**	0.054**
	(0.020)	(0.022)	(0.023)	(0.022)	(0.023)
Incentivized Outcome Index	0.168***	0.183***	0.189***	0.179***	0.192***
	(0.043)	(0.047)	(0.048)	(0.046)	(0.049)
B. Student Achievement					
State Math	0.139***	0.160***	0.123***	0.155***	0.126***
	(0.020)	(0.023)	(0.018)	(0.023)	(0.019)
State Reading	0.146***	0.169***	0.130***	0.163***	0.133***
	(0.020)	(0.023)	(0.018)	(0.022)	(0.018)
At or Above Proficient in Math	0.070***	0.081***	0.062***	0.078***	0.064***
	(0.011)	(0.013)	(0.010)	(0.013)	(0.010)
Advanced in Math	0.006	0.007	0.005	0.006	0.005
	(0.005)	(0.005)	(0.004)	(0.005)	(0.004)
At or Above Proficient in Reading	0.063***	0.073***	0.056***	0.071***	0.058***
	(0.011)	(0.013)	(0.010)	(0.013)	(0.010)
Advanced in Reading	0.008*	0.009*	0.007*	0.009*	0.007*
	(0.005)	(0.005)	(0.004)	(0.005)	(0.004)
Academic Achievement Index	0.138***	0.159***	0.122***	0.154***	0.125***
	(0.018)	(0.020)	(0.016)	(0.020)	(0.016)
C. Behavior and Motivation					
Attendance Rate	0.184	0.211	0.164	0.205	0.168
	(0.202)	(0.232)	(0.180)	(0.226)	(0.184)
Work Hard in School	-0.005	-0.005	-0.006	-0.005	-0.006
	(0.020)	(0.022)	(0.023)	(0.021)	(0.023)

Push Self in School	0.007	0.008	0.006	0.008	0.006
	(0.016)	(0.018)	(0.014)	(0.018)	(0.014)
Intrisic Motivation Index	0.075**	0.084**	0.063**	0.082**	0.064**
	(0.034)	(0.038)	(0.029)	(0.038)	(0.029)
Behavior Index	0.097**	0.105**	0.108**	0.103**	0.110**
	(0.046)	(0.050)	(0.051)	(0.049)	(0.052)

Notes: This table reports ITT and 2SLS estimates of the effects of our aligned incentives experiment in DC on direct outcomes and indirect outcomes such as test scores, survey responses, attendance and intrinsic motivation. In Column (2), the instrumented variable is Ever Treated (payment periods) which is 0 if the student did not receive treatment and 1 if the student ever recieved a positive payment in the either year . In Column (3), the instrumented variable is Periods Treated (payment periods) which is 0 if the student did not receive treatment or a fraction between 0 and 1 in the first year and 0 and 2 in the second year to determine fraction of payment periods for which the student received positive payments over two years. In Column (4), the instrumented variable is Ever Treated (attendance) which is 0 if the student never attended a treatment school and 1 if the student ever attended a treatment school in either year. In Column (5), the instrumened variable is Days Treated (attendance) which is 0 if the number of days the student attended a treatment school is 0 and a fraction between 0 and 1 in the first year and 0 and 2 to determine the fraction of the two-year experiment for which a student attended a treatment school. All dependent variables are defined analogously to those in Table 4A. All regressions follow the pooled controlled specification from Column (6) in Table 4A. Outcome variables that do not have a pooled specification in Table 4A follow the controlled specification from Column (4). In that case, the instrumented variables are restricted to their values in the first year, i.e. Ever Treated is a one if the student was ever treated in the first year and Periods/Days Treated range from 0 to 1 to determine the fraction of the first year of treatment that a student was treated. Standard errors are robust to heteroskedasticity. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table 2B: Mean Effect Sizes (Two Stage Least Squares Estimates) in Houston

		Payment Periods		Attendance	
	ITT	Ever Treated	Periods Treated	Ever Treated	Days Treated
	(1)	(2)	(3)	(4)	(5)
A. Incentivized Outcomes					
Parent Conferences Attended	1.546***	1.547***	1.656***	1.580***	1.649***
	(0.101)	(0.100)	(0.106)	(0.102)	(0.106)
Objectives Mastered	1.083***	1.087***	1.192***	1.102***	1.147***
	(0.032)	(0.032)	(0.033)	(0.032)	(0.033)
Incentivized Outcome Index	1.170***	1.170***	1.253***	1.196***	1.247***
	(0.046)	(0.045)	(0.047)	(0.046)	(0.048)
B. Student Achievement					
State Math	0.076***	0.076***	0.083***	0.077***	0.080***
	(0.025)	(0.026)	(0.028)	(0.026)	(0.027)
State Reading	-0.039	-0.039	-0.043	-0.039	-0.041
	(0.027)	(0.027)	(0.029)	(0.027)	(0.028)
Aligned State Math	0.112***	0.113***	0.123***	0.114***	0.119***
	(0.029)	(0.029)	(0.031)	(0.029)	(0.030)
Unaligned State Math	0.031	0.031	0.034	0.031	0.032
	(0.030)	(0.030)	(0.033)	(0.030)	(0.031)
Stanford 10 Math	0.026	0.027	0.029	0.027	0.028
	(0.022)	(0.022)	(0.025)	(0.023)	(0.024)
Stanford 10 Reading	-0.044*	-0.045**	-0.049**	-0.045**	-0.047**
	(0.023)	(0.023)	(0.025)	(0.023)	(0.024)
Stanford 10 Science	-0.085***	-0.086***	-0.094***	-0.086***	-0.090***
	(0.028)	(0.028)	(0.030)	(0.028)	(0.029)
Stanford 10 Social Studies	-0.055**	-0.055**	-0.061**	-0.056**	-0.058**
	(0.025)	(0.025)	(0.027)	(0.025)	(0.026)
Meets Minimum Math Standard	0.026**	0.027**	0.029**	0.027**	0.028**
	(0.012)	(0.012)	(0.014)	(0.013)	(0.013)
Math Commended Performance	0.016	0.016	0.017	0.016	0.017
	(0.015)	(0.015)	(0.016)	(0.015)	(0.015)
Meets Minimum Reading Standard	0.005	0.005	0.005	0.005	0.005
	(0.013)	(0.013)	(0.014)	(0.013)	(0.014)
Reading Commended Performance	-0.004	-0.004	-0.005	-0.004	-0.005
	(0.015)	(0.015)	(0.016)	(0.015)	(0.015)
Incentivized Achievement Index	0.053**	0.053**	0.058**	0.054**	0.056**
	(0.021)	(0.021)	(0.023)	(0.021)	(0.022)
Non-Incentivized Achievement Index	-0.059***	-0.060***	-0.065***	-0.060***	-0.063***
	(0.019)	(0.019)	(0.021)	(0.019)	(0.020)
C. Survey Outcomes					

Parents check Homework more	0.069***	0.069***	0.076***	0.071***	0.074***
	(0.025)	(0.025)	(0.027)	(0.025)	(0.027)
Student prefers Math to Reading	0.091***	0.091***	0.101***	0.093***	0.098***
	(0.023)	(0.023)	(0.025)	(0.024)	(0.025)
Parent asks about Math more than Rdg.	0.116***	0.116***	0.124***	0.118***	0.123***
	(0.028)	(0.028)	(0.030)	(0.028)	(0.030)
Survey Outcome Index	0.257***	0.256***	0.277***	0.266***	0.277***
	(0.092)	(0.091)	(0.098)	(0.094)	(0.098)
D. Behavior and Motivation					
Behavioral Offense	-0.014	-0.014	-0.016	-0.014	-0.015
	(0.011)	(0.011)	(0.012)	(0.011)	(0.012)
Attendance Rate	-0.019	-0.020	-0.022	-0.020	-0.021
	(0.112)	(0.115)	(0.126)	(0.113)	(0.119)
Intrinsic Motivation Index	-0.077	-0.077	-0.084	-0.079	-0.082
	(0.065)	(0.065)	(0.071)	(0.066)	(0.069)
Behavior Index	-0.071	-0.071	-0.079	-0.073	-0.077
	(0.063)	(0.062)	(0.068)	(0.063)	(0.066)

Notes: This table reports ITT and 2SLS estimates of the effects of our aligned incentives experiment in Houston on direct outcomes and indirect outcomes such as test scores, survey responses, attendance and intrinsic motivation. In Column (2), the instrumented variable is Ever Treated (payment periods) which is 0 if the student did not receive treatment and 1 if the student recieved a positive payment in any pay period. In Column (3), the instrumented variable is Periods Treated (payment periods) which is 0 if the student did not receive treatment or a fraction between 0 and 1 to determine fraction of payment periods for which the student received positive payments. In Column (4), the instrumented variable is Ever Treated (attendance) which is 0 if the student never attended a treatment school and 1 if the student ever attended a treatment school. In Column (5), the instrumented variable is Days Treated (attendance) which is 0 if the number of days the student attended a treatment school is 0 and a fraction between 0 and 1 to determine the fraction of the year for which student attended a treatment school otherwise. All dependent variables are defined analogously to Table 4B. All regressions follow the controlled specification described in Column (2) from Table 4B. Standard errors are robust to heteroskedasticity. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table 3A: Mean Effect Sizes (Intent to Treat Estimates) by Subsample in DC

	Whole	*Gender*		*Race*			*Free Lunch*		*Math Quintile*		*Reading Quintile*	
	Sample	Male	Female	Black	Hispanic	White	Yes	No	Bottom	Top	Bottom	Top
A. Incentivized Outcomes												
Behavioral Offense	-0.032***	-0.041***	-0.024**	-0.034***	0.002	-0.069	-0.034***	-0.025	-0.056***	-0.011	-0.063***	0.022
	(0.008)	(0.013)	(0.011)	(0.009)	(0.015)	(0.052)	(0.010)	(0.017)	(0.021)	(0.020)	(0.021)	(0.021)
N	9322	4675	4647	7934	870	357	6663	2584	1807	1429	1805	1392
p-value:			*0.330*			*0.085*		*0.657*		*0.119*		*0.004*
GPA	0.118***	0.136***	0.092**	0.112***	0.130	0.328	0.116***	0.112*	0.125**	0.105	0.140***	0.079
	(0.026)	(0.036)	(0.038)	(0.027)	(0.090)	(0.364)	(0.029)	(0.059)	(0.053)	(0.088)	(0.053)	(0.094)
N	5802	2902	2900	4950	518	232	4120	1636	1116	914	1117	892
p-value:			*0.390*			*0.809*		*0.948*		*0.840*		*0.562*
Complete Homework	0.087***	0.102***	0.078***	0.074***	0.278***	0.409*	0.089***	0.073**	0.064	0.134***	0.035	0.055
	(0.016)	(0.023)	(0.022)	(0.016)	(0.060)	(0.223)	(0.018)	(0.032)	(0.040)	(0.044)	(0.040)	(0.049)
N	5251	2491	2760	4390	522	229	3696	1528	845	895	874	844
p-value:			*0.444*			*0.001*		*0.677*		*0.233*		*0.753*
Arrive on Time	0.058***	0.054*	0.060**	0.042*	0.221***	2.434***	0.057**	0.060	-0.021	0.095	0.015	0.123*
	(0.021)	(0.030)	(0.029)	(0.022)	(0.079)	(0.742)	(0.024)	(0.046)	(0.053)	(0.063)	(0.052)	(0.070)
N	3350	1606	1744	2750	375	149	2349	986	518	599	533	555
p-value:			*0.893*			*0.000*		*0.950*		*0.148*		*0.201*
Behavior Not a Problem	0.048**	0.058**	0.036	0.032	0.253***	-0.759	0.055**	0.033	0.034	0.018	0.061	-0.024
	(0.020)	(0.029)	(0.029)	(0.022)	(0.073)	(0.644)	(0.023)	(0.046)	(0.051)	(0.064)	(0.050)	(0.066)
N	3331	1596	1735	2738	373	147	2346	970	513	596	530	557
p-value:			*0.599*			*0.004*		*0.663*		*0.844*		*0.290*
B. Student Achievement												
State Math	0.139***	0.158***	0.119***	0.128***	0.135**	-0.186	0.118***	0.216***	0.184***	0.046	0.066	-0.004
	(0.020)	(0.030)	(0.027)	(0.022)	(0.067)	(0.328)	(0.023)	(0.044)	(0.049)	(0.051)	(0.050)	(0.054)
N	9022	4478	4544	7654	852	356	6439	2517	1718	1414	1721	1374
p-value:			*0.330*			*0.607*		*0.048*		*0.050*		*0.334*
State Reading	0.146***	0.211***	0.077***	0.134***	0.219***	-0.032	0.120***	0.229***	0.241***	0.113**	0.067	0.068
	(0.020)	(0.030)	(0.026)	(0.021)	(0.070)	(0.376)	(0.023)	(0.043)	(0.050)	(0.054)	(0.049)	(0.053)
N	9033	4485	4548	7674	847	356	6453	2512	1727	1414	1725	1374
p-value:			*0.001*			*0.440*		*0.023*		*0.078*		*0.991*
At or Above Proficient in Math	0.070***	0.085***	0.056***	0.066***	0.088**	0.229**	0.055***	0.115***	0.025*	0.049	0.008	0.022
	(0.011)	(0.015)	(0.017)	(0.012)	(0.044)	(0.093)	(0.013)	(0.025)	(0.014)	(0.031)	(0.017)	(0.037)
N	9022	4478	4544	7654	852	356	6439	2517	1718	1414	1721	1374
p-value:			*0.201*			*0.176*		*0.030*		*0.462*		*0.730*
Advanced in Math	0.006	0.001	0.008	0.004	0.030	-0.268	0.006	0.006	0.003	-0.023	0.003	-0.033
	(0.005)	(0.006)	(0.007)	(0.005)	(0.022)	(0.322)	(0.005)	(0.013)	(0.003)	(0.034)	(0.005)	(0.029)
N	9022	4478	4544	7654	852	356	6439	2517	1718	1414	1721	1374
p-value:			*0.479*			*0.340*		*0.961*		*0.438*		*0.204*
At or Above Proficient in Reading	0.063***	0.056***	0.071***	0.057***	0.105**	0.245**	0.039***	0.145***	0.064***	0.061*	0.028**	-0.007
	(0.011)	(0.015)	(0.017)	(0.012)	(0.045)	(0.117)	(0.013)	(0.025)	(0.016)	(0.037)	(0.013)	(0.030)
N	9033	4485	4548	7674	847	356	6453	2512	1727	1414	1725	1374
p-value:			*0.513*			*0.153*		*0.000*		*0.932*		*0.286*
Advanced in Reading	0.008*	0.013**	0.005	0.006	0.041**	0.055	0.003	0.018	0.004	0.004	0.001	0.007

	(0.005)	(0.006)	(0.007)	(0.005)	(0.017)	(0.366)	(0.005)	(0.014)	(0.004)	(0.029)	(0.002)	(0.036)
N	9033	4485	4548	7674	847	356	6453	2512	1727	1414	1725	1374
p-value:			*0.392*			*0.118*		*0.326*		*0.991*		*0.860*
C. Behavior and Motivation												
Attendance Rate	0.184	0.346	0.009	-0.089	3.441***	0.947	0.264	-0.219	-0.010	0.821	-0.051	1.330*
	(0.202)	(0.284)	(0.293)	(0.212)	(0.652)	(3.062)	(0.232)	(0.412)	(0.527)	(0.568)	(0.527)	(0.685)
N	9322	4675	4647	7934	870	357	6663	2584	1807	1429	1805	1392
p-value:			*0.407*			*0.000*		*0.304*		*0.279*		*0.107*
Work Hard in School	-0.005	0.057*	-0.067**	-0.016	0.140*	1.242	0.002	-0.047	0.008	0.051	0.001	-0.081
	(0.020)	(0.030)	(0.028)	(0.021)	(0.079)	(0.944)	(0.023)	(0.044)	(0.050)	(0.061)	(0.050)	(0.068)
N	3361	1613	1748	2772	371	144	2358	987	537	592	548	550
p-value:			*0.002*			*0.052*		*0.321*		*0.585*		*0.318*
Push Self in School	0.007	0.021	-0.004	-0.003	0.141**	-0.212	0.013	0.004	0.081*	0.024	0.114***	-0.012
	(0.016)	(0.024)	(0.023)	(0.017)	(0.064)	(0.370)	(0.019)	(0.036)	(0.042)	(0.050)	(0.040)	(0.054)
N	5113	2414	2699	4260	518	230	3601	1487	794	892	806	845
p-value:			*0.436*			*0.070*		*0.824*		*0.370*		*0.057*
Intrisic Motivation Index	0.075**	0.142***	0.007	0.069*	0.307**	-0.924	0.030	0.212***	0.025	0.069	0.033	-0.036
	(0.034)	(0.051)	(0.047)	(0.037)	(0.124)	(0.561)	(0.040)	(0.070)	(0.094)	(0.097)	(0.093)	(0.101)
N	4401	2066	2335	3661	443	196	3086	1297	653	789	694	742
p-value:			*0.050*			*0.025*		*0.022*		*0.738*		*0.609*

Notes: This table reports ITT estimates of the effects of the experiment in DC on all direct and indirect outcomes for a variety of subsamples. All dependent variables are defined analogously to those in Table 4A. All regressions follow the pooled controlled specification described in Column (6) from Table 4A. Outcome variables that do not have a pooled specification in Table 4A follow the controlled specification from Column (4). *** = significant at 1 percent level, ** = significant at 5 percent level, and * = significant at 10 percent level.

Appendix Table 3B: Mean Effect Sizes (Intent to Treat Estimates) by Subsample in Houston

	Whole	*Gender*		*Race*		*Free Lunch*		*Math Quintile*		*Specialized Teacher*	
	Sample	Male	Female	Black	Hispanic	Yes	No	Bottom	Top	Yes	No
A. Incentivized Outcomes											
Parent Conferences Attended	1.546***	1.640***	1.389***	1.167***	1.492***	1.665***	1.621***	1.381***	2.046***	1.485***	1.679***
	(0.101)	(0.152)	(0.140)	(0.298)	(0.134)	(0.180)	(0.190)	(0.238)	(0.309)	(0.165)	(0.306)
N	2053	1020	1032	519	1421	588	659	389	265	1415	638
p-value:			*0.211*		*0.299*		*0.860*		*0.063*		*0.562*
Objectives Mastered	1.083***	1.015***	1.159***	0.946***	1.081***	1.138***	1.053***	0.697***	1.751***	0.878***	1.376***
	(0.032)	(0.047)	(0.044)	(0.054)	(0.046)	(0.058)	(0.057)	(0.048)	(0.115)	(0.044)	(0.097)
N	3292	1726	1565	853	2275	928	1072	687	413	2291	1001
p-value:			*0.022*		*0.051*		*0.280*		*0.000*		*0.000*
B. Student Achievement											
State Math	0.076***	0.067*	0.069*	0.052	0.041	0.090*	-0.029	0.001	0.179**	0.003	0.346***
	(0.025)	(0.036)	(0.037)	(0.060)	(0.033)	(0.046)	(0.045)	(0.051)	(0.078)	(0.035)	(0.092)
N	3153	1640	1512	827	2169	900	1030	659	418	2235	918
p-value:			*0.969*		*0.869*		*0.059*		*0.043*		*0.000*
State ELA	-0.039	-0.020	-0.059	0.084	-0.055*	-0.011	-0.164***	-0.127*	0.044	-0.115***	0.268***
	(0.027)	(0.038)	(0.039)	(0.074)	(0.033)	(0.052)	(0.049)	(0.065)	(0.078)	(0.038)	(0.091)
N	3128	1619	1508	818	2152	899	1021	653	417	2220	908
p-value:			*0.464*		*0.079*		*0.029*		*0.076*		*0.000*
Aligned State Math	0.112***	0.107***	0.109***	0.087	0.070*	0.114**	0.020	0.022	0.112**	0.036	0.399***
	(0.029)	(0.041)	(0.041)	(0.081)	(0.037)	(0.055)	(0.053)	(0.085)	(0.046)	(0.042)	(0.125)
N	3153	1640	1512	827	2169	900	1030	659	418	2235	918
p-value:			*0.963*		*0.841*		*0.204*		*0.327*		*0.005*
Unaligned State Math	0.031	0.003	0.046	0.105	-0.025	0.032	-0.078	-0.027	0.113**	-0.054	0.373***
	(0.030)	(0.041)	(0.044)	(0.080)	(0.038)	(0.055)	(0.056)	(0.088)	(0.048)	(0.042)	(0.119)
N	3153	1640	1512	827	2169	900	1030	659	418	2235	918
p-value:			*0.469*		*0.131*		*0.149*		*0.147*		*0.001*
Stanford 10 Math	0.026	0.047	-0.003	-0.003	-0.002	-0.011	-0.018	-0.014	0.181**	-0.084***	0.137*
	(0.022)	(0.033)	(0.030)	(0.054)	(0.028)	(0.042)	(0.040)	(0.044)	(0.071)	(0.032)	(0.073)
N	3337	1750	1586	875	2298	938	1090	701	417	2293	1044
p-value:			*0.254*		*0.991*		*0.903*		*0.014*		*0.004*
Stanford 10 ELA	-0.044*	-0.033	-0.063**	-0.102*	-0.027	-0.095**	-0.083**	-0.126**	0.161**	-0.116***	0.016
	(0.023)	(0.033)	(0.031)	(0.058)	(0.028)	(0.042)	(0.040)	(0.049)	(0.067)	(0.030)	(0.075)
N	3338	1752	1585	877	2297	939	1090	702	417	2294	1044
p-value:			*0.492*		*0.234*		*0.839*		*0.000*		*0.095*
Stanford 10 Science	-0.085***	-0.092**	-0.082**	-0.208***	-0.107***	-0.119**	-0.186***	-0.057	-0.104	-0.287***	0.205**
	(0.028)	(0.039)	(0.039)	(0.068)	(0.035)	(0.052)	(0.051)	(0.065)	(0.074)	(0.040)	(0.089)
N	3334	1751	1582	874	2297	936	1089	700	417	2292	1042
p-value:			*0.851*		*0.181*		*0.347*		*0.615*		*0.000*
Stanford 10 Social Studies	-0.055**	-0.031	-0.082**	-0.095	-0.085***	-0.074	-0.155***	-0.066	0.051	-0.202***	0.115
	(0.025)	(0.036)	(0.034)	(0.060)	(0.031)	(0.047)	(0.044)	(0.057)	(0.074)	(0.033)	(0.077)
N	3334	1750	1583	873	2298	936	1089	700	417	2291	1043
p-value:			*0.290*		*0.870*		*0.198*		*0.185*		*0.000*
Meets Minimum Math Standard	0.026**	0.027	0.021	0.009	0.018	0.023	-0.004	-0.016	0.012	0.010	0.113**

	(0.012)	(0.017)	(0.018)	(0.036)	(0.016)	(0.023)	(0.024)	(0.043)	(0.011)	(0.018)	(0.051)
N	3153	1640	1512	827	2169	900	1030	659	418	2235	918
p-value:			*0.825*		*0.826*		*0.402*		*0.519*		*0.049*
Math Commended Performance	0.016	0.018	0.006	-0.017	0.009	0.055**	-0.042	-0.000	0.083**	-0.030	0.097**
	(0.015)	(0.021)	(0.021)	(0.033)	(0.019)	(0.027)	(0.026)	(0.017)	(0.036)	(0.020)	(0.043)
N	3153	1640	1512	827	2169	900	1030	659	418	2235	918
p-value:			*0.696*		*0.482*		*0.008*		*0.027*		*0.006*
Meets Minimum Reading Standard	0.005	0.028	-0.023	-0.004	0.014	0.027	-0.031	-0.084**	0.019	-0.054***	0.106**
	(0.013)	(0.019)	(0.019)	(0.034)	(0.017)	(0.024)	(0.025)	(0.039)	(0.019)	(0.019)	(0.048)
N	3128	1619	1508	818	2152	899	1021	653	417	2220	908
p-value:			*0.052*		*0.612*		*0.090*		*0.014*		*0.001*
Reading Commended Performance	-0.004	-0.003	-0.004	0.044	-0.002	-0.003	-0.051*	0.001	0.096**	-0.006	0.039
	(0.015)	(0.020)	(0.022)	(0.037)	(0.019)	(0.029)	(0.026)	(0.025)	(0.049)	(0.020)	(0.045)
N	3128	1619	1508	818	2152	899	1021	653	417	2220	908
p-value:			*0.973*		*0.258*		*0.209*		*0.066*		*0.346*
C. Survey Outcomes											
Parents check Homework more	0.069***	0.032	0.099***	0.084	0.030	0.002	0.113***	0.173***	0.148**	0.127***	0.092
	(0.025)	(0.038)	(0.033)	(0.099)	(0.033)	(0.053)	(0.043)	(0.063)	(0.071)	(0.047)	(0.091)
N	2315	1192	1123	563	1621	630	774	491	285	1529	786
p-value:			*0.176*		*0.593*		*0.094*		*0.777*		*0.722*
Student prefers Math to Rdg	0.091***	0.089***	0.106***	0.028	0.085***	0.139***	0.073*	0.178***	0.012	0.072	-0.010
	(0.023)	(0.033)	(0.034)	(0.076)	(0.029)	(0.050)	(0.040)	(0.064)	(0.063)	(0.044)	(0.083)
N	2356	1209	1146	571	1650	640	785	498	291	1560	796
p-value:			*0.716*		*0.468*		*0.284*		*0.047*		*0.371*
Parent asks about Math > Rdg	0.116***	0.079*	0.141***	0.208**	0.122***	0.094*	0.161***	0.016	0.153*	0.141***	0.108
	(0.028)	(0.043)	(0.039)	(0.083)	(0.037)	(0.055)	(0.051)	(0.069)	(0.083)	(0.044)	(0.094)
N	1909	947	961	473	1332	547	609	348	256	1330	579
p-value:			*0.268*		*0.319*		*0.351*		*0.164*		*0.739*
D. Behavior and Motivation											
Attendance 10-11, percent	-0.019	-0.103	0.031	0.375	-0.142	-0.026	-0.096	-0.005	-0.245	-0.187	0.270
	(0.112)	(0.172)	(0.147)	(0.302)	(0.126)	(0.180)	(0.189)	(0.251)	(0.225)	(0.198)	(0.333)
N	3428	1796	1631	914	2334	965	1113	721	418	2354	1074
p-value:			*0.547*		*0.105*		*0.783*		*0.455*		*0.228*
Behavioral Offense	-0.014	-0.006	-0.017	-0.043	-0.002	-0.028	0.004	0.030	-0.068**	-0.010	-0.009
	(0.011)	(0.017)	(0.013)	(0.037)	(0.012)	(0.020)	(0.021)	(0.028)	(0.029)	(0.016)	(0.041)
N	3428	1796	1631	914	2334	965	1113	721	418	2354	1074
p-value:			*0.605*		*0.283*		*0.251*		*0.011*		*0.991*
Intrinsic Motivation Index	-0.077	-0.153	-0.038	-0.238	-0.090	-0.076	0.058	0.167	0.339	-0.038	0.309
	(0.065)	(0.098)	(0.090)	(0.175)	(0.086)	(0.134)	(0.110)	(0.155)	(0.220)	(0.113)	(0.193)
N	2137	1098	1038	510	1506	583	713	429	265	1417	720
p-value:			*0.375*		*0.432*		*0.419*		*0.483*		*0.108*

Notes: This table reports ITT estimates of the effects of the experiment in Houston on all direct and indirect outcomes for a variety of subsamples. All dependent variables are defined analogously to those in Table 4B. All regressions follow the controlled specification described in Column (2) from Table 4B. *** = significant at 1 percent level, ** = significant at 5 percent level, and * = significant at 10 percent level.

Appendix Table 4A: Attrition in DC

	Control Means			Randomization Controls			Full Controls		
	2008-2009	2009-2010	Pooled	2008-2009	2009-2010	Pooled	2008-2009	2009-2010	Pooled
	(1)	(2)	(3)	(4)	(5)	(6)			
A. Incentivized Outcomes									
Behavioral Offense	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
				6039	3283	9322	6039	3283	9322
GPA	0.044	—	—	0.007	—	—	0.006	—	—
				(0.008)			(0.007)		
				6039			6039		
Complete Homework	0.414	0.394	0.407	-0.095***	-0.029	-0.070***	-0.136***	-0.067***	-0.109***
				(0.015)	(0.020)	(0.012)	(0.015)	(0.020)	(0.012)
				6039	3283	9322	6039	3283	9322
Arrive on Time	0.428	—	—	-0.089***	—	—	-0.126***	—	—
				(0.015)			(0.016)		
				6039			6039		
Behavior Not a Problem	0.432	—	—	-0.092***	—	—	-0.128***	—	—
				(0.015)			(0.016)		
				6039			6039		
Incentivized Outcome Index	0.477	—	—	-0.082***	—	—	-0.120***	—	—
				(0.015)			(0.016)		
				6039			6039		
B. Student Achievement									
State Math	0.028	0.029	0.028	0.012**	0.013	0.012**	0.010*	0.012	0.011**
				(0.006)	(0.008)	(0.005)	(0.005)	(0.008)	(0.004)
				6039	3283	9322	6039	3283	9322
State Reading	0.027	0.021	0.025	0.016***	0.019**	0.017***	0.015***	0.018***	0.016***
				(0.006)	(0.008)	(0.005)	(0.005)	(0.007)	(0.004)
				6039	3283	9322	6039	3283	9322
Academic Achievement Index	0.031	0.029	0.030	0.012**	0.013	0.012**	0.011**	0.013	0.011**
				(0.006)	(0.008)	(0.005)	(0.006)	(0.008)	(0.005)
				6039	3283	9322	6039	3283	9322
C. Behavior and Motivation									
Attendance Rate	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
				6039	3283	9322	6039	3283	9322
Work Hard in School	0.424	—	—	-0.089***	—	—	-0.128***	—	—
				(0.015)			(0.015)		
				6039			6039		
Push Self in School	0.428	0.413	0.423	-0.085***	-0.048**	-0.070***	-0.121***	-0.086***	-0.107***
				(0.015)	(0.020)	(0.012)	(0.016)	(0.020)	(0.012)
				6039	3283	9322	6039	3283	9322
Intrisic Motivation Index	0.525	0.454	0.500	-0.081***	-0.032	-0.062***	-0.112***	-0.066***	-0.095***
				(0.015)	(0.020)	(0.012)	(0.016)	(0.021)	(0.013)
				6039	3283	9322	6039	3283	9322
Behavior Index	0.553	—	—	-0.081***	—	—	-0.111***	—	—
				(0.015)			(0.016)		
				6039			6039		

Notes: This table reports ITT estimates of the effects of our incentives experiment in DC on whether a student is missing various test scores and survey responses. Each attrition measure is coded as one if a given student does not have valid values or survey responses for that outcome and zero otherwise, where dependent variables are defined analogously to Table 4A. The randomization controls and fully controlled specifications are identical to those outlined in Table 4A. Standard errors are robust to heteroskedasticity. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table 4B: Attrition in Houston

	Control Mean	Baseline	Controlled
	(1)	(2)	(3)
A. Incentivized Outcomes			
Parent Conferences Attended	0.535	-0.291***	-0.331***
		(0.015)	(0.015)
		3428	3428
Objectives Mastered	0.047	-0.017**	-0.020***
		(0.007)	(0.007)
		3428	3428
Incentivized Outcome Index	0.549	-0.307***	-0.348***
		(0.015)	(0.015)
		3428	3428
B. Student Achievement			
State Math	0.074	0.011	0.005
		(0.008)	(0.008)
		3428	3428
State Reading	0.082	0.008	0.001
		(0.008)	(0.008)
		3428	3428
Aligned State Math	0.074	0.011	0.005
		(0.008)	(0.008)
		3428	3428
Unaligned State Math	0.074	0.011	0.005
		(0.008)	(0.008)
		3428	3428
Stanford 10 Math	0.024	0.005	0.004
		(0.006)	(0.006)
		3428	3428
Stanford 10 Reading	0.024	0.005	0.004
		(0.006)	(0.006)
		3428	3428
Stanford 10 Science	0.027	0.002	0.001
		(0.006)	(0.006)
		3428	3428
Stanford 10 Soc. Studies	0.027	0.002	0.001
		(0.006)	(0.006)
		3428	3428
Incentivized Achievement Index	0.081	0.011	0.004
		(0.008)	(0.009)
		3428	3428
Non-Incentivized Achievement Index	0.093	0.004	-0.003
		(0.009)	(0.009)
		3428	3428
C. Survey Outcomes			
Parents check Homework more	0.443	-0.267***	-0.309***
		(0.014)	(0.014)
		3428	3428
Student prefers Math to Reading	0.435	-0.273***	-0.321***
		(0.014)	(0.013)
		3428	3428
Parent asks about Math more than Rdg.	0.570	-0.279***	-0.313***
		(0.016)	(0.016)
		3428	3428
Survey Outcome Index	0.721	-0.314***	-0.347***
		(0.015)	(0.015)
		3428	3428
D. Attendance and Motivation			

Attendance Rate	0.000	0.000	0.000
		(0.000)	(0.000)
		3428	3428
Behavioral Offense	0.000	0.000	0.000
		(0.000)	(0.000)
		3428	3428
Intrinsic Motivation Index	0.493	-0.263***	-0.306***
		(0.015)	(0.015)
		3428	3428
Behavior Index	0.493	-0.263***	-0.306***
		(0.015)	(0.015)
		3428	3428
E. Post-Treatment Outcomes (t+2)			
State Math	0.322	0.030**	0.029**
		(0.012)	(0.012)
		2568	2568
State Reading	0.321	0.038***	0.032***
		(0.012)	(0.012)
		2568	2568
Stanford 10 Math	0.293	0.017*	0.021*
		(0.010)	(0.011)
		2568	2568
Stanford 10 Reading	0.293	0.013	0.018*
		(0.010)	(0.011)
		2568	2568
Stanford 10 Science	0.295	0.016	0.021*
		(0.010)	(0.011)
		2568	2568
Stanford 10 Soc. Studies	0.294	0.014	0.019*
		(0.010)	(0.011)
		2568	2568

Notes: This table reports ITT estimates of the effects of our aligned incentives experiment in Houston on whether a student is missing various test scores and survey responses. Each attrition measure is coded as a one if a given student does not have valid values or survey respones for that outcome and a zero otherwise, where each outcome is defined analogously to those in Table 4B. Baseline and controlled specifications are equivalent to those described in Table 4B. Models with outcome variables taken in 2012-13 additionally include grade in 2012-13 fixed-effects. Standard errors are robust to heteroskedasticity. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table 5A: Attrition-Bounded Estimates in DC

	2008-2009		2009-2010		Pooled	
	ITT	Lee Bound	ITT	Lee Bound	ITT	Lee Bound
	(1)	(2)	(3)	(4)	(5)	(6)
A. Incentivized Outcomes						
Behavioral Offense	-0.086***	-0.086***	0.051***	0.051***	-0.032***	-0.032***
	(0.011)	(0.011)	(0.013)	(0.013)	(0.008)	(0.008)
N	6039	6039	3283	3283	9322	9322
p-value		—		—		—
GPA	0.118***	0.112***	—	—	—	—
	(0.026)	(0.026)				
N	5802	5786				
p-value		0.009				
Complete Homework	0.100***	-0.050**	0.072***	0.014	0.087***	-0.023
	(0.020)	(0.020)	(0.026)	(0.026)	(0.016)	(0.016)
N	3441	2965	1810	1683	5251	4664
p-value		0.000		0.000		0.000
Arrive on Time	0.058***	-0.109***	—	—	—	—
	(0.021)	(0.021)				
N	3350	2909				
p-value		0.000				
Behavior Not a Problem	0.048**	-0.163***	—	—	—	—
	(0.020)	(0.019)				
N	3331	2883				
p-value		0.000				
Incentivized Outcome Index	0.168***	-0.229***	—	—	—	—
	(0.043)	(0.041)				
N	3079	2659				
p-value		0.000				
B. Student Achievement						
State Math	0.154***	0.124***	0.109***	0.067**	0.139***	0.104***
	(0.025)	(0.024)	(0.034)	(0.033)	(0.020)	(0.020)
N	5846	5820	3176	3159	9022	8978
p-value		0.000		0.000		0.000
State Reading	0.179***	0.123***	0.080**	0.021	0.146***	0.090***
	(0.025)	(0.024)	(0.033)	(0.031)	(0.020)	(0.019)
N	5844	5805	3189	3163	9033	8969
p-value		0.000		0.000		0.000
Academic Achievement Index	0.164***	0.134***	0.087***	0.052*	0.138***	0.107***
	(0.022)	(0.021)	(0.029)	(0.028)	(0.018)	(0.017)
N	5828	5800	3171	3152	8999	8955
p-value		0.000		0.000		0.000
C. Behavior and Motivation						
Attendance Rate	0.344	0.344	-0.171	-0.171	0.184	0.184
	(0.251)	(0.251)	(0.335)	(0.335)	(0.202)	(0.202)
N	6039	6039	3283	3283	9322	9322
p-value		—		—		—
Work Hard in School	-0.005	-0.170***	—	—	—	—

	(0.020)	(0.020)				
N	3361	2913				
p-value		0.000				
Push Self in School	0.007	-0.133***	0.009	-0.124***	0.007	-0.131***
	(0.021)	(0.021)	(0.027)	(0.028)	(0.016)	(0.017)
N	3338	2915	1775	1612	5113	4537
p-value		0.000		0.000		0.000
Intrisic Motivation Index	0.073	-0.352***	0.077	-0.189***	0.075**	-0.295***
	(0.045)	(0.042)	(0.054)	(0.049)	(0.034)	(0.032)
N	2766	2374	1635	1510	4401	3889
p-value		0.000		0.000		0.000
Behavior Index	0.097**	-0.348***	—	—	—	—
	(0.046)	(0.045)				
N	2603	2215				
p-value		0.000		0.000		0.000

Notes: Column (2) reports attrition bounded estimates of the effects of our aligned incentives experiment in DC on all direct and indirect outcomes. All dependent variables are defined analogously to those in Table 4A. If treatment students were more likely to have valid index measures, the highest performing treatment students were dropped from the attrition bounded regressions. If treatment students were less likely to have valid outcome measures, the lowest performing control students were dropped from the attrition bounded regressions. Specifications are the same as in Table 4A, Columns (4)-(6). Standard errors are robust to heteroskedasticity. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table 5B: Attrition-Bounded Estimates in Houston

	Observed ITT	Attrition-Bounded ITT	p-value
	(1)	(2)	(3)
A. Incentivized Outcomes			
Parent Conferences Attended	1.546***	0.182*	
	(0.101)	(0.095)	0.000
	2053	1492	
Objectives Mastered	1.083***	1.004***	
	(0.032)	(0.029)	0.000
	3292	3258	
Incentivized Outcome Index	1.170***	0.536***	
	(0.046)	(0.040)	0.000
	2027	1437	
B. Student Achievement			
State Math	0.076***	0.065**	
	(0.025)	(0.025)	0.007
	3153	3144	
State Reading	-0.039	-0.041	
	(0.027)	(0.027)	0.196
	3128	3126	
Aligned State Math	0.112***	0.097***	
	(0.029)	(0.028)	0.005
	3153	3144	
Unaligned State Math	0.031	0.012	
	(0.030)	(0.029)	0.006
	3153	3144	
Stanford 10 Math	0.026	0.017	
	(0.022)	(0.022)	0.008
	3337	3330	
Stanford 10 Reading	-0.044*	-0.054**	
	(0.023)	(0.022)	0.009
	3338	3331	
Stanford 10 Science	-0.085***	-0.086***	
	(0.028)	(0.028)	0.269
	3334	3332	
Stanford 10 Soc. Studies	-0.055**	-0.058**	
	(0.025)	(0.025)	0.161
	3334	3332	
Incentivized Achievement Index	0.053**	0.046**	
	(0.021)	(0.021)	0.013
	3129	3122	
Non-Incentivized Achievement Index	-0.059***	-0.068***	
	(0.019)	(0.019)	0.019
	3098	3092	
C. Survey Outcomes			
Parents check Homework more	0.069***	-0.191***	
	(0.025)	(0.017)	0.000
	2315	1791	
Student prefers Math to Reading	0.091***	-0.125***	
	(0.023)	(0.025)	0.000
	2356	1812	
Parent asks about Math more than Rdg.	0.116***	-0.285***	

	(0.028)	(0.022)	0.000
	1909	1379	
Survey Outcome Index	0.257***	-0.596***	
	(0.092)	(0.083)	0.000
	1453	865	
D. Behavior and Motivation			
Attendance Rate	-0.019	-0.019	
	(0.112)	(0.112)	—
	3428	3428	
Behavioral Offense	-0.014	-0.014	
	(0.011)	(0.011)	—
	3428	3428	
Intrinsic Motivation Index	-0.077	-0.736***	
	(0.065)	(0.060)	0.000
	2137	1618	
Behavior Index	-0.071	-0.724***	
	(0.063)	(0.059)	0.000
	2137	1618	
E. Post-Treatment Outcomes (t+2)			
State Math	-0.025	-0.025	
	(0.035)	(0.035)	—
	2297	2297	
State Reading	-0.082***	-0.090***	
	(0.029)	(0.029)	0.057
	2290	2286	
Stanford 10 Math	-0.021	-0.037	
	(0.029)	(0.028)	0.002
	2409	2398	
Stanford 10 Reading	-0.076***	-0.095***	
	(0.028)	(0.027)	0.001
	2414	2400	
Stanford 10 Science	-0.046	-0.060**	
	(0.031)	(0.030)	0.014
	2401	2392	
Stanford 10 Soc. Studies	-0.065**	-0.085***	
	(0.030)	(0.030)	0.001
	2409	2397	

Notes: Column (2) reports attrition bounded estimates of the effects of our aligned incentives experiment in Houston on all outcome measures. All dependent variables are defined analogously to those in Table 4B. If treatment students were more likely to have valid outcome measures, the highest performing treatment students were dropped from the attrition bounded regressions. If treatment students were less likely to have valid outcome measures, the lowest performing control students were dropped from the attrition bounded regressions. Regressions follow the fully controlled specification as Table 4B, Column (2). Models with outcome variables taken in 2012-13 additionally include grade in 2012-13 fixed-effects. Standard errors are robust to heteroskedasticity. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table 6A: Mean Effect Sizes (Intent to Treat Estimates) in DC School-Level Clustering and School-Level Regressions

	Full Controls		
	2008-2009	2009-2010	Pooled
	(1)	(2)	(3)
Panel I. School-level Clustering			
A. Incentivized Outcomes			
Behavioral Offense	-0.086	0.051	-0.032
	(0.067)	(0.031)	(0.039)
	6039	3283	9322
GPA	0.118	—	—
	(0.083)		
	5802		
Complete Homework	0.100***	—	—
	(0.025)		
	3441		
Arrive on Time	0.058**	—	—
	(0.022)		
	3350		
Behavior Not a Problem	0.048*	—	—
	(0.027)		
	3331		
Incentivized Outcome Index	0.168**	—	—
	(0.078)		
	3079		
B. Student Achievement			
State Math	0.154*	0.109	0.139
	(0.090)	(0.114)	(0.096)
	5846	3176	9022
State Reading	0.179**	0.080	0.146*
	(0.084)	(0.085)	(0.082)
	5844	3189	9033
At or Above Proficient in Math	0.081*	0.051	0.070
	(0.040)	(0.050)	(0.043)
	5846	3176	9022
Advanced in Math	0.003	0.012	0.006
	(0.010)	(0.009)	(0.008)
	5846	3176	9022
At or Above Proficient in Reading	0.083**	0.029	0.063*
	(0.034)	(0.038)	(0.034)
	5844	3189	9033
Advanced in Reading	0.006	0.010	0.008
	(0.007)	(0.014)	(0.009)
	5844	3189	9033
Academic Achievement Index	0.164*	0.087	0.138
	(0.085)	(0.097)	(0.088)
	5828	3171	8999

C. Behavior and Motivation			
Attendance Rate	0.344	-0.171	0.184
	(1.405)	(0.953)	(1.163)
	6039	3283	9322
Work Hard in School	-0.005	—	—
	(0.020)		
	3361		
Push Self in School	0.007	—	—
	(0.016)		
	3338		
Intrisic Motivation Index	0.073	0.077	0.075
	(0.051)	(0.064)	(0.047)
	2766	1635	4401
Behavior Index	0.097	—	—
	(0.114)		
	2603		
Panel II. School-level Regressions			
A. Incentivized Outcomes			
Pct with Behavioral Offense	-0.078*	0.052	-0.013
	(0.039)	(0.033)	(0.031)
	34	34	68
GPA	0.156	—	—
	(0.130)		
	34		
Complete Homework	0.139	—	—
	(0.096)		
	34		
Arrive on Time	0.089	—	—
	(0.086)		
	34		
Behavior Not a Problem	0.062	—	—
	(0.071)		
	34		
Incentivized Outcome Index	0.116	—	—
	(0.182)		
	34		
B. Student Achievement			
State Math	0.122	0.007	0.081
	(0.152)	(0.154)	(0.074)
	34	34	68
State Reading	0.196	0.068	0.126**
	(0.151)	(0.129)	(0.062)
	34	34	68
Pct At or Above Proficient in Math	0.062	-0.026	0.022
	(0.075)	(0.078)	(0.037)

	34	34	68
Pct Advanced in Math	-0.025	0.022	0.003
	(0.030)	(0.014)	(0.012)
	34	34	68
Pct At or Above Proficient in Reading	0.095	0.004	0.047
	(0.079)	(0.055)	(0.035)
	34	34	68
Pct Advanced in Reading	0.017	0.028	0.024**
	(0.019)	(0.023)	(0.010)
	34	34	68
Academic Achievement Index	0.156	0.036	0.101
	(0.151)	(0.136)	(0.067)
	34	34	68
C. Behavior and Motivation			
Attendance Rate	2.286*	-0.815	0.994
	(1.080)	(1.662)	(0.707)
	34	34	68
Work Hard in School	0.014	—	—
	(0.066)		
	34		
Push Self in School	-0.010	—	—
	(0.036)		
	34		
Intrisic Motivation Index	0.093	-0.064	0.040
	(0.138)	(0.097)	(0.068)
	34	34	68
Behavior Index	0.204	—	—
	(0.186)		
	34		

Notes: Panel I reports student-level ITT estimates of the effects of our aligned incentives experiment in DC on various outcomes. All dependent variables are defined analagously to those in Table 4A, and all specifications are the same as those found in Columns (4)-(6) in Table 4A. Standard errors are clustered at the school level and are robust to heteroskedasticity. Panel II reports school-level ITT estimates of the effects of our incentives experiment on various outcomes. Test score outcomes and controls are collapsed on the mean at the school level. In Panel B, regressions include school-level controls for the percentage of students in each demographic category, as well as school level means of previous test scores and their squares. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table 6B: Mean Effect Sizes (Intent to Treat Estimates) in Houston
School-Level Clustering and School-Level Regressions

	Full	Math Quintile		
	Sample	Bottom	Top	p-value
Panel I. School-level Clustering				
A. Incentivized Outcomes				
Parent Conferences Attended	1.546***	1.381***	2.046***	
	(0.103)	(0.154)	(0.313)	0.053
	2053	389	265	
Objectives Mastered	1.083***	0.697***	1.751***	
	(0.138)	(0.099)	(0.178)	0.000
	3292	687	413	
Incentivized Outcome Index	1.170***	0.844***	1.675***	
	(0.099)	(0.097)	(0.199)	0.000
	2027	382	262	
B. Student Achievement				
State Math	0.076	0.001	0.179**	
	(0.060)	(0.068)	(0.078)	0.011
	3153	659	418	
State Reading	-0.039	-0.127*	0.044	
	(0.049)	(0.070)	(0.070)	0.021
	3128	653	417	
Aligned State Math	0.112	0.022	0.112**	
	(0.071)	(0.134)	(0.045)	0.405
	3153	659	418	
Unaligned State Math	0.031	-0.027	0.113**	
	(0.057)	(0.086)	(0.055)	0.073
	3153	659	418	
Stanford 10 Math	0.026	-0.014	0.181***	
	(0.047)	(0.048)	(0.053)	0.001
	3337	701	417	
Stanford 10 Reading	-0.044	-0.126**	0.161*	
	(0.050)	(0.049)	(0.080)	0.000
	3338	702	417	
Stanford 10 Science	-0.085	-0.057	-0.104	
	(0.054)	(0.093)	(0.074)	0.593
	3334	700	417	
Stanford 10 Soc. Studies	-0.055	-0.066	0.051	
	(0.050)	(0.073)	(0.088)	0.170
	3334	700	417	
Meets Min Math Std.	0.026	—	—	
	(0.020)			
	3,153			
Math Commended Perf.	0.016	—	—	
	(0.024)			
	3,153			
Meets Min Reading Std.	0.005	—	—	
	(0.015)			
	3,128			
Reading Commended Perf.	-0.004	—	—	
	(0.019)			
	3,128			
Incentivized Achievement Index	0.053	-0.021	0.180***	
	(0.052)	(0.056)	(0.061)	0.000
	3129	653	417	
Non-Incentivized Achievement Index	-0.059	-0.109*	0.031	
	(0.045)	(0.064)	(0.063)	0.029
	3098	647	416	
C. Survey Outcomes				

Parents check Homework more	0.069***	0.173***	0.148**	
	(0.023)	(0.053)	(0.065)	0.664
	2315	491	285	
Student prefers Math to Reading	0.091***	0.178***	0.012	
	(0.024)	(0.040)	(0.047)	0.004
	2356	498	291	
Parent asks about Math more than Rdg.	0.116***	0.016	0.153***	
	(0.023)	(0.041)	(0.056)	0.008
	1909	348	256	
Survey Outcome Index	0.257**	0.489**	0.674***	
	(0.099)	(0.218)	(0.205)	0.410
	1453	261	206	
D. Behavior and Motivation				
Attendance Rate	-0.019	-0.005	-0.245	
	(0.132)	(0.238)	(0.155)	0.250
	3428	721	418	
Behavioral Offense	-0.014	0.030	-0.068**	
	(0.014)	(0.033)	(0.029)	0.009
	3428	721	418	
Intrinsic Motivation Index	-0.077	0.167	0.339*	
	(0.061)	(0.125)	(0.186)	0.452
	2137	429	265	
Behavior Index	-0.071	-0.099	0.010	
	(0.083)	(0.140)	(0.257)	0.636
	2137	429	265	
E. Post-Treatment Outcomes (t+2)				
State Math	-0.023	-0.042	0.331***	
	(0.050)	(0.055)	(0.118)	0.000
	2297	484	314	
State Reading	-0.080***	-0.133**	0.053	
	(0.030)	(0.052)	(0.070)	0.017
	2290	477	315	
Stanford 10 Math	-0.018	-0.060	0.344***	
	(0.048)	(0.063)	(0.104)	0.000
	2408	517	315	
Stanford 10 Reading	-0.072*	-0.166**	0.134*	
	(0.042)	(0.070)	(0.076)	0.000
	2413	519	315	
Stanford 10 Science	-0.043	-0.038	0.049	
	(0.042)	(0.076)	(0.098)	0.285
	2400	515	315	
Stanford 10 Soc. Studies	-0.062*	-0.145**	-0.026	
	(0.036)	(0.063)	(0.093)	0.109
	2408	516	315	
Meets Min Math Std.	-0.021	—	—	
	(0.030)			
	1,557			
Math Commended Perf.	0.003	—	—	
	(0.008)			
	2,286			
Meets Min Reading Std.	-0.042***	—	—	
	(0.016)			
	2,290			
Reading Commended Perf.	-0.024**	—	—	
	(0.009)			
	2,290			
Panel II. School-level Regressions				
A. Incentivized Outcomes				
Parent Conferences Attended	1.753***	1.514***	1.683***	

	(0.346)	(0.344)	(0.413)	0.472
	46	45	42	
Objectives Mastered	1.221***	0.644***	1.685***	
	(0.289)	(0.135)	(0.320)	0.000
	50	50	47	
Incentivized Outcome Index	1.150**	0.937***	1.538**	
	(0.392)	(0.157)	(0.451)	0.003
	46	45	42	
B. Student Achievement				
State Math	0.082	0.003	0.278	
	(0.117)	(0.102)	(0.156)	0.003
	50	50	47	
State Reading	-0.018	-0.083	0.193	
	(0.075)	(0.115)	(0.158)	0.005
	50	50	47	
Aligned State Math	0.171	0.005	0.172	
	(0.134)	(0.176)	(0.125)	0.135
	50	50	47	
Unaligned State Math	0.036	-0.033	0.198*	
	(0.112)	(0.175)	(0.096)	0.027
	50	50	47	
Stanford 10 Math	0.036	0.133	0.208	
	(0.139)	(0.114)	(0.133)	0.393
	50	50	47	
Stanford 10 Reading	-0.060	0.010	0.234	
	(0.118)	(0.096)	(0.253)	0.091
	50	50	47	
Stanford 10 Science	-0.019	0.077	0.013	
	(0.107)	(0.183)	(0.180)	0.622
	50	50	47	
Stanford 10 Soc. Studies	-0.132	0.072	0.172	
	(0.123)	(0.126)	(0.229)	0.444
	50	50	47	
Pct Meets Minimum Math Standard	0.033	—	—	
	(0.040)			
	50			
Pct Math Commended Performance	0.009	—	—	
	(0.046)			
	50			
Pct Meets Minimum Reading Standard	0.017	—	—	
	(0.023)			
	50			
Pct Reading Commended Performance	0.010	—	—	
	(0.031)			
	50			
Incentivized Achievement Index	0.053	0.054	0.243	
	(0.113)	(0.095)	(0.137)	0.023
	50	50	47	
Non-Incentivized Achievement Index	-0.072	-0.005	0.113	
	(0.083)	(0.122)	(0.172)	0.266
	50	50	47	
C. Survey Outcomes				
Parents check Homework more	0.070	0.130	0.635**	
	(0.049)	(0.093)	(0.115)	0.000
	40	40	37	
Student prefers Math to Reading	0.168	0.176**	0.022	
	(0.092)	(0.065)	(0.235)	0.010
	40	40	37	
Parent asks about Math more than Rdg.	0.104	-0.001	0.396**	

	(0.065)	(0.072)	(0.127)	0.000
	46	45	42	
Survey Outcome Index	0.398	0.046	2.593***	
	(0.000)	(0.704)	(0.032)	0.000
	39	38	35	
D. Behavior and Motivation				
Attendance Rate	0.092	0.205	-0.134	
	(0.339)	(0.252)	(0.319)	0.098
	50	50	47	
Percent with Behavioral Offenses	-0.002	-0.022	-0.160***	
	(0.034)	(0.036)	(0.035)	0.000
	50	50	47	
Intrinsic Motivation Index	-0.054	0.166	0.770	
	(0.311)	(0.591)	(0.668)	0.014
	40	39	37	
Behavior Index	-0.140	0.232	-0.053	
	(0.109)	(0.279)	(0.793)	0.167
	40	39	37	
E. Post-Treatment Outcomes (t+2)				
State Math	-0.060	-0.010	0.361	
	(0.082)	(0.124)	(0.243)	0.005
	50	50	46	
State Reading	-0.094**	0.034	0.023	
	(0.031)	(0.118)	(0.258)	0.936
	50	50	46	
Stanford 10 Math	-0.099	0.044	0.251	
	(0.097)	(0.120)	(0.293)	0.170
	50	50	46	
Stanford 10 Reading	-0.123	0.108	0.074	
	(0.086)	(0.133)	(0.279)	0.819
	50	50	46	
Stanford 10 Science	-0.103	0.079	0.037	
	(0.077)	(0.104)	(0.214)	0.716
	50	50	46	
Stanford 10 Soc. Studies	-0.130	0.036	-0.017	
	(0.077)	(0.109)	(0.258)	0.695
	50	50	46	
Pct Meets Minimum Math Standard	-0.045	—	—	
	(0.051)			
	50			
Pct Math Commended Performance	-0.000	—	—	
	(0.009)			
	50			
Pct Meets Minimum Reading Standard	-0.060**	—	—	
	(0.024)			
	50			
Pct Reading Commended Performance	-0.038***	—	—	
	(0.011)			
	50			

Notes: Panel I reports student-level ITT estimates of the effects of our aligned incentives experiment in Houston on various outcomes for a variety of subsamples in the indicated year. All dependent variables are defined analogously to those in Table 4B, and all specifications are the same as those in Table 4B Column (2). Models with outcome variables taken in 2012-13 additionally include grade in 2012-13 fixed effects. Standard errors are clustered at the school level and are robust to heteroskedasticity. Panel II reports school-level ITT estimates of the effects of our aligned incentives experiment in Houston. Test score outcomes and controls are collapsed on the mean at the school level for the indicated subsample. In Panel B, regressions include school-level controls for the percentage of students in each demographic category, and matched-pair fixed effects. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table 7: Cost Benefit Analysis

	Cost per student per year (Treatment)	Cost per student per year (Control)	Gains per year in Reading	Standard Error	Sample Size	Gains per year in Math	Standard Error	Sample Size	IRR (%)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Early Childhood									
Head Start Impact Study	$ 9,049.56	$ 2,851.82	0.188	0.064	4,667	0.135	0.071	4,667	9.26
Charter Schools									
Apollo ES	$ 326.93	$ 0.00	0.072	0.039	75,474	0.184	0.060	75,474	36.10
Apollo SS	$ 1,691.75	$ 0.00	-0.012	0.022	95,395	0.146	0.031	95,395	19.28
Harlem Children's Zone ES	$ 19,272.00	$ 12,443.00	0.114	0.095	748	0.191	0.116	748	10.84
Harlem Children's Zone MS	$ 19,272.00	$ 12,443.00	0.047	0.033	1,449	0.229	0.037	1,449	11.92
SEED Schools	$ 39,135.19	$ 20,449.98	0.201	0.086	221	0.218	0.082	221	8.66
Teacher Incentives									
Talent Transfer Initiative	$ 594.65	$ 0.00	0.125	0.050	3,201	0.110	0.060	3,327	28.66
Teacher Certification									
Teach For America	$ 3,359.16	$ 0.00	0.030	0.040	1,800	0.150	0.040	1,800	10.94
Class Size									
Tenessee STAR	$ 4,607.94	$ 0.00	0.133	0.033	11,600	0.107	0.033	11,600	8.55
Managed Professional Development									
Success for All	$ 772.45	$ 0.00	0.090	0.060	2,108	-	-	-	13.97
Tutoring									
Experience Corps	$ 788.03	$ 0.00	0.075	0.067	881	-	-	-	13.81
Curriculum									
Enhanced Reading Opportunities	$ 1,899.84	$ 0.00	0.110	0.037	2,244	0.070	0.035	2,668	22.26
Financial Incentives									
DC	$ 971.07	$ 0.00	0.146	0.020	9,033	0.123	0.020	9,022	31.77
Houston	$ 635.71	$ 94.85	-0.049	0.027	3,153	0.088	0.026	3,128	14.68
Coshocton Incentive Program	$ 69.08	$ 0.00	0.010	0.045	873	0.133	0.049	873	51.08
New York, Dallas, Chicago	$ 323.47	$ 0.00	-0.008	0.018	26,873	0.008	0.022	27,049	-

Notes: This table presents a summary of the costs, treatment effects, and calculated internal rates of return (IRRs) for DC, Houston, and 14 other major education interventions as summarized in Fryer (*forthcoming*). We have included the experiments that mark major education policy interventions for which we could find reliable cost estimates. The IRR is the discount rate that sets the cost of each intervention equal to the discounted stream of future income benefits associated with increased student achievement. For additional details on the calculation for each experiment, please see Online Appendix D.

Appendix Table 8: Mean Effect Sizes (Intent to Treat Estimates) in HISD; Retakes and Highest Scores

	First Score (Main)		Highest Score		Precedence to Retakes	
	Baseline Controls	Full Controls	Baseline Controls	Full Controls	Baseline Controls	Full Controls
	(1)	(2)	(3)	(4)		
State Math 10-11	0.071***	0.076***	0.074***	0.075***	0.077***	0.078***
	(0.024)	(0.025)	(0.024)	(0.026)	(0.024)	(0.026)
	3153	3153	3153	3153	3153	3153
State Reading 10-11	-0.061**	-0.039	-0.072***	-0.055**	-0.067***	-0.053*
	(0.026)	(0.027)	(0.026)	(0.027)	(0.026)	(0.027)
	3128	3128	3129	3129	3129	3129
State Math 12-13	-0.037	-0.023	-0.037	-0.029	-0.039	-0.032
	(0.033)	(0.034)	(0.034)	(0.035)	(0.034)	(0.035)
	2297	2297	2286	2286	2286	2286
State Reading 12-13	-0.091***	-0.080***	-0.089***	-0.080***	-0.093***	-0.084***
	(0.028)	(0.029)	(0.028)	(0.029)	(0.028)	(0.029)
	2290	2290	2290	2290	2290	2290

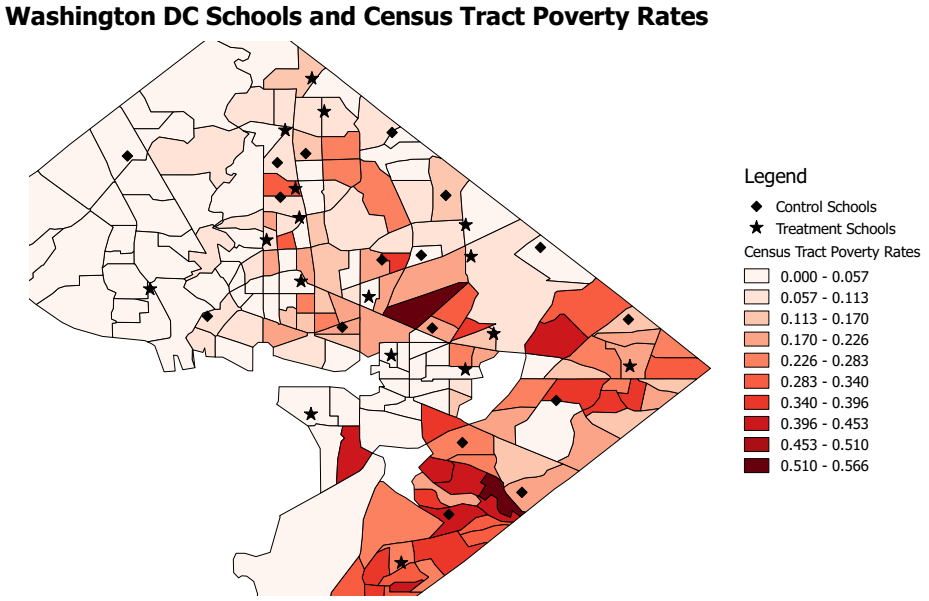
Notes: This table reports ITT estimates of the effects of our aligned incentives experiment in Houston on all outcomes. All specifications are identical to those in Table 4B. For outcomes measured in 2012-13, the specification additionally includes grade in 2012-13 fixed effects. In Columns (1) and (2), precedence is given to the first non-missing score. In Columns (3) and (4), test score variables are the highest score attained by each student (of duplicate test score entries, on-time tests, and retakes). In Columns (5) and (6), test score variables are a student's retest score unless it is missing, in which case it is the on-time score. Standard errors are robust to heteroskedasticity. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table 9: Mean Effect Sizes on Subscores of the Intrinsic Motivation Inventory

	Washington DC			Houston
	2008-2009	2009-2010	Pooled	2010-2011
	(1)	(2)	(3)	(4)
Standardized Intrinsic Motivation Index	0.073	0.077	0.075**	-0.077
	(0.045)	(0.054)	(0.034)	(0.065)
	2766	1635	4401	2137
Enjoy Schoolwork	0.031	0.018	0.027*	-0.001
	(0.020)	(0.027)	(0.016)	(0.030)
	3094	1759	4853	2307
Schoolwork is Fun	0.050**	0.026	0.039**	-0.006
	(0.021)	(0.027)	(0.017)	(0.030)
	3064	1747	4811	2297
Schoolwork is Not Boring	-0.013	-0.047*	-0.023	-0.042
	(0.022)	(0.028)	(0.017)	(0.031)
	3048	1734	4782	2256
Schoolwork Holds Attention	0.030	-0.005	0.020	-0.009
	(0.022)	(0.028)	(0.017)	(0.031)
	3006	1730	4736	2277
Schoolwork is Interesting	0.017	0.010	0.012	-0.010
	(0.021)	(0.027)	(0.016)	(0.031)
	3027	1729	4756	2288
Schoolwork is Enjoyable	0.026	-0.003	0.014	-0.036
	(0.021)	(0.028)	(0.017)	(0.030)
	3027	1730	4757	2268
Thinking about Schoolwork Brings Enjoyment	0.004	0.046*	0.020	-0.048*
	(0.021)	(0.027)	(0.017)	(0.029)
	3071	1751	4822	2261

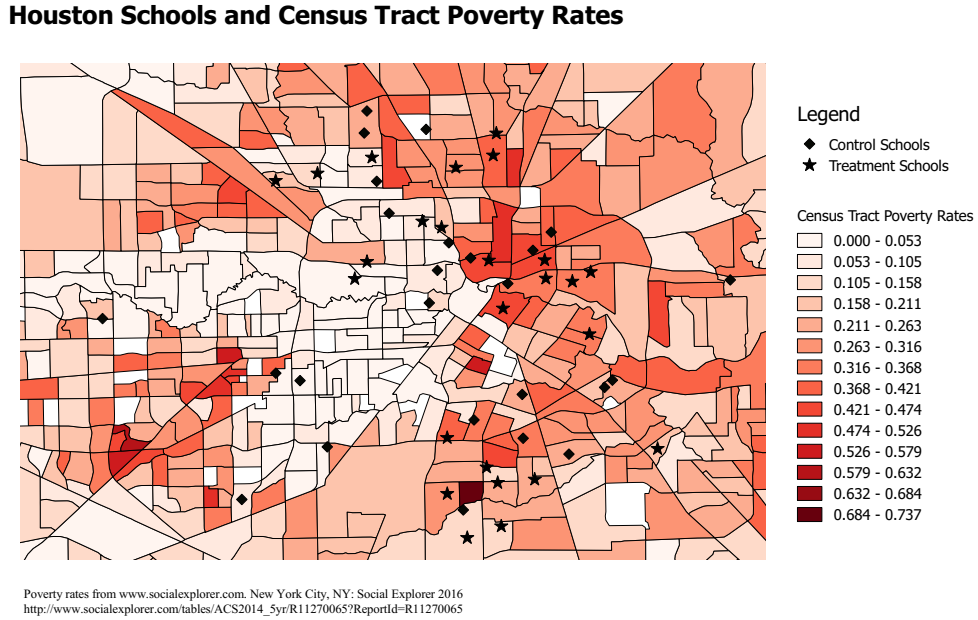
Notes: This table reports ITT estimates of the effects of our incentives experiment in DC and Houston on subscores of the Intrinsic Motivation Index. DC regressions follow the fully controlled specifications in Columns(4)-(6) of Table 4A, and Houston regressions follow the controlled specifications in Column (2) of Table 4B. The overall index measure in row one has been standardized to have a mean of zero and standard deviation one across all survey responses in each year. The subscores have been made into binary indicators for answering above the median on the original 5-option scaled responses. Standard errors are robust to heteroskedasticity. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Figure 1A:



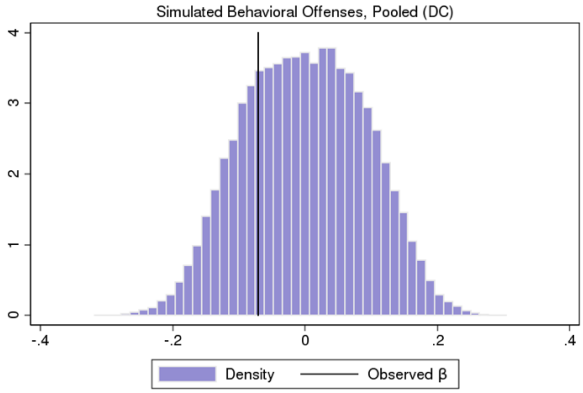
Poverty rates from www.socialexplorer.com. New York City, NY: Social Explorer 2016
http://www.socialexplorer.com/tables/ACS2014_5yr/R11270108?ReportId=R11270108

Appendix Figure 1B:

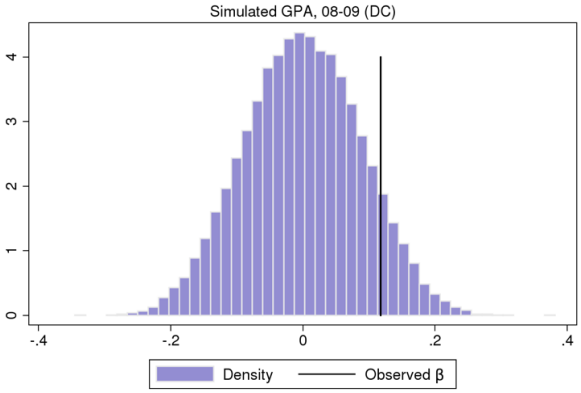


Poverty rates from www.socialexplorer.com. New York City, NY: Social Explorer 2016
http://www.socialexplorer.com/tables/ACS2014_5yr/R11270065?ReportId=R11270065

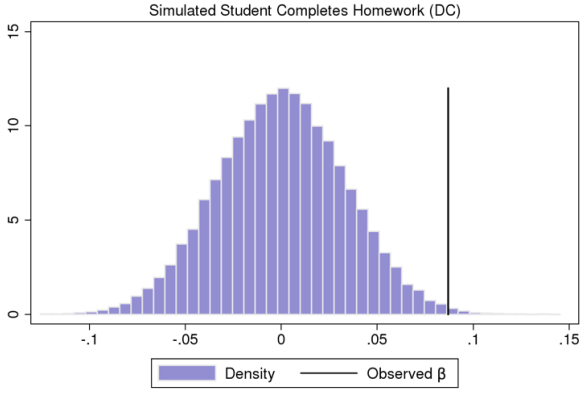
Appendix Figure 2A: Permutation Tests in DC



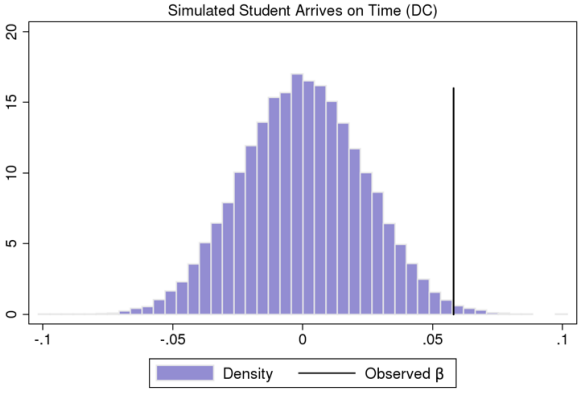
N = 50,000 Replications. **P-value** (two-tailed) = 0.486



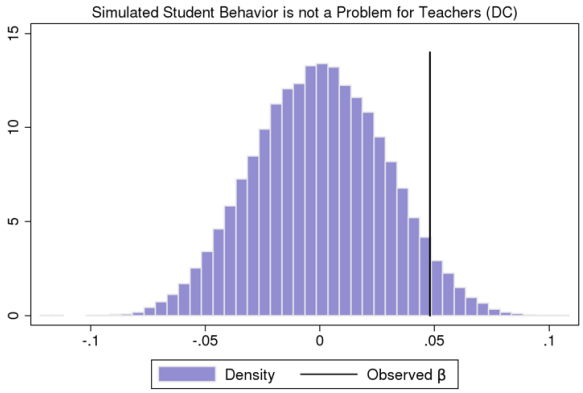
N = 50,000 Replications. **P-value** (two-tailed) = 0.185



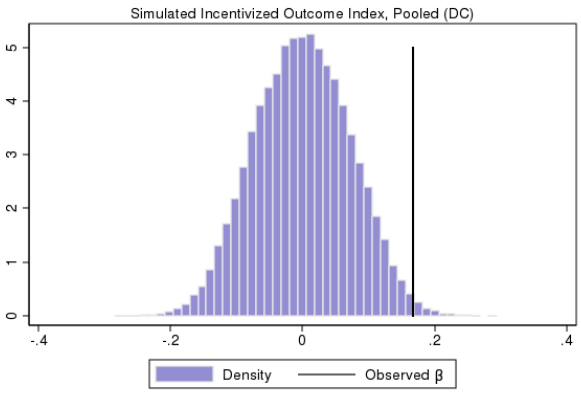
N = 50,000 Replications. **P-value** (two-tailed) = 0.008



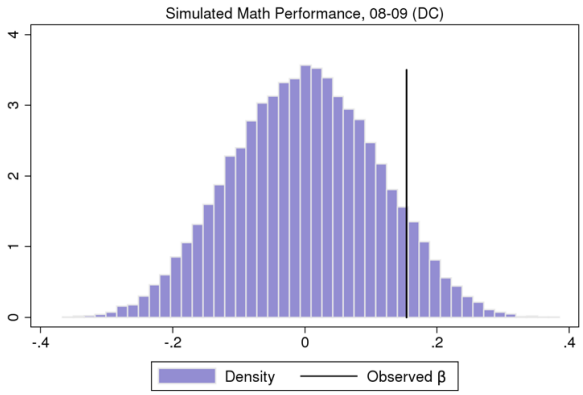
N = 50,000 Replications. **P-value** (two-tailed) = 0.012



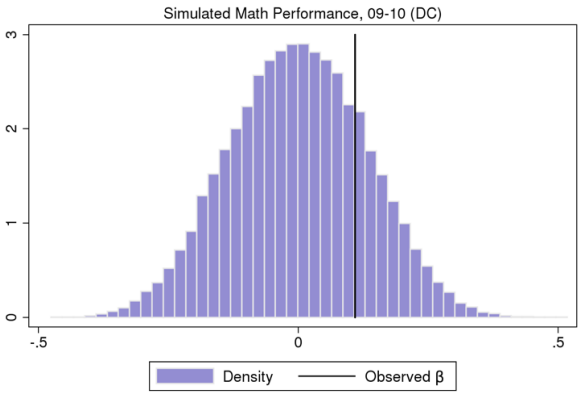
N = 50,000 Replications. **P-value** (two-tailed) = 0.094



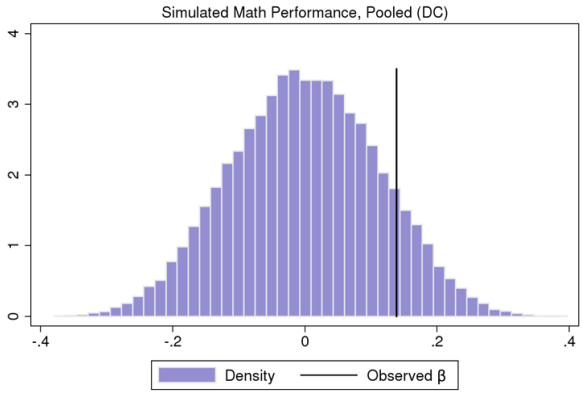
N = 50,000 Replications. **P-value** (two-tailed) = 0.015



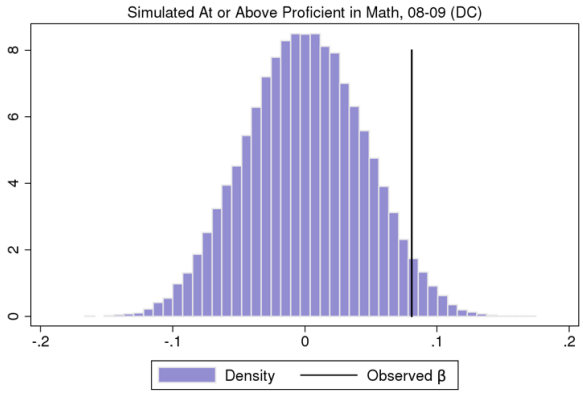
N = 50,000 Replications. **P-value** (two-tailed) = 0.169



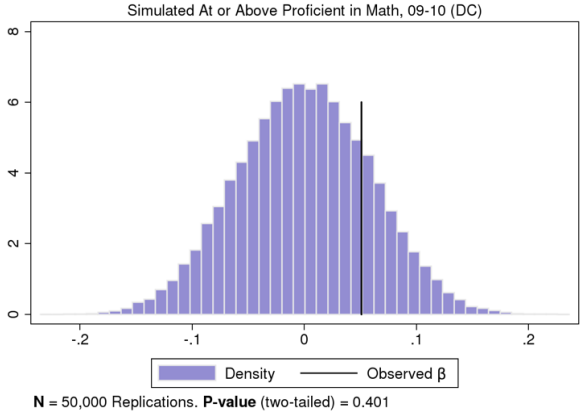
N = 50,000 Replications. **P-value** (two-tailed) = 0.422



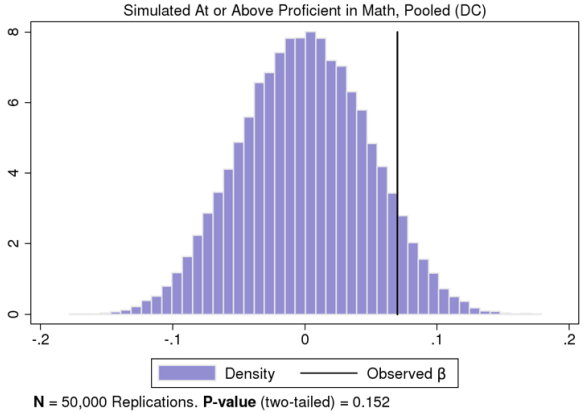
N = 50,000 Replications. **P-value** (two-tailed) = 0.224



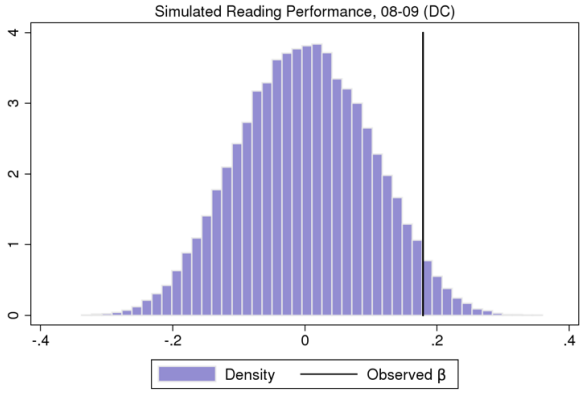
N = 50,000 Replications. **P-value** (two-tailed) = 0.069



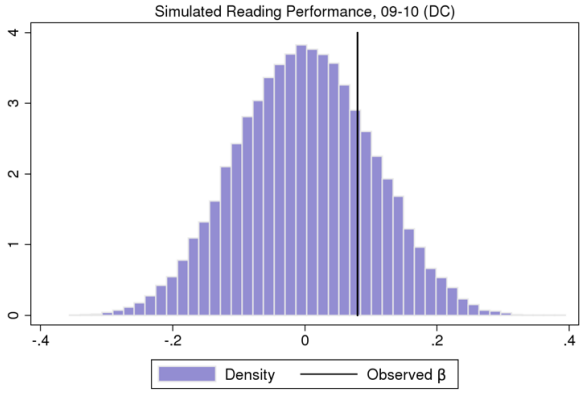
N = 50,000 Replications. **P-value** (two-tailed) = 0.401



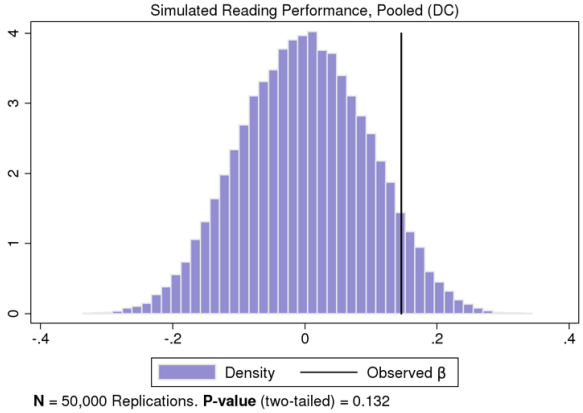
N = 50,000 Replications. **P-value** (two-tailed) = 0.152



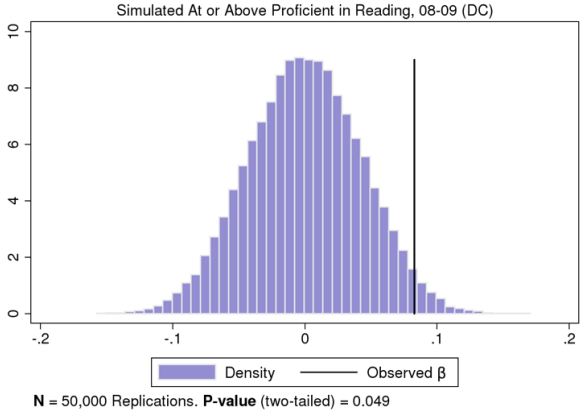
N = 50,000 Replications. **P-value** (two-tailed) = 0.068



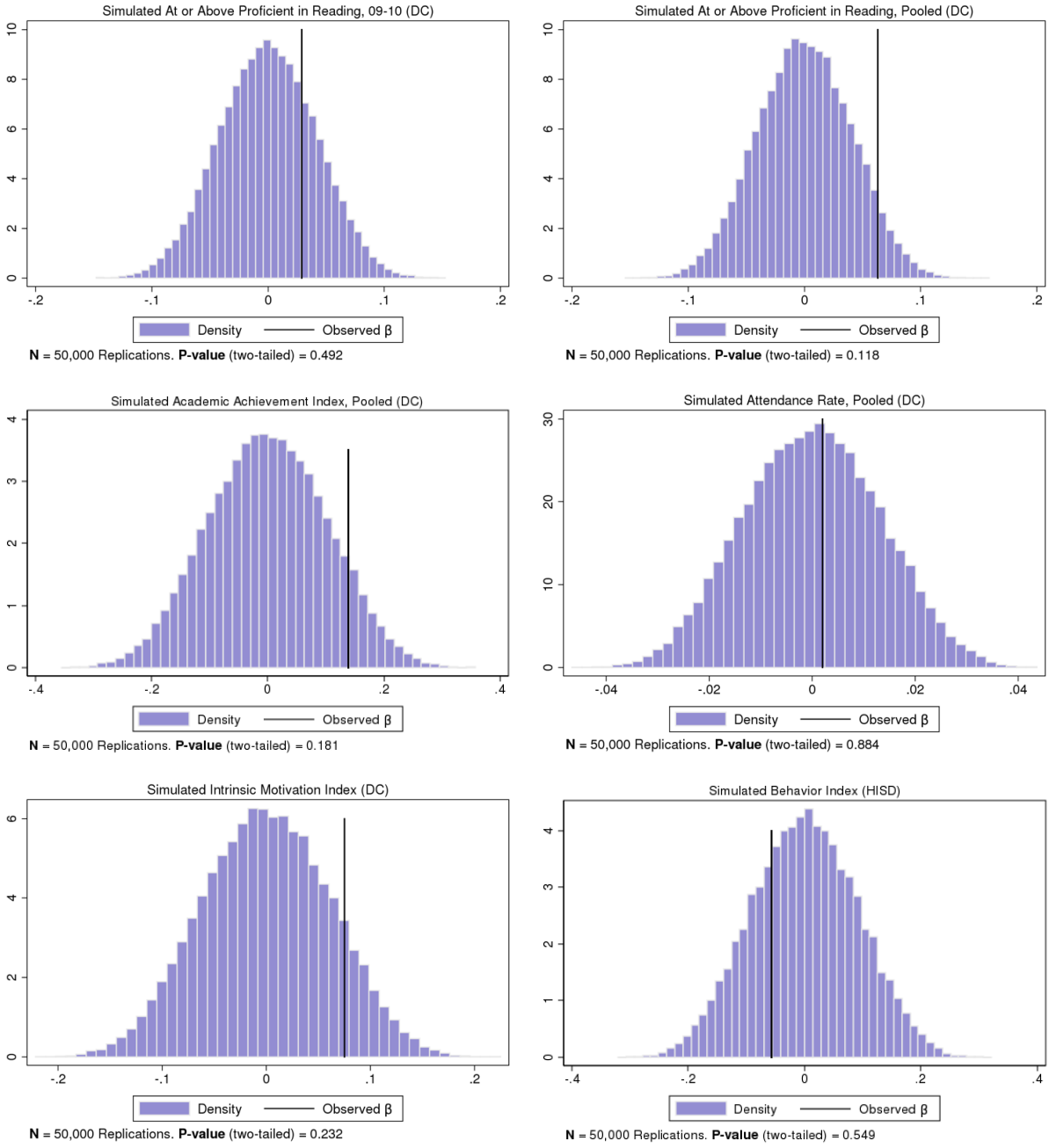
N = 50,000 Replications. **P-value** (two-tailed) = 0.443



N = 50,000 Replications. **P-value** (two-tailed) = 0.132

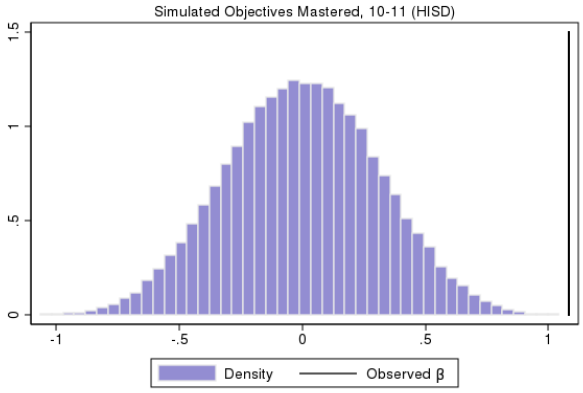


N = 50,000 Replications. **P-value** (two-tailed) = 0.049

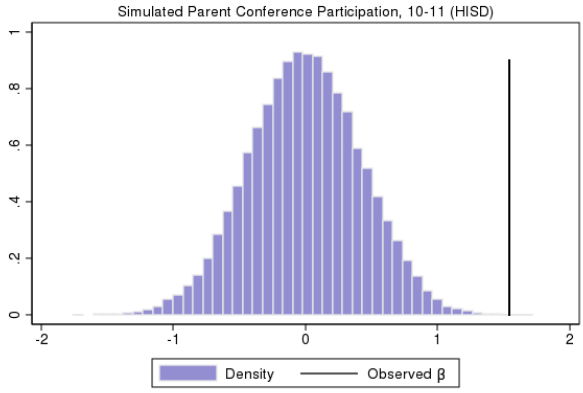


Note: This figure displays results of permutation tests (Rosenbaum 1988) in DC. We re-randomized 50,000 times following the same procedure as the original randomization: 17 schools were randomly assigned to treatment and 17 to control. We re-ran the regressions with the new, fake treatment assignments and recorded the new betas on treatment. Each graph plots the actual observed betas against the distribution of simulated betas.

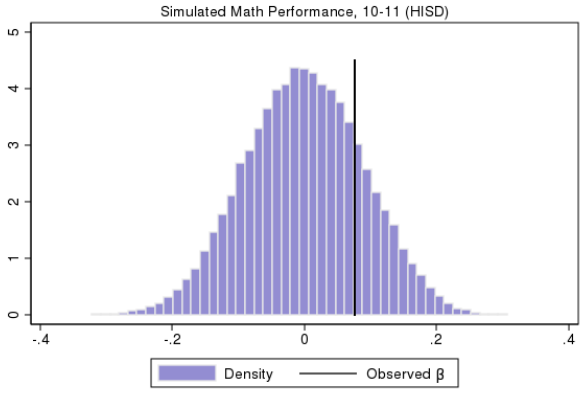
Appendix Figure 2B: Permutation Tests in Houston



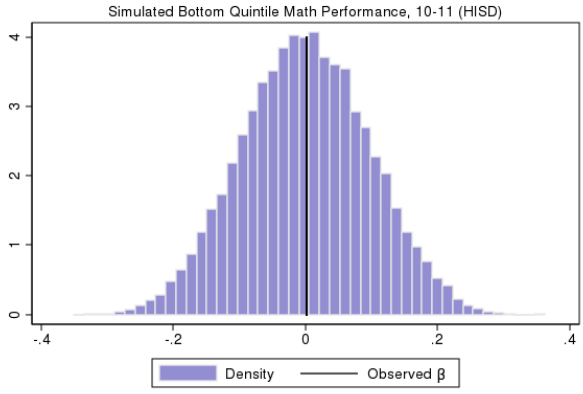
N = 50,000 Replications. **P-value** (two-tailed) = 0.000



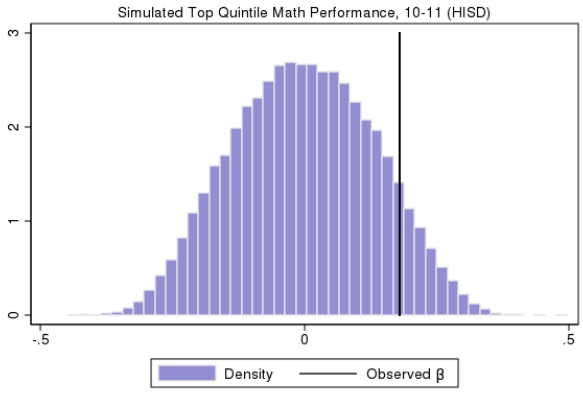
N = 50,000 Replications. **P-value** (two-tailed) = 0.000



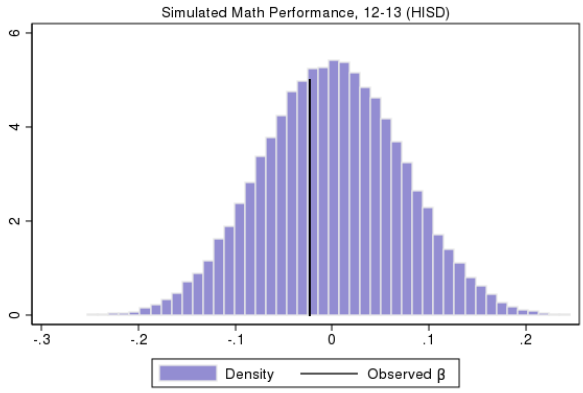
N = 50,000 Replications. **P-value** (two-tailed) = 0.407



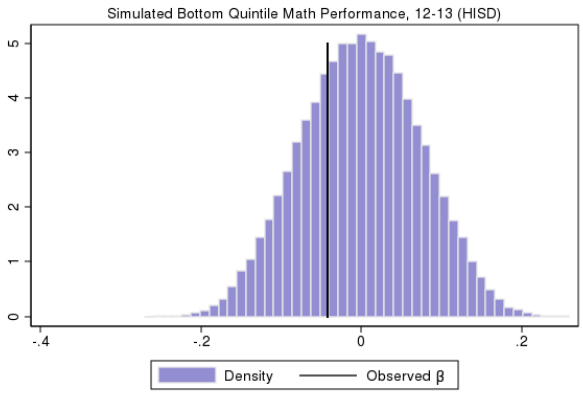
N = 50,000 Replications. **P-value** (two-tailed) = 0.993



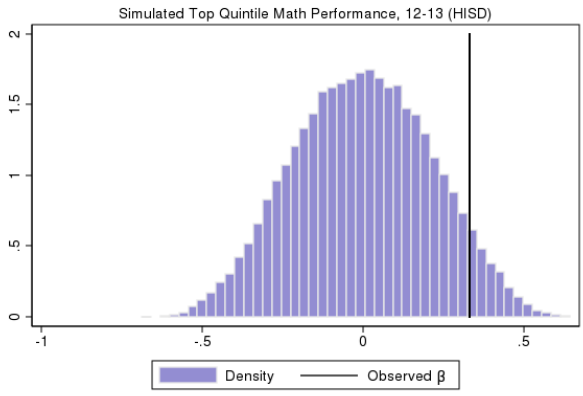
N = 50,000 Replications. **P-value** (two-tailed) = 0.193



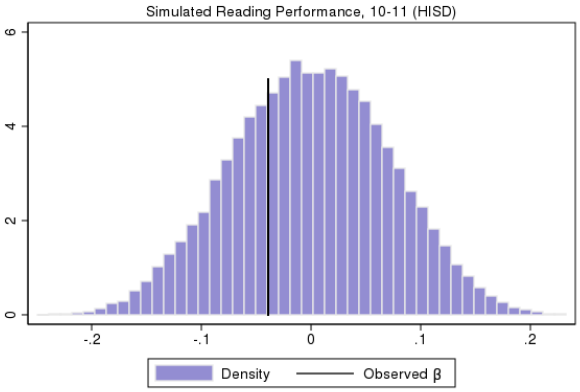
N = 50,000 Replications. **P-value** (two-tailed) = 0.756



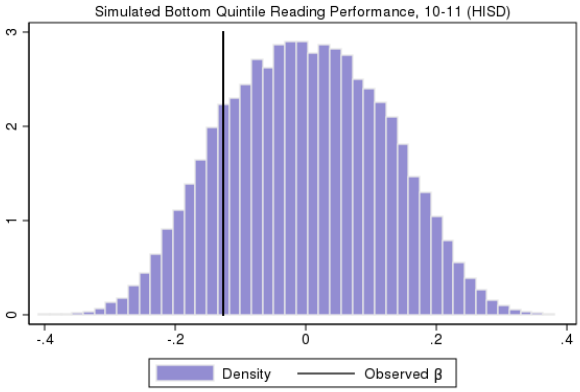
N = 50,000 Replications. **P-value** (two-tailed) = 0.586



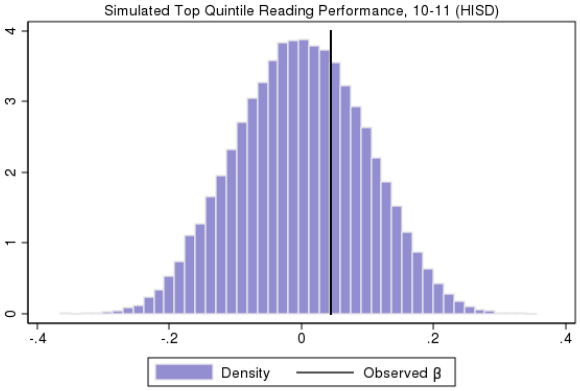
N = 50,000 Replications. **P-value** (two-tailed) = 0.123



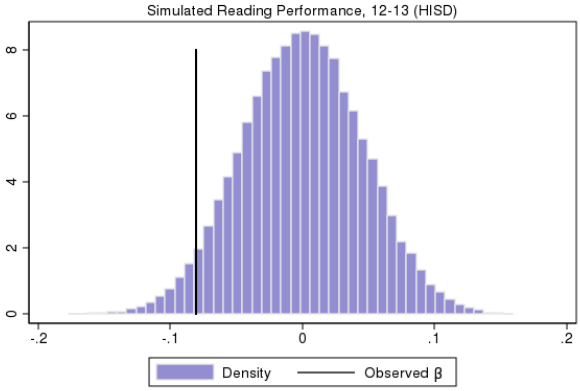
N = 50,000 Replications. **P-value** (two-tailed) = 0.605



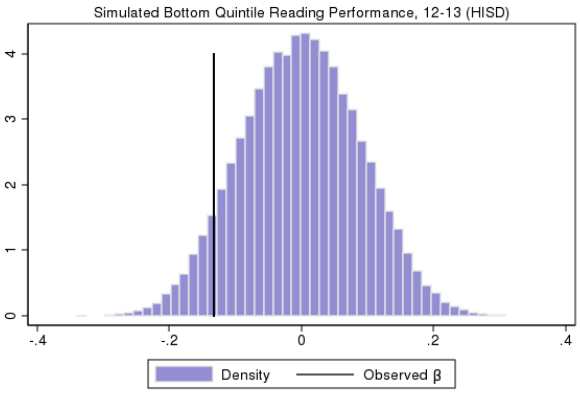
N = 50,000 Replications. **P-value** (two-tailed) = 0.334



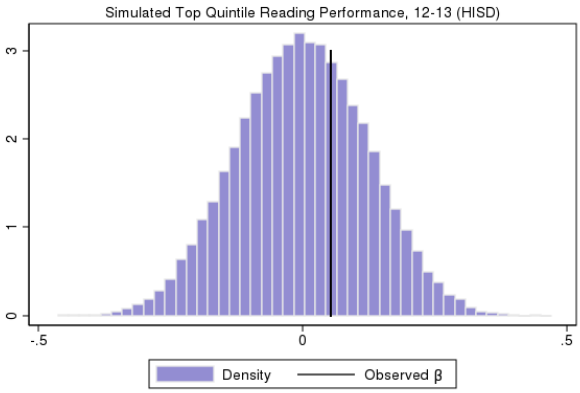
N = 50,000 Replications. **P-value** (two-tailed) = 0.664



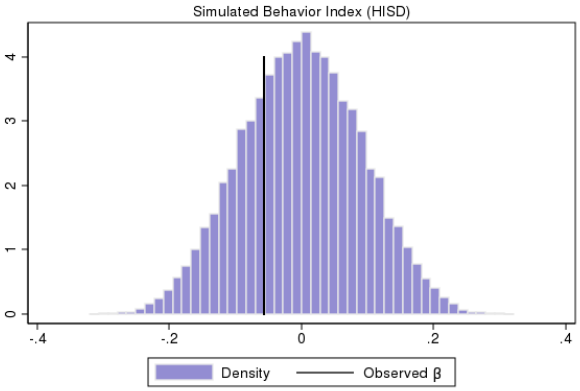
N = 50,000 Replications. **P-value** (two-tailed) = 0.076



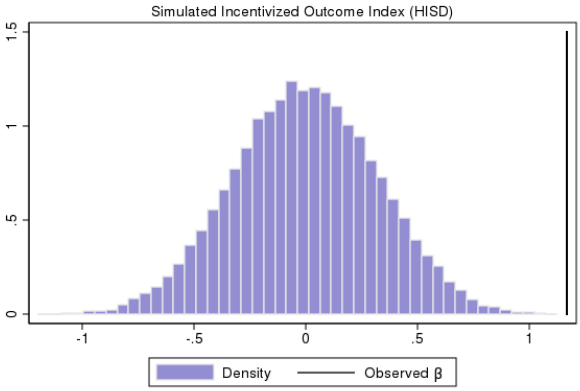
N = 50,000 Replications. **P-value** (two-tailed) = 0.141



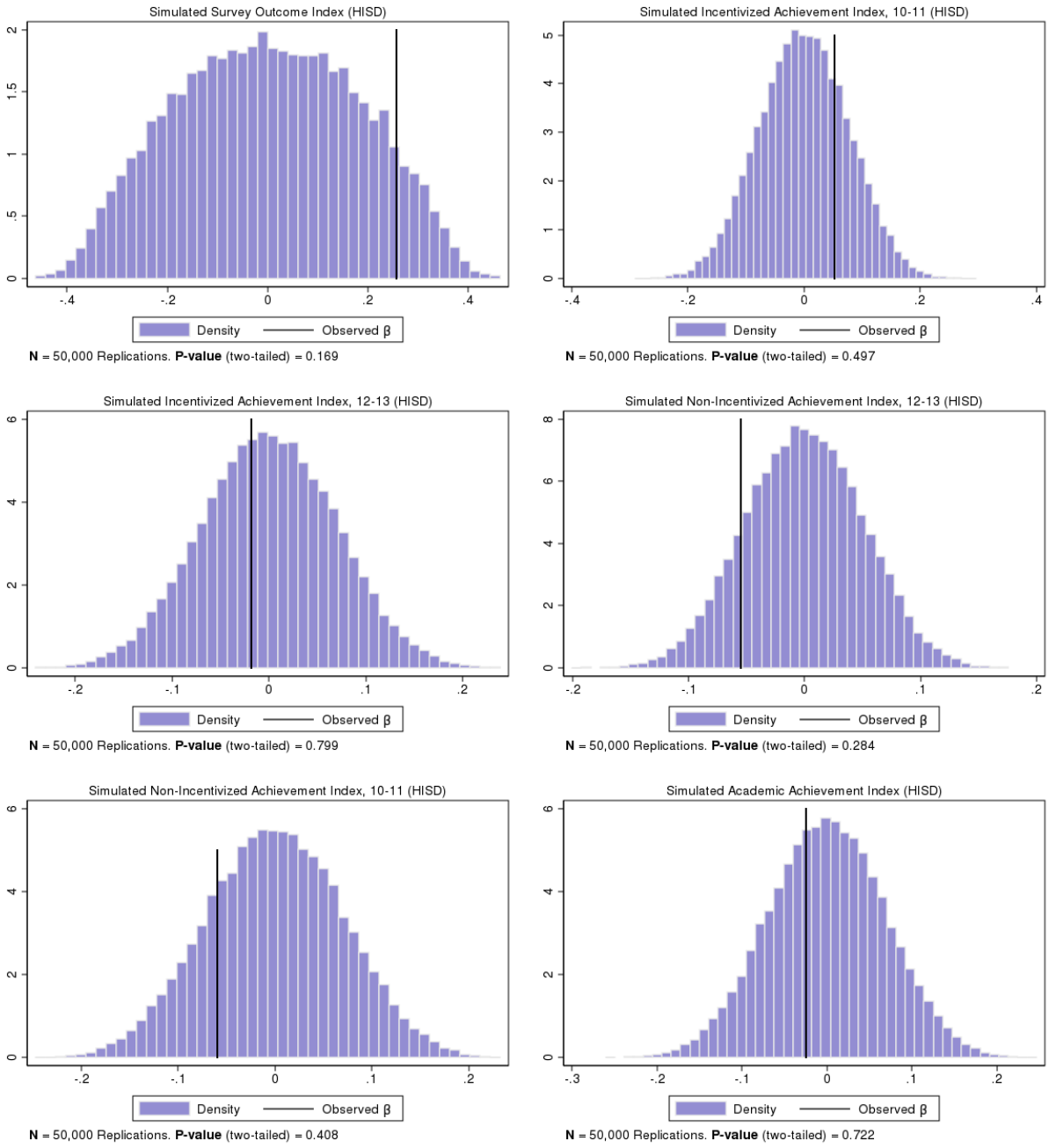
N = 50,000 Replications. **P-value** (two-tailed) = 0.675



N = 50,000 Replications. **P-value** (two-tailed) = 0.549



N = 50,000 Replications. **P-value** (two-tailed) = 0.000



Note: This figure displays results of permutation tests (Rosenbaum 1988) in Houston. We re-randomized the sample 50,000 times between matched pairs at the school level, just like the original randomization. We re-ran the regressions with the new, fake treatment assignments and recorded the new betas on treatment. Each graph plots the actual observed betas against the distribution of simulated betas.