

NBER WORKING PAPER SERIES

THE EFFECT OF MENTORING ON SCHOOL ATTENDANCE AND ACADEMIC OUTCOMES:
A RANDOMIZED EVALUATION OF THE CHECK & CONNECT PROGRAM

Jonathan Guryan
Sandra Christenson
Ashley Cureton
Ijun Lai
Jens Ludwig
Catherine Schwarz
Emma Shirey
Mary Clair Turner

Working Paper 27661
<http://www.nber.org/papers/w27661>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2020

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through grant R305A120809 to Northwestern University, by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) of the National Institutes of Health, through grant 5R01HD067500, and by the William T. Grant Foundation. The opinions expressed are those of the authors and do not represent views of the Institute for Education Sciences, the U.S. Department of Education, NICHD or the William T. Grant Foundation. This research could not have been completed without the hard work and support of Madeline Aden, Roseanna Ander, Eric Betzold, Philip Cook, Gretchen Cusick, Sarah Emmons, Megan Ferrier, Nathan Hess, Rachael Maguire, Farah Mallah, Angie Pohl, Letisha Smith, Karen Stout, Veronica Wald and Michael Weis. We are also very thankful to everyone who worked with us to support this research project at the Chicago Public Schools and SGA Youth and Family Services, including David Moskowitz, Carrie Bergen, Drew Beres, Molly Burke, Joshua Gray, Kimberlyn McNutt, Paige Ponder, Erin Starkey and Cynthia Wright. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2020 by Jonathan Guryan, Sandra Christenson, Ashley Cureton, Ijun Lai, Jens Ludwig, Catherine Schwarz, Emma Shirey, and Mary Clair Turner. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

The Effect of Mentoring on School Attendance and Academic Outcomes: A Randomized Evaluation of the Check & Connect Program

Jonathan Guryan, Sandra Christenson, Ashley Cureton, Ijun Lai, Jens Ludwig, Catherine Schwarz, Emma Shirey, and Mary Clair Turner

NBER Working Paper No. 27661

August 2020

JEL No. I20

ABSTRACT

In response to budget problems, many urban school systems reduced resources for getting students to come to school, like truancy officers. Chicago, for instance, went from 150 truancy officers down to, in 1991, a total of zero. Is that a good idea? We explore here the effects of increased support by a pro-social adult, or “social capital,” delivered through a structured student monitoring and mentoring program called Check & Connect (C&C). We carried out a large-scale randomized controlled trial with C&C in partnership with the Chicago Public Schools (CPS) to students in grades 1-8. Program participation decreased absences in grades 5-7 by 4.2 days, or 22.9 percent, but with no detectable effects on students in grades 1-4. We also did not find statistically significant effects on learning outcomes such as test scores or GPA, or any detectable spillovers to other students within the schools where the program was administered. The modest impacts per dollar spent, compared to previous evidence on either low-cost “nudges” or relatively intensive, higher-cost interventions, raise the possibility that, for very disadvantaged students, there may be decreasing but then increasing returns to program intensity for the problem of student disengagement.

Jonathan Guryan
Northwestern University
Institute for Policy Research
2040 Sheridan Road
Evanston, IL 60208
and NBER
j-guryan@northwestern.edu

Jens Ludwig
Harris School of Public Policy
University of Chicago
1307 East 60th Street
Chicago, IL 60637
and NBER
jludwig@uchicago.edu

Sandra Christenson
University of Minnesota
Educational Psychology 56
E River Pkwy Minneapolis,
MN 55455
chris002@umn.edu

Catherine Schwarz
University of Chicago
catherinemschwarz@gmail.com

Emma Shirey
University of Chicago
emmacshirey@gmail.com

Ashley Cureton
Johns Hopkins University
School of Education
2800 North Charles Street
Baltimore, MD 21218
acureto4@jhu.edu

Mary Clair Turner
Northwestern University
MaryClair.Turner@u.northwestern.edu

Ijun Lai
Mathematica
ijun.lai@gmail.com

INTRODUCTION

A key social-policy priority in the United States is to improve high school graduation rates for disadvantaged urban youth. Each of the nation's 10 largest public-school districts has a graduation rate below 80 percent (Common Core of Data, 2015). In Chicago, the site of our study, the four-year graduation rate is 77 percent overall (Common Core of Data, 2015) and about 57 percent for African American males (Healey, Nagaoka, & Michelman, 2014). Whatever increases we have seen in high school graduation rates over time have not been enough to keep pace with the changing demand in the labor market (Goldin & Katz, 2010; Murnane, 2013). Given the strong relationship between graduation and a range of other outcomes such as crime and health (Cutler & Lleras-Muney, 2010; Meara, Richards, & Cutler, 2008), inequality in schooling contributes to inequality in many other domains of American life.

While the decision to drop out of high school has received a great deal of attention, the problems that lead to dropout almost always start much earlier: with chronic school absences, or truancy. In Chicago, the site of this study, almost half of high school students miss at least 10 percent of the school year (over three weeks of missed school), a common threshold for chronic absence (Allensworth & Evans, 2016). These problems begin much earlier still: 12.9 percent of elementary students missed over four weeks of school (Jackson, Marx, & Richards, 2012), a group that is at greatly elevated risk of future high school dropout (Cook et al., 2016; Schoeneberger, 2011).¹ Similar patterns are seen in almost every major urban school system.

Unfortunately, very little is currently known about modifiable risk and protective factors that contribute to truancy, much less about the most effective possible remedies. While school

¹ Truancy is also associated with drug and alcohol use, early initiation of sexual activity, teenage pregnancy, and crime (Allensworth & Easton, 2007; Dryfoos, 1990; Hallfors et al., 2002; Huizinga & Jacob-Chien, 1998; Tait, 2004). For example, data from Illinois state prisons found that of the 182 male youth incarcerated in three medium-security youth prisons, 74 percent had previously been labeled chronically truant (Jackson & Marx, 2013).

districts have developed a wide range of policies and administrative systems to enforce truancy laws since schooling became compulsory in the late 19th century, few if any of these efforts have ever been subject to rigorous evaluation. Almost all of this research is observational and may confound the causal effects of truancy prevention programs with those of hard-to-measure attributes associated with either program implementation decisions (for school- or district-level analyses) or selection into program participation (for student-level analyses).

Perhaps partly in response to the field's limited understanding about the value of truancy prevention, such efforts often receive low priority in education policy decisions. For example, in the 1970s Chicago employed 150 truancy officers to serve the city's roughly 600 elementary, middle and high schools (WBEZ, 2014). Responding to budget pressures in 1991, in order to save \$4 million the Chicago school board reduced the number of truancy officers to ... zero.² That raises the question: Is this type of de-emphasis on truancy prevention a good idea or a bad idea?

The contribution of this paper is to carry out one of the few large-scale randomized controlled trials (RCTs) of a promising intervention to reduce truancy. The intervention we study here focuses on one important risk / protective factor identified by previous observational studies: social capital. Dating back at least to Coleman (1988), social scientists have thought the level of support children have from adults is strongly correlated with schooling outcomes. In many of our nation's most distressed urban areas it is challenging for adults to invest as much time and attention in children's outcomes as might be required because poverty, irregular (or long) work schedules, crime, transportation problems, child care challenges, and untreated

² Similar financial restrictions have led to reduced numbers of truancy officers in Santa Rosa, CA and Las Vegas (Los Angeles Times, 2003; Las Vegas Review-Journal, 2009).

mental or physical health problems make doing so difficult. These problems may be exacerbated when households have just a single adult that must handle all of the family's responsibilities.

The specific intervention we test seeks to supplement and support the social capital that parents can provide by randomly assigning students within the Chicago Public Schools (CPS) to receive Check & Connect (C&C). This program is a school-based, structured mentoring program that is designed to promote student engagement through relationship building, problem solving, and persistence for marginalized students. C&C has four components: (1) a mentor who works with individual students and their families, (2) regular check-ins by the mentor, (3) timely personalized interventions to reestablish student connection to school and learning, and (4) engagement with parents.

Two cohorts of students were assigned to the program to determine its impact on student attendance and achievement. The first cohort consisted of 487 participating students in 23 randomly selected elementary schools on the south and west sides of Chicago, who received C&C services during academic year (AY) 2011-12 and 2012-13 (including summers). The second cohort included 348 students in nine of the original 23 schools who received C&C services during AY 2013-14 and 2014-15 (including summers); 70 of the participating students in the second cohort were randomly selected from among the participating students from the first cohort. In general, students were eligible to receive C&C services if they had between 10 and 35 absences the previous school year. (More details are below). To detect spillover effects on non-participating students, we randomized at three different levels: the school, the grade, and the student. We used student-level, longitudinal administrative data collected by CPS to measure the program's impact on attendance, grades, and standardized tests scores.

Our findings suggest that C&C generated statistically significant reductions in absences for middle school-aged students, but not for elementary school-aged students. Based on estimates of the effects of treatment on the treated (TOT), we find that participation in C&C decreased student absences among students who began the program in grades 5-7 by a statistically significant 4.2 days, or 22.9 percent relative to the control complier mean. We do not find statistically significant effects of participating in C&C among students who began the program in grades 1-4. Across the two cohorts, the effect of participating in C&C was about 50 percent larger in the second year of the intervention than the first. While this difference was not statistically significant, it suggests that the development of relationships between the mentor and student could be an important mechanism through which the mentoring program is effective. We did not find statistically significant effects of C&C on academic learning outcomes such as achievement test scores or grade point average, although the size of our 95 percent confidence intervals do not allow us to rule out modest impacts on these outcomes.

While C&C was effective in that it improved key outcomes for a target population of students, it is useful to compare C&C to other interventions that have an impact on student engagement outcomes, as interventions vary in cost, complexity of implementation, and the mechanisms through which they affect change in student behavior and outcomes. As implemented in this project, the C&C program cost about \$1700 per student per year. This translates to a cost of approximately \$400 per incremental day of attendance brought about by the intervention for the 5th-7th grade students.

By way of comparison, other recent interventions attempted to improve attendance by leveraging insights from behavioral science to “nudge” people through information provision. This sort of intervention has the benefit of being automatized, low-cost, and scalable. For

example, Rogers and Feller (2018) and Robinson et al. (2018) find that a mail-based intervention providing parents of frequently-absent students with information about their students' attendance record decreases absences by about one day, at a cost of \$6-\$11 per additional attendance-day. It is clear this type of intervention operates through a different set of mechanisms than a more comprehensive and costly person-centered intervention like C&C, which suggests that a broader set of outcomes besides attendance may be required for evaluating relative effectiveness and cost-effectiveness. Other intensive interventions with costs similar to C&C have been shown to be effective at improving student engagement, even when school attendance was not the primary focus of the intervention (Heller et al., 2017) and to simultaneously produce large gains in other outcomes like school engagement, high school graduation, and delinquency.

It may ultimately turn out that for very economically-disadvantaged student populations, like those in our Chicago study, there may be decreasing but then increasing returns to program intensity to address the problems of attendance and school disengagement. That is, the impacts per dollar spent may be non-linear: potentially highest from the least-intensive *and* most-intensive programs compared to those of “medium” intensity. This is an important hypothesis for future research to examine.³

The next section reviews previous studies and relevant literature on the causes of absenteeism and attempts to combat it. Section three provides a detailed description of the C&C model. Section four discusses our experimental design including a description of how randomization was carried out for each cohort. Section five reviews our data for this study, as

³ A similar program of research seeks to understand the effectiveness of different levels of intensity for interventions for promoting college attendance and persistence for low-income students; see for example Bettinger and Evans (2019), Hyman (2019), and Oreopoulos and Ford (2019).

well as descriptive statistics and balance tests. Section six describes the analysis plan. The results are discussed in section seven, and we conclude in section eight.

PRIOR STUDIES

In order to understand what policies and programs might reduce student absenteeism, it is helpful first to understand what causes it. Balfanz and Byrnes (2012) categorize absent students by their agency and decisions to attend school, distinguishing between those who cannot attend (due to illness or housing instability), those who refuse to attend (to avoid bullying or unsafe conditions), and those who choose not to attend (because they are uninterested in school). The risk and protective factors that contribute to students falling into different categories can be school-based or within the family or community (see for example Chang and Romero, 2008).

Surely these different factors often interact in their effects. For example, a child with a sick younger sibling may stay home to provide care if the parent is unable to get off work to provide care themselves. Many affluent parents who have the advantage of working in more accommodating jobs, or being able to afford paid child care, would be able to send the child to school in the same case of sibling illness. Should we attribute the absence then to illness, or workplace problems, or unaffordable child care, or something else?

Recognizing these complexities, much of the research that has been done around student absenteeism argues that illness or health issues are the primary barrier to school attendance (Ehrlich et al., 2014; Kearney, 2008). This conclusion is often drawn from information collected through school administrative data, which may not be designed to detect more nuanced factors driving student absenteeism, especially those that occur outside of the school setting. However, further analyses reveal the significant role out-of-school economic and family circumstances may play in absenteeism. For example, one national survey found that children in single-mother

families are twice as likely as children in two-parent families to report missing at least 11 days of school the previous school year for health-related reasons (Bloom, Jones, & Freeman, 2013) and an analysis of administrative data from six states found students living in poverty are also more likely to be chronically absent (Balfanz & Byrnes, 2012).

Similarly, absences that stem from school refusal behavior are hard to quantify, as such behavior may be indicated by tardiness or incomplete absences, the definitions of which vary by district (Kearney, 2008). One study found that truancy rates jump for students who are transitioning school levels (elementary school to middle school and middle school to high school), which may indicate absences that are driven by student reluctance or anxiety about adjusting to a new environment, peers, and schedule (Garrison, 2006).

Students attending urban schools are also more likely to miss school. In a national survey of eighth grade students, those attending urban or city schools were more likely to report being absent three or more times in the past month than eighth grade students attending rural or suburban schools (Child Trends Databank, 2015). One study found that schools in high-poverty urban areas have up to one-third of their students chronically absent (Balfanz & Byrnes, 2012).

High levels of truancy clustered around certain grades reflect a pattern of absences that seem to evolve with the age of the student. A meta-analysis of administrative data from Oregon, Nebraska, Florida, and West Virginia found that chronic absenteeism goes down in third and fourth grades before sharply increasing in middle school, especially for students in 6th – 8th grades (Balfanz & Byrnes, 2012). This pattern makes clear that rather than being a static condition, absenteeism is often caused by multiple and shifting barriers to attendance.

Interventions to address student absenteeism often target one or only a few specific barriers to attendance. Clinical or medically-based interventions are sometimes deployed to

target youth with anxiety-based problems through pharmacotherapy or cognitive-behavioral strategies. Some work to influence the out-of-school environment of the student—by providing earlier family-school engagement or after-school programs. Others focus on providing additional professional development to teachers working with at-risk youth (Kearney, 2008).⁴

Some of these programs yield encouraging results in previous (mostly observational) studies: a review of absenteeism-prevention programs found that alternative education programs and behavioral programs may have positive impacts on attendance, academic performance, and graduation (Klima, Miller, & Nunlist, 2009). Conditional cash transfers to incentivize school attendance in Colombia have been shown to be effective in improving student outcomes (Angrist et al., 2002; Angrist, Bettinger, & Kremer 2006; Barrera-Osorio et al., 2008). However, providing financial incentives for improved attendance seems to be less common in the United States than outside of it. The results of the Opportunity NYC conditional cash transfer program, including the limited impacts on children’s attendance and other schooling outcomes, may have contributed to dampened enthusiasm for this approach in the U.S. (Riccio et al., 2013).

One potential limitation of many strategies is that they fail to support the personal relationships that are often vital for success within a socialized system, referred to initially by Coleman (1988) as “social capital.” Social capital exists in the relations between actors and, much in the same way that physical and financial capital do, facilitates productive activity. Coleman (1988) divides social capital into three forms: obligations and expectations, information channels, and social norms. Each of these forms provides a structure that promotes action and

⁴ Oreopolous et al. (2017) evaluate a program that combined mentoring, academic and other supports with the goal of improving a broad set of academic outcomes. The program, called Pathways to Education, included “mentoring, daily tutoring, and group activities, combined with intermediate and long-term incentives to reinforce a minimum degree of mandatory participation” and found positive effects on high school graduation and postsecondary enrollment rates.

may be targeted to support a specific behavior, like attending school. It is possible that mentoring programs designed to engage at-risk students may form and utilize social capital in a way that helps a student to change their behavior. Mentoring programs may develop trusted relationships that create perceived obligations by the student. Mentors may also provide information and influence social norms.

There is some evidence of the effectiveness of mentoring programs. Randomized controlled trials of Big Brothers Big Sisters, a national community-based and school-based mentoring program, have found significant impacts on improving academic achievement (Grossman et al., 2012; Bayer, Grossman, & DuBois, 2015; Herrera et al., 2007; Schwartz, et al., 2011) and decreasing unexcused absences (Grossman et al., 2012; Herrera et al., 2007; Schwartz et al., 2011). Other mentoring programs have also been found to improve socioemotional outcomes, including self-reported measures of depression (Herrera, DuBois, & Grossman, 2013), peer connectedness and self-esteem (Karcher, 2008), and pro-social behavior (Schwartz et al., 2011). One challenge with these studies is that they typically rely on self-reported outcomes. This may confound the effect of the intervention on actual behavior and outcomes with the possibility that youth assigned to mentors may be less willing to report socially undesirable outcomes for fear of disappointing their mentor (known in the survey research literature as “social desirability bias”).

THE CHECK & CONNECT (C&C) PROGRAM

In response to Chicago Public School concerns about truancy, and the decision several decades earlier to phase out all truancy officers for budget reasons, we visited the U.S. Department of Education’s What Works Clearinghouse (WWC) for what the available evidence suggests is best practice for improving school attendance. WWC suggested one intervention that

seemed particularly promising: Check & Connect, developed at the University of Minnesota by Sandra Christenson and other colleagues from the University of Minnesota, as well as various school personnel.

C&C is a structured mentoring program that aims to reduce the number of days that students miss school and to increase students' engagement with academic activities when they are in school. C&C has been implemented in several school districts around the United States, beginning in the Minneapolis Public Schools. C&C typically targets students who are at risk of disengagement or dropping out of school—often measured by high rates of absenteeism or poor academic performance—and assigns them to a mentor, who is typically an in-school staff member. Mentors are asked to monitor the attendance and school performance of the students on their caseload; serve as case managers, connecting students to social service and school-based resources that the mentors think might help the student to overcome barriers to school attendance; and develop relationships with the students on their caseload.

C&C is *standardized* in the sense that there is a manual and a set of training materials that can be used to implement the program, but it is also *adaptive* in the sense that mentors are encouraged to assess why different students are not coming to school and tailor the ways they intervene with students to match what they think students need.

The C&C manual and training directs mentors to support student engagement through two primary channels (Christenson, Stout, and Pohl, 2012). The “Check” component centers on monitoring student performance—tracking attendance, grades, and behavior referrals—for signs of disengagement. Mentors then deliver personalized interventions to students designed to boost engagement as part of the “Connect” piece. These interventions are supposed to be based on information the mentor has about the student's school engagement level and family

circumstances and to be shaped around available school and community resources. Mentors are also encouraged to connect with families of students, to partner with parents to increase student engagement, and to function as liaisons between home and school. In this study, mentors formally met with students, one-on-one or in small groups, an average of five times a month. Mentors also had informal check-ins with students through encounters that took place at recess, in the lunchroom, and hallway, among other school environments. On average, they connected with guardians through home visits or phone twice a month, although the level of family engagement varied substantially by mentor.

While there have been several previous randomized controlled trials of C&C, the most encouraging results to date have been limited to studies focusing on students receiving special education services (Sinclair, Christenson & Thurlow, 2005; Sinclair et al., 1998).⁵ The two RCTs that have been carried out with more general samples of high school students to date have yielded more mixed results. For example, Maynard, Kjellstrand and Thompson (2014) find some evidence for impacts on grades and disciplinary referrals, but no detectable impacts on the key outcome C&C was designed to change: absences. Heppen et al. (2017) found no detectable effects on any of the key student outcomes they examine, including educational attainment, academic performance, or various types of engagement such as attendance. However, this could be due to limitations of the studies themselves, which include small samples (just N=134 treatment-group students in Maynard et al. and N=276 treatment students in Heppen et al.) and in the case of Maynard et al., high rates of sample attrition (27 percent) that differed for treatment

⁵ C&C has also been implemented in a number of places outside of Minneapolis, but not structured to be studied like RCTs. For example, Tulsa and San Diego used C&C to support students at risk of discontinuing school or to serve youth with disabilities; Florida, Missouri, and Utah have developed state-wide initiatives that implement C&C in schools that have a high rate of students at risk of disengagement. The program has also been delivered to juvenile offenders and post-secondary students, although these contexts have been less common.

vs. control. In the case of Heppen et al., a high percentage of students was so credit deficient prior to C&C intervention from 9th grade performance that it was not possible for them to graduate on time. These results may indicate that C&C is more suitable for elementary or middle school youth than older students.⁶

The current study builds on this existing work by focusing on students earlier in their schooling careers (elementary and middle school), where beneficial impacts may be more likely, and on general samples of students who have exhibited previous absenteeism, not just restricted to those eligible for special education services. We also examine two full cohorts of program participants to try to improve statistical power through larger sample sizes relative to some previous research. Specifically, C&C was implemented in the CPS by a social service agency called SGA Youth and Family Services (SGA). Mentors were selected and hired by SGA to work as full-time C&C mentors. SGA initially hired 15 mentors to work in 23 CPS schools, and when mentors quit or were fired SGA hired replacements. SGA also employed a full-time project manager who served as the supervisor of the C&C mentors. The SGA project manager oversaw the work of the mentors, organized and led weekly meetings of the mentors, and provided guidance and feedback to mentors about how to work most effectively with the students. In addition, a project manager within CPS oversaw the implementation of the C&C program, oversaw the SGA project manager, and helped to collect data on participation and implementation. Once or twice each year, consultants from the research team at the University of Minnesota conducted training sessions with the C&C mentors to provide professional development and guidance on how to implement the C&C program with fidelity.

⁶ In addition, C&C's effect on elementary school students has been evaluated in the past (Anderson et al. 2004; Lehr, Sinclair, & Christenson, 2004), these studies were not experimental and have relied on comparing student outcomes to baseline measures, rather than a randomized control group.

Based on this previous work and conversations with CPS about the context and characteristics of the students to be targeted for intervention, the current study involved intervention for two years. Two cohorts of students received C&C services for two years each, and a small subset of students received the program for all four years.

EXPERIMENTAL DESIGN

In order to identify the causal effects of C&C on not just participating students but also others within the school, our research design involved random assignment of schools, and grades within schools, and students within grades. The specific eligibility criteria we used to identify students for cohort 1 (who participated AY 2011-12 and 2012-13) were somewhat different from those for cohort 2 (for whom the intervention took place during AY 2013-14 and 2014-15). Students who participated were assigned a C&C mentor and remained with that mentor for the full two years unless the mentor quit or was fired (in which case they were assigned a new mentor), or if the student moved too far away for it to be feasible for the mentor to continue providing services.

Cohort 1 Random Site Selection and Random Assignment

The cohort 1 randomization design is depicted in Figure 1. Random assignment took place in three steps. First, in collaboration with CPS, we went through a process of choosing 69 schools serving grades K-8 to be a part of the study.⁷ We tried to choose schools that were broadly representative of the district in terms of the demographic and socioeconomic characteristics of the students, and we wanted to ensure that the schools had a large enough group of students in each grade with absences in the range of 10-35 days in the prior year. We also took geographic location into consideration because in cohort 1, the C&C mentors were to

⁷ A total of 70 schools were invited to participate. One school declined.

be initially assigned to two schools each. We wanted to ensure that these schools were geographically close enough together to enable mentors to travel back and forth between them regularly. The 69 selected schools are mostly on the south and west sides of Chicago, in neighborhoods that range from some of the very poorest in the city and the country, to some with moderate poverty levels by a Chicago benchmark (but still quite high-poverty within the national distribution). The free or reduced-price lunch rates ranged from 71.2 to 99.8 percent in 2010-11 for the schools that were selected for the cohort 1 study.

We placed those 69 schools into groups of three, matching on geographic location, student race and ethnic demographics, and school-level absence rates, and randomly selected one school from within each group of three to be the location for treatment in the cohort 1 study. We conducted this first round of school-level random selection to allow for estimates of spillover effects on control students within the schools where C&C was implemented.

Then within each of the 23 cohort 1 study schools, we randomly selected five grades between the first and seventh to offer the program.⁸ The remaining two grades were in the control group and were intended to help identify spillover effects under the assumption that spillover effects might be more pronounced within grades than across grades.

Among the five selected grades we then placed students into groups of three, matched based on baseline absences, and randomly selected one of the three students to be offered

⁸ In CPS, the K-8 schools are called elementary schools even though they include students in grades that would commonly be in middle schools or junior high schools. Most elementary schools in CPS include grades K-8, and most high schools include grades 9-12. There are very few middle schools in CPS. We chose not to offer the program to students who would be in 8th grade in the first year of the study because the program was planned to last for two school years, and we thought it would be logistically difficult for mentors to follow students from their elementary school to a high school.

treatment. Students who had 10 to 35 absences in the prior year were eligible to be selected for the offer to participate in the C&C program.⁹

The students selected for treatment were sorted within school and grade in descending order based on baseline absences and were approached and offered the chance to participate in C&C in that order. In our analysis we always include all students randomly selected for treatment regardless of whether they were approached and offered the chance to participate. Thus, the ordering of students for the offer of treatment does not bias our results. The ordering did induce students with higher baseline absence rates to be more likely to be compliers, which means that if there are heterogeneous treatment effects the estimated effects of participating are for students with baseline absence rates somewhat towards the higher end of the 10 to 35 range. We include all students, not just those offered treatment, because we re-randomized some cohort 1 treatment group students into treatment and control groups in cohort 2 in a way that was not conditional on having been offered treatment in cohort 1. Though restricting the analysis to randomization blocks that include students who were offered treatment could generate an unbiased estimate of the treatment effect in cohort 1, it is necessary to include all randomized students from cohort 1 in the analysis to maintain the validity of randomization for cohort 2.

From the list of students randomly selected for treatment, students were offered the chance to participate until the predetermined caseload for the mentor serving each school was filled. Schools were put into two categories based on the school's enrollment. A total of 15

⁹ Because the size of the randomization blocks were not all multiples of three, some of the "triples" had 2, 4, or 5 students in them. In the main analyses reported below, we ignore the grouping of students into triples because, aside from the existence of the handful that had 2, 4 or 5 in them, within each student triplet the probability of being selected for treatment was constant at 1/3. If each block of students had exactly 3 students, the triple fixed effects would therefore be uncorrelated with treatment assignment and omitting them would not cause bias. In practice, while this is not literally the case, because there are few blocks that do not have exactly 3 students, excluding the triple fixed effects does not change the point estimates appreciably. We report results including triple fixed effects in an appendix table..

mentors were either assigned to one larger school – large enough to support a caseload of 30 plus a comparison group – or to two smaller schools. Mentors assigned to a larger school began the school year with a caseload of 30 students all in a single school, while mentors assigned to two schools began with caseloads of 15 in each of the two schools for a total of 30 students.

Cohort 2 Random Site Selection and Random Assignment

In the summer between the end of cohort 1 and the beginning of cohort 2, we conducted a second round of random assignment. The cohort 2 randomization design is depicted in Figure 2. There was enough funding to support nine mentors and based on feedback from the mentors and from CPS, we decided to assign each of the mentors to a caseload from a single school for cohort 2. Nine of the 15 mentors were invited to continue. Of the original 23 schools in the cohort 1 study, several were closed as a part of school closings that occurred at the end of the 2012-13 school year. The nine schools where the nine returning mentors primarily worked were selected to continue the program for cohort 2.

Within the cohort 2 study schools, we placed students into five randomization blocks. Three of the randomization blocks were for students who had been in the cohort 1 study, one block for students who had been assigned to treatment in cohort 1, one for students who had been assigned to control in cohort 1 who had baseline absences between 10-35, and one for students who had been assigned to control in cohort 1 who had baseline absences outside the 10-35 range. This will allow us to experimentally test whether getting four years of participation generates larger effects than two years of participation. The program was offered to students who were in grades 1-7 in the first year of cohort 1; students in both cohort 1 and cohort 2 studies were therefore in grades 1-5 in the first year of cohort 1, and in grades 3-7 in the first year of cohort 2. To fill in the two earlier grades for cohort 2, so that we would also have data on 1st and 2nd grade

students in cohort 2 (and so cover a similar range of grades as in cohort 1), we created a randomization block of students who were in 1st and 2nd grade in the first year of cohort 2. We also created a block of students who were new to the cohort 2 schools since the randomization for cohort 1. Since the probability of selection into treatment was not equal across all randomization blocks in cohort 2, we include randomization block fixed effects in all cohort 2 models. Finally, we offered principals at the schools the opportunity to nominate students to be in the study. Students nominated by principals were placed in their own randomization block within each school and subject to random assignment.¹⁰

DATA, DESCRIPTIVE STATISTICS, AND TESTS OF BASELINE BALANCE

In this section we describe the student-level school records we use to measure baseline characteristics and outcomes, document the level of disadvantage among the students in our study sample, and confirm that random assignment appears to have been carried out correctly.

Data

The data for this study are drawn from longitudinal student-level records from the Chicago Public Schools (CPS) for AY 2010-2011 through 2014-2015, and program participation data collected by C&C mentors and a CPS project manager. The CPS data include demographics, attendance, enrollment, misconduct, and achievement outcomes. The demographic data include each student's birth date, race / ethnicity, eligibility for free and reduced-price lunch, and an indicator for having a learning disability (indicated by having an Individualized Education Plan, or IEP).

¹⁰ In a few cases, principals submitted lists of nominated students after randomization had already taken place for their school. In these cases, we followed the initial random assignment for nominated students who were already subject to random assignment in one of the other randomization blocks, and then created a randomization block consisting of the nominated students who were not in one of the other randomization blocks.

Because a critical focus of C&C is to reduce student absences, measures of attendance and absences are the primary outcome variables in the analyses. The data include measures of attendance and absences: days present, meaning the number of enrolled days a student attended school over the school year; days absent, meaning the number of enrolled days a student was absent over the school year; percent present, meaning the percentage of enrolled days a student was present over the school year; and membership days, meaning the total number of days the student was officially enrolled in a CPS school. Membership days are the sum of days present and days absent, but do not necessarily equal the total number of school days in the CPS school year. Students can move in the middle of the school year and leave CPS, and it is also possible that students might not accumulate membership days for a short period when they transfer from one school to another within CPS.

Our analyses of achievement outcomes use annual grade point average (GPA) and math and reading test scores. The test score data come from two different sources because CPS administered multiple tests for elementary school students over the study period.¹¹

The C&C program manager at CPS maintained records of which students were approached and offered the chance to participate in C&C and which of those students participated in the program. We use this information to measure the C&C participation rates used

¹¹ The first test for which we have reading and math test score data is the Illinois Standards Achievement Test (ISAT), which was state-mandated during the 2011-12, 2012-13, and 2013-14 school years. The ISAT was administered in math and reading among 3rd through 8th graders annually in the spring during each of these school years, but was discontinued after the 2013-14 school year. The ISAT measures individual student achievement relative to the Common Core State Standards. During the first three academic years of the study, ISAT scores were used to determine Adequate Yearly Progress for schools as part of the district and state school accountability systems. The second standardized test we analyze is the Measures of Academic Progress (MAP), published by the Northwest Evaluation Association (NWEA). The MAP was administered in each of the four years of the study in math and reading to students in grades 3 through 8, and additionally to students in grade 2 in the final two years of the study. The MAP assessment is a computerized, adaptive test that changes which questions to ask students based on an estimate of ability level as indicated by responses to previous questions. For all four study years MAP scores were used as part of the growth component of teachers' evaluations, and for the 2014-15 school year the MAP scores replaced ISAT scores in the state school accountability system.

to estimate the effects of participation (treatment on the treated) described below. Finally, we conducted semi-structured interviews with students and their parents to understand the explicit and implicit reasons for student absences. We attempted to contact a random sample of 76 students and a parent or guardian from the full C&C intervention group and interviewed 51 of the 76 student/parent pairs.

Descriptive Statistics and Balance Tests

Tables 1 and 2 present baseline descriptive statistics and randomization balance tests for the students in the C&C schools. We focus for now (and in our initial impact analysis) on just those grades selected to receive C&C, so that the control group consists of students in treatment grades who were randomly assigned to the control condition.

Table 1 shows that for cohort 1, in the year before random assignment (2010-11), control and treatment students attended on average 150.2 and 151.0 out of a possible 170 days of school.¹² Nearly 60 percent of students in the study sample were African-American and just under 40 percent were Hispanic. On average, students failed 0.70 and 0.68 courses in control and treatment, respectively, in the year prior to the study, and the average grade point average was 2.21 in the control group and 2.24 in the treatment group. About 10-11 percent of the students had documented learning disabilities, and between 15-16 percent of the students were old for grade. The p-value on the F-test that all variables in Table 1 are jointly the same on average for the treatment and control group, which comes from regressing treatment assignment against these variables together with school-by-grade fixed effects, is 0.775 without missing data

¹² As a reminder, total days present and total days absent do not sum to the length of the school year because some students were not enrolled as a CPS student for the entire school year, possibly because they moved out of the district. For this reason, in the analyses below, we report effects on days present and days absent separately and analyze the treatment effect on the sum of days present and days absent, which is called membership days.

indicators and 0.840 with missing indicators. This is consistent with the idea that random assignment was indeed random.

Table 2 presents data for cohort 2 for the year prior to random assignment (2012-13). In the year before random assignment, on average control and treatment students attended 159.2 and 159.9 out of a possible 181 days of school.¹³ Relative to the cohort 1 study sample, there were fewer African-American (46 versus 57 percent) and more Hispanic students (48 versus 39 percent) in the cohort 2 study schools. The F-test statistic for the test of whether the mean values for all variables in Table 2 are equal for treatment and control groups yields a p-value of 0.650 (0.145 with missing indicators).

The information gathered from the interviews with students and their parents or guardians provides some insight into the reasons for absences for students in both cohorts. Typical reasons for absences included suspensions, illnesses, unreliable transportation, safety, and familial factors (such as needing to stay home to watch younger siblings). Overall, most families reported multiple factors and competing demands that influence school attention.

Table 3 shows participation rates for students assigned to treatment and control in each cohort. For cohort 1, during the first year of the program about 50 percent of those assigned to the treatment group participated in C&C (this is approximately equal to the figure we use to scale up our intent to treat estimates below to estimates of treatment on the treated). To get a sense for why students do not participate, we note again that not all students assigned to the treatment group were approached and offered treatment. There was a predetermined number of funded spots in the program and we could not predict how many students would accept the offer to

¹³ The CPS school year was extended from 170 to 181 days in 2012-13, the second year of the cohort 1 intervention and the baseline year for cohort 2. In 2013-14 and 2014-15, years one and two of the cohort 2 intervention there were 178 and 180 total days of school in CPS.

participate, so we randomly assigned more students to treatment than could be served and offered assigned students the chance to participate until we filled the program spots.

All students who were randomly assigned were included in the analysis sample, and students were analyzed based on their treatment assignment, not based on whether they were approached and offered conditional on being assigned to treatment. Among those who were actually approached and offered the chance to participate in C&C, 77 percent participated in cohort 1 and 79 percent participated in cohort 2. We can see that the offer rates were somewhat lower for cohort 2 than cohort 1, driven by the fact that some students left the treatment schools between the time of the original cohort 1 random assignment and the cohort 2 random assignment two years later. These students were not available to be offered treatment, but were considered treatment students in the analysis if they were randomly assigned to treatment in cohort 2, and were effectively never-takers in the cohort 2 analysis. Table 3 also shows that there was essentially no treatment “crossover” (no controls formally received C&C).¹⁴

Table 4 presents a descriptive analysis of which students chose to participate from among those offered the chance. The table shows results from linear probability regressions of an indicator for participation on baseline variables and a full set of randomization block fixed effects. The regressions only include students assigned to treatment. In cohort 1, students with more baseline absences and lower baseline GPA were more likely to participate conditional on being selected for treatment. This is most likely due to the fact that we rank-ordered randomization blocks (or “triples”) to approach students for the program, so those triples with

¹⁴ No students assigned to control were assigned a Check & Connect mentor or were offered a chance to formally participate in the program. Since the mentors worked in the schools every day, it is possible that they interacted with control students in a way that we are not able to measure. One nice feature of the experimental design is that because randomization occurred both at the school and student level, we are able to experimentally test whether there were spillover effects on control students in treatment schools.

more baseline absences were approached first during cohort 1. This affects who the compliers were, since we analyze the data based on randomly assigned treatment status – and random assignment occurred within all triples, even if the treatment students within the triple were not offered the program. In cohort 2, when we randomly ordered the triples to approach for the program, none of the baseline variables were significantly related to participation.

ANALYSIS PLAN

Estimating Program Impacts

Our goal is to estimate the causal effect of participating in the C&C program on student attendance and academic outcomes. We rely on the random assignment of the offer to participate in the program to identify the causal effect. We present two types of estimates: intent to treat (ITT) and treatment on the treated (TOT). The ITT estimate comes from estimating equation (1):

$$(1) \quad Y_{it} = \pi_0 + \pi_1 Z_{1i} + \pi_2 X_{0i} + \pi_3 B_i + \varepsilon_{it}$$

where Y_{it} is an outcome for student i measured after random assignment in year $t \in (1, 2)$ of the program, Z_{1i} is an indicator for student-level random assignment to treatment, or in other words and indicator for the student having been randomly assigned to the treatment condition within a treatment grade, B_i is a set of school effects for cohort 1 and a set of school and randomization block fixed effects for cohort 2, ε_{it} is a random error term, and X_{0i} is a set of baseline controls measured prior to random assignment that includes days present, days absent, GPA, course failures, indicators for gender, race / ethnicity, age, old for grade, and presence of a learning disability. Our estimate of π_1 corresponds to the standard estimate for the effect of offering students the intervention, or the “intent to treat” (ITT), which comes from comparing treatment and control students within the treatment grades at the same school.

Since the offer to participate was not extended to all students who were randomly assigned to treatment, and because some students offered the chance to participate declined, the ITT is likely to understate the magnitude of the effect of participating in C&C. To estimate the effect of having a C&C mentor, we use random assignment as an instrument for participation (Angrist, Imbens and Rubin, 1996; Bloom, 1984). This recovers the effect of the treatment on the treated (TOT) because no students assigned to the control group were assigned to a C&C mentor. We estimate the TOT by applying two-stage least squares to equations (2) and (3):

$$(2) \quad D_{it} = \rho_0 + \rho Z_{1i} + \rho X_{0i} + \rho B_i + \kappa_{it}$$

$$(3) \quad Y_{it} = \beta_0 + \beta D_{it} + \beta X_{0i} + \beta B_i + \nu_{it}$$

By using random assignment to be offered the treatment condition (Z_{1i}) as an instrument for participation (D_{it}), the TOT estimate is identified by conditional random assignment. The TOT does not compare participants to non-participants – that comparison would be biased because participants are different on average than non-participants. Rather, the TOT compares students randomly assigned to treatment, regardless of whether they were invited to participate and regardless of whether they chose to participate, to control students (the ITT), and scales this comparison by the participation rate among the treatment group to recover the effect of receiving treatment on those who participated.

We examine the possibility of within grade and treatment school spillover in section 7 (E), for cohort 1. The treatment grade and school spillover estimate come from equation (4), excluding treatment assigned students:

$$(4) \quad Y_{it} = \gamma_0 + \gamma_2 Z_{2i} + \gamma_3 Z_{3i} + \gamma_3 X_{0i} + \gamma_4 B_i + \zeta_{it}$$

where the numeric subscripts on the Z 's correspond to the three levels of random assignment (student, grade, and school) so that Z_{2i} and Z_{3i} are, respectively, indicators for being in a grade and school that was randomly assigned for treatment.

Our estimate for γ_2 in equation (4) represents the difference in average outcomes between control students in treatment versus control grades, and so captures any “spillover” effects of the C&C intervention on other students within the grade. For example, if C&C mentors help to reduce absences among chronically absent students, those students might be less likely to induce their friends to skip school. Or as noted above, the C&C mentor might wind up inadvertently working with control group students if they accompany treatment-group friends to their mentor visits. B_i is a set of school randomization blocks fixed effects.

Our estimate for γ_2 also tells us how “off” the standard ITT estimate is in the case of spillovers, in the sense that if we used random assignment of grades within treatment schools to treatment versus control, we could generate an unbiased estimate for the effects of offering students C&C by comparing treatment students in treatment grades to control students in control grades, which is $\pi_1 + \gamma_2$. The first term in this case can be thought of as the bias that results from using control students within treatment grades as a comparison group in the standard ITT estimator. Of course, it is also logically possible that there could be spillovers of the C&C mentor across grades, not just within grades. We can examine that possibility in our data by comparing control students in control grades in treatment schools to students in control schools (γ_3), since the 23 cohort 1 schools were randomly selected from a group of 69 schools. We return to this in our discussion of the empirical results below. We limit our spillover analysis to Cohort 1 because some of the Cohort 1 treatment schools were closed at the end of Cohort 1, and Cohort

2 randomization occurred within a non-random subset of Cohort 1 treatment schools. It was therefore impossible to construct comparable control-school comparison samples for Cohort 2.

Given our randomized experimental design, perhaps the main threats to valid inference with our study come from sample attrition and from the fact that we examine a number of different outcomes – that is, multiple testing concerns. We discuss our tests for sample attrition below. To address multiple hypothesis testing, we present family-wise error rates (FWER) in our main results tables using the bootstrapping procedure described in (Heller et al., 2017). For a given family of outcomes (defined in detail immediately below), we randomly permute the treatment assignment variables Z_{li} within randomization “blocks,” re-estimate the ITT effect in the permuted data, record the distribution of p-values that results, and re-permute the data and repeat this procedure 100,000 times. We then calculate our FWER-adjusted p-values by examining where for instance the most significant p-value in our data (with the true treatment assignment variable values) falls within the distribution of p-values for the most significant “effect” within the family of outcomes in our permuted data. This accounts for the correlation across outcomes in our data, and of outcomes across students within schools and grades.

RESULTS

Overall Effects on School Attendance

Table 5 presents our main results for our days absent outcome. Since most of our tables have the same structure, we take a moment to describe this structure in some detail. Each “cell” or panel within the table presents the results of a separate regression. Each row in the table presents results separately by cohort, with results from cohort 1 in the first row, cohort 2 in the second row and a sample that pools cohort 1 and 2 together (for improved statistical power) in the third row. The first column presents the results using data just from students who were in

grades 1-4, the second column presents the results for students in grades 5-7, and the third column presents the results from all students in grades 1-7 pooled together. For each separate regression we present the control group mean, the ITT effect and its standard error, the TOT effect and its standard error, and the average outcome for those controls who would have participated in C&C had they been offered treatment (the “control complier mean,” or CCM, from Katz, Kling and Liebman, 2001). Underneath each ITT estimate we present the FWER-adjusted p-value, defining our “family” of tests as each hypothesis test that shows up in Table 5 (or 16 in total).

Before turning to the estimated treatment effects, it is interesting to note a pattern in absences among the control group. Control group absences declined from the baseline year to the two program years. For cohort 1, absences among the control group declined from 16.1 in the baseline year to an average of 14.1 per year over the two program years. For cohort 2, absences among the control group declined from 14.0 in the baseline year to an average of 11.5 per year over the two program years. This decline is likely a reversion to the mean since the study sample was selected based on having absences within a high range in the baseline year. In the results presented below, the treatment effects are measured relative to the control group and therefore account for any mean reversion from baseline to subsequent years.

Focusing first on days absent, the ITT estimate for cohort 1 pooled across all grades is a 0.63-day reduction in absences, and the estimated TOT effect is a reduction of 1.28 days. These pair-wise comparisons are significant at the usual 5 percent cutoff but once we account for the number of tests we have carried out, there is a FWER-adjusted p-value of 0.23.

However, these average effects mask substantial heterogeneity by age of the child. For children in grades 1-4, the ITT and TOT effects are both close to zero. In contrast, for students in

grades 5-7 we see a TOT effect of 4.22 fewer absences; that is, C&C participants in these grades get nearly a week more school than their control group counterparts each year. The results for students in grades 5-7 are statistically significant even accounting for the number of tests, with a FWER p-value $<.01$. We can also reject the null hypothesis that the effects are the same for students in grades 1-4 versus grades 5-7, as shown in the rightmost column of the table. Another way to see that these results are not a fluke, besides our FWER-adjusted p-values, is that the findings replicate nearly identically in cohort 2, as shown in the second row.

Tables 6 and 7 repeat this analysis for our alternative measures of school attendance, namely days present (Table 6) and percent days present (Table 7). The results are qualitatively similar although somewhat less precisely estimated for days present than for the other outcomes. As mentioned above, days present and days absent do not sum to the total number of days in the school year because some students either leave the CPS district or have days when they are not officially enrolled at any CPS school while they are transitioning from one school to another. Each day a student is officially enrolled in a CPS school he or she accumulates what is called a membership day. A student's total membership days are equal to the sum of his days absent and days present. To check to make sure the estimated effects on school absences are not a result of reducing the number of days students are enrolled in school, we separately examine the effect of participation in C&C on membership days and find no statistically significant effect for either cohort 1 or cohort 2 or for both cohorts pooled together. When looking at the percent of membership days present (Table 7), we see a similar increase in percentage of days present.

One natural question is why the results look so different for relatively younger children (grades 1-4) versus older students (grades 5-7). One hypothesis is that what appears to be treatment heterogeneity by age is actually the result of differences in other student

characteristics. To address this possibility, in Appendix Table 7 we pool students in grades 1-4 and 5-7, and run a linear probability model that has as the dependent variable an indicator for being in grades 5-7 with all of our baseline characteristics (other than age) as explanatory variables. We then take the predicted probability estimated from this model and re-weight the data using inverse probability weighting, so that the re-weighted average baseline characteristics for grades 1-4 and 5-7 are the same on average. This procedure essentially up-weights young students who look more like older students in terms of their baseline characteristics, and down-weights older students who look too unlike younger students. The main findings of substantial and significant heterogeneity in program effects by age persist.

Effects on the Distribution of Days Absent

Thus far, we have presented results on the average number of days absent or present. The pattern of effects throughout the absences distribution may be informative of the mechanism by which C&C mentors help to reduce absenteeism. For example, C&C mentors may be most effective at identifying and remedying situations in which students are at risk for extreme absenteeism, or they may be effective at reducing absences a small amount for all students.

Table 8 examines where in the attendance distribution these effects are concentrated, by re-estimating our main TOT specification but now using as the dependent variable an indicator for whether a student's number of absences exceeds different thresholds (specifically, 5, 10, 15 or 20 days in total). The C&C effects are (when expressed as a share of the control complier mean) proportionately much larger among the higher-absence students (absent more than 10 days). We show this analysis for our two different age groups (grades 1-4 versus 5-7) and find that, similar to our main results on average days absent, the effects at higher points in the absence distribution are also driven by impacts on older students (those in grades 5-7). For

example, among grade 5-7 students in cohort 1, the estimated reduction in likelihood of missing 10 or more days is -0.125, equal to 16.8 percent of the control complier mean of 0.743. For the same students, the effects on missing 20 or more days is -.109, equal to 33.9 percent of the control complier mean of 0.322.

Effects by Intervention Duration

One motivation for designing the intervention to be for two years was the hypothesis that the effectiveness of mentoring depends on the strength of the relationship between the student and the mentor, and the idea that relationships take time to develop, especially for students who are showing signs of disengagement. Thus far, we have presented results that pool the two years within each cohort. In contrast, the results presented in Table 9 show the treatment effects on days absent separately for the first and second year of the intervention. Table 9 shows TOT estimates and breaks the results out by grade grouping, and by cohort (refer to Table 9a in the Appendix for ITT estimates, and Appendix Table 9b for cumulative treatment effects over the two years).

For the 5th-7th grade students, we find that within both cohorts C&C reduced absences by (roughly) 50 percent more in the second year of the program than in the first. Though the pattern of results is suggestive that the effects are larger in the second year of the program than the first, the differences in treatment effects between the first and second year of the program are not quite statistically significant (this is somewhat sensitive to whether we allow the effects of the baseline covariates to be different across cohorts or constrain them to be the same).

A subtlety raised by these estimates comes from the fact that in our main exhibits so far, we include the cohort 1 treatment students who were re-randomized into treatment and control groups in cohort 2 as part of the overall cohort 2 treatment group. If there are growing impacts

with time in the intervention, treating these students as part of the regular cohort 2 treatment group could inflate the first and second year impacts. To address this possibility, we conduct our analysis in Appendix Table 5b again, excluding these students from the cohort 2 analysis; the results are qualitatively similar to our overall analyses. For completeness we also show the results for students from cohort 1 who were randomly assigned to treatment versus control in cohort 2. In principle we could learn something about how the effects for the third and fourth years compare to the first and second years (as shown in Table 5), but in practice there are just too few cohort 1 re-randomized students like this to make the comparison informative, in light of the large standard errors around this sub-group estimate.

Overall these estimates are consistent with, though obviously not definitive proof of, the hypothesis that it may take some time for mentors and students to develop relationships, and that the strength of this relationship may be an important mediator of the effectiveness of the mentoring program. A related possibility is that it takes time for the C&C mentors to learn what is causing each individual student to be missing school, and that as the mentors recognize how they can most effectively help the student or intervene, the program's effect on absences grows. These are important questions for future research.

Effects on Academic Outcomes

We turn now to the question of whether the C&C mentoring program generated improvements in students' academic outcomes. Among mentoring programs, C&C is notable in its focus on attempting to engage students in school. In addition to the primary goal of increasing student attendance, C&C mentors are tasked with improving communications between students and teachers and helping students to overcome things that make it hard for them to succeed while in school. The C&C mentors in our study varied in the strategies they used to help students.

Some acted as tutors from time to time, while others drew on their backgrounds in social work and acted as case managers, advocates, and counselors.

One might hypothesize that the increase in attendance induced by the C&C program may have generated improved academic outcomes. At the same time, given the emphasis on school engagement beyond attendance and the ways C&C mentors were asked to track grades and academic performance alongside attendance, the mentors may have had a direct effect on academic outcomes over and above any effect that operated through improved attendance.

Tables 10 and 11 present the effect on academic outcomes, including grade point average (GPA), course failures, and math and reading standardized test scores. Table 10 presents the results pooling all grades, and Table 11 shows the TOT results separated out by age group.¹⁵ Regardless of whether the estimates are separated by age or not, we do not find significant effects of C&C on students' GPA or course failures. Point estimates for these outcomes are close to zero, never statistically significant, and based on the 95-percent confidence intervals we can rule out increases in the range of 0.1 to 0.15 on a standard 4-point GPA scale (or, for course failures, of a decrease any larger than about -0.13 relative to control mean of 0.47).

We also report treatment effects on reading and math achievement test scores. These results are shown in the rows labeled "Standardized MAP" and "Standardized Official Test." We present results for two test-score dependent variables because CPS changed the test that was used for accountability purposes during the period we study – the Illinois Standard Achievement Test (ISAT) from 2011-12 through 2013-14, and the Measures of Academic Progress (MAP) in 2014-15 – but administered the MAP test throughout the time period. The "Standardized MAP" dependent variable uses the MAP score in each year, and the "Standardized Official Test"

¹⁵ Refer to Tables 11 and 12 in the Appendix for ITT.

dependent variable uses the ISAT score in the years it was used for accountability and the MAP score in the year it was used for accountability. As is common, test scores are missing for some students in some years. To assess whether test score missingness is associated with treatment assignment, we also estimate models with indicators for having a valid test score as the dependent variable. These results are shown in the rows labeled “Valid MAP Data” and “Valid Official Data.”

Taken at face value and setting aside for the moment the size of the standard errors, the estimated effects on Standardized MAP and Standardized Official Test seem to indicate, if anything, a slight decline in test scores as a result of C&C. But the table also shows that C&C increases the likelihood that students have a valid test recorded in the CPS data, presumably because they are now more likely to attend school on testing days.

To see how this selection related to missing test scores influences our results, we implement an imputation procedure for the missing data and estimate a median regression (see e.g. Brown, 1984; Chandra, 2000; and Mulligan and Rubinstein, 2008 for the use of similar procedures to account for sample selection bias). We assume that the students who would not show up for testing in the control condition but who would if assigned to treatment have below-median test scores. We therefore impute for every student with missing test scores the lowest recorded value to the test we see in our data. Then we re-estimate a quantile regression that examines the C&C effect on the median test score (the row called “standardized MAP, imputed”), and compare that to the results of a quantile regression that uses only data for those students with non-missing test scores. If we look for example at the “pooled cohorts” panel of Table 10, the estimated effect of C&C assignment (TOT) on the median reading score of just those with valid scores is a statistically significant 0.026 (standard error 0.016). If we impute

below-median values for those with missing scores, the estimated effect on the median is now much smaller and no longer close to statistically significant, equal to -0.007 (standard error 0.037). We see a similar attenuation of the estimated effect when we impute below-median values for missing math scores. Based on the Lee (2009) bounds estimates, the estimated effects range from 0.150 to -0.167 for standardized math MAP scores and 0.163 to -0.154 for standardized reading MAP scores for the pooled cohorts and grades.¹⁶ From this we conclude that there is no reliable evidence in our data that C&C affected test scores, either for the better or for the worse.

Spillover Effects

To be able to examine possible spillover effects of C&C, we randomized school and grade to treatment versus control conditions, and then within treatment grades randomly assigned students. Because of the randomization of grade to treatment versus control, when we analyze spillovers for grade sub-groups (students in grades 1-4 versus grades 5-7 separately), we use only those treatment schools where the control grade was randomly assigned to be among 1-4 to test spillovers for students in grades 1-4, and similarly for when we test for spillovers among grade 5-7 students we only use treatment schools where the control grade was randomly assigned to be in the grade 5-7 range.

Table 12 shows that there is balance in baseline characteristics for the control-group students in treatment versus control grades within the treatment schools, as well as between control students in the treatment versus control schools. While a few pairwise comparisons are statistically significant, the F-test statistics for the joint hypothesis of no difference in the full set of baseline characteristics are not close to statistically significant.

¹⁶ The bound estimates are based on Lee (2009) bounds. No covariates were included to tighten the range.

Tables 13 and 14 show that there are no statistically significant spillovers, either overall or separately for grades 1-4 or 5-7, for either school attendance and engagement measures, or on our measures of academic achievement. There is a suggestive hint of a possible increase in absences by 1.63 days for students within treatment grades assigned to the control condition ($p < .01$), but this result is not statistically significant once we account for the number of hypothesis tests carried out in Table 13 (FWER p -value = 0.116). No other estimates for spillover effects for any other cohorts or grade levels or outcomes approach statistical significance, either.

CONCLUSION

In this paper we report the results of one of the few large-scale RCTs of a policy effort to improve student school attendance. The intervention, Check & Connect, seeks to supplement the social capital parents can provide children to support them in school by assigning students to an in-school mentor. The intervention relies on providing youth with a fair amount of one-on-one time with an adult whose full-time job is to deliver the program; the cost per participant per year equals approximately \$1,700, or \$3,400 total over two years.

The program seems to be effective in improving student attendance, particularly for middle-school-aged (i.e. grades 5-8) children. Theories of child development may help explain these results. As children enter middle childhood and adolescence, they are more likely to seek and obtain autonomy from their parents/guardians while simultaneously developing more meaningful relationships with peers and non-familial adults (Bandura, 1964; Eccles, 1999; Erikson, 1950). This suggests that Check & Connect may have a larger impact on older youth via two potential avenues: older children have more agency around school attendance decisions and are more open to building relationships with and seeking guidance from mentors.

While the program leads to significant improvements in its primary target outcome – attendance – we find no detectable impacts on academic outcomes. Moreover, given the costs of the program, the size of the impacts on absences raises the question of whether this intervention is the most cost-effective way to improve student attendance. One way to measure cost-effectiveness is by the cost per day of attendance increased. Our calculations suggest C&C costs approximately \$400 per day of improved attendance. By way of comparison, another program implemented in CPS during our study called the Chicago Attendance Project—a mail-based, large-scale intervention that informed CPS guardians of how many absences their student had accumulated that semester—only incrementally decreased absences, but did so at about \$6-\$11 per day (Rogers and Feller [2018], Robinson et al. [2018]).

Another relevant comparison is youth mentoring programs in the U.S. that use volunteer mentors rather than paid, full-time employees. While these programs generally show few statistically significant impacts on academic and attendance outcomes (Bernstein et al., 2009; Herrera, et al., 2007), the Big Brothers Big Sisters of America (BBBA) program has seen marginal decreases in unexcused absences (Schwartz et al., 2011). Interestingly, BBBA costs are similar to those of C&C – they average \$1500 per student per year (Laura and John Arnold Foundation, 2017). One-to-one mentoring programs tend to be more expensive than other types of mentoring models, regardless of whether volunteer or paid mentors are used (Garringer, McQuillin, & McDaniel, 2017). Future research could explore whether less-expensive mentoring models, such as cross-age peer mentoring or group mentoring, could decrease costs while still providing intensive mentoring services.

One lesson from these results may be that lower-cost, less-intensive interventions may provide the most cost-effective way to improve school attendance, particularly for marginal

improvements. However, higher-cost interventions may be necessary to reduce absences by more than small amounts. One possibility is to implement a multi-tiered system of supports, where lower-cost attendance monitoring interventions are applied for all students, but higher-cost interventions focused on building social capital are provided for chronically truant students. Our results also suggest that helping students to overcome the causes of truancy may not be enough to significantly improve learning. School-based academic interventions that improve the quality of instruction or the learning environment in school more generally are surely, however, complementary with interventions that increase school attendance. Further research should focus on the most effective and cost-effective ways to improve learning and academic achievement for students who are currently missing substantial amounts of school.

REFERENCES

- Allen, B.M. & Fryer, Jr., R.G. (2011). The power and pitfalls of education incentives. The Hamilton Project. Retrieved July 17, 2020, from http://scholar.harvard.edu/files/fryer/files/092011_incentives_fryer_allen_paper2.pdf.
- Allensworth, E.M. & Easton, J.Q. (2007). What matters for staying on-track and graduating in Chicago public high schools. Consortium on Chicago School Research at the University of Chicago.
- Allensworth, E., Evans, S. (2016). Tackling absenteeism in Chicago. *Phi Delta Kappan*, 98, 16–21.
- Anderson, A.R., Christenson, S.L., Sinclair, M.F., & Lehr, C.A. (2004). Check & Connect: The importance of relationships for promoting engagement with school. *School Psychology*, 42, 95–113.
- Angrist, J., Bettinger, E., Bloom, E., King, E., & Kremer, M. (2002). Vouchers for private schooling in Colombia: Evidence from a randomized natural experiment. *American Economic Association*, 92, 1535–1559. Retrieved July 17, 2020, from <http://pubs.aeaweb.org/doi/pdfplus/10.1257/000282802762024629>.
- Angrist, J., Bettinger, E., & Kremer, M. (2006). Long-term educational consequences of secondary school vouchers: Evidence from administrative records in Colombia. *American Economic Review*, 96, 847–862.
- Angrist, J.D., Imbens, G.W., & Rubin, D.B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 81, 444–455.
- Balfanz, R., & Byrnes, V. (2012). *Chronic Absenteeism: Summarizing what we know from nationally available data*. Baltimore: Johns Hopkins University Center for Social

- Organization of Schools. Retrieved July 17, 2020, from http://new.every1graduates.org/wp-content/uploads/2012/05/FINALChronicAbsenteeismReport_May16.pdf.
- Bandura, A. (1964). The stormy decade: Fact or fiction? *Psychology in the Schools*, 1, 224–231.
- Barrera-Osorio, F., Bertrand, M., Linden, L.L., & Perez-Calle, F. (2008). Conditional cash transfers in education design features, peer and sibling effects: Evidence from a randomized experiment in Colombia (NBER Working Paper No. 13890). Cambridge, MA: National Bureau of Economic Research.
- Bayer, A., Grossman, J.B., & DuBois, D.L. (2015). Using volunteer mentors to improve the academic outcomes of underserved students: The role of relationships. *Journal of Community Psychology*, 43, 408–429.
- Bernstein, L., Rappaport, C., Olsho, L., Hunt, D., & Levin, M. (2009). Impact evaluation of the U.S. Department of Education’s Student Mentoring Program. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Bettinger, E.P., & Evans, B.J. (2019). College guidance for all: A randomized experiment in pre-college advising. *Journal of Policy Analysis and Management*, 38, 579–599.
- Bloom, B.S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13, 4–16.
- Bloom, B., Jones, L.I., & Freeman, G. (2013). Summary health statistics for U.S. children: National health interview survey, 2012. National Center for Health Statistics. Vital

- Health Stat, 10. Retrieved July 17, 2020, from http://www.cdc.gov/nchs/data/series/sr_10/sr10_258.pdf.
- Brown, C. (1984). Black-white earnings ratios since the Civil Rights Act of 1964: The importance of labor market dropouts. *Quarterly Journal of Economics*, 95, 31–44.
- Chandra, A. (2000). Labor-market dropouts and the racial wage gap: 1940-1990. *American Economic Review*, 90, 333–338.
- Chang, H.N., & Romero, M. (2008). Present, engaged, and accounted for: The critical importance of addressing chronic absence in the early grades. National Center for Children in Poverty, Mailman School of Public Health at Columbia University, New York, NY. Retrieved July 17, 2020, from http://www.nccp.org/publications/pdf/text_837.pdf
- Child Trends Databank (2015). Student absenteeism. Retrieved July 17, 2020, from <http://www.childtrends.org/?indicators=student-absenteeism>.
- Christenson, Stout, & Pohl. (2012). Check & Connect: A comprehensive student engagement intervention: Implementing with Fidelity. Institute on Community Integration.
- Coleman, J.S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, 95, S95–S120.
- Common Core of Data (2010). Number of students and selected high school dropout and completion statistics for the 100 largest public elementary and secondary school districts in the United States and jurisdictions: School year 2009-10 [Table]. National Center for Education Statistics. Retrieved July 17, 2020, from https://nces.ed.gov/ccd/tables/2009-10_Dropout_and_completer_data_for_100_LSD_table.asp.

- Cook, P.J, Crowley, D.M., Dodge, K., & Gearing, M.E. (2016). Primary school truancy: Risk factors and consequences for academic success. Duke University working paper.
- Cutler, D.M., & Lleras-Muney, A. (2010). Understanding differences in health behaviors by education. *Journal of Health Economics* 29, 1–28.
- Dryfoos, J.G. (1990). *Adolescents at Risk: Prevalence and Prevention*. New York, NY: Oxford University Press.
- Eccles, J.S. (1999). The development of children ages 6 to 14. *The future of children*, 9, 30–44.
- Ehrlich, S.B., Gwynne, J.A., Pareja, A.S., & Allensworth, E.M. (2014). Preschool attendance in Chicago Public Schools: Relationship with learning outcomes and reasons for absences. University of Chicago Consortium on Chicago School Research. Retrieved July 17, 2020, from <http://www.attendanceworks.org/wordpress/wp-content/uploads/2014/06/CCSR-Pre-K-Attendance-Full-Report-May-2014-revised.pdf>.
- Erickson, E.H. (1950). *Childhood and society*. New York: Norton.
- Finn, J.D. (1989). Withdrawing from school. *Review of Educational Research*, 59, 117–142. Retrieved July 17, 2020, from <https://pdfs.semanticscholar.org/221e/ec4c94a3484f6a025d2477b65f00c481c541.pdf>.
- Gage, N.A., Sugai, G., Lunde, K., & DeLoreto, L. (2013). Truancy and zero tolerance in high school: Does policy align with practice? *Education and Treatment of Children*, 36, 117–138. Retrieved July 17, 2020, from <http://ne.glr.org/wp-content/uploads/2014/11/attendance-and-behavior.pdf>.
- Garringer, M., McQuillin, S., & McDaniel, H. (2017). *Examining Youth Mentoring Services Across America: Findings from the 2016 National Mentoring Program Survey*. Boston,

- MA: MENTOR. Retrieved July 17, 2020, from https://www.mentoring.org/new-site/wp-content/uploads/2017/07/Mentor-Survey-Report_FINAL_small.pdf.
- Garrison, A.H. (2006). "I missed the bus": School grad transition, the Wilmington Truancy Center, and reasons youth don't go to school. *Youth Violence and Juvenile Justice*, 4, 204–212. Retrieved July 17, 2020, from <http://yvj.sagepub.com/content/4/2/204.full.pdf+html>.
- Goldin, C. & Katz, L.F. (2010). *The race between education and technology*. Belknap Press.
- Grossman, J.B., Chan, C.S., Schwartz, S.E.O., & Rhodes, J.E. (2012). The test of time in school-based mentoring: The role of relationship duration and re-matching on academic outcomes. *American Journal of Community Psychology*, 49, 43–54. Retrieved from <http://rhodeslab.org/files/ToTime2.pdf>.
- Hallfors, D., Vevea, J. L., Iritani, B., Cho, H., Khatapoush, S., & Saxe, L. (2002). Truancy, grade point average, and sexual activity: A meta-analysis of risk indicator for youth substance use. *Journal of Social Health*, 72, 205–211.
- Healey, K., Nagaoka, J., & Michelman, V. (2014). The Educational attainment of Chicago Public Schools students: A focus on four-year college degrees. Retrieved July 17, 2020, from <https://consortium.uchicago.edu/sites/default/files/publications/Fast%20Facts%20Brief.pdf>.
- Heckman, J.J., & LaFontaine, P.A. (2007). *The American high school graduation rate: trends and levels* (NBER Working Paper No. 13670). Cambridge, MA: National Bureau of Economic Research.

- Heller, S., Shah, A., Guryan, J., Ludwig, J., Mullainathan, S., & Pollack, H. (2017). Thinking, fast and slow? Some field experiments to reduce crime and dropout in Chicago. *Quarterly Journal of Economics*, 132, 1–54.
- Heppen, J.B., Zeiser, K., O’Cummings, M., Holtzman, D., Christenson, S., & Pohl, A. (2017). Efficacy of the Check & Connect mentoring program for at-risk general education high school students. *Journal of Research on Educational Effectiveness*, 11, 56–85.
- Herrera, C., DuBois, D.L., & Grossman, J.B. (2013). *The role of risk: Mentoring experiences and outcomes for youth with varying risk profiles*. New York, NY: A Public/Private Ventures project distributed by MDRC. Retrieved July 17, 2020, from http://www.mdrc.org/sites/default/files/Role%20of%20Risk_Final-web%20PDF.pdf.
- Herrera, C., Grossman, J.B., Kauh, T.J., Feldman, A.F., McMaken, J., & Jucovy, L.Z. (2007). *Making a difference in schools: The Big Brothers Big Sisters school-based mentoring impact study*. Philadelphia: Public/Private Ventures. Retrieved July 17, 2020, from <http://files.bigsister.org/file/Making-a-Difference-in-Schools.pdf>.
- Huizinga, D., and Jacob-Chien, C. (1998). The contemporaneous co-occurrence of serious and violent juvenile offending and other problem behavior. In *Serious & Violent Juvenile Offenders: Risk Factors and Successful Interventions*, by R. Loeber and D.P. Farrington. Thousand Oaks, CA: Sage Publications, Inc., p. 57.
- Hyman, J. (2019). Can light-touch college-going interventions make a difference? Evidence from a statewide experiment in Michigan. *Journal of Policy Analysis and Management*, 39, 159–190.

- Jackson, D., Marx, G., & Richards, A. (2012, November 11). An empty-desk epidemic. Chicago Tribune. Retrieved July 17, 2020, from <http://www.chicagotribune.com/ct-met-truancy-mainbar-20121111-story.html>
- Jackson, D., and Marx, G. (2013, February 19). Prison data, court files show link between school truancy and crime. Chicago Tribune. Retrieved from <http://www.chicagotribune.com/investigations/ct-xpm-2013-02-19-ct-met-prison-truancy-20130219-story.html>.
- Karcher, M.J. (2008). The study of mentoring in the learning environment (SMILE): A randomized evaluation of the effectiveness of school-based mentoring. *Prevention Science*, 9, 99–113.
- Katz, L.F., Kling, J.R., and Liebman, J.B. (2001). Moving to opportunity in Boston: Early results of a randomized mobility experiment. *Quarterly Journal of Economics*, 116, 607–654.
- Kearney, C.A. (2008). School absenteeism and school refusal behavior in youth: A contemporary review. *Clinical Psychology Review*, 28, 451–471.
- Klima, T., Miller, M., & Nunlist, C. (2009) What works? Targeted truancy and dropout programs in middle and high school. Olympia; Washington State Institute for Public Policy, Document No. 09-06-2201. Retrieved July 17, 2020, from http://www.wsipp.wa.gov/ReportFile/1045/Wsipp_What-Works-Targeted-Truancy-and-Dropout-Programs-in-Middle-and-High-School_Full-Report.pdf.
- Las Vegas Review-Journal (2009, May 16). School district workers uneasy because attendance officers might be ‘bumped.’ Retrieved July 17, 2020, from <https://www.reviewjournal.com/news/school-district-workers-uneasy-because-attendance-officers-might-be-bumped/>.

- Laura and John Arnold Foundation. (2017). Evidence Summary for Big Brothers Big Sisters. Social Programs that Work Review: November 2017. Retrieved July 17, 2020, from <https://evidencebasedprograms.org/document/big-brothers-big-sisters-evidence-summary/>.
- Lehr, C.A., Sinclair, M.F., & Christenson, S.L. (2004). Addressing student engagement and truancy prevention during the elementary school years: A replication study of the Check & Connect model. *Education for Students Placed at Risk*, 9, 279–301.
- Los Angeles Times (2003, September 20). School district reassigns its truant officers.
- Maynard, B.R., Kjellstrand, E.K., Thompson, A.M. (2014). Effects of Check and Connect on attendance, behavior and academics: A randomized effectiveness trial. *Research on Social Work Practice*, 24, 296–309.
- Meara, E.R, Richards, S., & Cutler, D.M. (2008). The gap gets bigger: Changes in mortality and life expectancy, by education, 1981–2000. *Health Affairs*, 27, 350–360
- Mulligan, C.B., & Rubinstein, Y. (2008). Selection, investment, and women’s relative wages over time. *Quarterly Journal of Economics*, 123, 1061–1110.
- Murnane, R.J. (2013). U.S. high school graduation rates: Patterns and explanations. (NBER Working Paper No. 18701). Cambridge, MA: National Bureau of Economic Research.
- Oreopoulos, P., Brown, R.S., & Lavecchia, A.M. (2017). Pathways to education: An integrated approach to helping at-risk high school students. *Journal of Political Economy*, 125, 947–984.
- Oreopoulos, P., & Ford, R. (2019). Keeping college options open: A field experiment to help all high school seniors through the college application process. *Journal of Policy Analysis and Management*, 38, 426–454.

- Riccio, J.A., Dechausey, N., Miller, C., Nunez, S., Verma, N., and Yang, E. (2013). Conditional cash transfers in New York City: The continuing story of the Opportunity-NYC Family Rewards demonstration. NYC: MDRC Corporation.
- Rogers, T., & Feller, A. (2018). Reducing student absences at scale by targeting parents' misbeliefs. *Nature Human Behaviour*, 2, 335–342.
- Robinson, C.D., Lee, M.G., Dearing, E., and Rogers, T. (2018). Reducing student absenteeism in the early grades by targeting parental beliefs. *American Education Research Journal*, 55, 1163–1192.
- Schoeneberger, J.A. (2011). Longitudinal attendance patterns: Developing high school dropouts. *The Clearinghouse: A Journal of Educational Strategies, Issues and Ideas*, 85, 7–14.
- Schwartz, S.E.O., Rhodes, J.E., Chan, C.S. & Herrera, C. (2011). The impact of school-based mentoring on youths with different relational profiles. *Developmental Psychology*, 47, 450–462. Retrieved July 17, 2020, from <http://www.rhodeslab.org/files/RelationshipProfiles.pdf>.
- Sinclair, M.F., Christenson, M.L., & Thurlow, M.L. (2005). Promoting school completion of urban secondary youth with emotional or behavioral disabilities. *Exceptional Children*, 71, 465–482.
- Sinclair, M.F., Christenson, S.L., Evelo, D.L., & Hurley, C.M. (1998). Dropout prevention for youth with disabilities: Efficacy of a sustained school engagement procedure. *Exceptional Children*, 65, 7–21.
- Tait, C. (2004). *Strategies for adolescence: Turning it around for middle-school students*. Clemson, SC: National Dropout Prevention Center.
- U.S. Department of Education (1996). *Manual to combat truancy*.

WBEZ. (2014, June 04). What the heck happened to Chicago's truancy officers? Retrieved from <https://www.wbez.org/stories/what-the-heck-happened-to-chicagos-truancy-officers/ff4ca74d-e067-43cd-b2b1-e71cc3e7bb68>.

Figure 1. Cohort 1 Random Assignment (Spring 2011).

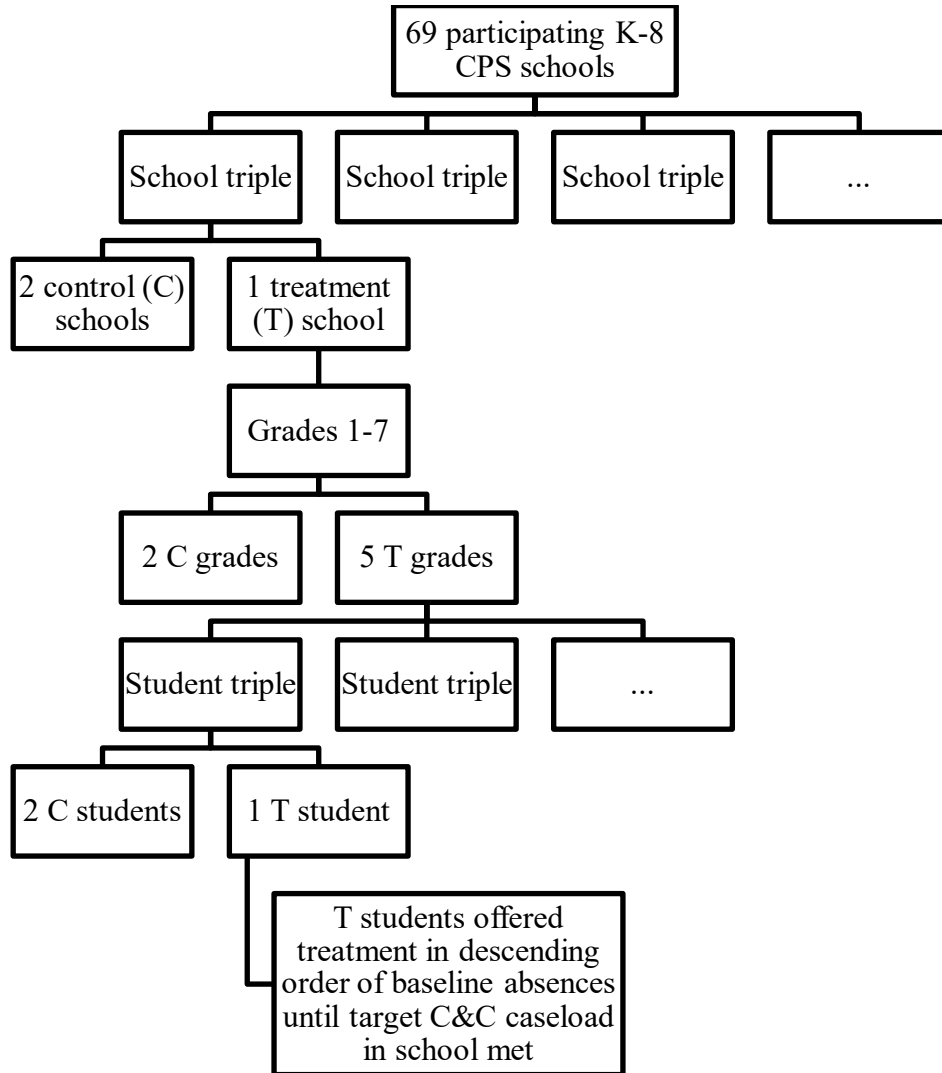


Figure 2. Cohort 2 Random Assignment (Summer 2013).

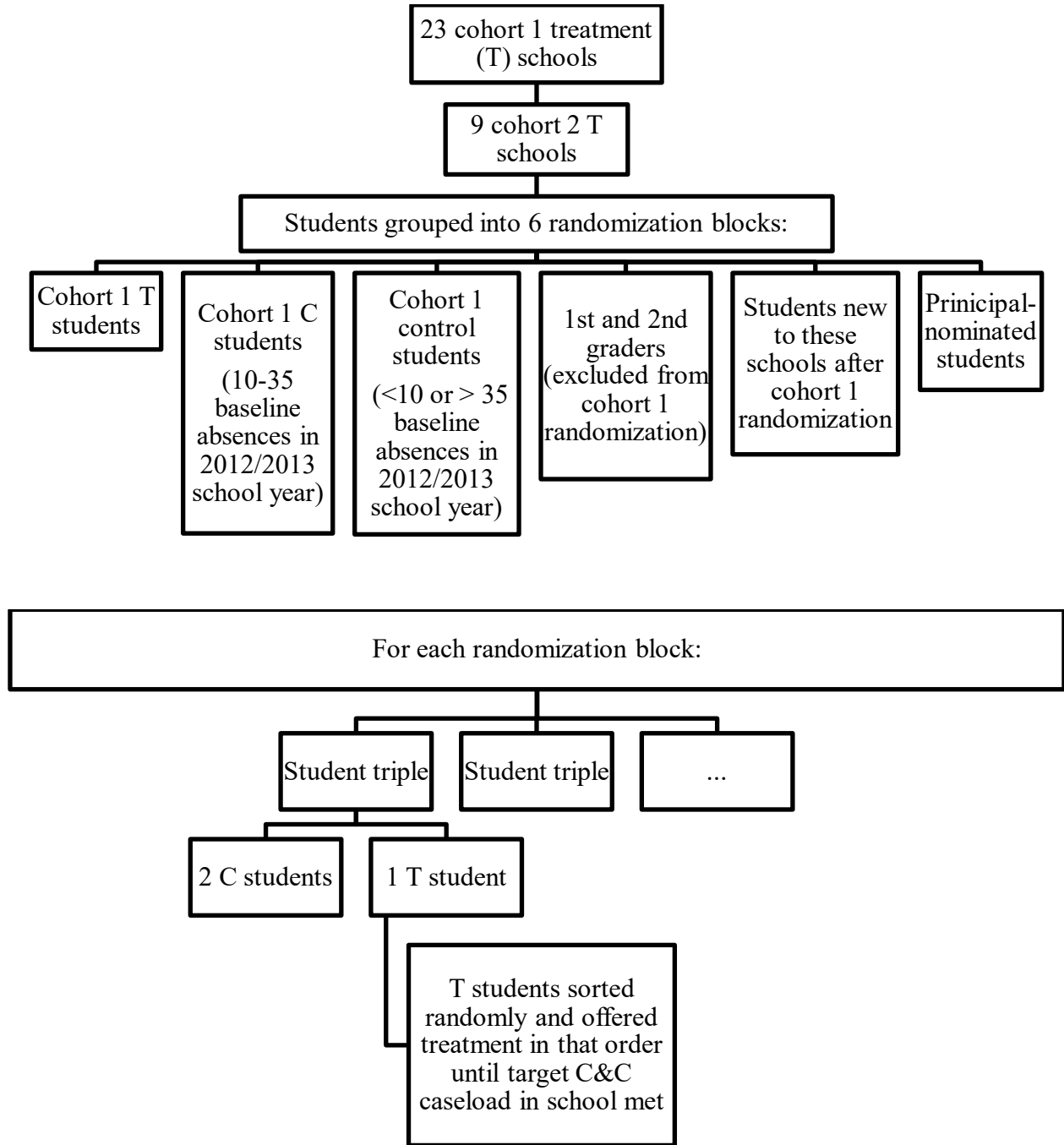


Table 1. Cohort 1 student baseline characteristics.

	Control	Treatment
Students (n)	1846	933
Days Present in 2010-11 SY	150.18	150.95
Days Absent in 2010-11 SY	16.13	16.28
Percent Male	0.53	0.53
Age	8.74	8.72
Percent Old for Grade	0.16	0.15
Percent Black	0.57	0.57
Percent Hispanic	0.39	0.4
Percent Learning Disability	0.11	0.1
Number of Course Failures in 2010-11 SY	0.7	0.68
GPA in 2010-11 SY	2.21	2.24
P-value on F-test		
Without missing data indicators		p=.775
With missing data indicators		p=.84
<p><i>Note:</i> The length of the CPS school year changed during the study. The number of days in the school year, by year, was: 2010-11,170; 2011-12, 170; 2012-13, 181; 2013-14, 178; 2014-15, 180. In the study sample, days present and days absent do not sum to total days in the school year because some students were not enrolled as CPS students or the full school year. P-value on null hypothesis that average values are jointly the same for treatment and control group comes from regressing treatment assignment indicator against the baseline covariates, with imputed missing values and a missing indicator for observations that had missing values.</p>		
<p>* p<0.10; ** p<0.05; *** p<0.01.</p>		

Table 2. Cohort 2 student baseline characteristics.

Baseline Characteristics	Control	Treatment
Students (n)	1111	1039
Days Present in 2012-13 SY	159.23	159.94
Days Absent in 2012-13 SY	14.01	13.86
Percent Male	0.53	0.53
Age	8.53	8.39
Percent Old for Grade	0.12	0.1
Percent Black	0.46	0.47
Percent Hispanic	0.49	0.47
Percent Learning Disability	0.05	0.06
Number of Course Failures in 2012-13 SY	0.41	0.44
GPA in 2012-13 SY	2.29	2.35
P-value on F-test		
Without missing data indicators		p=.65
With missing data indicators		p=.145
<p><i>Note:</i> The length of the CPS school year changed during the study. The number of days in the school year, by year, was: 2010-11,170; 2011-12, 170; 2012-13, 181; 2013-14, 178; 2014-15, 180. In the study sample, days present and days absent do not sum to total days in the school year because some students were not enrolled as CPS students or the full school year. P-value on null hypothesis that average values are jointly the same for treatment and control group comes from regressing treatment assignment indicator against the baseline covariates, with imputed missing values and a missing indicator for observations that had missing values.</p>		
<p>* p<0.10; ** p<0.05; *** p<0.01.</p>		

Table 3. Participation rates.

	Assigned to Treatment	Assigned & Approached	Control
Cohort 1			
Year 1	0.5	0.77	0
Year 2	0.46	0.72	0
Either Year	0.52	0.81	0
N	933	601	1846
Cohort 2			
Year 1	0.31	0.79	0
Year 2	0.32	0.8	0
Either Year	0.33	0.84	0
N	1039	415	1111
<i>Note:</i> This table reports the participation rates for three groups, students who were randomly assigned to treatment, students who were both assigned to treatment and approached or offered treatment, and students who were assigned to the control group. The set of students included in each group is constant across years within each cohort.			

Table 4. Regression of participation dummy variable against baseline characteristics.

Baseline Characteristics	Cohort 1	Cohort 2
Outcome	Participation (either year)	Participation (either year)
Baseline Days Present	0	0
Baseline Excused Absences	0.023***	0.004 *
Baseline Unexcused Absences	0.024***	0.001
Baseline GPA	-0.090***	0.003
Baseline Course Failures	-0.023	0.006
Dummy for Male	-0.012	0.004
Age	0.021 **	0.01
Old for Grade	-0.082 *	-0.08
Dummy for Black	0.236 **	-0.051
Dummy for Hispanic	0.178 **	-0.044
Dummy for Learning Disability	0.069	0.036
Dummy for Missing Baseline Grade Data	-0.244 **	0.136
Dummy for Missing Baseline Attendance Data	0	0.002
N	933	1039
R-squared	0.633	0.664
<p><i>Note:</i> Both models include school fixed effects. The model for cohort 2 includes additional randomization block fixed effects to account for differential probability of selection for treatment based on whether the student participated in cohort 1 and which grade the student was in during cohort 2.</p>		
<p>* p<0.10; ** p<0.05; *** p<0.01.</p>		

Table 5. Main outcome – Days absent.

	Grades 1 - 4				Grades 5 – 7				All Grades				H0: Grades 1-4 = Grades 5-7 p-value	
	CM	ITT	TOT	CCM	CM	ITT	TOT	CCM	CM	ITT	TOT	CCM	ITT	TOT
Cohort 1														
Treatment	12.818	0.329	0.696	13.118	16.369	-2.290***	-4.224***	18.447	14.107	-0.633**	-1.275**	15.252	<0.001	<0.001
Standard Error		(0.3613)	(0.7635)			(0.5897)	(1.0771)			(0.3184)	(0.6374)			
FWER p-value		0.5817				0.0052				0.2297				
N		3397	3397			1938	1938			5335	5335			
Cohort 2														
Treatment	10.352	-0.472	-2.066	12.979	14.454	-1.918**	-4.001**	16.577	11.465	-0.844**	-2.881***	14.469	0.086	0.367
Standard Error		(0.3357)	(1.4486)			(0.7821)	(1.5791)			(0.3277)	(1.1040)			
FWER p-value		0.4286				0.1365				0.1204				
N		3076	3076			1078	1078			4154	4154			
Pooled Cohorts														
Treatment	11.807	-0.062	-0.177	12.779	15.771	-2.149***	-4.129***	17.649	13.109	-0.740***	-1.826***	14.797	<0.001	<0.001
Standard Error		(0.2458)	(0.6915)			(0.4675)	(0.8809)			(0.2284)	(0.5581)			
FWER p-value		0.8214				0.0005				0.0281				
N		6473	6473			3016	3016			9489	9489			
* p<0.10; ** p<0.05; *** p<.01.														
CM = Control Mean, ITT = Intent to Treat, CCM = Control Complier Mean, TOT = Treatment on Treated.														
Family-wise error rate (FWER) is the probability that a given family of tests (defined for three grade groups, grades 1-4, 5-7 and pooled 1-7; and three cohort groups, cohort 1, cohort 2, and pooled cohort 1 and 2 total; so 9 families of outcomes are shown in the table) includes at least one false positive result. The FWER p-value is based on 100,000 iterations of randomly re-assigning the treatment grade and treatment-condition-within-grade indicators in our sample.														
The samples are for both cohorts, with each student having one observation for each year of treatment.														
The regression covariates include: baseline days present, baseline excused absences, baseline unexcused absences, baseline GPA, baseline course failures, baseline age, baseline old for grade, dummy for male, dummy for black, dummy for Hispanic, dummy for learning disability, and dummy for missing baseline data.														
We include randomization block groups which are school-by-grade-by-group. In cohort 1, these blocks are only school-by-grade. For cohort 2, we added block groups based on a student's original treatment assignment in cohort 1.														
We cluster at the student level because students have multiple observations, one for each year of treatment.														

Table 6. Main outcome – Days present.

	Grades 1 - 4				Grades 5 - 7				All Grades				H0: Grades 1-4 = Grades 5-7 p-value	
	CM	ITT	TOT	CCM	CM	ITT	TOT	CCM	CM	ITT	TOT	CCM	ITT	TOT
Cohort 1														
Treatment	157.848	0.068	0.144	158.557	154.53	3.155***	5.819***	152.997	156.644	1.244**	2.506**	156.24	0.016	0.021
Standard Error		(0.7594)	(1.5993)			(1.0285)	(1.8567)			(0.6111)	(1.2208)			
FWER p-value		0.9966				0.0389				0.27				
N		3397	3397			1938	1938			5335	5335			
Cohort 2														
Treatment	165.084	-0.027	-0.116	165.7	161.013	1.792	3.74	161.202	163.98	0.511	1.743	163.58	0.207	0.339
Standard Error		(0.7294)	(3.1512)			(1.2572)	(2.5177)			(0.6395)	(2.1523)			
FWER p-value		0.27				0.9966				0.8177				
N		3076	3076			1078	1078			4154	4154			
Pooled Cohorts														
Treatment	160.815	0.114	0.323	161.251	156.554	2.643***	5.080***	156.35	159.416	0.973**	2.401**	159.115	0.008	0.023
Standard Error		(0.5223)	(1.4689)			(0.8013)	(1.4942)			(0.4402)	(1.0735)			
FWER p-value		0.9833				0.0234				0.2073				
N		6473	6473			3016	3016			9489	9489			
* p<0.10; ** p<0.05; *** p<.01.														
CM = Control Mean, ITT = Intent to Treat, CCM = Control Complier Mean, TOT = Treatment on Treated.														
Family-wise error rate (FWER) is the probability that a given family of tests (defined for three grade groups, grades 1-4, 5-7 and pooled 1-7, and three cohort groups, cohort 1, cohort 2, and pooled cohort 1 and 2 total, so 9 families of outcomes are shown in the table) includes at least one false positive result. The FWER p-value is based on 100,000 iterations of randomly re-assigning the treatment grade and treatment-condition-within-grade indicators in our sample.														
The samples are for both cohorts, with each student having one observation for each year of treatment.														
The regression covariates include: baseline days present, baseline excused absences, baseline unexcused absences, baseline GPA, baseline course failures, baseline age, baseline old for grade, dummy for male, dummy for black, dummy for Hispanic, dummy for learning disability, and dummy for missing baseline data.														
We include randomization block groups which are school-by-grade-by-group. In cohort 1, these blocks are only school-by-grade. For cohort 2, we added block groups based on a student's original treatment assignment in cohort 1.														
We cluster at the student level because students have multiple observations, one for each year of treatment.														

Table 7. Main outcome – Percent days present.

	Grades 1 - 4				Grades 5 - 7				All Grades				H0: Grades 1-4 = Grades 5-7 p-value	
	CM	ITT	TOT	CCM	CM	ITT	TOT	CCM	CM	ITT	TOT	CCM	ITT	TOT
Cohort 1														
Treatment	92.391	-0.152	-0.322	92.214	90.238	1.443***	2.661***	89.122	91.609	0.438**	0.882**	90.968	0	0
Standard Error		(0.2173)	(0.4588)			(0.3675)	(0.6681)			(0.1944)	(0.3887)			
FWER p-value		0.7069				0.0037				0.1441				
N		3397	3397			1938	1938			5335	5335			
Cohort 2														
Treatment	94.005	0.258	1.128	92.629	91.573	1.283***	2.677***	90.194	93.345	0.526***	1.795***	91.603	0.041	0.222
Standard Error		(0.1979)	(0.8522)			(0.4653)	(0.9389)			(0.1923)	(0.6466)			
FWER p-value		0.4764				0.0764				0.0764				
N		3076	3076			1078	1078			4154	4154			
Pooled Cohorts														
Treatment	93.053	0.055	0.156	92.515	90.655	1.380***	2.652***	89.596	92.265	0.491***	1.210***	91.291	0	0
Standard Error		(0.1466)	(0.4122)			(0.2878)	(0.5405)			(0.1376)	(0.3358)			
FWER p-value		0.743				0.0004				0.011				
N		6473	6473			3016	3016			9489	9489			
* p<0.10; ** p<0.05; *** p<.01.														
CM = Control Mean, ITT = Intent to Treat, CCM = Control Complier Mean, TOT = Treatment on Treated.														
Family-wise error rate (FWER) is the probability that a given family of tests (defined for three grade groups, grades 1-4, 5-7 and pooled 1-7, and three cohort groups, cohort 1, cohort 2, and pooled cohort 1 and 2 total, so 9 families of outcomes are shown in the table) includes at least one false positive result. The FWER p-value is based on 100,000 iterations of randomly re-assigning the treatment grade and treatment-condition-within-grade indicators in our sample.														
The samples are for both cohorts, with each student having one observation for each year of treatment.														
The regression covariates include: baseline days present, baseline excused absences, baseline unexcused absences, baseline GPA, baseline course failures, baseline age, baseline old for grade, dummy for male, dummy for black, dummy for Hispanic, dummy for learning disability, and dummy for missing baseline data.														
We include randomization block groups which are school-by-grade-by-group. In cohort 1, these blocks are only school-by-grade. For cohort 2, we added block groups based on a student's original treatment assignment in cohort 1.														
We cluster at the student level because students have multiple observations, one for each year of treatment.														

Table 8. Days absent greater than 5, 10, 15, 20 by age (pooled years).

	Grades 1 - 4			Grades 5 - 7			H0: Grades 1-4 = Grades 5-7 p-value
	Control Mean	Treatment Students	Control Complier Mean	Control Mean	Treatment Students	Control Complier Mean	Treatment Students
Cohort 1							
Absent >5dys	0.846	-0.018 (0.0303)	0.872	0.895	-0.05 (0.0317)	0.932	0.477
Absent >10dys	0.579	0.001 (0.0406)	0.629	0.677	-0.125*** (0.0467)	0.743	0.043
Absent >15dys	0.326	0.053 (0.0383)	0.338	0.433	-0.131*** (0.0468)	0.49	0.002
Absent >20dys	0.179	0.045 (0.0321)	0.184	0.273	-0.109*** (0.0419)	0.322	0.004
Cohort 2							
Absent >5dys	0.692	0.011 (0.0773)	0.746	0.864	-0.136*** (0.0514)	0.958	0.113
Absent >10dys	0.422	-0.11 (0.0781)	0.523	0.58	-0.089 (0.0665)	0.647	0.839
Absent >15dys	0.232	-0.096 (0.0654)	0.328	0.352	-0.091 (0.0641)	0.393	0.956
Absent >20dys	0.128	-0.076 (0.0508)	0.213	0.216	-0.129** (0.0509)	0.28	0.457
Pooled Cohorts							
Absent >5dys	0.783	-0.004 (0.0322)	0.818	0.885	-0.076*** (0.0270)	0.933	0.087
Absent >10dys	0.515	-0.03 (0.0368)	0.569	0.646	-0.116*** (0.0380)	0.709	0.105
Absent >15dys	0.288	0.01 (0.0334)	0.314	0.408	-0.121*** (0.0377)	0.456	0.009
Absent >20dys	0.158	0.008 (0.0270)	0.183	0.255	-0.118*** (0.0327)	0.305	0.003
<p><i>Note:</i> Treatment students and treatment grade columns show treatment on the treated estimates. Each row shows results from a separate regression, with a fixed effect for grade 5-7 status and separate treatment-grade and within-treatment-grade-treatment-assignment indicators for grades 1-4 and 5-7. Standard errors clustered at the student-level. For intent to treat estimates, see Appendix Table 8a.</p> <p>* p<0.10; ** p<0.05; *** p<0.01.</p> <p>The samples are for both cohorts, with each student having one observation for each year of treatment.</p>							

The regression covariates include: baseline days present, baseline excused absences, baseline unexcused absences, baseline GPA, baseline course failures, baseline age, baseline old for grade, dummy for male, dummy for black, dummy for Hispanic, dummy for learning disability, and dummy for missing baseline data.

We include randomization block groups which are school-by-grade-by-group. In cohort 1, these blocks are only school-by-grade. For cohort 2, we added block groups based on a student's original treatment assignment in cohort 1.

We cluster at the student level because students have multiple observations, one for each year of treatment.

Table 9. Year effects (outcome – days absent).

	Grades 1 - 4			Grades 5 - 7		
	Control Mean	Treatment Students	Control Complier Mean	Control Mean	Treatment Students	Control Complier Mean
Cohort 1						
Year 1	12.299	0.871	12.753	14.335	-3.660***	16.942
		(0.8336)			(0.9820)	
Year 2	13.383	0.524	13.493	18.553	-4.997***	20.25
		(0.9919)			(1.5325)	
H0:Y1=Y2; p-value		0.435			0.202	
Cohort 2						
Year 1	10.609	-2.029	12.942	13.294	-2.837*	15.044
		(1.7392)			(1.6818)	
Year 2	10.08	-1.977	12.889	15.674	-5.190**	18.137
		(1.6726)			(2.0186)	
H0:Y1=Y2; p-value		0.506			0.875	
Pooled Cohorts						
Year 1	11.61	-0.034	12.546	14.012	-3.368***	16.201
		(0.7804)			(0.8467)	
Year 2	12.02	-0.265	12.962	17.648	-5.041***	19.286
		(0.8660)			(1.2210)	
H0:Y1=Y2; p-value		0.891			0.283	
<p><i>Note:</i> Treatment students and treatment grade columns show treatment on the treated estimates. Each row is a separate regression. For intent to treat estimates, see Appendix Table 9a. For cumulative effects for individuals who were in the treatment group in cohort 1 and re-randomized in cohort 2, see Appendix Table 9b.</p>						
<p>* p<0.10; ** p<0.05; *** p<0.01.</p>						
<p>The samples are for both cohorts, with each student having one observation for each year of treatment.</p>						

The regression covariates include: baseline days present, baseline excused absences, baseline unexcused absences, baseline GPA, baseline course failures, baseline age, baseline old for grade, dummy for male, dummy for black, dummy for Hispanic, dummy for learning disability, and dummy for missing baseline data.

We include randomization block groups which are school-by-grade-by-group. In cohort 1, these blocks are only school-by-grade. For cohort 2, we added block groups based on a student's original treatment assignment in cohort 1.

We cluster at the student level because students have multiple observations, one for each year of treatment.

Table 10. Academic impacts (pooled years).

Cohort 1			
	Control Mean	Treatment Students	Control Complier Mean
GPA	2.295	-0.04 (0.0530)	2.225
Course Failures	0.534	0.002 (0.0719)	0.601
Math Scores			
Standardized MAP*	0	-0.152** (0.0770)	0.052
Standardized Official Test*	0	-0.039 (0.0686)	-0.019
Valid MAP Data	0.929	0.039** (0.0191)	0.921
Valid Official Data	0.957	0.003 (0.0141)	0.976
Standardized MAP (imputed)	-1.344	-0.047 (0.0581)	-1.357
Reading Scores			
Standardized MAP*	0	-0.025 (0.0752)	-0.007
Standardized Official Test*	0	-0.066 (0.0638)	-0.024
Valid MAP Data	0.928	0.038** (0.0193)	0.927
Valid Official Data	0.958	0.009 (0.0137)	0.971
Standardized MAP (imputed)	-1.318	-0.005 (0.0569)	-1.333
Cohort 2			
	Control Mean	Treatment Students	Control Complier Mean
GPA	2.67	-0.01	2.454

		(0.0862)	
Course Failures	0.357	-0.037	0.437
		(0.1135)	
Math Scores			
Standardized MAP*	0	0	-0.055
		(0.1114)	
Standardized Official Test*	0	-0.021	-0.062
		(0.1008)	
Valid MAP Data	0.948	0.005	0.956
		(0.0230)	
Valid Official Data	0.952	0.013	0.961
		(0.0213)	
Standardized MAP (imputed)	-0.225	-0.004	-0.208
		(0.0428)	
Reading Scores			
Standardized MAP*	0	-0.084	0.01
		(0.1120)	
Standardized Official Test*	0	-0.064	-0.067
		(0.0988)	
Valid MAP Data	0.939	0.012	0.942
		(0.0250)	
Valid Official Data	0.94	0.029	0.936
		(0.0232)	
Standardized MAP (imputed)	-0.242	-0.076*	-0.167
		(0.0426)	
Pooled Cohorts			
	Control Mean	Treatment Students	Control Complier Mean
GPA	2.435	-0.026	2.319
		(0.0454)	
Course Failures	0.468	-0.014	0.532
		(0.0608)	
Math Scores			

Standardized MAP*	0	-0.085	0.01
		(0.0655)	
Standardized Official Test*	0	-0.033	-0.037
		(0.0570)	
Valid MAP Data	0.939	0.024	0.937
		(0.0150)	
Valid Official Data	0.955	0.007	0.969
		(0.0119)	
Standardized MAP (imputed)	-0.921	-0.026	-0.876
		(0.0377)	
Reading Scores			
Standardized MAP*	0	-0.047	-0.009
		(0.0653)	
Standardized Official Test*	0	-0.066	-0.044
		(0.0546)	
Valid MAP Data	0.934	0.026*	0.932
		(0.0158)	
Valid Official Data	0.951	0.016	0.957
		(0.0122)	
Standardized MAP (imputed)	-0.911	-0.007	-0.87
		(0.0370)	
* The official test used for accountability purposes switched after year 1 of cohort 2, but the MAP was given all four years that cover both cohorts.			
<i>Note:</i> Treatment students and treatment grade columns show treatment on the treated estimates. For intent to treat estimates, see Appendix Table 10.			
* p<0.10; ** p<0.05; *** p<.01.			
The samples are for both cohorts, with each student having one observation for each year of treatment.			
Standardized MAP and Standardized Official Test refer to test scores standardized to have mean zero and standard deviation one. Valid MAP Data and Valid Official Data are binary variables that indicate whether valid test score data is available for the student. Standardized MAP (imputed) is a variable that includes imputed values when valid test scores are not available. The imputation procedure is described in the text.			
The regression covariates include: baseline days present, baseline excused absences, baseline unexcused absences, baseline GPA, baseline course failures, baseline age, baseline old for grade, dummy for male, dummy for black, dummy for Hispanic, dummy for learning disability, and dummy for missing baseline data.			
We include randomization block groups which are school-by-grade-by-group. In cohort 1, these blocks are only school-by-grade. For cohort 2, we added block groups based on a student's original treatment assignment in cohort 1.			
We cluster at the student level because students have multiple observations, one for each year of treatment.			

Table 11. Academic impacts, by age group (pooled years).

Cohort 1							
	Grades 1 – 4			Grades 5 - 7			H0: Grades 1-4 = Grades 5-7 p-value
	Control Mean	Treatment Students	Control Complier Mean	Control Mean	Treatment Students	Control Complier Mean	Treatment Students
GPA	2.413	-0.027 (0.0721)	2.35	2.087	-0.096 (0.0712)	2.077	0.498
Course Failures	0.427	0.04 (0.0891)	0.443	0.722	-0.025 (0.1157)	0.808	0.655
<u>Math Scores</u>							
Standardized MAP*	0	-0.069 (0.0985)	0.019	0	-0.261** (0.1145)	0.088	0.204
Standardized Official Test*	0	0.005 (0.1022)	0.027	0	-0.069 (0.0892)	-0.063	0.582
Valid MAP Data	0.939	0.023 (0.0237)	0.946	0.911	0.054* (0.0312)	0.893	0.429
Valid Official Data	0.959	-0.007 (0.0214)	0.986	0.956	0.005 (0.0187)	0.973	0.682
Standardized MAP (imputed)	-1.337	0 (0.0739)	-1.359	-1.358	-0.114 (0.1008)	-1.357	
<u>Reading Scores</u>							
Standardized MAP*	0	-0.018 (0.1011)	-0.005	0	-0.031 (0.1029)	-0.015	0.93
Standardized Official Test*	0	-0.045 (0.0940)	0.064	0	-0.071 (0.0832)	-0.107	0.841
Valid MAP Data	0.934	0.019 (0.0246)	0.947	0.916	0.056* (0.0304)	0.908	0.338
Valid Official Data	0.96	-0.012 (0.0215)	0.988	0.956	0.02 (0.0175)	0.963	0.245
Standardized MAP (imputed)	-1.23	0 (0.0681)	-1.243	-1.473	0.052 (0.1025)	-1.535	
Cohort 2							
	Grades 1 – 4			Grades 5 - 7			H0: Grades 1-4 = Grades 5-7 p-value

	Control Mean	Treatment Students	Control Complier Mean	Control Mean	Treatment Students	Control Complier Mean	Treatment Students
GPA	2.791	0.045 (0.1227)	2.477	2.344	-0.047 (0.0996)	2.376	0.56
Course Failures	0.216	0.129 (0.1298)	0.199	0.739	-0.312* (0.1748)	0.817	0.043
Math Scores							
Standardized MAP*	0	0.144 (0.1593)	-0.122	0	-0.102 (0.1410)	-0.059	0.248
Standardized Official Test*	0	0.036 (0.1401)	-0.085	0	-0.021 (0.1366)	-0.107	0.772
Valid MAP Data	0.949	-0.02 (0.0346)	0.974	0.947	0.037 (0.0267)	0.933	0.197
Valid Official Data	0.667	-0.015 (0.0299)	0.885	0.943	0.050* (0.0274)	0.927	0.11
Standardized MAP (imputed)	-0.237	0.111** (0.0490)	-0.279	-0.192	-0.199** (0.0855)	-0.079	
Reading Scores							
Standardized MAP*	0	-0.049 (0.1654)	-0.003	0	-0.07 (0.1388)	-0.035	0.923
Standardized Official Test*	0	0.01 (0.1418)	-0.091	0	-0.114 (0.1272)	-0.08	0.517
Valid MAP Data	0.937	-0.001 (0.0382)	0.942	0.943	0.034 (0.0274)	0.94	0.455
Valid Official Data	0.653	0.011 (0.0334)	0.844	0.945	0.046* (0.0268)	0.935	0.412
Standardized MAP (imputed)	-0.249	-0.002 (0.0477)	-0.261	-0.223	-0.111 (0.0901)	-0.102	
Pooled Cohorts							
	Grades 1 – 4			Grades 5 - 7			H0: Grades 1-4 = Grades 5-7 p-value
	Control Mean	Treatment Students	Control Complier Mean	Control Mean	Treatment Students	Control Complier Mean	Treatment Students
GPA	2.567	-0.001 (0.0634)	2.407	2.166	-0.082 (0.0588)	2.211	0.348
Course Failures	0.341	0.072 (0.0733)	0.346	0.727	-0.12 (0.0979)	0.786	0.116

Math Scores							
Standardized MAP*	0	0.025 (0.0888)	-0.036	0	-0.194** (0.0907)	0.027	0.084
Standardized Official Test*	0	0.021 (0.0840)	-0.033	0	-0.059 (0.0745)	-0.071	0.474
Valid MAP Data	0.944	0.006 (0.0208)	0.955	0.928	0.046** (0.0204)	0.915	0.169
Valid Official Data	0.794	-0.01 (0.0179)	0.927	0.952	0.02 (0.0156)	0.958	0.214
Standardized MAP (imputed)	-0.886	0.075* (0.0450)	-0.937	-0.994	-0.112 (0.0712)	-0.849	
Reading Scores							
Standardized MAP*	0	-0.025 (0.0915)	-0.013	0	-0.034 (0.0847)	-0.045	0.945
Standardized Official Test*	0	-0.026 (0.0810)	-0.009	0	-0.077 (0.0702)	-0.108	0.638
Valid MAP Data	0.936	0.012 (0.0224)	0.939	0.929	0.045** (0.0203)	0.925	0.269
Valid Official Data	0.786	-0.002 (0.0191)	0.909	0.953	0.029* (0.0147)	0.954	0.207
Standardized MAP (imputed)	-0.827	0.001 (0.0422)	-0.835	-1.083	-0.079 (0.0743)	-0.862	
* The official test used for accountability purposes switched after year 1 of cohort 2, but the MAP was given all four years that cover both cohorts.							
<i>Note:</i> treatment students and treatment grade columns show treatment on the treated estimates. For intent to treat estimates, see Appendix Table 11.							
* p<0.10; ** p<0.05; *** p<.01.							
CM = Control Mean, ITT = Intent to Treat, CCM = Control Complier Mean, TOT = Treatment on Treated.							
The samples are for both cohorts, with each student having one observation for each year of treatment.							
Standardized MAP and Standardized Official Test refer to test scores standardized to have mean zero and standard deviation one. Valid MAP Data and Valid Official Data are binary variables that indicate whether valid test score data is available for the student. Standardized MAP (imputed) is a variable that includes imputed values when valid test scores are not available. The imputation procedure is described in the text.							
The regression covariates include: baseline days present, baseline excused absences, baseline unexcused absences, baseline GPA, baseline course failures, baseline age, baseline old for grade, dummy for male, dummy for black, dummy for Hispanic, dummy for learning disability, and dummy for missing baseline data.							
We include randomization block groups which are school-by-grade-by-group. In cohort 1, these blocks are only school-by-grade. For cohort 2, we added block groups based on a student's original treatment assignment in cohort 1.							
We cluster at the student level because students have multiple observations, one for each year of treatment.							

Table 12. Cohort 1 balance check between schools and grades, excluding treatment students.

	Control Grade	Treatment Grade (excluding treatment students)		Control School	Treatment School (excluding treatment grade)	
Students (n)	1112	1846		6614	1112	
Baseline Present data	150.36	150.18		150.37	150.36	
Baseline Absent data	16.1	16.13		15.69	16.1	*
Percent Male	0.53	0.53		0.53	0.53	
Age	9.48	8.74	**	8.97	9.48	
Percent Old for Grade	0.22	0.16	**	0.18	0.22	
Percent Black	0.6	0.57		0.58	0.6	
Percent Hispanic	0.37	0.39		0.39	0.37	
Percent Learning Disability	0.12	0.11		0.11	0.12	
Number of Course Failures in 2010-11 SY	0.65	0.7		0.62	0.65	
GPA in 2010-11 SY	2.23	2.21		2.24	2.23	
P-value on F-test						
Without missing data indicators		p=.515		p=.209		
With missing data indicators		p=.48		p=.213		
<p><i>Note:</i> The length of the CPS school year changed during the study. The number of days in the school year, by year, was: 2010-11,170; 2011-12, 170; 2012-13, 181; 2013-14, 178; 2014-15, 180. In the study sample, days present and days absent do not sum to total days in the school year because some students were not enrolled as CPS students or the full school year. P-value on null hypothesis that average values are jointly the same for treatment and control group comes from regressing treatment grade and school assignment indicator against the baseline covariates, with imputed missing values and a missing indicator for observations that had missing values. All exclude treatment assigned students. Standard error is clustered at the grade-school level.</p>						
<p>* p<0.10; ** p<0.05; *** p<0.01.</p>						

Table 13. Spillover effects cohort 1 - Main outcomes (excluding treatment students).

	Grades 1 - 4			Grades 5 – 7			All Grades		
	Control Mean	Treatment Grade	Treatment Schools	Control Mean	Treatment Grade	Treatment Schools	Control Mean	Treatment Grade	Treatment Schools
Days Absent	12.532	-0.154	0.465	14.864	1.626***	0.105	13.513	0.321	0.298
Standard Error		(0.4602)	(0.4311)		(0.5699)	(0.4102)		(0.3557)	(0.2929)
FWER p-value		0.9959	0.8859		0.1155	0.9959		0.9203	0.8925
Days Present	158.584	0.202	-0.62	155.672	-1.798*	1.216*	157.305	-0.713	0.071
Standard Error		(0.7748)	(0.6798)		(0.9850)	(0.7017)		(0.6276)	(0.5059)
FWER p-value		0.9959	0.9203		0.5366	0.5664		0.8725	0.9959
Percent Present	92.598	0.098	-0.271	91.11	-1.026***	0.07	91.98	-0.233	-0.146
Standard Error		(0.2750)	(0.2589)		(0.3720)	(0.2775)		(0.2219)	(0.1829)
FWER p-value		0.9955	0.8925		0.1361	0.9959		0.8925	0.9279
N	9703			5014			18448		
* p<0.10; ** p<0.05; *** p<.01.									
Family-wise error rate (FWER) is the probability that a given family of tests (defined for three grade groups, grades 1-4, 5-7 and pooled 1-7, and three cohort groups, cohort 1, cohort 2, and pooled cohort 1 and 2 total, so 9 families of outcomes are shown in the table) includes at least one false positive result. The FWER p-value is based on 100,000 iterations of randomly re-assigning the treatment grade and treatment-condition-within-grade indicators in our sample.									
The samples are for both cohorts, with each student having one observation for each year of treatment.									
The regression covariates include: baseline days present, baseline excused absences, baseline unexcused absences, baseline GPA, baseline course failures, baseline age, baseline old for grade, dummy for male, dummy for black, dummy for Hispanic, dummy for learning disability, and dummy for missing baseline data.									
We include randomization school-level block groups. Standard errors clustered at the grade-school.									
For segmented grade analysis, grades 1-4 and 5-7, we limited the sample to triple school blocks which include treatment schools within the same grade group.									

Table 14. Spillover effects - Academic outcomes (exclude treatment students).

	Grades 1 - 4			Grades 5 - 7			All Grades		
	CM	Treatment Grade	Treatment School	CM	Treatment Grade	Treatment School	CM	Treatment Grade	Treatment School
GPA	2.361	0.027	0.013	2.02	0.06	0.121	2.214	0.036	0.033
		(0.0559)	(0.0466)		(0.0913)	(0.0847)		(0.0503)	(0.0462)
		1.0000	1.0000		1.0000	0.9971		1.0000	1.0000
Course Failures	0.442	-0.034	0.017	0.839	-0.103	-0.074	0.607	-0.046	-0.016
		(0.0618)	(0.0575)		(0.1358)	(0.1342)		(0.0713)	(0.0698)
		1.0000	1.0000		1.0000	1.0000		1.0000	1.0000
Math Scores									
Standardized MAP*	0.006	-0.094	-0.009	0.039	0.086	-0.116	0	-0.03	-0.055
		(0.0620)	(0.0546)		(0.0838)	(0.0738)		(0.0549)	(0.0505)
		0.9947	1.0000		1.0000	0.9919		1.0000	0.9999
Standardized Official Test*	0.008	-0.057	-0.058	0.04	-0.013	-0.099	0	-0.049	-0.076
		(0.0817)	(0.0645)		(0.0922)	(0.0788)		(0.0594)	(0.0529)
		1.0000	1.0000		1.0000	0.9994		1.0000	0.9971
Valid MAP Data	0.904	-0.001	0.031*	0.904	-0.007	0.012	0.902	0.002	0.024**
		(0.0180)	(0.0170)		(0.0204)	(0.0156)		(0.0119)	(0.0110)
		1.0000	0.9661		1.0000	1.0000		1.0000	0.8365
Valid Official Data	0.942	0.021	0.007	0.938	0	0.016	0.939	0.008	0.014
		(0.0145)	(0.0115)		(0.0165)	(0.0123)		(0.0101)	(0.0087)
		0.9961	1.0000		1.0000	0.9994		1.0000	0.9883
Standardized MAP (imputed)	-1.47	-0.026	0.111	-2.233	0.158	0.093	-1.755	-0.021	0.123
		(0.1842)	(0.1734)		(0.1566)	(0.1007)		(0.1376)	(0.1273)
Reading Scores									
Standardized MAP*	0.02	-0.031	-0.015	0.007	0.087	-0.026	0	0.007	-0.018
		(0.0655)	(0.0588)		(0.0854)	(0.0688)		(0.0518)	(0.0484)
		1.0000	1.0000		1.0000	1.0000		1.0000	1.0000
Standardized Official Test*	0.018	-0.019	0.012	0.001	0.034	-0.027	0	-0.005	-0.006
		(0.0672)	(0.0591)		(0.0856)	(0.0703)		(0.0518)	(0.0467)
		1.0000	1.0000		1.0000	1.0000		1.0000	1.0000
Valid MAP Data	0.911	0.04	-0.021	0.907	-0.003	0.013	0.907	0.024	-0.004
		(0.0343)	(0.0344)		(0.0180)	(0.0153)		(0.0214)	(0.0212)
		0.9998	1.0000		1.0000	1.0000		0.9999	1.0000

Valid Official Data	0.946	0.016	0.009	0.941	-0.004	0.016	0.941	0.005	0.015*
		(0.0141)	(0.0105)		(0.0168)	(0.0124)		(0.0099)	(0.0084)
		0.9998	1.0000		1.0000	0.9992		1.0000	0.9642
Standardized MAP (imputed)	-1.41	0.276	-0.008	-1.855	0.033	0.248	-1.577	0.076	0.143
		(0.2202)	(0.1931)		(0.2329)	(0.1672)		(0.1646)	(0.1645)
N	5751			5014			13890		
* The official test used for accountability purposes switched after year 1 of cohort 2, but the MAP was given all four years that cover both cohorts.									
We include randomization school-level block groups. Standard errors clustered at the grade-school.									
Family-wise error rate (FWER) p-value is provided below the standard error. FWER is the probability that a given family of tests (defined for three grade groups, grades 1-4, 5-7 and pooled 1-7, so 30 families of outcomes (excluding imputed values) are shown in the table) includes at least one false positive result. The FWER p-value is based on 100,000 iterations of randomly re-assigning the treatment school and treatment-grade-within-school indicators in our sample.									
For segmented grade analysis, grades 1-4 and 5-7, we limited the sample to triple school blocks which include treatment schools within the same grade group.									
The regression covariates include: baseline days present, baseline excused absences, baseline unexcused absences, baseline GPA, baseline course failures, baseline age, baseline old for grade, dummy for male, dummy for black, dummy for Hispanic, dummy for learning disability, and dummy for missing baseline data.									
CM = Control Mean									
* p<0.10; ** p<0.05; *** p<.01 .									

Appendix Table 2a. Cohort 1 students who stayed in study schools and were randomized again in cohort 2.

Baseline Characteristics	Control	Treatment
Students (n)	231	345
Days Present in 2012-13 SY	166.05	166.31
Baseline Absent data	13.61	13.79
Percent Male	0.57	0.53
Age	9.93	9.91
Percent Old for Grade	0.16	0.16
Percent Black	0.51	0.51
Percent Hispanic	0.46	0.44
Percent Learning Disability	0.08	0.1
Missing GPA data in year 1213	0.01	0
Number of Course Failures in 2012-13 SY	0.55	0.56
GPA in 2012-13 SY	2.39	2.4
P-value on F-test		
Without missing data indicators		p=.921
With missing data indicators		p=.093
* p<0.10; ** p<0.05; *** p<0.01.		

Appendix Table 2b. Cohort 2 new entrants.

Baseline Characteristics	Control	Treatment
Students (n)	880	694
Days Present in SY12-13	157.44	156.78
Days Absent in SY12-13	14.11	13.89
Percent Male	0.52	0.53
Age	8.16	7.63
Percent Old for Grade	0.11	0.07
Percent Black	0.45	0.44
Percent Hispanic	0.5	0.48 *
Percent Learning Disability	0.05	0.03
Number of Course Failures in 2012-13 SY	0.37	0.38
GPA in 2012-13 SY	2.27	2.32
P-value on F-test		p=.176
Notes: * p<0.10; ** p<0.05; *** p<0.01.		

Appendix Table 3. Participation rates.

	Year 1		Year 2		Either Year	
	Treatment	Control	Treatment	Control	Treatment	Control
Cohort 1	0.5	0	0.46	0	0.52	0
Cohort 2	0.31	0	0.32	0	0.33	0
Cohort 1 participants who were also randomized in Cohort 2						
Cohort 1	0.52	0	0.52	0	0.54	0
Cohort 2	0.61	0	0.58	0	0.61	0

Appendix Table 5a. Main impacts with triple fixed effects (pooled years).

	Control Mean	Intent to Treat	Treatment on the Treated	Control Complier Mean	Unadjusted p-value	FWER
Cohort 1						
Days Absent	14.107	-0.512*	-1.022*	14.999	0.0895	0.3963
		(0.3015)	(0.6004)			
Days Present	156.644	1.123*	2.241*	156.505	0.0680	0.3492
		(0.6150)	(1.2236)			
Percent Present	91.609	0.376**	0.750**	91.099	0.0449	0.2794
		(0.1872)	(0.3727)			
Membership Days	170.751	0.611	1.219	171.504	0.2673	0.6224
		(0.5504)	(1.0972)			
N	5335					
Cohort 2						
Days Absent	11.465	-0.768***	-2.632***	14.221	0.0051	0.0817
		(0.2742)	(0.9388)			
Days Present	163.98	0.222	0.763	164.56	0.7176	0.9262
		(0.6150)	(2.1047)			
Percent Present	93.345	0.458***	1.569***	91.828	0.0059	0.0830
		(0.1662)	(0.5675)			
Membership Days	175.445	-0.546	-1.87	178.781	0.3462	0.6578
		(0.5790)	(1.9920)			
N	4154					
Pooled Cohorts						
Days Absent	13.109	-0.666***	-1.635***	14.606	0.0014	0.0337
		(0.2078)	(0.5087)			
Days Present	159.416	0.805*	1.975*	159.541	0.0667	0.3492
		(0.4388)	(1.0734)			
Percent Present	92.265	0.443***	1.087***	91.414	0.0006	0.0184
		(0.1282)	(0.3136)			
Membership Days	172.525	0.139	0.34	174.147	0.7285	0.9262
		(0.3993)	(0.9797)			
N	9489					

* p<0.10; ** p<0.05; *** p<.01.

The samples are for both cohorts, with each student having one observation for each year of treatment.

The regression covariates include: baseline days present, baseline excused absences, baseline unexcused absences, baseline GPA, baseline course failures, baseline age, baseline old for grade, dummy for male, dummy for black, dummy for Hispanic, dummy for learning disability, and dummy for missing baseline data.

We include randomization school-level block groups. Standard errors clustered at the grade-school.

Appendix Table 5b. Main outcome – Days absent (excluding students who were in cohort 1 treatment and re-randomized in cohort 2).

	Grades 1 - 4				Grades 5 – 7				All Grades				H0: Grades 1-4 = Grades 5-7 p-value	
	CM	ITT	TOT	CCM	CM	ITT	TOT	CCM	CM	ITT	TOT	CCM	ITT	TOT
Cohort 2														
Treatment	10.371	-0.387	-2.045	13.398	14.514	-2.078**	-5.020**	17.314	11.441	-0.758**	-3.130**	14.849	0.063	0.266
Standard Error		(0.3432)	(1.7888)			(0.8530)	(1.9918)			(0.3388)	(1.3812)			
N		2866	2866			891	891			3757	3757			
* p<0.10; ** p<0.05; *** p<.01.														
CM = Control Mean, ITT = Intent to Treat, CCM = Control Complier Mean, TOT = Treatment on Treated.														
The samples are for cohort 2, excluding students who were in cohort 1 treatment and re-randomized in cohort 2 to avoid conflating 2 and 4 year effects. Each student has one observation for each year of treatment.														
The regression covariates include: baseline days present, baseline excused absences, baseline unexcused absences, baseline GPA, baseline course failures, baseline age, baseline old for grade, dummy for male, dummy for black, dummy for Hispanic, dummy for learning disability, and dummy for missing baseline data.														
We include randomization block groups which are school-by-grade-by-group. In cohort 1, these blocks are only school-by-grade. For cohort 2, we added block groups based on a student's original treatment assignment in cohort 1.														
We cluster at the student level because students have multiple observations, one for each year of treatment.														

Appendix Table 5c. 4-Year effect for treatment students in Cohort 1, re-randomized in Cohort 2 (outcome – days absent).

	Grades 1 - 4				Grades 5 - 7				All Grades			
	CM	ITT	TOT	CCM	CM	ITT	TOT	CCM	CM	ITT	TOT	CCM
Cohort 2												
Treatment	9.86	-1.313	-1.562	11.651	13.905	-2.29	-2.79	15.795	11.9	-1.724	-2.035	13.399
Standard Error		(0.9244)	(1.0429)			(2.5801)	(2.9588)			(1.2807)	(1.4658)	
N		210	210			187	187			397	397	
* p<0.10; ** p<0.05; *** p<.01.												
CM = Control Mean, ITT = Intent to Treat, CCM = Control Complier Mean, TOT = Treatment on Treated.												
The samples are for both cohorts, with each student having one observation for each year of treatment.												
The regression covariates include: baseline days present, baseline excused absences, baseline unexcused absences, baseline GPA, baseline course failures, baseline age, baseline old for grade, dummy for male, dummy for black, dummy for Hispanic, dummy for learning disability, and dummy for missing baseline data.												
We include randomization block groups which are school-by-grade-by-group. In cohort 1, these blocks are only school-by-grade. For cohort 2, we added block groups based on a student's original treatment assignment in cohort 1.												
We cluster at the student level because students have multiple observations, one for each year of treatment.												

Appendix Table 6. Main impacts by age (pooled years with triple fixed effects).

	Grades 1 - 4		Grades 5 - 7		H0: Grades 1-4 = Grades 5-7 p-value	Grades 1 - 4		Grades 5 - 7		H0: Grades 1-4 = Grades 5-7 p-value
	Control Mean	Intent to Treat	Control Mean	Intent to Treat		Control Complier Mean	Treatment on the Treated	Control Complier Mean	Treatment on the Treated	
Cohort 1										
Days Absent	12.818	0.44 (0.3354)	16.369	-2.209*** (0.5803)	0	12.888	0.926 (0.7081)	18.271	-4.047*** (1.0609)	0
Days Present	157.848	-0.04 (0.7619)	154.53	3.069*** (1.0460)	0.03	158.785	-0.084 (1.6032)	153.193	5.623*** (1.8979)	0.086
Percent Present	92.391	-0.21 (0.2043)	90.238	1.406*** (0.3677)	0.003	92.335	-0.442 (0.4309)	89.208	2.576*** (0.6708)	0.014
Membership Days	170.667	0.4 (0.7159)	170.899	0.86 (0.8656)	0.412	171.673	0.842 (1.5056)	171.464	1.575 (1.5800)	0.573
N	3397		1938			3397		1938		
Cohort 2										
Days Absent	10.352	-0.470 * (0.2737)	14.454	-1.473 ** (0.6470)	0	12.964	-2.052 * (1.1939)	15.668	-3.092 ** (1.3572)	0
Days Present	165.084	-0.116 (0.7179)	161.013	0.704 (1.1347)	0.03	166.092	-0.508 (3.1333)	163.464	1.478 (2.3738)	0.086
Percent Present	94.005	0.249 (0.1694)	91.573	0.966 ** (0.3767)	0.003	92.672	1.085 (0.7372)	90.844	2.028 ** (0.7922)	0.014
Membership Days	175.437	-0.587 (0.6932)	175.467	-0.769 (1.0171)	0.412	179.056	-2.56 (3.0359)	179.132	-1.615 (2.1421)	0.573
N	3076		1078			3076		1078		
Pooled Cohorts										
Days Absent	11.807	-0.03 (0.2174)	15.771	-1.978*** (0.4414)	0	12.686	-0.083 (0.6125)	17.304	-3.784*** (0.8410)	0
Days Present	160.815	0.01 (0.5232)	156.554	2.301*** (0.7958)	0.03	161.546	0.028 (1.4744)	157.028	4.402*** (1.5059)	0.086
Percent Present	93.053	0.035 (0.1332)	90.655	1.268*** (0.2761)	0.003	92.571	0.1 (0.3752)	89.822	2.426*** (0.5256)	0.014
Membership Days	172.623	-0.02 (0.4966)	172.325	0.323 (0.6649)	0.412	174.232	-0.055 (1.3996)	174.332	0.618 (1.2693)	0.573
N	6473		3016			6473		3016		

* p<0.10; ** p<0.05; *** p<.01.

The samples are for both cohorts, with each student having one observation for each year of treatment.

The regression covariates include: baseline days present, baseline excused absences, baseline unexcused absences, baseline GPA, baseline course failures, baseline age, baseline

old for grade, dummy for male, dummy for black, dummy for Hispanic, dummy for learning disability, and dummy for missing baseline data.

We include randomization school-level block groups. Standard errors clustered at the grade-school.

For segmented grade analysis, grades 1-4 and 5-7, we limited the sample to triple school blocks which include treatment schools within the same grade group.

Appendix Table 7. Days absent by age, weighted by probability of being in older grades (pooled years).

	Grades 1 - 4				Grades 5 - 7				H0: Grades 1-4 = Grades 5-7 p-value	
	CM	ITT	TOT	CCM	CM	ITT	TOT	CCM	ITT	TOT
Cohort 1										
Treatment	12.738	0.515	1.102	12.854	15.809	-2.122***	-4.142***	19.221	0.001	0.001
		(0.3977)	(0.8492)			(0.6642)	(1.2698)			
Cohort 2										
Treatment	10.52	-0.419	-1.697	13.326	14.408	-1.667**	-3.492**	15.747	0.13	0.391
		(0.3484)	(1.3958)			(0.7571)	(1.5581)			
Pooled Cohorts										
Treatment	11.883	0.02	0.056	12.784	15.142	-1.972***	-4.048***	17.014	0	0.001
		(0.2603)	(0.7144)			(0.4890)	(0.9811)			
* p<0.10; ** p<0.05; *** p<.01.										
CM = Control Mean, ITT = Intent to Treat, CCM = Control Complier Mean, TOT = Treatment on Treated.										
<p><i>Note:</i> To conduct weighting, we pool together students in grades 1-4 and 5-7, and run a linear probability model that has as the dependent variable an indicator for being in grades 5-7 with all of our baseline characteristics (other than age) as explanatory variables. We then take the predicted probability that results from this model and re-weight the data using inverse probability weighting, so that the re-weighted average baseline characteristics for grades 1-4 and 5-7 are the same on average. This procedure up-weights young students who look more like older students in terms of their baseline characteristics, and down-weights older students who look too unlike younger students.</p>										
The samples are for both cohorts, with each student having one observation for each year of treatment.										
The regression covariates include: baseline days present, baseline excused absences, baseline unexcused absences, baseline GPA, baseline course failures, baseline age, baseline old for grade, dummy for male, dummy for black, dummy for Hispanic, dummy for learning disability, and dummy for missing baseline data.										
We include randomization school-level block groups. Standard errors clustered at the grade-school.										
For segmented grade analysis, grades 1-4 and 5-7, we limited the sample to triple school blocks which include treatment schools within the same grade group.										

Appendix Table 8a. Days absent greater than 5, 10, 15, 20 by age (pooled years, ITT estimates).

	Grades 1 – 4		Grades 5 - 7		H0: Grades 1-4 = Grades 5-7 p-value
	Control Mean	Treatment Students	Control Mean	Treatment Students	Treatment Students
Cohort 1					
Absent >5dys	0.846	-0.009 (0.0144)	0.895	-0.027 (0.0173)	0.418
Absent >10dys	0.579	0 (0.0193)	0.677	-0.068*** (0.0255)	0.034
Absent >15dys	0.326	0.025 (0.0181)	0.433	-0.071*** (0.0257)	0.002
Absent >20dys	0.179	0.021 (0.0152)	0.273	-0.059** (0.0230)	0.004
Cohort 2					
Absent >5dys	0.692	0.003 (0.0179)	0.864	-0.065** (0.0252)	0.028
Absent >10dys	0.422	-0.025 (0.0181)	0.58	-0.043 (0.0326)	0.636
Absent >15dys	0.232	-0.022 (0.0151)	0.352	-0.043 (0.0318)	0.536
Absent >20dys	0.128	-0.017 (0.0117)	0.216	-0.062** (0.0253)	0.106
Pooled Cohorts					
Absent >5dys	0.783	-0.002 (0.0114)	0.885	-0.040*** (0.0142)	0.036
Absent >10dys	0.515	-0.011 (0.0131)	0.646	-0.060*** (0.0200)	0.037
Absent >15dys	0.288	0.004 (0.0119)	0.408	-0.063*** (0.0200)	0.004
Absent >20dys	0.158	0.003 (0.0096)	0.255	-0.061*** (0.0174)	0.001
<i>Note:</i> treatment students columns show intent to treat estimates.					
* p<0.10; ** p<0.05; *** p<.01.					
The samples are for both cohorts, with each student having one observation for each year of treatment.					
The regression covariates include: baseline days present, baseline excused absences, baseline unexcused absences, baseline GPA, baseline course failures, baseline age, baseline old for grade, dummy for male, dummy for black, dummy for Hispanic, dummy for learning disability, and dummy for missing baseline data.					

We include randomization school-level block groups. Standard errors clustered at the grade-school.

For segmented grade analysis, grades 1-4 and 5-7, we limited the sample to triple school blocks which include treatment schools within the same grade group.

Appendix Table 8b. Days absent greater than 5, 10, 15, 20 (pooled years).

Attendance Indicators	Control Mean	Treatment Students	Control Complier Mean
Cohort 1			
Absent >5dys	0.864	-0.03 (0.0223)	0.895
Absent >10dys	0.615	-0.051 (0.0310)	0.676
Absent >15dys	0.365	-0.023 (0.0299)	0.401
Absent >20dys	0.213	-0.019 (0.0257)	0.242
Cohort 2			
Absent >5dys	0.739	-0.046 (0.0502)	0.83
Absent >10dys	0.465	-0.104* (0.0539)	0.577
Absent >15dys	0.264	-0.102** (0.0472)	0.363
Absent >20dys	0.152	-0.103*** (0.0374)	0.245
Pooled Cohorts			
Absent >5dys	0.817	-0.034 (0.0222)	0.865
Absent >10dys	0.558	-0.068** (0.0270)	0.628
Absent >15dys	0.327	-0.047* (0.0254)	0.376
Absent >20dys	0.19	-0.046** (0.0212)	0.236
<p><i>Note:</i> treatment students and treatment grade columns show treatment on the treated estimates. Each row is a separate regression. * p<0.10; ** p<0.05; *** p<.01.</p>			
<p>The samples are for both cohorts, with each student having one observation for each year of treatment.</p>			
<p>The regression covariates include: baseline days present, baseline excused absences, baseline unexcused absences, baseline GPA, baseline course failures, baseline age, baseline old for grade, dummy for male, dummy for black, dummy for Hispanic, dummy for learning disability, and dummy for missing baseline data.</p>			
<p>We include randomization school-level block groups. Standard errors clustered at the grade-school.</p>			
<p>For segmented grade analysis, grades 1-4 and 5-7, we limited the sample to triple school blocks which include treatment schools within the same grade group.</p>			

Appendix Table 8c. Days absent greater than 5, 10, 15, 20 (pooled years, ITT estimates).

Attendance Indicators	Control Mean	Treatment Students
Cohort 1		
Absent >5dys	0.864	-0.015 (0.0111)
Absent >10dys	0.615	-0.025 (0.0154)
Absent >15dys	0.365	-0.011 (0.0149)
Absent >20dys	0.213	-0.009 (0.0128)
Cohort 2		
Absent >5dys	0.739	-0.014 (0.0149)
Absent >10dys	0.465	-0.031* (0.0159)
Absent >15dys	0.264	-0.030** (0.0140)
Absent >20dys	0.152	-0.030*** (0.0111)
Pooled Cohorts		
Absent >5dys	0.817	-0.014 (0.0091)
Absent >10dys	0.558	-0.027** (0.0110)
Absent >15dys	0.327	-0.019* (0.0104)
Absent >20dys	0.19	-0.019** (0.0087)
Note: treatment students column shows intent to treat estimates.		
* p<0.10; ** p<0.05; *** p<.01.		
The samples are for both cohorts, with each student having one observation for each year of treatment.		
The regression covariates include: baseline days present, baseline excused absences, baseline unexcused absences, baseline GPA, baseline course failures, baseline age, baseline old for grade, dummy for male, dummy for black, dummy for Hispanic, dummy for learning disability, and dummy for missing baseline data.		
We include randomization school-level block groups. Standard errors clustered at the grade-school.		
For segmented grade analysis, grades 1-4 and 5-7, we limited the sample to triple school blocks which include treatment schools within the same grade group.		

Appendix Table 9a. Year effects.

	Outcome = Days Present							
	Grades 1 – 4				Grades 5 - 7			
	Control Mean	Intent to Treat	Treatment on the Treated	Control Complier Mean	Control Mean	Intent to Treat	Treatment on the Treated	Control Complier Mean
Cohort 1								
Year 1	152.497	0.196	0.418	152.553	151.689	2.473**	4.493**	150.87
		(0.9480)	(1.2085)			-2.0037	(2.1312)	
Year 2	163.661	-0.073	-0.154	164.98	157.582	3.989***	7.477***	155.117
		(0.9794)	(1.4384)			-2.0335	(2.6376)	
H0:Y1=Y2; p-value		0.828	0.825			0.365	0.327	
Cohort 2								
Year 1	163.698	-0.161	-0.735	164.843	161.98	-0.324	-0.703	164.414
		(0.9437)	(1.5411)			-4.2141	(3.1567)	
Year 2	166.558	0.165	0.687	166.38	159.995	4.080**	8.134***	158.048
		(0.9719)	(1.6709)			-3.9476	(3.0928)	
H0:Y1=Y2; p-value		0.795	0.79			0.028	0.024	
Pooled Cohorts								
Year 1	157.062	0.073	0.211	157.328	154.881	1.439	2.772	156.071
		(0.6689)	(0.9506)			-1.9017	(1.7629)	
Year 2	164.856	0.17	0.473	165.306	158.341	4.055***	7.762***	156.4
		(0.6803)	(1.1059)			-1.8607	(2.0383)	
H0:Y1=Y2; p-value		0.912	0.915			0.042	0.037	
Outcome = Percent of Days Present								
	Grades 1 – 4				Grades 5 - 7			
	Control Mean	Intent to Treat	Treatment on the Treated	Control Complier Mean	Control Mean	Intent to Treat	Treatment on the Treated	Control Complier Mean
Cohort 1								
Year 1	92.425	-0.157	-0.335	92.128	91.314	1.076***	1.954***	90.17

		(0.2410)	(0.3564)			(0.5104)	(0.6325)	
Year 2	92.355	-0.151	-0.318	92.318	89.081	1.896***	3.553***	87.859
		(0.2821)	(0.5186)			(0.5866)	(0.9516)	
H0:Y1=Y2; p-value		0.985	0.977			0.088	0.069	
Cohort 2								
Year 1	93.798	0.241	1.105	92.632	92.261	0.802	1.743*	91.204
		(0.2418)	(0.5095)			(1.0759)	(1.0400)	
Year 2	94.225	0.272	1.137	92.64	90.849	1.801***	3.591***	89.204
		(0.2342)	(0.6145)			(0.9502)	(1.1550)	
H0:Y1=Y2; p-value		0.908	0.978			0.101	0.121	
Pooled Cohorts								
Year 1	92.984	0.047	0.136	92.454	91.608	0.961***	1.850***	90.617
		(0.1688)	(0.2892)			(0.4799)	(0.5370)	
Year 2	93.127	0.056	0.155	92.6	89.637	1.884***	3.606***	88.411
		(0.1839)	(0.4021)			(0.5030)	(0.7460)	
H0:Y1=Y2; p-value		0.964	0.972			0.015	0.014	
Outcome = Days Absent, ITT Estimates								
	Grades 1-4			Grades 5-7				
	Control Mean	Intent to Treat	Treatment on Treated	Control Mean	Intent to Treat	Treatment on Treated		
Cohort 1								
Year 1	12.299	0.408	0.871	14.335	-2.015***	-3.660***		
		(0.3933)	(0.8336)		(0.5469)	(0.9820)		
Year 2	13.383	0.25	0.524	18.553	-2.666***	-4.997***		
		(0.4770)	(0.9919)		(0.8360)	(1.5325)		
H0:Y1=Y2; p-value		0.444	0.435		0.175	0.202		
Cohort 2								
Year 1	10.609	-0.443	-2.029	13.294	-1.305	-2.837*		
		(0.3902)	(1.7392)		(0.8244)	(1.6818)		

Year 2	10.08	-0.473	-1.977	15.674	-2.604**	-5.190**
		(0.4117)	(1.6726)		(1.0731)	(2.0186)
H0:Y1=Y2; p-value		0.558	0.506		0.996	0.875
Pooled Cohorts						
Year 1	11.61	-0.012	-0.034	14.012	-1.749***	-3.368***
		(0.2744)	(0.7804)		(0.4533)	(0.8467)
Year 2	12.02	-0.095	-0.265	17.648	-2.634***	-5.041***
		(0.3165)	(0.8660)		(0.6586)	(1.2210)
H0:Y1=Y2; p-value		0.888	0.891		0.304	0.283
<i>Note:</i> treatment students columns show intent to treat estimates.						
* p<0.10; ** p<0.05; *** p<.01.						
The samples are for both cohorts, with each student having one observation for each year of treatment.						
The regression covariates include: baseline days present, baseline excused absences, baseline unexcused absences, baseline GPA, baseline course failures, baseline age, baseline old for grade, dummy for male, dummy for black, dummy for Hispanic, dummy for learning disability, and dummy for missing baseline data.						
We include randomization school-level block groups. Standard errors clustered at the grade-school.						
For segmented grade analysis, grades 1-4 and 5-7, we limited the sample to triple school blocks which include treatment schools within the same grade group.						

Appendix Table 9b. Year effects (outcome – cumulative days absent across two years).

	Grades 1 - 4			Grades 5 - 7		
	Control Mean	Treatment Students	Control Complier Mean	Control Mean	Treatment Students	Control Complier Mean
Cohort 1						
Years 1 and 2	25.067	1.713	12.168	32.264	-7.910***	22.049
		(-1.5169)			(-2.1713)	
Cohort 2						
Years 1 and 2	20.273	-4.782*	15.918	28.454	-7.880**	20.525
		(-2.8656)			(-3.1601)	
Pooled Cohorts						
Year 1	23.106	-0.378	13.105	31.067	-7.888***	21.391
		(-1.3790)			(-1.7864)	
<i>Note:</i> treatment students and treatment grade columns show treatment on the treated estimates. Each row is a separate regression.						
* p<0.10; ** p<0.05; *** p<.01.						
The samples are for both cohorts, with each student having one observation for each year of treatment.						
The regression covariates include: baseline days present, baseline excused absences, baseline unexcused absences, baseline GPA, baseline course failures, baseline age, baseline old for grade, dummy for male, dummy for black, dummy for Hispanic, dummy for learning disability, and dummy for missing baseline data.						
We include randomization school-level block groups. Standard errors clustered at the grade-school.						
For segmented grade analysis, grades 1-4 and 5-7, we limited the sample to triple school blocks which include treatment schools within the same grade group.						

Appendix Table 10. Academic impacts (pooled years, ITT estimates).

Cohort 1			
	Control Mean	Treatment Students	FWER Adjusted P-Value
GPA	2.295	-0.02 (0.0266)	0.9987
Course Failures	0.534	0.001 (0.0362)	1
Math Scores			
Standardized MAP*	0	-0.073** (0.0372)	0.5466
Standardized Official Test*	0	-0.021 (0.0360)	0.9994
Valid MAP Data	0.929	0.019** (0.0096)	0.5053
Valid Official Data	0.957	0.001 (0.0073)	1
Reading Scores			
Standardized MAP*	0	-0.012 (0.0366)	0.9999
Standardized Official Test*	0	-0.034 (0.0334)	0.9881
Valid MAP Data	0.928	0.019** (0.0097)	0.5466
Valid Official Data	0.958	0.004 (0.0071)	0.9993
Cohort 2			
	Control Mean	Treatment Students	1
GPA	2.67	-0.003 (0.0259)	0.9999
Course Failures	0.357	-0.011 (0.0341)	
Math Scores			
Standardized MAP*	0	0 (0.0355)	1
Standardized Official Test*	0	-0.008 (0.0373)	1
Valid MAP Data	0.948	0.002 (0.0073)	0.9993

Valid Official Data	0.952	0.005 (0.0078)	
Reading Scores			0.9987
Standardized MAP*	0	-0.027 (0.0360)	0.9993
Standardized Official Test*	0	-0.023 (0.0368)	0.9994
Valid MAP Data	0.939	0.004 (0.0079)	0.9548
Valid Official Data	0.94	0.011 (0.0085)	
Pooled Cohorts			
	Control Mean	Treatment Students	
GPA	2.435	-0.01 (0.0188)	1
Course Failures	0.468	-0.006 (0.0251)	0.9478
Math Scores			
Standardized MAP*	0	-0.033 (0.0257)	0.9994
Standardized Official Test*	0	-0.015 (0.0258)	0.8192
Valid MAP Data	0.939	0.009 (0.0058)	0.9994
Valid Official Data	0.955	0.003 (0.0053)	0.9992
Reading Scores			
Standardized MAP*	0	-0.018 (0.0258)	0.9628
Standardized Official Test*	0	-0.03 (0.0248)	0.7628
Valid MAP Data	0.934	0.010* (0.0062)	0.9396
Valid Official Data	0.951	0.007 (0.0055)	0.9987
* The official test used for accountability purposes switched after year 1 of cohort 2, but the MAP was given all four years that cover both cohorts.			
<i>Note:</i> treatment students columns show intent to treat estimates.			
Family-wise error rate (FWER) is the probability that a given family of tests (defined for pooled 1-7, and three cohort groups, cohort 1, cohort 2, and pooled			

cohort 1 and 2 total, so 30 families (excluding imputed values) of outcomes are shown in the table) includes at least one false positive result. The FWER p-value is based on 100,000 iterations of randomly re-assigning the treatment-condition-within-grade indicators in our sample.

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

The samples are for both cohorts, with each student having one observation for each year of treatment.

The regression covariates include: baseline days present, baseline excused absences, baseline unexcused absences, baseline GPA, baseline course failures, baseline age, baseline old for grade, dummy for male, dummy for black, dummy for Hispanic, dummy for learning disability, and dummy for missing baseline data.

We include randomization school-level block groups. Standard errors clustered at the grade-school.

For segmented grade analysis, grades 1-4 and 5-7, we limited the sample to triple school blocks which include treatment schools within the same grade group.

Appendix Table 11. Academic impacts, by age group (pooled years, ITT estimates).

Cohort 1					
	Grades 1 - 4		Grades 5 - 7		H0: Grades 1-4 = Grades 5-7 p-value
	Control Mean	Treatment Students	Control Mean	Control Mean	Treatment Students
GPA	2.413	-0.013 (0.0346)	2.087	-0.052 (0.0393)	0.45
Course Failures	0.427	0.019 (0.0427)	0.722	-0.014 (0.0639)	0.668
Math Scores					
Standardized MAP*	0	-0.032 (0.0457)	0	-0.137** (0.0605)	0.163
Standardized Official Test*	0	0.003 (0.0509)	0	-0.038 (0.0497)	0.565
Valid MAP Data	0.939	0.011 (0.0114)	0.911	0.029* (0.0170)	0.382
Valid Official Data	0.959	-0.003 (0.0104)	0.956	0.003 (0.0102)	0.686
Reading Scores					
Standardized MAP*	0	-0.008 (0.0470)	0	-0.016 (0.0554)	0.912
Standardized Official Test*	0	-0.022 (0.0467)	0	-0.039 (0.0463)	0.802
Valid MAP Data	0.934	0.009 (0.0119)	0.916	0.030* (0.0166)	0.301
Valid Official Data	0.96	-0.006 (0.0105)	0.956	0.011 (0.0096)	0.238
Cohort 2					
	Grades 1 - 4		Grades 5 - 7		H0: Grades 1-4 = Grades 5-7 p-value
	Control Mean	Treatment Students	Control Mean	Control Mean	Treatment Students
GPA	2.791	0.011 (0.0289)	2.344	-0.022 (0.0494)	0.56
Course Failures	0.216	0.03 (0.0304)	0.739	-0.150* (0.0869)	0.049
Math Scores					
Standardized MAP*	0	0.036	0	-0.049	0.291

		(0.0405)		(0.0699)	
Standardized Official Test*	0	0.011	0	-0.01	0.794
		(0.0436)		(0.0676)	
Valid MAP Data	0.949	-0.005	0.947	0.018	0.154
		(0.0087)		(0.0132)	
Valid Official Data	0.667	-0.003	0.943	0.024*	0.069
		(0.0069)		(0.0135)	
Reading Scores					
Standardized MAP*	0	-0.012	0	-0.034	0.789
		(0.0420)		(0.0692)	
Standardized Official Test*	0	0.003	0	-0.055	0.455
		(0.0444)		(0.0632)	
Valid MAP Data	0.937	0	0.943	0.016	0.318
		(0.0097)		(0.0136)	
Valid Official Data	0.653	0.002	0.945	0.022*	0.199
		(0.0077)		(0.0133)	
Pooled Cohorts					
	Grades 1 - 4		Grades 5 - 7		H0: Grades 1-4 = Grades 5-7 p-value
	Control Mean	Treatment Students	Control Mean	Control Mean	Treatment Students
GPA	2.567	0	2.166	-0.043	0.268
		(0.0229)		(0.0312)	
Course Failures	0.341	0.026	0.727	-0.063	0.127
		(0.0264)		(0.0520)	
Math Scores					
Standardized MAP*	0	0.008	0	-0.097**	0.055
		(0.0302)		(0.0465)	
Standardized Official Test*	0	0.008	0	-0.031	0.448
		(0.0332)		(0.0398)	
Valid MAP Data	0.944	0.002	0.928	0.023**	0.092
		(0.0070)		(0.0105)	
Valid Official Data	0.794	-0.003	0.952	0.01	0.184
		(0.0058)		(0.0083)	

Reading Scores					
Standardized MAP*	0	-0.009	0	-0.017	0.871
		(0.0312)		(0.0438)	
Standardized Official Test*	0	-0.01	0	-0.04	0.544
		(0.0322)		(0.0375)	
Valid MAP Data	0.936	0.004	0.929	0.023**	0.141
		(0.0075)		(0.0105)	
Valid Official Data	0.786	-0.001	0.953	0.015*	0.12
		(0.0062)		(0.0078)	
* The official test used for accountability purposes switched after year 1 of cohort 2, but the MAP was given all four years that cover both cohorts.					
* p<0.10; ** p<0.05; *** p<.01.					
The samples are for both cohorts, with each student having one observation for each year of treatment.					
The regression covariates include: baseline days present, baseline excused absences, baseline unexcused absences, baseline GPA, baseline course failures, baseline age, baseline old for grade, dummy for male, dummy for black, dummy for Hispanic, dummy for learning disability, and dummy for missing baseline data.					
We include randomization school-level block groups. Standard errors clustered at the grade-school.					
For segmented grade analysis, grades 1-4 and 5-7, we limited the sample to triple school blocks which include treatment schools within the same grade group.					

Appendix Table 12. Cohort 1 balance check between schools and grades, excluding treatment students.

Grades 1-4						
	Control Grade	Treatment Grade (excluding treatment students)		Control School	Treatment School (excluding treatment grade)	
Students (n)	617	910		3532	617	
Baseline Present data	150.34	148.77	*	149.89	150.34	
Baseline Absent data	16.02	16.4		15.7	16.02	
Percent Male	0.52	0.52		0.52	0.52	
Age	7.97	7.25	**	7.51	7.97	*
Percent Old for Grade	0.16	0.08	**	0.12	0.16	
Percent Black	0.6	0.55		0.57	0.6	
Percent Hispanic	0.37	0.4		0.4	0.37	**
Percent Learning Disability	0.1	0.09		0.09	0.1	
Number of Course Failures in 2010-11 SY	0.48	0.51		0.52	0.48	
GPA in 2010-11 SY	2.45	2.43		2.38	2.45	
P-value on F-test						
Without missing data indicators		p=.248		p=.073		
With missing data indicators		p=.27		p=.079		
Grades 5-7						
	Control Grade	Treatment Grade (excluding treatment students)		Control School	Treatment School (excluding treatment grade)	
Students (n)	495	402		1689	495	
Baseline Present data	150.39	152.78	**	151.5	150.39	
Baseline Absent data	16.2	15.63		15.58	16.2	**
Percent Male	0.54	0.58		0.54	0.54	
Age	11.36	11.13		11.28	11.36	
Percent Old for Grade	0.29	0.25		0.26	0.29	
Percent Black	0.6	0.55		0.59	0.6	
Percent Hispanic	0.37	0.4		0.36	0.37	

Percent Learning Disability	0.16	0.13		0.15	0.16	
Number of Course Failures in 2010-11 SY	0.84	1.15		0.74	0.84	
GPA in 2010-11 SY	2	1.93		2.11	2	
P-value on F-test						
Without missing data indicators		p=.049		p=.6951		
With missing data indicators		p=.065		p=.369		
<i>Note:</i> The length of the CPS school year changed during the study. The number of days in the school year, by year, was: 2010-11,170; 2011-12, 170; 2012-13, 181; 2013-14, 178; 2014-15, 180. In the study sample, days present and days absent do not sum to total days in the school year because some students were not enrolled as CPS students or the full school year. P-value on null hypothesis that average values are jointly the same for treatment and control group comes from regressing treatment grade and school assignment indicator against the baseline covariates, with imputed missing values and a missing indicator for observations that had missing values. All exclude treatment assigned students. Standard error is clustered at the grade-school level.						
Limited sample to triple school blocks which include treatment schools with control grades within the same grade group						
* p<0.10; ** p<0.05; *** p<0.01.						

Appendix Table 13. Spillover effects Cohort 1 - Main outcomes (excluding treatment students).

	Grades 1 - 4			Grades 5 - 7			All Grades		
	Control Mean	Treatment Grade	Treatment Schools	Control Mean	Treatment Grade	Treatment Schools	Control Mean	Treatment Grade	Treatment Schools
Days Absent	12.532	-0.154	0.465	14.864	1.626**	0.105	13.513	0.321	0.298
Standard Error		(0.3850)	(0.3255)		(0.6795)	(0.4875)		(0.3214)	(0.2710)
FWER p-value		0.9939	0.8027		0.3084	0.9963		0.9165	0.8865
Days Present	158.584	0.202	-0.62	155.672	-1.798	1.216	157.305	-0.713	0.071
Standard Error		(0.8188)	(0.6873)		(1.1075)	(0.8661)		(0.6082)	(0.5182)
FWER p-value		0.9963	0.9296		0.7155	0.8027		0.8727	0.9963
Percent Present	92.598	0.098	-0.271	91.11	-1.026**	0.07	91.98	-0.233	-0.146
Standard Error		(0.2277)	(0.1918)		(0.4387)	(0.3155)		(0.2006)	(0.1690)
FWER p-value		0.9928	0.8027		0.33	0.9963		0.8727	0.9296
N	9703			5014			18448		
* p<0.10, ** p<0.05, *** p<.01.									
Family-wise error rate (FWER) is the probability that a given family of tests (defined for three grade groups, grades 1-4, 5-7 and pooled 1-7, and three cohort groups, cohort 1, cohort 2, and pooled cohort 1 and 2 total, so 9 families of outcomes are shown in the table) includes at least one false positive result. The FWER p-value is based on 100,000 iterations of randomly re-assigning the treatment grade and treatment-condition-within-grade indicators in our sample.									
The samples are for both cohorts, with each student having one observation for each year of treatment.									
The regression covariates include: baseline days present, baseline excused absences, baseline unexcused absences, baseline gpa, baseline course failures, dummy for male, dummy for black, dummy for Hispanic, dummy for learning disability, and dummy for missing baseline data.									

We include randomization school-level block groups. Cluster at the student level because students have multiple observations, one for each year of treatment.

For segmented grade analysis, grades 1-4 and 5-7, we limited the sample to triple school blocks which include treatment schools within the same grade group.

Appendix Table 14. Spillover effects - Academic outcomes (exclude treatment students).

	Grades 1 - 4			Grades 5 - 7			All Grades		
	Control Mean	Treatment Grade	Treatment School	Control Mean	Treatment Grade	Treatment School	Control Mean	Treatment Grade	Treatment School
GPA	2.361	0.027	0.013	2.02	0.06	0.121***	2.214	0.036	0.033
		(0.0365)	(0.0303)		(0.0455)	(0.0348)		(0.0259)	(0.0222)
		1.0000	1.0000		1.0000	0.7859		1.0000	0.9999
Course Failures	0.442	-0.034	0.017	0.839	-0.103	-0.074	0.607	-0.046	-0.016
		(0.0436)	(0.0376)		(0.0710)	(0.0587)		(0.0367)	(0.0324)
		1.0000	1.0000		0.9999	1.0000		1.0000	1.0000
<u>Math Scores</u>									
Standardized MAP*	0.006	-0.094*	-0.009	0.039	0.086	-0.116**	0	-0.03	-0.055*
		(0.0499)	(0.0421)		(0.0614)	(0.0477)		(0.0358)	(0.0308)
		0.9994	1.0000		1.0000	0.9859		1.0000	0.9996
Standardized Official Test*	0.008	-0.057	-0.058	0.04	-0.013	-0.099**	0	-0.049	-0.076***
		(0.0492)	(0.0389)		(0.0540)	(0.0409)		(0.0322)	(0.0272)
		1.0000	0.9999		1.0000	0.9859		0.9999	0.9552
Valid MAP Data	0.904	-0.001	0.031**	0.904	-0.007	0.012	0.902	0.002	0.024***
		(0.0142)	(0.0125)		(0.0194)	(0.0150)		(0.0104)	(0.0092)
		1.0000	0.9841		1.0000	1.0000		1.0000	0.9712
Valid Official Data	0.942	0.021*	0.007	0.938	0	0.016*	0.939	0.008	0.014**
		(0.0114)	(0.0096)		(0.0120)	(0.0093)		(0.0071)	(0.0064)
		0.9994	1.0000		1.0000	0.9998		1.0000	0.9939
Standardized MAP (imputed)	-1.47	-0.026	0.111*	-2.233	0.158**	0.093**	-1.755	-0.021	0.123***
		(0.0726)	(0.0657)		(0.0673)	(0.0420)		(0.0524)	(0.0467)
<u>Reading Scores</u>									
Standardized MAP*	0.02	-0.031	-0.015	0.007	0.087	-0.026	0	0.007	-0.018
		(0.0507)	(0.0426)		(0.0586)	(0.0450)		(0.0347)	(0.0298)

		1.0000	1.0000		0.9999	1.0000		1.0000	1.0000
Standardized Official Test*	0.018	-0.019	0.012	0.001	0.034	-0.027	0	-0.005	-0.006
		(0.0474)	(0.0376)		(0.0532)	(0.0392)		(0.0315)	(0.0261)
		1.0000	1.0000		1.0000	1.0000		1.0000	1.0000
Valid MAP Data	0.911	0.040**	-0.021	0.907	-0.003	0.013	0.907	0.024**	-0.004
		(0.0158)	(0.0142)		(0.0190)	(0.0150)		(0.0111)	(0.0100)
		0.9824	0.9999		1.0000	1.0000		0.9965	1.0000
Valid Official Data	0.946	0.016	0.009	0.941	-0.004	0.016*	0.941	0.005	0.015**
		(0.0112)	(0.0093)		(0.0119)	(0.0092)		(0.0070)	(0.0063)
		1.0000	1.0000		1.0000	0.9997		1.0000	0.9847
Standardized MAP (imputed)	-1.41	0.276***	-0.008	-1.855	0.033	0.248***	-1.577	0.076	0.143**
		(0.0788)	(0.0646)		(0.0931)	(0.0688)		(0.0606)	(0.0567)
N		5751		5014		13890			
* The official test used for accountability purposes switched after year 1 of cohort 2, but the MAP was given all four years that cover both cohorts.									
We include randomization school-level block groups. Cluster at the student level because students have multiple observations, one for each year of treatment.									
For segmented grade analysis, grades 1-4 and 5-7, we limited the sample to triple school blocks which include treatment schools within the same grade group.									
Family-wise error rate (FWER) p-score is provided below the standard error. FWER is the probability that a given family of tests (defined for three grade groups, grades 1-4, 5-7 and pooled 1-7, so 30 families of outcomes (excluding imputed values) are shown in the table) includes at least one false positive result. The FWER p-value is based on 100,000 iterations of randomly re-assigning the treatment school and treatment-grade-within-school indicators in our sample.									
* p<0.10; ** p<0.05; *** p<.01 .									