THE DYNAMIC EFFECTS OF A SUMMER LEARNING PROGRAM ON BEHAVIORAL ENGAGEMENT IN SCHOOL

Jaymes Pyne
(corresponding author)
Graduate School of Education
Stanford University
Stanford, CA 94305
pyne@stanford.edu

Erica Messner
Graduate School of Education
Stanford University
Stanford, CA 94305
ericamessner.em@gmail.com

Thomas S. Dee
Graduate School of Education
Stanford University
Stanford, CA 94305
and
National Bureau of Economic Research
Cambridge, MA 02138
tdee@stanford.edu

Abstract

Evidence that student learning declines or stagnates during summers has motivated an interest in programs providing intensive summer instruction. However, existing literature suggests that such programs have modest effects on achievement and no impact on measures of engagement in school. In this quasi-experimental study, we present evidence on the impact of a comprehensive and mature summer learning program that serves low-income middle school students and features unusual academic breadth, including a robust and well-designed social-emotional learning curriculum. Our results indicate that this program led to substantial reductions in unexcused absences, chronic absenteeism, and suspensions and a modest gain in English language arts test scores. We find evidence that the gains in behavioral engagement are dynamic, growing over time and with additional summers of participation.

https://doi.org/10.1162/edfp_a_00368

1. INTRODUCTION

Prominent and widely discussed studies claim that learning drops during summer months. For example, adaptive and vertically scaled assessment data from over 3.4 million elementary and middle school students shows that median student learning falls in summer by "one to two months in reading and a little more than one to three months of school-year learning in math" (Kuhfeld 2019, p. 27). Although debates on the size of those losses continue,[1] evidence that student learning often falls (or at least stagnates) during the summer has motivated a broad and longstanding interest in the design features, impact, and cost-effectiveness of summer learning programs. The COVID-19 pandemic's disruptive influence on schooling—through lost learning and social interaction—amplifies the need to understand how summer learning opportunities best support students, like recent interest in scaling up summer "vacation academies" for children who would not otherwise recoup lost instruction (Schueler 2020). Such innovative summer learning programs could provide students suffering from reduced academic and social-emotional learning much-needed structured time to regain ground lost during the COVID-19 pandemic.

The narrow design and targeting of prior programs highlight areas for further study. For example, because summer learning losses may be more consequential for socioeconomically disadvantaged students (see Downey, von Hippel, and Broh 2004; von Hippel, Workman, and Downey 2018; Quinn and Le 2018), most summer learning programs studied to date target their services to students from disadvantaged socioeconomic backgrounds and feature several hours a day of academic instruction, offered five days a week over a period of about five weeks. An extensive experimental and quasi-experimental evaluation literature suggests that these intensive programs affect student achievement modestly (i.e., effect sizes of 0.10 or less) and have no effects on important non-test outcomes like chronic absenteeism and suspension from school.[2] These programs rarely target middle school students, for whom summer learning loss is particularly large (Kuhfeld 2019). The program curricula in most existing studies also focus on just one or two subjects (typically, reading and, sometimes, mathematics). Instead, summer programs could provide a broad and vertically aligned curriculum addressing both the academic and social-emotional needs of students that help them feel more attached to learning and, by association, more engaged in school. Exploring such design features not found in prior studies can clarify how to structure impactful summer learning opportunities.

In this study, we evaluate Aim High, a comprehensive voluntary summer learning program promoting academic achievement and behavioral engagement among socioeconomically disadvantaged students. Students in the San Francisco Unified School District (SFUSD) are eligible to attend Aim High the summer prior to sixth grade through the summer before ninth grade. The program meets seven hours a day, five days a week, for five weeks. Its academically rich curriculum includes three core

1. For example, recent evidence (e.g., Hippel and Hamrock 2019; Kuhfeld 2019) suggests that evidence for such summer learning loss is unclear, because early evidence in support of the claim suffered from psychometric flaws, like the confounding influence of test-form changes and not using vertically aligned scales.
2. Both Lauer et al. (2006) and Kim and Quinn (2013) provide meta-analytic reviews of program evaluations in this domain. Because we focus on programs that provide general academic enrichment during the summer, we do not focus on remedial "summer school" programs.

subjects (i.e., mathematics, humanities, and science) delivered through both classroom instruction and project-based learning. Aim High also implements a social-emotional learning (SEL) curriculum in a course called “Issues and Choices.” This SEL curriculum features lessons scaffolded from summer-to-summer on topics such as fostering a growth mindset, understanding social identity, building community through mindfulness, challenging stereotypes, and advocating against bullying. Additionally, participants’ enthusiasm for the summer learning program seems clear, considering attendance rates are between 88 percent to 92 percent in the summers we observe.

Together, these features suggest that understanding the program’s effects would be a novel and useful contribution to the extensive literature on summer learning programs. Aim High appears to be unique among summer learning programs in offering both unusual academic breadth and an explicit SEL curriculum. These novel features of the Aim High curriculum serve as timely interventions, administered at crucial developmental and contextual moments in students’ lives, to recursively help them become engaged and invested in school over time (Walton and Wilson 2018). Motivated by these multifaceted design features, our outcome measures include not only test scores but also important measures of students’ behavioral engagement in school (i.e., chronic absenteeism and suspensions), measured over multiple years preceding and following when students first participate. Few prior summer learning program studies (i.e., Chaplin and Capizzano 2006; Augustine et al. 2016; Lynch and Kim 2017) consider such behavioral engagement measures, which strongly relate to academic achievement and school dropout (Finn 1989; Alexander, Entwisle, and Horsey 1997; Wald and Losen 2003; Gottfried 2010; Morris and Perry 2016; Gershenson, Jacknowitz, and Brannegan 2017).[3]

Other features of Aim High are novel as well. First, unlike programs evaluated in most prior studies, Aim High targets middle school students—for whom summer learning loss is particularly dramatic (Kuhfeld 2019). Second, Aim High is an unusually mature program and operates at some scale. Their summer program has operated for over thirty years in SFUSD, now serving around 1,000 students per summer at eight district sites. Third, both the multiyear design of Aim High and longitudinal data from SFUSD allow us to examine the longer-term effects of participating, and the “dosage” effects of attending over more than one summer.

To understand Aim High’s impacts on its participants, we use static, semidynamic, and dosage difference-in-differences (DD) designs that rely on longitudinal student-level data observed both before and after becoming eligible to participate in the summer learning program. One of the most important maintained assumptions of these designs is the “parallel trends” assumption, which rests on within-unit changes in outcomes over time among the untreated serving as valid counterfactuals to evaluate effects among the treated. Researchers often test the parallel trends assumption using an “event study” model, effectively comparing differences in trends between treated and

3. While much of this research linking absenteeism to later student outcomes is correlational, two instrumental-variable studies (Carlsson, Dahl, Öckert and Rooth 2015; Aucejo and Romano 2016) find that 10 missed days of instruction reduces achievement by 0.010 to 0.055 standard deviation. Other evidence suggests that chronic absenteeism and suspension also create negative externalities that harm the achievement of classmates (Perry and Morris 2014; Gottfried and Hutt 2019).

untreated units prior to an event. If event study trends in outcomes among treated and untreated groups do not vary significantly prior to anyone receiving the program, that evidence supports the parallel trends assumption.

In our case, the DD design leverages student-level longitudinal data, from first to eighth grade, and treats students' first participation in Aim High as an event whose impact we analyze. This student-level DD design controls unrestrictively for all unobserved and time-invariant determinants unique to each student. We rely on event study models to evaluate the parallel trends assumption, effectively comparing whether trends in targeted outcomes are equivalent between participants and nonparticipants, prior to any enrollment in the program. Our evidence supports that identifying assumption, suggesting comparable trends in attendance, disciplinary involvement, and English language arts (ELA) test scores among those who do and do not participate in Aim High after fifth grade. That means we find no evidence of bias through selection effects based on trends in outcomes prior to any enrollment in the summer program.

Additionally, Aim High participation appears to relate to longer-term improvements in several measures of behavioral engagement. Aim High participation over three summers is associated with a 58 percent reduction in chronic absenteeism relative to the sample mean of students who do not attend Aim High at all. We also find that those participating more than one summer (i.e., the majority of participants) experience a 37 percent reduction in the probability of being suspended. In contrast, its effects on California state assessment test scores are modest. We document a one-time improvement of 0.06 standard deviation on ELA assessments from participating, but no statistically significant impact on math performance, which we speculate could be due to several factors—including Common Core math changes during the period we study. Finally, we find substantial heterogeneity in these effects across student subgroups; reductions in chronic absenteeism and suspensions, and gains in ELA achievement, are prominent among boys and Latinx students.

2. SUMMER LEARNING OPPORTUNITIES AND STUDENT SUCCESS

The literature on summer learning loss (i.e., "summer slide") has motivated substantial interest in how to design, implement, and evaluate programs that extend structured learning opportunities for students through the summer (Alexander, Pitcock, and Boulay 2016). In general, existing evaluations of summer learning programs warrant modest and cautious optimism about their impact on student achievement. However, the design and targeting of those programs, including their focus on short-term standardized test scores as outcomes, suggest further study of how to structure summer learning opportunities to best support student success.

Two meta-analytic reviews illustrate the narrow focus of many summer learning program studies. The first (Lauer et al. 2006) evaluates eighteen early experimental studies of summer learning programs published between 1985 and 2003—fourteen of which examined reading outcomes and twelve of which examined math outcomes. They find that participating in a summer program leads to modest improvements in reading achievement ($d = 0.05$) and mathematics ($d = 0.09$). However, this early group of studies comes with several limitations of note. First, an open question in these early studies is whether interventions apply to broader populations of students; most only target students for remediation based on low prior achievement.[4] Second, most studies have small sample sizes and are implemented as controlled experiments, which may provide a poor guide to the impact of programs conducted in real-world settings and at a larger scale. Finally, most programs serve students in kindergarten, elementary school, or high school rather than in middle school, where summer learning loss appears particularly large (Kuhfeld 2019).

Another meta-analysis (Kim and Quinn 2013) summarizes evidence from forty-one literacy-focused summer initiatives situated in either the home or in classrooms. Seventeen of those studies (i.e., 40 percent) use experimental or rigorous quasi-experimental designs (e.g., regression discontinuity) and 82 percent of those seventeen studies focus exclusively on elementary-school students rather than middle-school students. Similar to Lauer et al. (2006), results suggest a modest treatment effect on reading achievement among low-income students ($d = .10$).[5]

Notably, prior studies mostly describe summer programs with a curricular focus (i.e., reading and/or math), much narrower than what students experience during the academic year. The narrow design features of prior summer programs (and their limited impact) suggest an important question: Might students uniquely benefit from a summer program that more closely parallels the breadth of the standard academic year, with intentional focus on social-emotional development? A program successfully providing such opportunities could shift the evaluation of summer learning in insightful ways, including examining longer-term effects and allowing for the cumulative effects of more than one summer of participation.

Such an alternative approach also implies that our evaluative lens should extend beyond test scores to other educationally relevant outcomes related to *behavioral engagement*, or students' participation in the work and social life of school. A long-standing research literature has recognized that the multifaceted dimensions of school engagement (e.g., behavioral versus emotional aspects of engagement) are important antecedents to longer-run educational success (Fredricks, Blumenfeld, and Paris 2004; McCarthy and Kuh 2006). Prominent contemporary education policies (and the corresponding measurement and reporting practices) now reflect the consensus view on the importance of such student outcomes. In particular, the federal Every Student Succeeds Act allows states to use measures of SEL as indicators in their school-accountability systems, in addition to test score achievement. However, because of concerns that survey-based SEL measures are currently "unreliable and unusable for accountability purposes" (Blad 2017), most state accountability systems instead rely on measuring chronic absenteeism, an important and more reliable indicator of behavioral engagement in school (Jordan and Miller 2017; Hough 2019). Measures of chronic absenteeism are still limited for our purposes, however; although chronic absenteeism functions as a good proxy for SEL skills (Holahan and Batey 2019), it is by no means a pure measure of those skills.

4. Similarly, several regression-discontinuity studies also focus on summer remediation programs targeted to low-performing students, and consistently found positive effects (Jacob and Lefgren 2004; Matsudaira 2008; Zvoch and Stevens 2011; Mariano and Martorell 2013).
5. A more recent addition to this literature (Zvoch and Robertson 2017) finds that random assignment of rising first grade students to a summer literacy program improves early literacy. By contrast, Lynch and Kim (2017) find that random assignment to an online summer mathematics program had no impact on math achievement.

Three prior studies of summer learning programs merit close attention because either they include dimensions of behavioral engagement among their outcomes and/or because their program design features an atypical breadth. In general, the results from these carefully designed studies are discouraging. First, a small-scale experimental evaluation of an online summer math program conducted by Lynch and Kim (2017) includes math-related behavioral and emotional engagement outcome measures. However, they find that, though the treatment increases students' summer participation in math activities, the program has no effect on math-related enjoyment or intrinsic motivation.

Two remaining noteworthy studies most closely resemble the Aim High program's curricular breadth. Chaplin and Capizzano (2006) examine the impact of Building Educated Leaders for Life (BELL), a summer program that seeks to "not only increase academic success . . . but also works to assist in social and emotional development by exposing program participants to positive role models, and by building self-esteem and encouraging parents to become more involved in their children's lives" (p. 4). The BELL program targeted children entering grades 1 through 7, in three cities. The randomized evaluation indicates that the program generated only a modest gain in reading achievement (i.e., equivalent to one month of learning) and had no effect on students' academic self-perceptions or parents' reports of positive or negative behaviors.

A RAND study (Augustine et al. 2016) examines the impact of the National Summer Learning Project (NSLP). This program consists of full-day programming, five days a week over five weeks, and focuses on both academics and "enrichment" activities (e.g., sports and arts). Each day features at least three hours of mathematics and ELA instruction, with no more than fifteen students per teacher. Though academic success is the primary focus, the randomized evaluation includes longitudinal outcome measures of behavioral engagement (i.e., attendance and suspensions) and SEL. The study found that NSLP led to modest, near-term math gains ($d = 0.08$) that faded out before the next summer. The NSLP program had no detectable effects on participants' future attendance, suspension, or social-emotional competencies.

In sum, the extensive literature on summer learning programs suggest that, as designed, they impact short-term achievement only modestly, and show no effects on dimensions of behavioral engagement, which recursively support more substantial and longer-term gains in educational success (Finn 1989; Walton and Wilson 2018).

3. THE AIM HIGH SUMMER LEARNING PROGRAM

The Aim High summer learning program is an independent nonprofit organization seeking to enhance the persistent behavioral engagement and academic achievement of middle school students, including those in SFUSD. Participating students attend seven hours a day over five days a week through five weeks at both independent sites and ones provided by SFUSD. Aim High is not a remediation program. Rather, its mission is to "create . . . life-changing opportunities during the summer and beyond"[6] for students from low-income families and neighborhoods, with the goals of building positive relationships with teachers, feeling a sense of belonging, and developing a lifelong love of learning. To enable such life-changing opportunities, the organization provides

6. Aim High, 2021, *About us*. Available http://aimhigh.org/about-us. Accessed 19 November 2021.

"engaging curriculum and pedagogy" and a "positive supportive culture" so participating students reorient their beliefs about schooling in a more positive direction (see Gardner Center [2016] for a complete discussion of Aim High's theory of change). Aim High also indicates that the program ensures students feel "seen," places a great deal of emphasis on strong and positive learning relationships, and strives to make sure that their educators relate to students and look like them (i.e., 80 percent of faculty are people of color).

Aim High has built a strong and sustained partnership with SFUSD district and school leaders, resulting in coordinated efforts to garner robust student participation and provide those participants with additional enrichment opportunities in the district during non-summer seasons. Typically, participating students at any Aim High site come from many different schools in the district, meaning participants have "Aim High friends" in addition to the friends they make at their traditional school. Aim High staff believe that this, along with the fact that over half of the organization's educators return from summer to summer, are the reasons about 70 percent of those enrolled participate in the program for two or more summers.

Aim High's curricular focus includes three traditional subjects: mathematics, humanities, and science; with state and nationally aligned standards that incorporate Common Core and Next Generation Science Standards. In general, Aim High uses SFUSD's curricular scope and sequence to ensure summer learning in the district aligns with each respective grade and subject. Every subject includes experiential and project-based learning through each summer, in addition to traditional in-class lessons.

Aim High has developed a fourth curricular focus on "Issues and Choices" to strengthen students' SEL and positive views of learning. Aim High's Issues and Choices coach, along with Aim High teachers and site directors, designed the Issues and Choices curriculum in 2013 by both borrowing from existing resources (e.g., from best practices in SEL learning from Aim High teachers over the years) and creating new content to build a comprehensive succession of lessons. The curriculum explicitly addresses SEL and behavioral engagement goals, including fostering a growth mindset, building awareness/ relationships in school, advocating against bullying, understanding identity, exploring social messages of gender, using mindfulness to build strong community, and empowering others to challenge stereotypes. Issues and Choices lessons are implemented by classroom instructors and have the same time and intensity of other courses—50 minutes a day over five days a week. The lessons are not scripted but rather present a content overview about the general topics to be covered by the teacher in each lesson (available in a separate online appendix that can be accessed on *Education Finance and Policy*'s Web site at https://doi.org/10.1162/edfp_a_00368).

Complementing a strong curricular focus on behavioral engagement in school is an emphasis on relationship-building through team teaching, which involves cooperation between lead teachers, teacher assistants, and interns. Aim High places high priority on selecting instructors who will build meaningful, positive relationships with students. To that end, the organization recruits Aim High alumni to serve as teacher assistants and emphasizes hiring lead teachers who contribute to diversity and live in or come from the communities they serve. Lead teachers and teacher assistants do not require teaching certifications, but Academic Coordinators who provide classroom teams with training and coaching are required to hold either a teaching certification or a graduate degree in an education-related field.

Instructors are allowed some latitude when choosing how to implement and present requisite lessons to students. Even so, the structure of Aim High instruction is quite consistent across classrooms and courses. Instructors in each of the four main subject areas teach students fifty minutes a day and benefit from both an Aim High Content Overview for a general pathway to follow over the summer and from large databases of lessons by Aim High instructors, shared across the organization and available at all times to shape curriculum. Students do not receive letter grades; rather, teacher teams provide each student with individualized narrative evaluations of their performance at the conclusion of each of their courses.

Students are first eligible to attend Aim High in the summer before their sixth-grade school year (i.e., as "rising sixth graders"). In the years we study, about 21 percent of students enter the summer before sixth grade, the majority (about 66 percent) enter the program the summer before they begin seventh grade, and 14 percent begin the summer before eighth grade. Fifty-seven percent participate for two summers and 14 percent participate for all three summers observed. Aim High sites span either four levels (rising sixth graders through rising ninth graders), or three levels (rising seventh through ninth). Historically, fewer sites offer four levels, which is why most students begin Aim High in the summer before seventh grade.

Before the application window, Aim High does a good deal of recruiting and marketing. The organization enlists alumni of the program to recruit at middle schools, present to classrooms, and attend summer opportunity fairs. There are also representatives who serve as liaisons for the organization at schools and encourage students to fill out applications to the program. Aim High representatives speak at parent events, Parent–Teacher Association meetings, or at community events such as church services. If possible, the organization sends letters to families of fifth-grade students, encouraging them to apply.

The application and admission processes are consistent across sites. Students and their parents or guardians fill out an application form, which staff use to assess need, diversity, interest, and commitment among applicants. Dimensions of need include family income at or near the federal poverty level, low parental education, family structure, and home stability. Within the structure of these admissions criteria, acceptance into Aim High is flexible. For example, site directors and staff accommodate students in unstable living situations who are harder to contact or whose parents are not involved in their education. In the summers of 2015 through 2017 in the district under study, Aim High accepted about 80 percent of those who applied to the program, who went on to attend any of thirty-four middle schools in the district following first participating. Typically, those not accepted were from higher socioeconomic backgrounds than those accepted. Those accepted are also much more likely to be black or Latinx, and more likely to be female, compared with those denied.

The median household annual income among participants is about $38,000—or roughly two fifths of the median household income citywide. Forty-eight percent of Aim High parents have a high school education or less. The parents of 71 percent of Aim High students report qualifying for free or reduced-price lunch while another 10 percent are unsure whether their children qualify for school meal subsidies. Across SFUSD, 52 percent of students participate in the district's free or reduced-price lunch program. Aim High also considers how applicants contribute to the program's diversity, by race, ethnicity, language, and ability. Twenty-six percent of participating students identify as Latinx, 10 percent black, 54 percent Asian, and 1 percent white. Across SFUSD, 31 percent of students identify as Latinx, 8 percent black, 35 percent Asian, 14 percent white, and 12 percent of another race or ethnicity.

Additionally, the 21 percent of participating students who begin Aim High in the summer before sixth grade are less likely to be black or Asian, and more likely to be Latinx, English language learners, or from parents with less than a high school diploma, than their peers who begin Aim High in the summers before seventh or eighth grades. Those who attend Aim High for only one summer versus two summers are less likely to be Asian and more likely to be female. The smaller set of students observed attending Aim High for three summers were increasingly likely to be male and to have a parent who did not graduate from high school (appendix table B4, available in the online appendix). Program sites are also diverse in terms of the schools from which students come. In the summers we study, the seven Aim High sites each represent between twenty and fifty-two different feeder schools in the district, and the average site serves students representing forty-one feeder schools.

4. THE CURRENT STUDY

We ask how Aim High affects participants' behavioral engagement and achievement. An evaluation of Aim High offers a novel and useful contribution to the literature on summer learning program effectiveness because of its design features (i.e., academic breadth and an SEL focus), its large-scale operations, and its relevant targeting on middle school students. Middle school students face many contextual and developmental challenges that affect their ongoing academic engagement and achievement in school (Eccles 2004; Rockoff and Lockwood 2010; West and Schwerdt 2012). During that challenging time in early adolescence, educational programming that boosts students' engagement and learning in school initiates recursive processes that lead to long-term school successes (Finn 1989; Walton and Wilson 2018). This means that attending Aim High a single summer may continue to benefit students academically and behaviorally over several years, serving as a catalyst for students' positive feelings about school. Furthermore, because Aim High programming has many thematic goals that are scaffolded from one summer to the next (e.g., middle school transition goal-setting among rising sixth graders, evaluating academic goals among rising seventh graders, successful high school transitions among rising eighth graders), students participating over multiple summers may be well-positioned to enjoy additional academic and behavioral benefits from the program. As we discuss below, the longitudinal data we use and the methods we apply allow us to provide novel evidence on the dynamic effects of program participation.

5. DATA AND METHODS

Data

The data come from Aim High and SFUSD, from the 2009–10 through 2017–18 school years. We define two intent-to-treat (ITT) samples using cohorts of fifth-grade students, since students first become eligible to participate in Aim High the summer following their fifth-grade year (i.e., as rising sixth graders). We use data on the two cohorts of fifth-grade students from the 2013–14 and 2014–15 school years when examining behavioral engagement outcomes (i.e., absence and suspension from school). This design strategy maximizes the benefits of the panel dataset, which includes annual observations for each of the students in these two cohorts during a conventional grade progression from grades 1 through 8 (i.e., using data from the school years from 2009–10 through 2017–18). Given the years of data available to us, this means we are able to observe Aim High participants and their peers five or six years prior to first participating in the program and two or three years after first participating. Consequently, these sample restrictions do not allow us to assess program effects on rising ninth graders.

Our two-cohort analytical sample consists of 7,908 students and 57,559 student-year observations for which we observe students at least 175 days in a given school year, from grades 1 through 8.[7] Student mobility into and out of the school district implies a somewhat unbalanced panel (i.e., we do not observe all students in each school year, nor for 175 or more days in a given school year). Such missingness potentially threatens the internal validity of our study. For example, if students with an unobserved propensity for poorer educational outcomes were more likely to remain in SFUSD because of Aim High, we would understate the true impact of the program (i.e., negative selection into treatment). However, we find evidence that missingness is conditionally random with respect to Aim High participation. Using our preferred panel-based specifications, we examine the "effect" of Aim High participation on missingness and find small and statistically insignificant effects (online appendix table B3). Additionally, in online appendix tables D2 through D5 we show that balanced-panel results (i.e., retaining only students observed in all grades 1 through 8), are statistically and substantively similar to results using an unbalanced panel design.

To evaluate the impacts of Aim High participation on state test scores, we also define a one-cohort ITT sample of 4,322 students (14,853 student-year observations) who sat for the ELA and math California Assessment of Student Performance and Progress (CAASPP) tests in fifth grade during the 2015–16 school year.[8] The CAASPP consists of Smarter Balanced Summative Assessments aligned to the Common Core State Standards. California began administering the CAASPP in the 2014–15 school year for students in grades 3 through 8. This means we observe test scores over each of four academic years (i.e., 2014–15 through 2017–18) in grades 4 through 7. As with the two-cohort sample, auxiliary regressions suggest that the missingness in these unbalanced panel data is conditionally random with respect to Aim High participation (online appendix table B3). We also find similar results when using the smaller sample of students with complete observations over the four-year study window (online appendix table D6).

We measure treatment status using Aim High records capturing each student's enrollment in the summer learning program, along with a range of personal identifiers.

7. This baseline sample excludes a small number of students ($N = 179$) who were in fifth grade for fewer than 175 days. In online appendix table B1, we show that these students were more likely to be absent, to be English language learners, to have lower test scores, and were less likely to be Asian. Our subgroup analyses allow us to explore the external-validity implications of this sample construction.
8. This sample definition excludes a small number of students who were enrolled but did not sit for the ELA and math state tests in their fifth-grade year (online appendix table B2).

Aim High shares their list of participants with SFUSD data managers, who link Aim High records both to randomized student identifiers and to district administrative data. Using these data, we construct a simple binary indicator equal to 1 for student-year observations from students who participate in Aim High during any previous summer (i.e., a "static" measure of treatment). Nearly 7 percent of our two-cohort ITT sample (i.e., 520 out of 7,908 students) participate in Aim High at least once. We also use the timing of Aim High participation to define less restrictive and flexibly dynamic measures of program participation. These include binary indicators for the academic year after the first summer of program participation and separate indicators for being one or two academic years after that first participation. These measures flexibly allow for the initial participation in Aim High to have effects that increase or decline over time. Additionally, we also include measures that consider dosage effects by constructing binary measures that identify student-year observations occurring after participating one, two, and three years.

We construct several student outcome measures using SFUSD administrative data, which contain student attendance as time enrolled, time present, and number of excused and unexcused absences. We calculate the absence rate by dividing the time the student is absent from school by the amount of time they are enrolled in the district (i.e., between 175 and 180 days in total). We do the same to calculate excused and unexcused absence rates. Finally, we calculate a chronically absent indicator to flag students who were absent for more than 10 percent of days enrolled during a given school year, so long as they were enrolled for 175 to 180 days.

SFUSD suspension data include records of each suspension incident (whether suspended in-school or out-of-school). From these records, we create a flag indicating whether a student is suspended at least once each school year. We measure academic achievement using annual student ELA and math assessments. Our dataset includes four years of ELA and math state test score data, from 2014–15 through 2017–18. For each grade and school year, we use state test scale scores to create standardized scores separately for ELA and math tests, within every grade level and school year, each of which has a mean of zero and a standard deviation of 1.

Our analyses also include several time-varying, student-level covariates based on the SFUSD administrative data. We construct binary indicators for English learner status, special education status, and foster care status. The final covariates in our model reflect the educational status of parents or guardians. SFUSD administrative data contain parent or guardian educational status, split into the following categories: not a high school graduate, high school graduate, some college, college graduate, and graduate school/postgraduate degree. We preserve these categories and create one "parent education" variable that indicates the highest level of education completed by any parent or guardian. We retain students whose parents or guardians do not report their education level with an additional category we call "not reported."

Research Design

We use student-year panel data from SFUSD to estimate the effects of Aim High participation on behavioral engagement (i.e., attendance and suspension from school) and academic achievement (i.e., ELA and math state test scores). We do so by comparing changes in these outcomes among those who participated in Aim High to outcomes of students who either never participated or had yet to participate in Aim High. This DD approach effectively compares the change in outcomes among treated students to the contemporaneous change among untreated students. A key assumption of DD models is that trends between the two groups proceed in parallel before exposure to the policy or program shock.

Our analyses begin with a basic "static" DD model, which assumes that Aim High participation leads to a constant, one-time change in a given student outcome. This specification takes the following form:

$$Y_{st} = \alpha_s + \gamma_t + \theta A_{st} + \boldsymbol{\beta}\mathbf{X}_{st} + \varepsilon_{st}, \tag{1}$$

where Y_{st} is outcome Y for student s at time t. α_s are student fixed effects, which account for all observed and unobserved time-invariant characteristics of each student. γ_t are fixed effects unique to each school year that account for common disturbances across all students in a given year. ε_{st} is presumed to be a mean-zero error term with clustering at the student level.[9] $\mathbf{X}_{st}$ is a vector of time-varying characteristics of students and their families, including their special education status, English language proficiency, parents' or guardians' highest level of education, and grade-level fixed effects.[10] θ is the coefficient of interest, representing the estimated effect of A_{st}, a binary indicator for whether a student participated in any summer prior before year t.

The static DD specification represented in equation 1 assumes a constant treatment effect over time. However, the character of the Aim High program suggests dynamic effects. For example, Aim High may begin a recursive cycle of improved behavioral engagement and achievement due to the SEL skills students learn while enrolled in the program. As those skills build recursively over the years, we would expect larger treatment effects in subsequent school years as those skills lead to ever-greater school success (see Finn 1989; Walton and Wilson 2018). Alternatively, the effects of Aim High could instead "fade out" in the years after initial participation. To test for time-varying treatment effects, we next use a semi-dynamic DD model that unrestrictively allows for treatment effects unique to the school year immediately after a student first participates, one year later, and two years later:

$$Y_{st} = \alpha_s + \gamma_t + \sum_{n=0}^{2} \delta_{-n} A_{s,t-n} + \boldsymbol{\beta}\mathbf{X}_{st} + \varepsilon_{st}. \tag{2}$$

In this model, the three coefficients of interest are represented by δ_n, which identify the effects of Aim High after the summer of a student's initial participation (i.e., $A_{s,t-0}$) as well as the current effect of having participated one year earlier (i.e., $A_{s,t-1}$) and two years earlier (i.e., $A_{s,t-2}$). We test the equivalence of these coefficients in the null hypothesis of a constant treatment effect:

$$H_0\colon \delta_0 = \delta_{-1} = \delta_{-2}.$$

9. We also examined the robustness of our inferences to standard errors that instead allow for clustering within schools. Because students attend different elementary and middle schools during our study window, we clustered on the first middle school attended. Those results are similar to the ones reported here.

10. We include time-varying covariates out of an abundance of caution for their influencing the effect of Aim High on our outcomes of interest. However, results are nearly identical when we do not control for this vector of covariates.

A "dosage" treatment effect model is also theoretically plausible if additional summers of Aim High participation allow students to build on the skills and knowledge developed over previous summers. Alternatively, if additional summers of participation were educationally redundant, the effects unique to additional summers of participation would be smaller. We examine this through a flexible dosage DD that takes the following form:

$$Y_{st} = \alpha_s + \gamma_t + \sum_{j=1}^{3} \pi_j A_{st}^{j} + \boldsymbol{\beta}\mathbf{X}_{st} + \varepsilon_{st}. \tag{3}$$

Here, the variables of interest are represented as π_j, which identify the effect of whether student s in school year t had just participated in Aim High a first, second, or third time (i.e., $j = 1$, 2, or 3). Notably, using unrestrictive dummy variables capturing a student's dosage implies that we are not imposing a functional form on the effects of additional years of participation. In supplementary models (see online appendix table C1), we simultaneously test dosage effects against passive lagged effects on each outcome.

Arguably, the most critical maintained assumption of this quasi-experimental approach is that the year-to-year outcome changes among comparison students (i.e., those without a change in treatment status) serve as valid counterfactuals for what would have changed among treatment students in the absence of treatment. This "parallel trends" assumption is fundamentally untestable. However, we can provide qualified evidence on the validity of this important assumption through unrestrictive event study specifications that allow us to examine whether treatment and comparison group students had similar year-to-year changes in outcomes prior to the onset of treatment. To the extent that this hypothesis is true, it is consistent with the parallel trends assumption. We examine this question in event study specifications of the following form:

$$Y_{st} = \alpha_s + \gamma_t + \sum_{\tau=1}^{4} \delta_\tau A_{s,t+\tau} + \sum_{n=0}^{2} \delta_{-n} A_{s,t-n} + \boldsymbol{\beta}\mathbf{X}_{st} + \varepsilon_{st}. \tag{4}$$

This event study specification effectively extends the semi-dynamic specification (equation 2). That is, the semi-dynamic models do not control for each year prior to participating in Aim High (i.e., "leads" of treatment adoption) and thus tacitly embed the parallel-trends assumption into the models. On the other hand, event study specifications do control for leads and allow for fixed effects unique to each year prior to participating in Aim High. That means the coefficients of interest are represented as δ_{-n} and δ_τ, which designate the effect for student s in year t of participation in Aim High n years in the future or τ years in the past. The reference category includes those never participating in Aim High and those in school five or six years prior to their first summer of participation in Aim High. To examine the assumption of parallel trends, we test whether, prior to their participation, treatment students have year-to-year changes in outcomes distinct from comparison students:

$$H_0\colon \delta_4 = \delta_3 = \delta_2 = \delta_1 = 0.$$

If H_0 is not rejected in this parametric test of pretreatment trends, it serves as strong evidence for maintaining the parallel trends assumption. Thus, while the semi-dynamic models are part of the confirmatory results, the event studies are robustness checks of those semi-dynamic results, in part to assess each model's underlying assumption of parallel trends. Results from these event studies can be found in online appendix C, including figure C1 and tables C2 and C3.

In the presence of the kinds of dynamic treatment effects we seek to model in our semi-dynamic and dosage specifications, a conventional static DD model can even apply negative weights to some treated observations. Several recent methodological studies have underscored how DD research designs like ours can sometimes reflect a tacit weighting that can be empirically consequential in the presence of treatment heterogeneity. For example, DD designs effectively upweight observations that have a higher conditional variance in the treatment indicator (i.e., those who change treatment status closer to the middle of our longitudinal window). In our context, this would imply that our static DD places more emphasis on program effects among those who enter before sixth grade relative to those who enter later. This property also implies that such DD estimates are sensitive to the time window used. To assess the empirical relevance of these concerns, we implemented procedures recently introduced by de Chaisemartin and D'Haultfoeuille (2020) by examining the weights implied by static DD specifications. We found they produced no negative values.

Even so, our results suggest that the effects of Aim High are clearly dynamic rather than static, leading us to emphasize interpreting the semi-dynamic and dosage DD models over the static models. Reliance on these models raises different but equally important issues about external validity, because not every Aim High student is observed over multiple post-treatment periods. That is, 57 percent of students participating in Aim High in our sample attend for two summers and 14 percent participate all three summers. Those attending multiple summers are more likely to be Asian and male (i.e., differences of roughly 10 percentage points, online appendix table B4). The smaller group of students who attended Aim High for three summers were also more likely to have parents who did not graduate high school. We note that these take-up patterns may have external-validity implications for the dosage estimates we present. We examine such heterogeneity in the estimated effects of Aim High in the results presented below.

6. RESULTS

Table 1 displays descriptive statistics of the unbalanced two- and one-cohort panel analytic samples, differentiating Aim High participants from comparison group students. Note that the descriptives we report here span pre- and post-program participation; we later parse those time frames in our event study design, which reports unrestrictive year-to-year changes between participants and nonparticipants, both before and after students have an opportunity to participate in Aim High. Our two-cohort analytic sample contains 7,908 unique students followed over multiple school years (columns 1 and 2 of table 1). The 520 unique students in the sample who ever participate in Aim High have lower rates of absence than nonparticipants across grades 1 through 8, by about 1.5 percentage points, (2.0 percent compared with 3.4 percent). Unexcused absence rate trends are similar. Non–Aim High students are labeled "chronically absent" at about 2.3 times the rate of Aim High students across this grade range. Aim High students' suspension rates are half that of other students from grades 1 through 8 (0.7 compared to 1.4 percent). Aim High and non–Aim High students' ELA and math test scores are similar across our four school years of data in each of the two cohorts.

Table 1. Descriptive Statistics

	Analytic Sample			
	Two-Cohort		One-Cohort	
Variable	Aim High ($N = 520$)	Never Aim High ($N = 7{,}388$)	Aim High ($N = 5248$)	Never Aim High ($N = 4{,}074$)
Outcome Measures Across All Grades: Mean (SD)				
Total absence rate	2.0 (3.6)	3.4 (5.1)	2.0 (2.9)	3.3 (4.2)
Unexcused absence rate	0.7 (2.3)	1.3 (3.6)	0.7 (1.8)	1.8 (2.9)
Chronically absent (0/1)	2.6 (15.9)	6.1 (23.9)	2.2 (14.7)	6.2 (24.1)
Suspended (0/1)	0.7 (8.5)	1.4 (11.6)	1.2 (11.0)	1.3 (11.4)
ELA score (standardized)[a]	0.04 (0.87)	0.06 (0.99)	−0.03 (0.91)	0.02 (1.00)
Math score (standardized)[a]	0.10 (0.92)	0.06 (0.98)	0.08 (0.95)	0.02 (1.00)
Baseline 5th-Grade Demographics: % (*N*)				
Female	49.0 (255)	48.1 (3,556)	48.8 (114)	49.4 (1,913)
Special education student	8.5 (44)	12.2 (907)	9.3 (23)	11.8 (449)
English learner	21.7 (113)	22.3 (1,646)	26.4 (61)	26.8 (1,022)
White	1.2 (6)	14.1 (1,044)	1.7 (4)	14.8 (565)
Black	8.3 (43)	8.0 (587)	15.1 (35)	7.1 (272)
Latinx	20.4 (106)	25.6 (1,892)	21.2 (49)	26.2 (1,002)
Asian	58.1 (302)	38.2 (2,822)	57.1 (132)	37.3 (1,426)
Multiracial or other race/ethnicity	5.2 (27)	7.7 (571)	4.3 (10)	11.3 (431)
Missing race/ethnicity	6.9 (36)	6.4 (472)	0.4 (1)	3.3 (124)
Parent Education				
Not high school graduate	16.4 (85)	13.8 (1,022)	12.1 (28)	12.4 (473)
High school graduate	25.4 (132)	18.0 (1,327)	24.7 (57)	13.6 (518)
Some college	17.9 (93)	13.6 (1,008)	16.0 (37)	13.0 (495)
College graduate or higher	14.4 (75)	28.9 (2,135)	12.1 (28)	27.0 (1,003)
Not reported	26.0 (135)	25.7 (1,896)	34.9 (81)	34.1 (1,301)

Notes: The intent-to-treat sample consists of two cohorts of students who are enrolled in the district in fifth grade for 175 to 180 days during the 2013–14 and 2014–15 school years. This analytic sample is an unbalanced panel of all students from that intent-to-treat (ITT) sample with full information and who are enrolled for 175 to 180 days in a given year from grades 1 through 8; $N = 7{,}908$ unique students (57,559 student-year observations), 520 of whom were ever in Aim High. Ninety-six percent of students show up in five or more grades in the sample. The one-cohort analytic sample used for California Assessment of Student Performance and Progress (CAASPP) test score outcomes is based on an ITT sample of fifth-grade students enrolled in the 2015-16 school year. The one-cohort analytic sample is an unbalanced panel that retains students from the ITT sample whose CAASPP test scores we observe in SFUSD's longitudinal data in a given year from grades 4 through 7, over the 2014–15 through 2017–18 school years; $N = 4{,}322$ unique students (14,853 student-year observations), 248 of whom were ever in Aim High. Eighty percent of students show up three or more grades in the sample. See online appendix B for attrition from the samples. ELA = English language arts.

[a]Students in the two-cohort sample only have CAASPP test scores in grades 5 through 8 and are not used to evaluate test score outcomes.

Both Aim High participants and their peers in the two-cohort sample are slightly more likely to be male, and the groups have similar proportions of English language learners. Aim High students are slightly less likely to be designated as special education compared with non–Aim High students (8.5 percent and 12.2 percent, respectively). White students constitute 14 percent of students who never participate in Aim High, but only about 1 percent of Aim High students. Black students are represented equally between Aim High students and their peers (about 8 percent) but fewer Latinx students participate in Aim High (20.4 percent compared with 25.6 percent). Fifty-eight percent of Aim High students are Asian, compared with 38 percent of non–Aim High students. Parents of Aim High students in the sample report lower levels of education than other parents. Far fewer Aim High students have parents with a college degree or higher than do their peers (14 percent of Aim High students compared with 29 percent of non–Aim High participants).

Our one-cohort sample contains 4,322 unique students followed over multiple school years, 248 of whom participated in Aim High for at least one summer we observe (columns 3 and 4 of table 1). Descriptive patterns in table 1 for the one-cohort sample are similar to those in the two-cohort sample, although certain differences between Aim High and non–Aim High students are more pronounced. For example, black students and those whose parents have no more than a high school diploma are less likely to participate in Aim High in the one-cohort sample compared with the two-cohort sample.

Excused and Unexcused Absences

We rely first on tracking students' absence rates over time to measure the effect of Aim High on behavioral engagement in school. We speculate that Aim High improves attendance by increasing a student's desire to come to school when they are not sick or experiencing some other emergency. To test this hypothesis, we first evaluate the effect of Aim High on the total absence rate. We then evaluate the program's effects for excused and unexcused absence rates separately, because unexcused absences are a stronger signal of student and family disengagement from school (Gottfried 2009; Fredricks et al. 2011; Gershenson, Jacknowitz, and Brannegan 2017; Pyne et al. 2021). To the extent that engagement is the mechanism through which Aim High affects absence rates, we should see improvements in unexcused absences more so than excused.

Table 2 displays the key results from the static, semi-dynamic, and dosage DD specifications for overall absence rate, excused absence rate, and unexcused absence rate dependent variables. These results indicate that Aim High participation leads to a substantial reduction in student absenteeism. The static DD model (column 1) suggests that Aim High reduces the absence rate by roughly one third of a percentage point ($b = -0.31$, SE $= 0.18$, $t = -1.68$, $p = 0.093$).

Four other features of these results are noteworthy. First, the reduced absences associated with Aim High concentrate mostly among unexcused absences. While the static DD specification indicates that Aim High does not have a statistically significant effect on excused absences, there is a statistically significant effect with respect to unexcused absences ($b = -0.33$, SE $= 0.14$, $t = -2.38$, $p = 0.017$). This heterogeneity is consistent with the hypothesis that Aim High participation promotes behavioral engagement in school.

Second, the results in table 2 consistently indicate dynamic effects of Aim High participation on overall and unexcused absences. Results in column 8 indicate that, after the first summer of participation, Aim High has small and statistically insignificant effects on unexcused absences. However, in the second school year after a student first participates, impact estimates are much larger and statistically significant ($b = -0.57$, SE $= 0.11$, $t = -5.40$, $p < 0.001$). In the third school year after a student's first participation, the estimated impact of Aim High participation grows again, to nearly three quarters of a percentage point ($b = -0.73$, SE $= 0.14$, $t = -5.19$, $p < 0.001$).

Table 2. The Estimated Effects of Aim High Participation on Total, Excused, and Unexcused Absence Rates

	Dependent Variables								
	Overall Absence Rate			Excused Absence Rate			Unexcused Absence Rate		
	Static	Semi-Dynamic	Dosage	Static	Semi-Dynamic	Dosage	Static	Semi-Dynamic	Dosage
Independent Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
After Aim High participation	-0.31^{*} (0.18)	–	–	0.02 (0.12)	–	–	-0.33^{**} (0.14)	–	–
First participated one summer prior	–	0.16 (0.28)	–	–	0.22 (0.19)	–	–	−0.06 (0.22)	–
First participated two summers prior	–	-0.76^{***} (0.14)	–	–	-0.19^{**} (0.09)	–	–	-0.57^{***} (0.11)	–
First participated three summers prior	–	-0.98^{***} (0.21)	–	–	-0.25^{*} (0.15)	–	–	-0.73^{***} (0.14)	–
After one summer of participation	–	–	0.17 (0.28)	–	–	0.23 (0.19)	–	–	−0.06 (0.21)
After two summers of participation	–	–	-0.81^{**} (0.12)	–	–	−0.16 (0.08)	–	–	-0.65^{***} (0.08)
After three summers of participation	–	–	-1.06^{***} (0.23)	–	–	−0.28 (0.17)	–	–	-0.78^{***} (0.14)
p-value (H_0: $\delta_0=\delta_{-1}=\delta_{-2}$)	–	<0.01	<0.01	–	0.04	0.05	–	0.02	0.01

Notes: The intent-to-treat (ITT) sample consists of two cohorts of students enrolled in fifth grade for 175 to 180 days during the 2013–14 and 2014–15 school years (see online appendix B for more information on attrition from the sample). This analytic sample is an unbalanced panel of all students from that ITT sample with full information and who are enrolled for 175–180 days in a given year from grades 1 through 8; $N = 7{,}908$ unique students (57,559 student-year observations), 520 of whom were ever in Aim High. Ninety-six percent of students show up in five or more grades in the sample. Estimates are derived from ordinary least squares multiple regression models. The dependent variable is the rate of absences during the school year (total, excused, or unexcused). All models include student, school year, and grade-level fixed effects and the following time-varying student-year controls: Special education, parent's highest education level, English language proficiency, and foster care status. Event study models support the parallel trends assumption and can be found in online appendix table C2. Balanced panel data results are very similar to those shown above and can be found in online appendix table D2. Standard errors, clustered at the student level, are in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

The "dosage" specifications return quite similar results. For example, the unexcused absence rate among those who participate in Aim High for three summers is three quarters of a percentage point lower than never-participating students (column 9). This statistically significant evidence suggests that the beneficial impact of Aim High participation grows monotonically over time and with additional exposure. Specifically, the *p*-values in the bottom row of table 2 indicate that the dosage model consistently rejects the assumption of a common treatment effect (i.e., as assumed in the static DD specifications).

The strong correspondence between results in the semi-dynamic and dosage specifications is not surprising. More than two thirds of Aim High participants participate in the summer program more than once, implying high collinearity between those participants who are two years from first participating in the program and those who are in the academic year immediately after the second summer of participation. More complicated DD specifications that simultaneously allow for both longer-term and dosage effects may help (e.g., being in the second year after first participation and being in the year immediately after participating a second time). Those ancillary analyses (online appendix table C1) suggest that dosage effects are more relevant than the recursive effects of earlier participation. For example, the estimated effect of a second year of participation on the rate of unexcused absences is larger than the effect of being in the second year after first participation (column 2, online appendix table C1). However, these differences are not always statistically significant, which warrants some agnosticism.

A third important feature of table 2 results concerns effect sizes. In terms of the percent reduction in attendance rate, the estimated benefits of Aim High participation are quite large. Non–Aim High students' unexcused absence rate is 1.5 percent. We find that three years after first participating, Aim High students' average absence rate is 1 percentage point lower (see column 3, table 2). This amounts to a 71 percent reduction in the overall absence rate (i.e., −1.06/1.5). However, framing this with respect to days of attendance suggests a more modest effect of nearly two days of additional attendance in a 180-day school year. That is the equivalent of about 0.035 and 0.016 standard deviation increments in math and reading test scores, respectively (see Carlsson et al. 2015; Aucejo and Romano 2016).

Finally, table 2 results are quite robust. Corresponding event study estimates in online appendix table C2 imply similar trends in attendance measures between Aim High participants and their peers in years before participating. Differences emerge only after participation. These patterns are consistent with the identifying parallel trends assumption of these DD specifications. We also find that school absence results are robust to several alternate model specifications. For example, excused and unexcused absence rate results hold when using count outcomes in negative binomial models (online appendix table D1). These results are also all similar when using the smaller balanced-panel data sample, which includes only students for whom we observe full information across all eight school years (online appendix tables D2, D3). All of these results are available in the online appendix.

Chronic Absenteeism

We next examine Aim High's impact on whether a student is chronically absent (i.e., missing 10 percent or more of days in the school year). This is a particularly salient outcome measure because missing a substantial number of school days both hinders student learning in the near term and implies academic disengagement that is likely to have pejorative long-run consequences. For these reasons, states frequently use chronic absenteeism in school accountability systems, in response to the Every Student Succeeds Act (Jordan and Miller 2017).

The results in table 3 indicate that Aim High participation substantially and consistently reduces the probability a student is chronically absent. For example, the static DD model in column 1 of table 3 indicates that Aim High students are 1.4 percentage points less likely to become chronically absent following their participation in Aim High, relative to students who never participated ($b = -0.014$, SE $= 0.006$, $t = -2.17$, $p = 0.03$). However, the results in columns 2 and 3 indicate that this static specification obscures the dynamic effects of program participation. As with the attendance rate results, the estimated benefits of Aim High participation on reducing chronic absenteeism grow monotonically larger both with the passage of time since first participating (column 2) and with additional years of participation (column 3).[11] More formally, the p-values

11. As with the attendance results, specifications that simultaneously allow for both lagged effects of first exposure and dosage effects suggest that dosage effects are particularly important (online appendix table C1, column 3).

Table 3. The Estimated Effects of Aim High on Probability of Chronic Absenteeism

Independent Variables	Static (1)	Semi-Dynamic (2)	Dosage (3)
After Aim High participation	-0.014^{**} (0.006)	–	–
First participated one summer prior	–	0.000 (0.008)	–
First participated two summers prior	–	-0.027^{***} (0.007)	–
First participated three summers prior	–	-0.039^{***} (0.013)	–
After one summer of participation	–	–	0.000 (0.008)
After two summers of participation	–	–	-0.035^{***} (0.005)
After three summers of participation	–	–	-0.048^{***} (0.010)
p-value (H_0: $\delta_0 = \delta_{-1} = \delta_{-2}$)	–	<0.001	<0.001

Notes: The intent-to-treat (ITT) sample consists of two cohorts of students enrolled in fifth grade for 175–180 days during the 2013–14 and 2014–15 school years (see online appendix B for more information on attrition from the sample). This analytic sample is an unbalanced panel of all students from that ITT sample with full information and who are enrolled for 175–180 days in a given year from grades 1 through 8; $N = 7{,}908$ unique students (57,559 student-year observations), 520 of whom were ever in Aim High. Ninety-six percent of students show up in five or more grades in the sample. Estimates derived from linear probability models. Alternate analyses also retaining students enrolled for fewer than 175 days in any school year and those using logistic regression yield very similar results. The dependent variable is a binary indicator of whether each student was chronically absent for the school year. All models include student, school year, and grade-level fixed effects and the following time-varying student-year controls: Special education, parent's highest education level, English language proficiency, and foster care status. Event study models support the parallel trends assumption and can be found in online appendix table C2. Balanced panel data results are very similar to those shown above and can be found in online appendix table D3. Standard errors, clustered at the student level, are in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$.

reported in the bottom row of table 3 indicate that the assumption of a constant treatment effect is rejected.

The estimated effect sizes implied by Aim High participation are substantial. For example, the dosage specification indicates that students who participated in Aim High for three years are 4.8 percentage points less likely to be chronically absent in eighth grade than never-participating peers. As a point of comparison, the rate of chronic absenteeism among eighth graders who never participate in Aim High is 8.3 percent. This implies a 58 percent reduction in chronic absenteeism with persistent participation in Aim High. These findings appear quite robust. Most notably, event study results (column 3, online appendix table C2) are consistent with the identifying parallel trends assumption of this research design, suggesting similar trends in chronic absenteeism between participants and nonparticipants in the years prior to any Aim High participation. We also find comparable results using logistic regression (online appendix table D1) and when only using a balanced-panel sample of students across all eight school years (online appendix table D3).

However, these distinctions are not often statistically meaningful given the high number of Aim High participants who attend more than one summer.

Table 4. The Estimated Effects of Aim High Participation on Probability of Suspension

Independent Variables	Static (1)	Semi-Dynamic (2)	Dosage (3)
After Aim High participation	−0.011** (0.004)	–	–
First participated one summer prior	–	−0.003 (0.006)	–
First participated two summers prior	–	−0.018*** (0.005)	–
First participated three summers prior	–	−0.020* (0.011)	–
After one summer of participation	–	–	−0.004 (0.006)
After two summers of participation	–	–	−0.027*** (0.003)
After three summers of participation	–	–	−0.017 (0.014)
p-value (H_0: $\delta_0 = \delta_{-1} = \delta_{-2}$)	–	0.121	0.001

Notes: The intent-to-treat (ITT) sample consists of two cohorts of students enrolled in fifth grade for 175–180 days during the 2013–14 and 2014–15 school years (see online appendix B for more information on attrition from the sample). This analytic sample is an unbalanced panel of all students from that ITT sample with full information and who are enrolled for 175–180 days in a given year from grades 1 through 8; $N = 7{,}908$ unique students (57,559 student-year observations), 520 of whom were ever in Aim High. Ninety-six percent of students show up in five or more grades in the sample. Estimates are derived from ordinary least squares linear probability models. The dependent variable is a binary indicator of whether the student is suspended one or more times during the school year. All models include student, school year, and grade-level fixed effects and the following time-varying student-year controls: Special education, parent's highest education level, English language proficiency, and foster care status. Event study models support the parallel trends assumption and can be found in online appendix table C2. Balanced panel data results are very similar to those shown above and can be found in online appendix table D4. Standard errors, clustered at the student level, are in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Suspension from School

We also consider Aim High's impact on students' probability of experiencing a suspension from school in each school year observed. Being suspended from school is consequential for a student's learning opportunities and for their future engagement in school and other social institutions (Kupchik and Catlaw 2015; Pyne 2019; Jacobsen 2020). However, the probability of suspension is likely to reflect other determinants such as the structural features of a school and district (e.g., policies around suspension) and the subjective, often culturally mediated, interpretations of behavior made by decision makers in schools (Okonofua, Walton, and Eberhardt 2016). This important contextual caveat may have relevance for extrapolating the results from this study to school settings with different disciplinary policies. Fortunately, it does not imply a clear internal-validity threat for inference based on our quasi-experimental approach.

Table 4 displays key results from DD specifications estimating the impact of Aim High on the probability of being suspended. The static DD specification (column 1) suggests just over 1 percentage point lower probability of becoming suspended in school years following participation in Aim High, compared with students who never participated ($b = -0.011$, SE $= 0.004$, $t = -2.53$, $p = 0.011$). The results in columns 2 and 3 again provide suggestive evidence that the estimated effects of Aim High are dynamic, since reductions in suspensions due to Aim High are concentrated in the second and third years after first participating and after a second year of participation.[12] The effect sizes implied in table 4 estimates are quite large. Specifically, given that 3 percent of those who never participated in Aim High became suspended at least once in a given school year while in middle school, this would amount to an estimated 37 percent reduction in the probability of suspension among nonparticipants. Finally, we note that the event study results (online appendix table C2) suggest similar trends in suspension probabilities between Aim High participants and nonparticipants in years before the program was available. We found similar results to those reported above when using logistic regression (online appendix table D1) and when using a balanced panel of student-by-year observations (online appendix table D4).

Academic Achievement

We now turn to the effects of Aim High participation on student achievement over time. Due to data limitations, we track only one cohort of students' state standardized test scores from the 2014–15 through the 2017–18 school years, from fourth through seventh grade. This means we observe students two or three school years prior to initially participating in Aim High, and one or two school years after participating. Below, we report unbalanced panel data on the effects of Aim High on ELA and mathematics state test scores, standardized within grade, school year, and test score subject.

Table 5 results indicate that students experience an average increase of about 0.06 standard deviation in ELA test scores in the year or years following participation in Aim High (column 1), compared with expected scores without participating ($b = 0.06$, SE $= 0.03$, $t = 1.95$, $p = 0.052$). A test-score impact this size is consistent with effects found in prior summer learning program evaluations (see Lauer et al. 2006; Kim and Quinn 2013) and are larger than what would be suggested by the program's effects on improved attendance as a mediator (based on estimates by Carlsson et al. 2015; Aucejo and Romano 2016). However, we find that these gains appear to be limited to ELA. The estimated effect of Aim High participation on math scores is smaller and statistically insignificant.

The dynamic specifications suggest that the ELA gains from Aim High come immediately after participating. However, the structure of the available test-score data (one cohort, four years) limits our capacity to examine dynamic treatment effects with statistical precision. A further limitation is that only 64 Aim High participants in the one-cohort sample began Aim High in the summer before sixth grade. Thus, only those 64 students could take a second summer of Aim High in the relatively short time frame that we observe ELA test scores.

The weakly significant estimated effects on ELA test scores appears robust. Event study specifications (online appendix table C3) are consistent with its internal validity. This finding is also similar in magnitude and statistically significant when using only data from a balanced panel of students (online appendix tables D6, D7) and when excluding the 64 students who first participated as rising sixth graders (online appendix table D8).

12. Regarding suspensions, our capacity to discriminate between lagged effects of first exposure and dosage effects seems limited (online appendix table C1). The reduction in suspensions appears concentrated among participants in Aim High twice or who are in their third year after first participating.

Table 5. The Estimated Effects of Aim High on Standardized Test Scores

	Dependent Variables					
	English Language Arts			Mathematics		
	Static	Semi-Dynamic	Dosage	Static	Semi-Dynamic	Dosage
Independent Variables	(1)	(2)	(3)	(4)	(5)	(6)
After Aim High participation	0.06* (0.03)	–	–	−0.01 (0.03)	–	–
First participated one summer prior		0.07** (0.03)	–	–	0.00 (0.03)	–
First participated two summers prior	–	0.02 (0.06)	–	–	−0.03 (0.06)	–
After one summer of participation	–	–	0.07** (0.03)	–	–	0.00 (0.03)
After two summers of participation	–	–	0.01 (0.06)	–	–	−0.03 (0.06)
p-value (H_0: $\delta_0 = \delta_{-1}$)	–	0.37	0.35	–	0.60	0.63

Notes: The intent-to-treat (ITT) sample consists of students enrolled in fifth grade during the 2015–16 school year (see online appendix B for more information on attrition from the sample). This analytic sample is an unbalanced panel of all students from that ITT sample with full test score information in a given year from grades 4 through 7, for school years 2014–15 through 2017–18; $N = 4{,}322$ unique students (14,853 student-year observations), 248 of whom were ever in Aim High. Eighty percent of students show up three or more grades in the sample. Estimates are derived from ordinary least squares multiple regression models. The dependent variables are the California Assessment of Student Performance and Progress (CAASPP) English language arts and mathematics test scores for each student in each school year. All models include student, school year, and grade-level fixed effects and the following time-varying student-year controls: Special education, parent's highest education level, English language proficiency, and foster care status. Event study models support the parallel trends assumption for English language arts but not mathematics test scores and can be found in online appendix table C3. Balanced panel data results are very similar to those shown above and can be found in online appendix table D6. Standard errors, clustered at the student level, are in parentheses $^{**}p < 0.05$, $^{*}p < 0.10$.

Effects by Racial, Ethnic, and Gender Subgroups

In table 6, we use static DD models to estimate the effects of Aim High on behavioral engagement and achievement by race/ethnicity and gender. By race/ethnicity, we only report results among black, Latinx, and Asian students, excluding reports among white, multiracial, and other racial and ethnic minority students due to their small cell counts of Aim High participants (see table 1). Across our outcome measures, results suggest that Latinx Aim High students experience the largest effects out of all reported racial subgroups on all engagement outcomes, while black and Asian students typically experience no statistically significant effects of participating. For example, these estimates indicate that, among Latinx students, Aim High reduced the probability of being chronically absent by 4.6 percentage points ($b = -0.046$, SE $= 0.021$, $t = -2.20$, $p = 0.028$) and the probability of being suspended by 3.9 percentage points ($b = -0.039$, SE $= 0.006$, $t = -6.18$, $p < 0.001$).

Gender subgroup analyses suggest that girls stand to benefit from Aim High participation more so than boys through reductions in their unexcused absence rates ($b = -0.48$, SE $= 0.11$, $t = -4.44$, $p < 0.001$), while boys stand to benefit more than girls in terms of chronic absenteeism ($b = -0.023$, SE $= 0.006$, $t = -3.58$, $p < 0.001$) and suspension from school ($b = -0.019$, SE $= 0.006$, $t = -3.00$, $p = 0.003$). Additionally, boys can expect to experience nearly a tenth of a standard deviation increase in ELA state test scores due to Aim High participation ($b = 0.09$, SE $= 0.04$, $t = 2.21$, $p = 0.027$), whereas the estimated effect among girls on those scores is effectively zero ($b = 0.01$, SE $= 0.05$, $t = 0.16$, $p = 0.870$).

Table 6. The Estimated Static Effects of Aim High on Engagement and Achievement, Overall and by Subgroup

		Race and Ethnicity[a]			Gender	
Dependent Variables	**Full Sample**	**Black**	**Latinx**	**Asian**	**Male**	**Female**
Two-Cohort Sample						
Total absence rate	−0.31	1.24	−1.53***	0.18	−0.31	−0.30
	(0.18)	(1.67)	(0.29)	(0.19)	(0.26)	(0.26)
Unexcused absence rate	−0.33***	0.64	−1.36***	−0.02	−0.17	−0.48***
	(0.14)	(1.61)	(0.19)	(0.07)	(0.24)	(0.11)
Chronic absenteeism	−0.014***	0.003	−0.046**	0.002	−0.023***	−0.005
	(0.006)	(0.043)	(0.021)	(0.004)	(0.006)	(0.012)
Suspension	−0.011***	0.029	−0.039***	−0.004	−0.019***	−0.002
	(0.004)	(0.046)	(0.006)	(0.002)	(0.006)	(0.006)
Student-year observations	57,559	4,462	14,211	23,532	29,831	27,728
Unique students	7,908	630	1,998	3,124	4,097	3,811
One-Cohort Sample						
State ELA test	0.06*	0.01	0.06	0.03	0.09**	0.01
	(0.03)	(0.10)	(0.08)	(0.04)	(0.04)	(0.05)
State mathematics test	−0.01	0.01	−0.04	0.01	−0.02	0.01
	(0.03)	(0.11)	(0.07)	(0.03)	(0.04)	(0.04)
Student-year observations	14,853	769	3,241	5,639	6,709	6,538
Unique students	4,322	206	871	1,469	1,784	1,754

Notes: The intent-to-treat (ITT) sample for absence and suspension outcomes consists of two cohorts of students enrolled in fifth grade for 175–180 days during the 2013–14 and 2014–15 school years. The intent-to-treat sample for California Assessment of Student Performance and Progress (CAASPP) English language arts (ELA) and mathematics test score outcomes consists of students who were in fifth grade in the 2014-15 school year. These analytic samples are unbalanced panels of all students from those ITT samples with full information in a given year. All models are static difference-in-differences models and include student, school year, and grade-level fixed effects along with the following time-varying student-year controls: Special education, parent's highest education level, English language proficiency and foster care status. Standard errors, clustered at the student level, are in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.10$

[a]We exclude reports among white, multiracial, and other racial and ethnic minority groups because very few students from these groups participate in Aim High (see table 1).

7. DISCUSSION

The growing evidence of summer learning loss, the interest in expanded instructional time, and developmental disruptions due to the COVID-19 pandemic motivate an ongoing interest in the design of effective summer learning opportunities. However, recent meta-analyses (e.g., Lauer et al. 2006; Kim and Quinn 2013) indicate that summer learning programs—which often feature a narrow, single- or two-subject curriculum—only modestly impact achievement and show no clear effects on social-emotional or behavioral engagement outcomes that are important antecedents to longer-run educational success. The large prior literature on summer learning programs also focuses mostly on short-run outcomes and not those that may grow recursively over time or accrue after additional summers of participation.

In this study, we examined a summer learning program, Aim High, which has several distinctive and noteworthy features. First, its design elements include both academic breadth and an explicit social-emotional curriculum that is vertically integrated across the middle school years (Issues and Choices, see online appendix A). Second, Aim High targets its program to middle school–aged students—a time when stark summer learning losses coincide with intensifying challenges of sustaining students' behavioral engagement. Third, Aim High is also an unusually mature program that operates at a fairly large scale in the SFUSD. The program's scale and maturity suggest that inferences provide a reliable guide to the likely impact of other large-scale initiatives to provide summer learning opportunities.

Finally, programmatic elements of Aim High appear to garner considerable enthusiasm among its participants compared with other similar summer learning programs. For example, the average SFUSD student enrolled in Aim High during the summers we study attends between 88 and 92 percent of the time. In the BELL program reported by Chaplin and Capizzano (2006), the attendance rate was 50–60 percent, while in the RAND experiment reported by Augustine and colleagues (2016), the attendance rate was 60–80 percent. Additionally, the latter study reported that only 30–50 percent of students had attendance rates of 80 percent or higher in the years they studied their program. That compares to 75 percent of Aim High students attending 80 percent or more of program days in one summer we study. Although the questions of whether students attended, or attended more often than others, are endogenous to the treatment, these high attendance reports speak to the Aim High program's ability to engage students in learning.

To evaluate causal impacts, longitudinal data from SFUSD and Aim High allow us to implement a quasi-experimental examination of the program's impact and to consider how effects vary both over subsequent years and with additional summers of participation. Our main finding is that Aim High participation implies substantial reductions in chronic absenteeism and suspensions, which we believe are two important proxies of adolescents' behavioral engagement in school rather than indicators of substantial learning gains. Available quasi-experimental evidence suggests that the learning loss associated with one day of missed instruction is quite modest—at about a 0.001 standard deviation decrease in math test score growth (see Carlsson et al. 2015; Aucejo and Romano 2016). This makes for a modest interpretation of learning gains from the two-day decrease in absences and 4 percent decrease in the proportion of students who fall below the threshold of chronic absenteeism as a result of the summer program. Even so, we believe the additional days of school attendance resulting from program participation, coming mostly from reductions in unexcused absences, are a reflection of increased behavioral engagement in schooling, and thus a greater commitment to learning, due to enrolling in Aim High.

This interpretation is consistent with our findings that Aim High has modest and weakly significant effects on ELA achievement (i.e., effect size = 0.06) and no effects on math achievement. The effects we report on behavioral engagement outcomes often appear dynamic, growing over time and with additional summers of participation. For example, our estimates imply that the probability of being chronically absent in eighth grade is 4.8 percentage points lower for students who participated in Aim High during all three of their middle school summers (i.e., a 58 percent reduction relative to eighth graders who never participated in Aim High). Similarly, we estimate that Aim High participation reduces the probability of being suspended by 1.1 percentage points (i.e., a 37 percent reduction relative to peers who never participated).

Not only do effects on behavioral engagement grow over time, but they also appear to show up several years after first participating. In semi-dynamic models, all relevant outcomes—including total absences, unexcused absences, probability of chronic absenteeism, and probability of suspension—show no statistically significant effects in the school year following the first summer of participation, and only show effects two school years after or later. Our supplemental models in online appendix table C1 seem to favor dosage over passive effects explaining this phenomenon in many cases, suggesting that those seeking to implement similar programs would do well to plan on investing in children attending the programs over multiple summers.

We also find that these effects are concentrated among boys and Latinx students. We speculate that we are unable to detect behavioral engagement effects among Asian Aim High participants due to floor effects. Asian middle school students who never participate in Aim High average a chronic absenteeism rate of 1.2 percent and a suspension rate of 0.5 percent—far below the average rates of non-Asian, nonparticipating middle school students. This leaves little room for the program to improve their behavioral engagement, although we do see descriptively that rates of both outcomes are higher among Asian students participating in Aim High. We are less clear about what drives the null effects among black participants, since floor effects are not a concern in this subgroup.

Regarding gender differences, past research suggests that boys are less engaged in school than girls beginning at school entry, and girls experience comparatively greater increases in behavioral engagement over elementary school (DiPrete and Jennings 2012; Downey, Workman, and von Hippel 2019; Pyne 2020). These trends become more pronounced when boys experience much higher levels of disciplinary involvement, disengagement, and withdrawal starting in middle school (DiPrete and Buchmann 2013). We believe the engagement-focused nature of Aim High impacts boys more than girls because there is more ground boys can regain.

Several puzzles remain concerning the program's uneven and modest impacts on academic achievement. For example, while the size of the ELA gain we observe is consistent with findings in prior studies (see Lauer et al. 2006; Kim and Quinn 2013), our data suggest those gains are only temporary, and we find no effects on math gains. The absence of a math effect is unexpected, given that folk wisdom in education research suggests similar interventions should have a greater impact on math achievement than reading or ELA achievement. Although we cannot be certain of why we observe these differences in effects between content areas, we can think of several possible explanations. First, the absence of math effects may reflect the fact that our achievement scores coincided with SFUSD's implementation of the new Common Core State Standards, and we have some indications that the summer learning program did not align immediately with the new pedagogical strategies that come with Common Core. It is also possible that the program's effects on math are simply more delayed than the data available to us can record. Or, more simply, Aim High may need to reconsider its strategies for improving participants' math (and, possibly, ELA) skills. Regardless, we believe these achievement results, although worthy of further study, do not overshadow the clear and robust effects of the program on students' behavioral engagement in school.

Finally, several additional caveats about our findings are worth underscoring. One is that the program has nontrivial costs. In examining Aim High's 2017 Form 990 filing with the Internal Revenue Service, we estimate that their total spending on each student per year is approximately $2,700. To the extent that the program benefits rely on participating in multiple summers, the relevant per-student costs would be correspondingly larger. In contrast, recent evidence from nudge-like interventions suggest they generate similar short-term improvements in attendance and discipline, at lower cost (Rogers and Feller 2018; Borman et al. 2019). However, these intervention studies only report attendance and disciplinary outcomes at the end of the initial treatment school year, so the comparable long-run benefits of these brief interventions are unclear. Second, the capacity of other districts to replicate the effects documented here is necessarily an open and empirical question. Regardless, our results provide novel, robust, and encouraging evidence that a summer learning program with a social-emotional curriculum can generate meaningful improvements in important measures of behavioral engagement and longer-run success. These results suggest that further innovations in and assessments of summer learning programs will be productive to support the educational potential of our nation's students.

ACKNOWLEDGMENTS

Authors are listed in descending alphabetical order. We thank Jorge Ruiz de Velasco, Liz Newman, and Nancy Mancini for comments on earlier versions of this paper and other assistance on the project. We also thank the Editors and three anonymous referees for their many helpful comments and suggestions. This work was supported by a grant to The John W. Gardner Center for Youth and their Communities (Gardner Center) by Aim High, a nonprofit, 501(c)3 organization that implements the summer learning program studied in this paper. The authors have analyzed and disseminated these results freely and independent of approval from Aim High, as stipulated in a Letter of Agreement between the Gardner Center and Aim High. None of the authors have a financial interest in Aim High.

REFERENCES

Alexander, Karl L., Doris R. Entwisle, and Carrie S. Horsey. 1997. From first grade forward: Early foundations of high school dropout. *Sociology of Education* 70(2): 87–107. 10.2307/2673158

Alexander, Karl, Sarah Pitcock, and Matthew Boulay. 2016. *The summer slide: What we know and can do about summer learning loss.* New York, NY: Teachers College Press.

Aucejo, Esteban M., and Teresa Foy Romano. 2016. Assessing the effect of school days and absences on test score performance. *Economics of Education Review* 55:70–87. 10.1016/j.econedurev.2016.08.007

Augustine, Catherine H., Jennifer S. McCombs, John F. Pane, Heather L. Schwartz, Jonathan Schweig, Andrew McEachin, and Kylie Siler-Evans. 2016. *Learning from summer: Effects of voluntary summer learning programs on low-income urban youth.* Santa Monica, CA: RAND Corporation.

Blad, Evie. 2017. No state will measure social-emotional learning under ESSA. Will that slow its momentum? *Education Week*. Available https://www.edweek.org/ew/articles/2017/10/04/no-state-will-measure-social-emotional-learning-under.html. Accessed 19 November 2021.

Borman, Geoffrey D., Christopher S. Rozek, Jaymes Pyne, and Paul Hanselman. 2019. Reappraising academic and social adversity improves middle-school students' academic achievement, behavior, and well-being. *Proceedings of the National Academy of Sciences USA* 116(33): 16286–16291. 10.1073/pnas.1820317116

Carlsson, Magnus, Gordon B. Dahl, Björn Öckert, and Dan-Olof Rooth. 2015. The effect of schooling on cognitive skills. *Review of Economics and Statistics* 97(3): 533–547. 10.1162/REST_a_00501

Chaplin, Duncan, and Jeffrey Capizzano. 2006. Impacts of a summer learning program: A random assignment study of Building Educated Leaders for Life (BELL). Washington, DC: Urban Institute.

de Chaisemartin, Clément, and Xavier D'Haultfœuille. 2020. Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review* 110(9): 2964–2996. 10.1257/aer.20181169

DiPrete, Thomas A., and Claudia Buchmann. 2013. The rise of women: The growing gender gap in education and what it means for American schools. New York, NY: Russell Sage Foundation.

DiPrete, Thomas A., and Jennifer L. Jennings. 2012. Social and behavioral skills and the gender gap in early educational achievement. *Social Science Research* 41(1): 1–15. 10.1016/j.ssresearch.2011.09.001

Downey, Douglas B., Paul T. von Hippel, and Beckett Broh. 2004. Are schools the great equalizer? School and non-school sources of inequality in cognitive skills. *American Sociological Review* 69(5): 613–635. 10.1177/000312240406900501

Downey, Douglas B., Joseph Workman, and Paul von Hippel. 2019. Socioeconomic, ethnic, racial, and gender gaps in children's social/behavioral skills: Do they grow faster in school or out? *Sociological Science* 6:446–466. 10.15195/v6.a17

Eccles, Jacquelynne S. 2004. Schools, academic motivation, and stage-environment fit. In *Handbook of adolescent psychology*, edited by Richard M. Lerner and Laurence Steinberg, pp. 125–153. Hoboken, NJ: John Wiley and Sons.

Finn, Jeremy D. 1989. Withdrawing from school. *Review of Educational Research* 59(2): 117–142. 10.3102/00346543059002117

Fredricks, Jennifer A., Phyllis C. Blumenfeld, and Alison H. Paris. 2004. School engagement: Potential of the concept, state of the evidence. *Review of Educational Research* 74(1): 59–109. 10.3102/00346543074001059

Fredricks, Jennifer A., Wendy McColskey, Jane Meli, Joy Mordica, Bianca Montrosse, and Kathleen Mooney. 2011. Measuring student engagement in upper elementary through high school: A description of 21 instruments. Issues and Answers. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. Retrieved from http://ies.ed.gov/ncee/edlabs.

Gardner Center. 2016. *Aim High: Closing the achievement and opportunity gaps through summer learning*. Stanford, CA: The John W. Gardner Center for Youth and their Communities.

Gershenson, Seth, Alison Jacknowitz, and Andrew Brannegan. 2017. Are student absences worth the worry in US primary schools? *Education Finance and Policy* 12(2): 137–165. 10.1162/EDFP_a_00207

Gottfried, Michael A. 2009. Excused versus unexcused: How sudent absences in elementary school affect academic achievement. *Educational Evaluation and Policy Analysis* 31(4): 392–415. 10.3102/0162373709342467

Gottfried, Michael A. 2010. Evaluating the relationship between student attendance and achievement in urban elementary and middle schools: An instrumental variables approach. *American Educational Research Journal* 47(2): 434–465. 10.3102/0002831209350494

Gottfried, Michael A., and Ethan L. Hutt. 2019. Introduction. In *Absent from school: Understanding and addressing student absenteeism*, edited by Michael A. Gottfried and Ethan L. Hutt, pp. 1–14. Cambridge, MA: Harvard Education Press.

Holahan, Cathy, and Brooklyn Batey. 2019. Measuring school climate and social and emotional learning and development: A navigation guide for states and districts. Washington, DC: Council of Chief State School Officers.

Hough, Heather. 2019. Roll call. In *Absent from school: Understanding and addressing student absenteeism*, edited by Michael A. Gottfried and Ethan L. Hutt, pp. 15–33. Cambridge, MA: Harvard Education Press.

Jacob, Brian A., and Lars Lefgren. 2004. Remedial education and student achievement: A regression-discontinuity analysis. *Review of Economics and Statistics* 86(1): 226–244. 10.1162/003465304323023778

Jacobsen, Wade C. 2020. School punishment and interpersonal exclusion: Rejection, withdrawal, and separation from friends. *Criminology* 58(1): 35–69. 10.1111/1745-9125.12227

Jordan, Phyllis W., and Raegen Miller. 2017. *Who's in: Chronic absenteeism under the Every Student Succeeds Act*. Washington, DC: FutureEd.

Kim, James S., and David M. Quinn. 2013. The effects of summer reading on low-income children's literacy achievement from kindergarten to grade 8: A meta-analysis of classroom and home interventions. *Review of Educational Research* 83(3): 386–431. 10.3102/0034654313483906

Kuhfeld, Megan. 2019. Surprising new evidence on summer learning loss. *Phi Delta Kappan* 101(1): 25–29. 10.1177/0031721719871560

Kupchik, Aaron, and Thomas J Catlaw. 2015. Discipline and participation the long-term effects of suspension and school security on the political and civic engagement of youth. *Youth and Society* 47(1): 95–124. 10.1177/0044118X14544675

Lauer, Patricia A., Motoko Akiba, Stephanie B. Wilkerson, Helen S. Apthorp, David Snow, and Mya L. Martin-Glenn. 2006. Out-of-school-time programs: A meta-analysis of effects for at-risk students. *Review of Educational Research* 76(2): 275–313. 10.3102/00346543076002275

Lynch, Kathleen, and James S. Kim. 2017. Effects of a summer mathematics intervention for low-income children: A randomized experiment. *Educational Evaluation and Policy Analysis* 39(1): 31–53. 10.3102/0162373716662339

Mariano, Louis T., and Paco Martorell. 2013. The academic effects of summer instruction and retention in New York City. *Educational Evaluation and Policy Analysis* 35(1): 96–117. 10.3102/0162373712454327

Matsudaira, Jordan D. 2008. Mandatory summer wchool and student achievement. *Journal of Econometrics* 142(2): 829–850. 10.1016/j.jeconom.2007.05.015

McCarthy, Martha, and George D. Kuh. 2006. Are students ready for college? What student engagement data say. *Phi Delta Kappan* 87(9): 664–669. 10.1177/003172170608700909

Morris, Edward W., and Brea L. Perry. 2016. The punishment gap: School suspension and racial disparities in achievement. *Social Problems* 63(1): 68–86. 10.1093/socpro/spv026

Okonofua, Jason A., Gregory M. Walton, and Jennifer L. Eberhardt. 2016. A vicious cycle: A social–psychological account of extreme racial disparities in school discipline. *Perspectives on Psychological Science* 11(3): 381–398. 10.1177/1745691616635592

Perry, Brea L., and Edward W. Morris. 2014. Suspending progress: Collateral consequences of exclusionary punishment in public schools. *American Sociological Review* 79(6): 1067–1087. 10.1177/0003122414556308

Pyne, Jaymes. 2019. Suspended attitudes: Exclusion and emotional disengagement from school. *Sociology of Education* 92(1): 9–82. 10.1177/0038040718816684

Pyne, Jaymes. 2020. Gender test score gaps under equal behavioral engagement. *Educational Researcher* 49(6): 459–464. 10.3102/0013189X20930203

Pyne, Jaymes, Eric Grodsky, Elizabeth Vaade, Bo McCready, Eric Camburn, and Dominique Bradley. 2021. The signaling power of unexcused absence from school. *Educational Policy*. 10.1177/08959048211049428

Quinn, David M., and Q. Tien Le. 2018. Are we trending to more or less between-group achievement inequality over the school year and summer? Comparing across ECLS-K cohorts. *AERA Open* 4(4). 10.1177/2332858418819995

Rockoff, Jonah E., and Benjamin B. Lockwood. 2010. Stuck in the middle: Impacts of grade configuration in public schools. *Journal of Public Economics* 94(11): 1051–1061. 10.1016/j.jpubeco.2010.06.017

Rogers, Todd, and Avi Feller. 2018. Reducing student absences at scale by targeting parents' misbeliefs. *Nature Human Behaviour* 2(5): 335–342. 10.1038/s41562-018-0328-1

Schueler, Beth. 2020. Summer "vacation academies" can narrow coronavirus learning gaps. *Education Next*. Available https://www.educationnext.org/summer-vacation-academies-narrow-coronavirus-learning-gaps-springfield. Accessed 19 November 2021.

von Hippel, Paul T., and Caitlin Hamrock. 2019. Do test score gaps grow before, during, or between the school years? Measurement artifacts and what we can know in spite of them. *Sociological Science* 6(3): 43–80. 10.15195/v6.a3

von Hippel, Paul T., Joseph Workman, and Douglas B. Downey. 2018. Inequality in reading and math skills forms mainly before kindergarten: A replication, and partial correction, of "Are schools the great equalizer?" *Sociology of Education* 91(4): 323–357. 10.1177/0038040718801760

Wald, Johanna, and Daniel J. Losen. 2003. Defining and redirecting a school-to-prison pipeline. *New Directions for Youth Development* 99:9–15. 10.1002/yd.51

Walton, Gregory M., and Timothy D. Wilson. 2018. Wise interventions: Psychological remedies for social and personal problems. *Psychological Review* 125(5): 617–655. 10.1037/rev0000115

West, Martin, and Guido Schwerdt. 2012. The middle school plunge: Achievement tumbles when young students change schools. *Education Next* 12(2): 62–68.

Zvoch, Keith, and Mark C. Robertson. 2017. Multivariate summer school effects. *Studies in Educational Evaluation* 55:145–152. 10.1016/j.stueduc.2017.10.003

Zvoch, Keith, and Joseph J. Stevens. 2011. Summer school and summer learning: An examination of the short-and longer term changes in student literacy. *Early Education and Development* 22(4): 649–675. 10.1080/10409289.2010.489891