ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 20-297

The Dynamic Effects of a Summer Learning Program on Behavioral Engagement in School

Jaymes Pyne
Stanford University

Erica Messner
Stanford University

Thomas S. Dee
Stanford University

The evidence that student learning declines sharply (or stagnates) during the summer has motivated a substantial interest in programs that provide intensive academic instruction during the summer. However, the existing literature suggests that such programs, which typically focus on just one or two subjects, have modest effects on students' achievement and no impact on measures of their engagement in school. In this quasi-experimental study, we present evidence on the educational impact of a unique and mature summer learning program that serves low-income middle school students and features unusual academic breadth and a social emotional curriculum with year-to-year scaffolding. Our results indicate that this program led to substantial reductions in unexcused absences, chronic absenteeism and suspensions and a modest gain in ELA test scores. We find evidence that the gains in behavioral engagement grow over time and with additional summers of participation. Our results also suggest that these effects were particularly concentrated among boys and Latinx students.

Acknowledgements - We thank Jorge Ruiz de Velasco and Liz Newman for contributing to the implementation of this study, and Jorge Ruiz de Velasco and Nancy Mancini for comments on this manuscript. The work was funded 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.

VERSION: September 2020

Suggested citation: Pyne, Jaymes, Erica Messner, and Thomas S. Dee. (2020). The Dynamic Effects of a Summer Learning Program on Behavioral Engagement in School. (EdWorkingPaper: 20-297). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/r8hy-wt83

One of the most important stylized facts in education research is that out-of-school factors make substantial contributions to students' learning trajectories. A prominent and widely discussed piece of evidence in support of this claim is the existence of large but variable drops in learning during the summer. For example, Kuhfeld (2019), using adaptive and vertically scaled assessment data from over 3.4 million elementary and middle school students, finds that the learning of the median student falls over the summer by "one to two months in reading and a little more than one to three months of school-year learning in math" (p. 27).[1] The evidence indicating that student learning often falls (or at least stagnates) during the summer has motivated a broad and long-standing interest in the design features, impact, and cost-effectiveness of summer learning programs. The ongoing schooling disruptions due to the COVID-19 pandemic have only amplified the need to understand how summer learning opportunities can best support students. For example, there is a new interest in scaling up summer "vacation academies" for children who would otherwise not make up lost school time (Schueler, 2020).

The summer learning programs studied to date typically target their services to students with disadvantaged socioeconomic backgrounds and feature several hours a day of academic instruction offered five days a week over a period of four to five weeks. An extensive experimental and quasi-experimental evaluation literature suggests that these intensive programs have only modest effects on subsequent measures of student achievement (i.e., effect sizes of 0.10 or less) and no effects on important non-test outcomes like chronic absenteeism and suspension from school.[2] However, the relatively narrow design and targeting of prior programs highlight areas for further study. For example, these programs frequently target elementary rather than the middle school students for whom summer learning loss is particularly large (Kuhfeld, 2019). The program curricula in existing studies also tend to focus on just one or perhaps 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 attached to schooling in general. The absence of these design features in prior

[1] Another long-standing and widely circulated claim is that socioeconomic differences in this summer learning loss contribute to the growing gaps in student achievement. However, recent evidence (e.g., von Hippel and Hamrock, 2019; Kuhfeld 2019) suggests that there is *not* clear evidence for such gaps in summer learning loss and that early evidence in support of this claim was due in part to psychometric flaws like not using vertically aligned scales and the confounding influence of test-form changes.
[2] Both Lauer et al. (2006) and Kim and Quinn (2013) provide meta-analytic reviews of program evaluations in this domain. Because our focus is on programs that provide general academic enrichment during the summer break, we do not focus on "summer school" programs that instead provide remediation to low-performing students.

studies suggest that there is still more to learn about how to structure impactful summer learning opportunities.

In this study, we evaluate the effectiveness of Aim High, a unique and voluntary summer learning program that seeks to promote the academic achievement and behavioral engagement of 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 roughly seven hours a day, five days a week, for five weeks. Its academically rich curriculum focuses on several core subjects (i.e., mathematics, humanities, and science) through both classroom instruction and project-based learning. Since 2014, Aim High has also implemented 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. These features of the Aim High program suggest that understanding its 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. Motivated by these unique 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), which few prior studies have considered and are strongly linked to a variety of poorer schooling outcomes such as lower academic achievement and school dropout (Alexander, Entwisle, & Horsey, 1997; Finn, 1989; Gershenson, Jacknowitz, & Brannegan, 2017; Gottfried, 2010; Morris & Perry, 2016; Rumberger, 1995; Wald & Losen, 2003).[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 both an unusually mature program and one that operates at some scale. Their summer program has operated within SFUSD for over 30 years. In recent years, the program has typically served around 1,000 SFUSD students per summer at eight district sites. Third, both the multi-year design of Aim High and the longitudinal data available through

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

our research partnership with SFUSD allow us to examine the longer-term effects of program participation and the "dosage" effects of participating over more than one summer.

To determine Aim High's impact on the educational outcomes of its participants, we rely on static, semi-dynamic, and dosage "difference-in-differences" (DD) specifications that take advantage of student-level panel data observed both before and after each student has had the opportunity to participate in the summer learning program. We find that program participation significantly reduces the subsequent probability of being chronically absent (i.e., a 58 percent reduction among eighth grade students who attend Aim High all three summers in middle school relative to the sample mean of eighth grade students who do not attend Aim High at all). We also find that Aim High reduces the probability of being suspended by 37 percent, an effect concentrated among Aim High students who participated during more than one middle school summer (i.e., the majority of participants). In contrast, the effects of Aim High on student scores on the California state assessments are more modest. Participation implies a one-time improvement on ELA assessments of 0.06 SD but does not have a statistically significant impact on math performance. Finally, we find substantial heterogeneity in these effects across student subgroups; the reductions in chronic absenteeism and suspensions, as well as the gain in ELA achievement, are particularly prominent among boys and among Latinx students.

In the following sections, we discuss the prior research evaluating the effects of summer learning programs on students and describe the Aim High summer learning program in detail. After presenting our research aims and the data and measures we use, we describe the quasi-experimental difference-in-differences models employed in these panel data to make our claims. We then present results sequentially by each outcome of interest as well as discuss critical robustness checks. We conclude by discussing our findings in the broader context of the literature we seek to inform.

Summer Learning Opportunities and Student Success

The character of students' learning trajectories during the summer reached prominence among education researchers with Haynes' (1978) seminal study of schoolchildren in Atlanta, Georgia. The study showed that low-income and African American students tend to keep pace with their high-income and white peers during the school year while falling behind in summer months when school is out (i.e., a phenomenon widely known as "summer slide"). These findings suggest that schools may not be the engines of inequality suggested by the contemporary work of others (e.g., Bowles & Gintis, 1976). Entwisle and Alexander's (1992) longitudinal study among Baltimore youth bolstered the summer slide hypothesis and gave it a new conceptual framing. They argue that, during school hours, the "faucet" of learning is on for all students regardless of socioeconomic background, because school presumably offers everyone roughly similar opportunities to learn. However, when school is not in session, the faucet is usually still on for students from families with greater resources while the faucet is generally off for other students from less advantaged families. Another important study by Downey, von Hippel and Broh (2004) replicated the earlier seasonal learning findings using nationally representative data. However, some more recent scholarship (von Hippel & Hamrock, 2019; Kuhfeld, 2019) suggests that the evidence for socioeconomic and racial gaps in summer learning loss was a measurement artifact. Regardless, even after accounting for such measurement concerns there is clear evidence that, in general, student learning declines substantially or at least stagnates for many students during the summer (Atteberry & McEachin, 2020).

This literature has motivated substantial interest in the design, implementation, and evaluation of programs that extend structured learning opportunities for students through the summer (Alexander, Pitcock, & Boulay, 2016). In general, the existing evaluations of summer learning programs provide some justification for modest and cautious optimism about their impact on student achievement. However, the design and targeting of these programs, including their focus on short-term standardized test scores as outcomes, suggests that there is still much to learn about how to structure summer learning opportunities to best support student success.

Two meta-analytic reviews illustrate the relatively narrow focus of many summer learning program studies. The first (Lauer et al., 2006) identifies 18 early experimental studies of summer learning programs published between 1985 and 2003—14 of which examined reading outcomes and 12 of which examined math outcomes. They find that participation in a summer program leads to modest improvements in reading achievement ($d = 0.05$) and in mathematics ($d = 0.09$). However, there are several notable limitations in this early group of interventions. First, generalizability is an open question in these early studies; most of these programs target only those students who were identified for remediation based on low prior achievement.[4] Second, the inferences based on smaller samples and implemented in the context of a controlled experiment may provide a poor guide to the impact of programs conducted in real-world settings and at a larger scale. Finally, nearly all of these

[4] Similarly, several regression-discontinuity (RD) studies have also focused on the impact of summer remediation programs targeted to low-performing students and consistently found positive effects (e.g., Jacob & Lefgren, 2004; Mariano & Martorell, 2013; Matsudaira, 2008; Zvoch & Stevens, 2011).

programs serve students in kindergarten, elementary school, or high school rather than in middle school where summer learning loss appears to be particularly large (Kuhfeld, 2019).

Another meta-analytic review (Kim & Quinn, 2013) summarizes evidence from 41 summer initiatives that are mostly literacy-focused and situated in either the home or in classrooms. Seventeen (i.e., 40 percent) of those studies employ experimental or rigorous quasi-experimental designs (e.g., regression discontinuity) and 82 percent of those 17 studies focus exclusively on elementary-school students rather than middle-school students. Similar to Lauer et al. (2006), results suggest a modest treatment effect (d = .10) on reading achievement among low-income students.[5]

Notably, the curricular focus of these summer programs (i.e., reading and/or math) is often quite narrow relative to what students experience during the academic year. These and other design features of prior summer programs (and their limited impact) suggest an important question: Might there be unique benefits to students from a summer program that parallels the breadth of the standard academic year more closely and also has an intentional focus on social-emotional development? This sort of alternative design for summer learning opportunities could also shift the evaluative frame of the corresponding research in potentially insightful ways. Programs that focus on supporting *behavioral engagement*—students' participation in the work and social life of school—suggest the importance of examining longer-term effects and allowing for the cumulative effects of more than one summer opportunity.

This alternative perspective also implies that our evaluative lens should extend beyond test scores to other educationally relevant outcomes related to behavioral engagement. A long-standing research literature has recognized that the multifaceted dimensions of school engagement (i.e., behavioral and emotional aspects) are important antecedents to longer-run educational success (Fredricks, Blumenfeld, & Paris, 2004; McCarthy & 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 (ESSA) allows states to use measures of "social and emotional learning" (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

[5] A more recent addition to this literature (Zvoch & Robertson, 2017) also finds that random assignment of rising first grade students to a summer literacy program improves indicators of early literacy. By contrast, Lynch and Kim (2017) find that random assignment to an online summer mathematics program had no impact on math achievement, even when coupled with the offer of a free laptop.

important and more reliable indicator of behavioral engagement in school (Hough, 2019; Jordan & Miller, 2017).

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 engagement outcome measures. However, they find that, though the treatment increased students' summer participation in math activities, the program has no effect on math-related enjoyment or intrinsic motivation.

The 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 by improving basic math and literacy skills, 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." 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 social skills or problem behaviors.

A more recent RAND study (Augustine et al., 2016) examines the impact of the National Summer Learning Project (NSLP). This program consists of full-day programming for 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 instruction in mathematics and language arts in small classes, with no more than 15 students per teacher. Though the primary focus of the NSLP is academic success, the randomized evaluation also includes social-emotional measures and data on the school behavioral engagement of students (i.e., attendance and suspensions) among the outcome measures—and measures these outcomes longitudinally. Augustine et al. (2016) find that NSLP led to a modest, near-term gain in math performance (d = 0.08) that faded out before the next summer. The report also indicated that it had no detectable effects on future attendance, suspensions, or on teacher-reported social-emotional competencies.

In sum, the extensive literature on summer learning programs suggests that, as designed, the programs have modest effects on short-term achievement but no effects on dimensions of behavioral engagement that can support more substantial and longer-term gains in educational success.

The *Aim High* Summer Learning Program

Aim High is a summer learning program that seeks to enhance both the persistent behavioral engagement and the academic achievement of entering and continuing middle school students in San Francisco Unified School District (SFUSD) and other surrounding communities. Participating students attend the program at both independent sites and sites provided by SFUSD for roughly seven hours a day over five days a week and five weeks. Aim High is an independent non-profit organization and is not a remediation program. Rather, its mission is to "create […] life-changing opportunities during the summer and beyond" with a focus on serving students from low-income families and neighborhoods. Aim High has built a strong and sustained partnership with SFUSD central administration and school leaders, resulting in coordinated efforts to garner SFUSD student participation and provide those participants with additional enrichment opportunities in the district during non-summer seasons. Descriptively, Aim High students enjoy higher ninth-grade rates of attendance, lower ninth-grade rates of disciplinary involvement, and higher graduation rates than average non-participants in the district.

Aim High's curricula focus on 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 is aligned with each respective grade and subject. Each of these traditional subjects includes experiential and project-based learning for each year of participation, in additional to traditional in-class lessons. Aim High has developed a fourth curricular focus on "Issues and Choices" to strengthen students' positive associations with schooling. Since the summer of 2014, Issues and Choices courses have offered a full curriculum explicitly addressing 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 a stronger community and empowering others to challenge stereotypes (see Appendix A for details).

Complementing a strong curricular focus on behavioral engagement in school is an emphasis on relationship-building through team teaching, which includes cooperation between lead teachers, teacher assistants and interns. When selecting classroom instructors, Aim High places a high priority on how well instructors will build meaningful, positive relationships with students. To that end, the organization regularly 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. A teaching certification is not required for lead teachers or teacher assistants, but a teaching certification or a graduate degree in an education-related field is required for the Academic Coordinator position that provides training and coaching for classroom teams.

Control over content taught in the classroom is loosely coupled with the structure of the organization, giving instructors some creativity and latitude for how they choose to implement and present requisite lessons to students. Although teachers have some creative control over content, the structure of instruction in Aim High is quite consistent across classrooms and courses. Instructors in each of the four main subject areas teach students for 50 minutes a day. Instructors of each subject benefit from 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 while the majority of Aim High students (about 66 percent) enter the program the summer before they begin seventh grade and 14 percent began the summer before eighth grade. Over two-thirds of Aim High participants participate during more than one of their middle school summers—57 percent of Aim High students participate for two summers and 14 percent participate for all three summers that we observe them eligible to do so. 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.

The process for application and admission into an Aim High summer cohort is consistent across sites. Students and their parents or guardians fill out an application form that staff use to assess need, diversity, interest, and commitment among applicants. Dimensions of need include considerations of 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 a flexible and subjective process. For example, site directors and staff make accommodations for students with unstable living conditions who are harder to contact or who have parents who are not as involved in their child's education. Between 2016 and 2019, the median household annual income for a SFUSD Aim High student is about $38,000 (i.e., roughly two-fifths of the household median income city-wide), and 48 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. The program also considers how applicants contribute to racial, ethnic, language, and ability diversity among Aim High participants. The SFUSD program serves a student population that is 26 percent Latinx, 10 percent African American, 54 percent Asian, and 1 percent white.

Data and Methods

Our research questions ask how participation in Aim High affects students' subsequent behavioral engagement and achievement. We view an examination of Aim High's educational impact as a novel and useful addition to the literature on summer learning programs because of its design features (i.e., academic breadth and an SEL focus), its large-scale operations, and its relevant targeting (i.e., middle school students). The longitudinal data and methods we describe below also allow us to provide unique evidence on the dynamic effects of program participation. Educational programming that boosts students' engagement and learning in school may initiate recursive processes that bring about further school successes (Finn, 1989; Walton & Wilson, 2018). This means involvement in Aim High for a single summer may continue to benefit students academically over several years. Furthermore, because Aim High programming has many thematic goals that are scaffolded from one summer to the next (e.g., goal-setting for entering middle school as a rising sixth grader, evaluating academic goals as a rising seventh grader, goal-setting for successful transition to high school as a rising eighth grader), students participating over multiple summers may be well-positioned to enjoy additional benefits from the program.

Data

The data we examine in this study come from a combination of Aim High and San Francisco Unified School District (SFUSD) administrative records. SFUSD is a diverse public-school district serving a student population that is 35 percent Asian, 31 percent Latinx, 14 percent white, 8 percent African American, and 12 percent of another race or ethnicity. Fifty-two percent of students participate in the district's free or reduced priced lunch program and 28 percent are English language learners. Our data include records from the 2009-10 through 2017-18 school years.

We define two intent-to-treat (ITT) samples using cohorts of fifth grade students. We focus on fifth graders because 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 absence and suspension from school. These are the first two cohorts of SFUSD students who experienced Aim High's full SEL curriculum roll-out in summer 2014, and this design strategy allows us to collect data on these students back to their first-grade school year and track them through their time in middle school. The two-cohort ITT sample consists of all SFUSD students who were enrolled in fifth grade at an SFUSD school for at least 175 days (n = 7,908 students).[6] Specifically, our panel dataset 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). Consequently, these sample restrictions do not allow us to assess program effects on "rising ninth graders."

Our two-cohort analytical sample consists of 57,599 student-year observations for which we observe students at least 175 days in a given school year. Student mobility into and out of the school district implies that this is a somewhat unbalanced panel (i.e., each student is not observed in each school year, nor for 175 or more days in a given school year). This missingness could threaten 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 might 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 examined the "effect" of Aim High participation on missingness and found small and statistically insignificant effects (Table B3). Additionally, in Appendix Tables D2 through D5 we show that balanced-panel results (i.e., retaining only students who we observe in every grade from 1 through 8), are statistically and substantively similar to results using an unbalanced panel design.

We also define a one-cohort ITT sample to evaluate the impacts of Aim High participation on state test scores as measured by the California Assessment of Student Performance and Progress (CAASPP) English language arts (ELA) and math assessments. 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 can observe test scores over each of four academic years (i.e., 2014-15 through 2017-18) in grades

[6] This baseline sample definition excludes a small number of students (n = 179) who were not enrolled in fifth grade for at least 175 days. In Table B1, we show that these students were more likely to be absent, to be English 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.

4 through 7. We define the one-cohort ITT sample as all SFUSD students who sat for the ELA and math state tests in fifth grade during the 2015-16 school year (N=4,322).[7] Our one-cohort analytical sample consists of 14,853 student-year observations. 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 (Table B3). We also find similar results when using the smaller sample of students with complete observations over the four-year study window (Table D6).

To measure treatment status, we rely on the fact that Aim High records each student's participation in their summer learning program along with a range of personal identifiers. The organization shares their list of participants with SFUSD data managers, who link Aim High records to randomized student identifiers generated by the district and to regularly collected 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 seven percent of our two-cohort ITT sample (i.e., 520 out of 7,908 students) participate in Aim High at least once. Roughly two percent of our student-year observations occur after a student has participated in Aim High. 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 instead allow for dosage effects by constructing binary measures that identify student-year observations that occur after one year, two years, and three years of participation.

We rely on SFUSD administrative datasets to construct several student outcome measures. SFUSD administrative datasets contain student attendance information recorded 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 ten percent of days enrolled during a given school year, so long as they were enrolled for 175 to 180 days in that school year.

[7] This sample definition based on baseline enrollment 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 (Table B2).

Suspension data provided by SFUSD include individual records of each time a student is suspended (whether suspended in-school or out-of-school). From these records, we create a flag to indicate whether a student is suspended at least once during a given school year. We measure academic achievement using annual student English language arts (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 z-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 one.

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., English language arts 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 "difference-in-differences" (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, 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. $\mathbf{X}_{st}$ 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.[8] θ is the coefficient of interest, representing the estimated effect of A_{st}, a binary indicator for whether a student participated in Aim High in any summer prior before year *t*.

The static DD specification represented in equation 1 assumes that the treatment effect is constant over time. However, the character of the Aim High program suggests that its effects could instead have dynamic features. For example, given its SEL focus, Aim High may begin a recursive cycle of improved behavioral engagement and achievement that implies larger treatment effects in subsequent years (see Finn, 1989). 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 employ 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 then test the equivalence of these coefficients of interest using the null hypothesis of a constant treatment effect:

$$H_0: \delta_0 = \delta_{-1} = \delta_{-2}$$

Another theoretically plausible way to conceptualize dynamic treatment effects in this context is with respect to "dosage." That is, additional summers of Aim High participation could have additional impact by building on the skills and knowledge developed over previous summers. Alternatively, if additional years 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 for the first, second, or third time (i.e., j=1, 2, or 3).

[8] 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.

Notably, the use of unrestrictive dummy variables to capture a student's dosage implies that we are not imposing a functional form on the effects of additional years of participation. In supplementary models, the results of which can be found in Table C1, we simultaneously test for the presence of dosage effects for each outcome versus passive lagged effects.

Arguably, the most critical maintained assumption of this quasi-experimental approach is that the year-to-year outcome changes among "control" students (i.e., those without a change in treatment status) provide a valid counterfactual for what would have changed for 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 control 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 through event-study specifications of the following form:

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

This event-study specification effectively extends the semi-dynamic specification (equation 2) to allow for fixed effects unique to each year prior to participating in Aim High (i.e., "leads" of treatment adoption). 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 in Aim High, treatment students have year-to-year changes in outcomes distinct from control students:

$$H_0: \delta_4 = \delta_3 = \delta_2 = \delta_1 = 0$$

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. Furthermore, 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 can even apply negative weights to some treated observations. To assess the empirical relevance of these concerns, we implemented the recommendations and procedure recently introduced by de Chaisemartin and D'Haultfoeuille (forthcoming). We examined the weights implied by our static DD specifications and found that they produced no negative values. Even so, as we describe below, our results suggest that the effects of Aim High clearly have dynamic rather than static properties, which leads us to emphasize interpretation of the semi-dynamic and dosage DD models over the static DD models.

Results

Table 1 displays descriptive statistics of the unbalanced two-cohort and one-cohort panel analytic samples, differentiating Aim High participants from all other students. Our two-cohort analytic sample contains 57,559 student-year observations and 7,908 unique students (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 non-participants across grades one through eight, by about one-and-a-half percentage points, (2.0 percent compared to 3.4 percent). These trends are similar for unexcused absence rate. 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 have half the rate of suspensions as other students from grades one through eight (0.7 compared to 1.4 percent). Aim High and non-Aim High students have largely similar ELA and math test scores in the four school years for which we have state test data for each of the two cohorts.

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 learners. Special education students have slightly lower representation in Aim High than among non-Aim High students (8.5 percent and 12.2 percent, respectively). White students make up 14 percent of the population of students who never participate in Aim High, but only about one percent of Aim High students. The Aim High population has a similar share of African American students as the non-Aim High group (about 8 percent) and a lower share of Latinx students (20.4 percent compared to 25.6 percent). Fifty-eight percent of Aim High students are Asian, compared to 38 percent of non-Aim High students. Parents of Aim High students in the sample report lower levels of education than those whose children do not participate in Aim High. The largest difference is between the proportions of students whose parents are college graduates or higher (14 percent of Aim High students compared to 29 percent of non-Aim High participants).

Our one-cohort sample contains 14,853 student-year observations and 4,322 unique students, 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, African American students and those whose parents have no more than a high school diploma have a more disproportionate representation in Aim High in the one-cohort sample compared to the two-cohort sample.

Excused and Unexcused Absences

To measure the effect of Aim High on behavioral engagement in school, we rely first on tracking students' absence rates over time. 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 engagement 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 have been shown to be a signal of student and family disengagement from school (Fredricks et al., 2011; Gershenson et al., 2017; Gottfried, 2009). Thus, 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 absences.

Table 2 displays the key results from the static, semi-dynamic, and dosage difference-in-differences (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. For example, the static DD model (column 1) suggests that Aim High reduces the absence rate by roughly a third of a percentage point (b=-0.31, SE=0.18, t=-1.68, p=0.093). Four other features of these results are particularly noteworthy. First, the reduced absences associated with Aim High participation are concentrated almost exclusively 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 is effective in promoting behavioral engagement.

Second, the results in Table 2 consistently indicate that the effects of Aim High participation on overall and unexcused absences have dynamic properties. 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's first participation, the estimated impact is 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). The specifications that instead model the dynamic effects of Aim High in terms of "dosage" return quite similar results. For example, relative to students who never participated, those who participate in Aim High for three summers have a rate of unexcused absences that is again three-quarters of a percentage point lower (column 9). This evidence that the beneficial impact of Aim High participation grows monotonically over time and with additional exposure is statistically significant. Specifically, the p-values in the bottom row of Table 2 show that the assumption of a common treatment effect (i.e., as assumed in the static DD specifications) is consistently rejected.

The strong correspondence between the results of the two dynamic specifications is not surprising. More than two-thirds of Aim High participants take part in the summer program more than once, meaning there is a high degree 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. Nonetheless, we have conducted ancillary analyses (see Table C1) based on more complicated DD specifications that simultaneously allowed for both longer-term and dosage effects (e.g., being in the second year after first participation and being in the year immediately after participating a second time). These results generally suggest that dosage effects are somewhat 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 substantially larger than the effect of being in the second year after first participation (column 2, Table C1). However, these differences are not always statistically significant, so some agnosticism is warranted.

A third important feature of the results in Table 2 concerns effect sizes. In terms of the percent reduction in attendance rate, the estimated benefits of Aim High participation are quite large. The average unexcused absence rate of middle school students who did not participate in Aim High is 1.5 percent. We find that, three years after their first participation, Aim High participants have an absence rate that is a full 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. This is the equivalent of about 0.35 and 0.16 percent of a standard deviation decrement in math and reading test scores, respectively (see Aucejo & Romano, 2016; Carlsson, Dahl, Öckert, & Rooth, 2015).

Finally, the results reported in Table 2 appear to be quite robust. Corresponding event-study estimates for these results in Appendix Table C2 imply similar trends in our attendance measures for Aim High participants and non-participants in the years *prior* to participation and differences that emerged *after* participation. These patterns are consistent with the identifying assumption of these DD specifications. We also find that the 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 (Appendix Table D1). These results are also all similar when using the smaller sample balanced-panel data, which includes only students for whom we can observe full information across all eight school years (Appendix Tables D2 and D3).

Chronic Absenteeism

In Table 3, we examine the impact of Aim High participation on the probability that a student is chronically absent (i.e., missing 10 percent or more of days in the school year). This is a uniquely 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, chronic absenteeism is a frequently used metric in the school-accountability systems states have recently developed and implemented under the Every Student Succeeds Act (Jordan & Miller, 2017).

The results in Table 3 consistently indicate that Aim High participation implies a substantial reduction of the probability a student is chronically absent. For example, the static DD model in column 1, 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 with respect to 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).[9] More formally, the p-values 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

[9] 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 (Table C1, column 3). However, these distinctions are not often statistically meaningful given the high number of Aim High participants who attend more than one summer.

percentage points less likely to be chronically absent in eighth grade. 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 that persistent participation in Aim High reduces chronic absenteeism by 58 percent. We also note that, as with the absence rate results, these findings appear quite robust. Most notably, the results of an event-study specification (column 3, Table C2) are consistent with the identifying assumption of this research design. Those results indicate that, in the years prior to Aim High participation, the trends in chronic absenteeism were similar among future Aim High participants and non-participants. We also find similar results in a logistic regression (Table D1) and when only using students for whom we can observe full information across all eight school years (Table D3).

Suspension from School

We also consider the impact of Aim High 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 their future engagement with school and other social institutions (Jacobsen, 2020; Kupchik & Catlaw, 2015; Pyne, 2019). However, we also note that 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, & 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 the inferences based on our quasi-experimental approach.

In Table 4, we present the key results from DD specifications that estimate the impact of Aim High participation on the probability of being suspended. The static DD specification (column 1) suggests that a student is just over one percentage point less likely to become suspended in school years following participation in Aim High compared to 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 somewhat suggestive evidence that the estimated effects of Aim High have dynamic properties. That is, the reductions in suspensions due to Aim High are concentrated in the second and third year after first participating as well as after a second year of participation.[10] The effect sizes implied by the estimates in Table 4 are quite large. Specifically, given that 3.0 percent of those who never participated in Aim High became

[10] With regard to suspensions, our capacity to discriminate between lagged effects of first exposure and dosage effects seems limited (Table C1). The reduction in suspensions appears concentrated among students who participated in Aim High twice or who are in their third year after first participating.

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 non-participants. Finally, we note that the event-study results (Table C2) indicate that suspension probabilities trended similarly among Aim High participants and non-participants in the years prior to the program becoming available. We found similar results to those reported above when using a logistic regression (Table D1) and when using a balanced panel of student-by-year observations (Table D4).

Academic Achievement

We now turn to the effects of Aim High participation on student achievement over time. Due to data limitations, we are able to 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 allows us to 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 English language arts and mathematics state test scores, standardized within grade, school year, and test score subject.

The results in Table 5 indicate that students experience an average increase of about six percent of a standard deviation in English language arts (ELA) test scores in the year or years following their participation in Aim High (column 1), compared to their expected score without Aim High participation (b=0.06, SE=0.03, t=1.95, p=0.052). A test-score impact of this magnitude is consistent with those found in prior studies of summer learning programs (see Kim & Quinn, 2013; Lauer et al., 2006) and are larger than what would be suggested by the program's effects on improved attendance as a mediator (based on estimates by Aucejo & Romano, 2016; Carlsson et al., 2015). However, we found 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 due to Aim High came immediately after participating. However, the structure of the available test-score data (one cohort, four years) limits our capacity to examine such 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 these 64 students are in a position to take a second summer of Aim High in the relatively short time frame that we observe ELA test scores.

The weakly significant estimated effect of Aim High participation on ELA scores appears to be a robust result. The evidence from an event-study specification (Table C3) is 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 (Tables D6 and D7) and are also robust when excluding the 64 students who first participated as rising sixth graders (Table D8).

Effects by Racial, Ethnic and Gender Subgroups

In Table 6, we employ 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 African American, Latinx and Asian students. We exclude reports among white, multiracial and other racial and ethnic minority students because very few students in those racial groups participate in Aim High (see Table 1). Across our outcome measures, these results suggest Latinx Aim High students experience the largest effects out of all reported racial subgroups as a result of Aim High participation on all engagement outcomes, while African American and Asian students typically experience no statistically significant effects of participation. 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 bump in English language arts 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).

Discussion

Several contemporary factors—the growing evidence of summer learning loss, the interest in expanded instructional time, and the 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., 2016; Kim and Quinn, 2013) indicate that summer learning programs, which often feature a narrow (i.e., single- or two-subject) curriculum, have only modest achievement effects and no clear effects on the social-emotional outcomes that are important antecedents to longer-run educational success. Furthermore, the large prior literature on summer learning programs tends to focus on short-run effects and not those effects 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. 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 Appendix A). The targeting of the Aim High program to middle school students may also be apt as this is a period when summer learning loss is particularly stark and the challenges of sustaining students' behavioral engagement intensify. Finally, Aim High is also an unusually mature program operating at a fairly large scale in the San Francisco Unified School District (SFUSD). Its scale and maturity suggest that inferences based on this program may provide a reliable guide to the likely impact of other large-scale initiatives to provide summer learning opportunities.

The availability of longitudinal administrative data from SFUSD and Aim High allowed us to implement a quasi-experimental examination of the program's impact and to consider how its effects varied over subsequent years and with additional summers of participation. Our main finding is that Aim High participation implies substantial reductions in our behavioral engagement proxies (e.g., chronic absenteeism and suspensions). We find that these effects often appear to grow 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 reduced the probability of being suspended by 1.1 percentage points (i.e., a 37 percent reduction relative to peers who never participated). In contrast, we found that Aim High had modest and weakly significant effects on ELA achievement (i.e., ES = 0.06) and no effects on math achievement.[11]

Interestingly, we also find that these effects tend to be concentrated among boys and Latinx students. We speculate that we see no detectable effects on behavioral engagement outcomes among Asian Aim High participants due to floor effects; among Asian middle school students never participating in Aim High, the chronic absenteeism rate is 1.2% and the suspension rate is 0.5%, far below the average rates of non-Asian nonparticipating middle school students. This leaves little room for the program to improve behavioral engagement, although descriptively we see that Asian students

[11] The ELA gain due to Aim High participation is consistent with those found in prior studies of summer learning programs (see Kim & Quinn, 2013; Lauer et al., 2006). However, we also note that our evidence suggests these gains are temporary. The absence of a math effect may reflect the fact that our achievement scores coincided with SFUSD's implementation of the new Common Core State Standards.

participating in Aim High have rates on both outcomes that are descriptively lower than the above figures. We are more agnostic about what drives the null effects among African American participants, since floor effects are not a concern for this subgroup. Regarding gender differences, past research suggests that boys are generally less engaged in school than girls beginning at school entry, and girls tend to experience increases in behavioral engagement over elementary school while boys experience decreases (DiPrete & Jennings, 2012; Downey, Workman, & von Hippel, 2019; Pyne, 2020). These trends in behavioral engagement become more pronounced when boys experience much higher levels of disciplinary sanctions and disengagement or withdrawal starting in middle school (DiPrete & Buchmann, 2013). It seems likely that the engagement-focused nature of Aim High is more impactful on boys than girls because there is more ground for boys to recover.

At least two caveats about our findings are worth underscoring. One is that the program has non-trivial costs. In examining Aim High's 2017 Form 990 filing with the Internal Revenue Service (IRS), we estimate that their total spending per student/year is approximately $2,700. To the extent that the program benefits rely on participating during multiple summers, the relevant costs would be correspondingly larger. In contrast, the recent evidence from nudge-like interventions suggest that it may be possible to generate similar short-term improvements in attendance and discipline at lower cost (Borman, Rozek, Pyne and Hanselman, 2019; Rogers & Feller, 2018), although 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 indicate that further innovations in and assessments of summer learning programs will be a productive endeavor for supporting the educational potential of students in the United States.

References

Alexander, K. L., Entwisle, D. R., & Horsey, C. S. (1997). From first grade forward: Early foundations of high school dropout. *Sociology of Education, 70*(2), 87-107.

Alexander, K. L., Pitcock, S., & Boulay, M. C. (2016). *The summer slide: What we know and can do about summer learning loss*. New York: Teachers College Press.

Atteberry, A., & McEachin, A. (2020). School's out: The role of summers in understanding achievement disparities. *American Educational Research Journal*. doi:10.3102/0002831220937285

Aucejo, E. M., & Romano, T. F. (2016). Assessing the effect of school days and absences on test score performance. *Economics of Education Review, 55*, 70-87.

Augustine, C. H., McCombs, J. S., Pane, J. F., Schwartz, H. L., Schweig, J., McEachin, A., & Siler-Evans, K. (2016). Effects of voluntary summer learning programs on low-income urban youth. Santa Monica, CA: RAND Corporation.

Blad, E. (2017). No state will measure social-emotional learning under ESSA. Will that slow its momentum? *Education Week*. Retrieved from https://www.edweek.org/ew/articles/2017/10/04/no-state-will-measure-social-emotional-learning-under.html.

Borman, G. D., Rozek, C. S., Pyne, J., & Hanselman, P. (2019). Reappraising academic and social adversity improves middle school students' academic achievement, behavior, and well-being. *Proceedings of the National Academy of Sciences*, *116*(33), 16286-16291.

Bowles, S., & Gintis, H. (1976). *Schooling in capitalist America: Educational reform and the contradictions of economic life*. New York: Basic Books.

Carlsson, M., Dahl, G. B., Öckert, B., & Rooth, D.-O. (2015). The effect of schooling on cognitive skills. *Review of Economics and Statistics, 97*(3), 533-547.

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

de Chaisemartin, C., & D'Haultfœuille, X. (forthcoming). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*.

DiPrete, T. A., & Buchmann, C. (2013). *The rise of women: The growing gender gap in education and what it means for American schools*. New York: Russell Sage Foundation.

DiPrete, T. A., & Jennings, J. L. (2012). Social and behavioral skills and the gender gap in early educational achievement. *Social Science Research, 41*d, 1-15.

Downey, D. B., von Hippel, P. T., & Broh, B. (2004). Are schools the great equalizer? School and non-school sources of inequality in cognitive skills. *American Sociological Review, 69*(5), 613-635.

Downey, D. B., Workman, J., & von Hippel, P. (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.

Entwisle, D. R., & Alexander, K. L. (1992). Summer setback: Race, poverty, school composition, and mathematics achievement in the first two years of school. *American Sociological Review, 57*(1), 72-84.

Finn, J. D. (1989). Withdrawing from school. *Review of Educational Research, 59*(2), 117-142.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*(1), 59-109.

Fredricks, J. A., McColskey, W., Meli, J., Mordica, J., Montrosse, B., and Mooney, K. (2011). Measuring student engagement in upper elementary through high school: A description of 21 instruments. (Issues & Answers Report, REL 2011–No. 098). 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.

Gershenson, S., Jacknowitz, A., & Brannegan, A. (2017). Are student absences worth the worry in US primary schools? *Education Finance and Policy, 12*(2), 137-165.

Gottfried, M. A. (2009). Excused versus unexcused: How student absences in elementary school affect academic achievement. *Educational Evaluation and Policy Analysis, 31*(4), 392-415.

Gottfried, M. 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. doi:10.3102/0002831209350494

Heyns, B. (1978). *Summer learning and the effects of schooling*. New York: Basic Books.

Hough, H. (2019). Role call. In M. A. Gottfried & E. L. Hutt (Eds.), *Absent from school: Understanding and addressing student absenteeism* (pp. 15-33). Cambridge, MA: Harvard Education Press.

Jacob, B., & Lefgren, L. (2004). Remedial education and student achievement: A regression - discontinuity analysis. *Review of Economics and Statistics, 86*(1), 226-224.

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

Jordan, P. W., & Miller, R. (2017). Who's in: Chronic absenteeism under the Every Student Succeeds Act. Washington, D.C.: FutureEd.

Kim, J. S., & Quinn, D. M. (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.

Kuhfeld, M. (2019). Surprising new evidence on summer learning loss. *Phi Delta Kappan, 101*(1), 25-29.

Kupchik, A., & Catlaw, T. J. (2015). Discipline and participation the long-term effects of suspension and school security on the political and civic engagement of youth. *Youth & Society, 47*(1), 95-124.

Lauer, P. A., Akiba, M., Wilkerson, S. B., Apthorp, H. S., Snow, D., & Martin-Glenn, M. L. (2006). Out-of-school-time programs: A meta-analysis of effects for at-risk students. *Review of Educational Research, 76*(2), 275-313.

Lynch, K., & Kim, J. S. (2017). Effects of a summer mathematics intervention for low-income children: A randomized experiment. *Educational Evaluation and Policy Analysis, 39*(1), 31-53.

Mariano, L. T., & Martorell, P. (2013). The academic effects of summer instruction and retention in New York City. *Educational Evaluation and Policy Analysis, 35*(1), 96-117.

Matsudaira, J. D. (2008). Mandatory summer school and student achievement. *Journal of Econometrics, 142*(2), 829-850.

McCarthy, M., & Kuh, G. D. (2006). Are students ready for college? What student engagement data say. *Phi Delta Kappan, 87*(9), 664-669.

Morris, E. W., & Perry, B. L. (2016). The punishment gap: School suspension and racial disparities in achievement. *Social Problems, 63*(1), 68-86. doi:10.1093/socpro/spv026

Okonofua, J. A., Walton, G. M., & Eberhardt, J. L. (2016). A vicious cycle: A social–psychological account of extreme racial disparities in school discipline. *Perspectives on Psychological Science, 11*(3), 381-398.

Perry, B. L., & Morris, E. W. (2014). Suspending progress: Collateral consequences of exclusionary punishment in public schools. *American Sociological Review, 79*(6), 1067-1087.

Pyne, J. (2019). Suspended attitudes: Exclusion and emotional disengagement from school. *Sociology of Education, 92*(1), 59-82.

Pyne, J. (2020). Gender test score gaps under equal behavioral engagement. *Educational Researcher, 49*(6): 459-464.

Rogers, T., & Feller, A. (2018). Reducing student absences at scale by targeting parents' misbeliefs. *Nature Human Behaviour, 2*(5), 335-342.

Rumberger, R. W. (1995). Dropping out of middle school: A multilevel analysis of students and schools. *American Educational Research Journal, 32*(3), 583-625.

Schueler, B. (2020). Summer "vacation academies" can narrow coronavirus learning gaps. Cambridge, MA: Education Next. Retrieved from https://www.educationnext.org/summer-vacation-academies-narrow-coronavirus-learning-gaps-springfield.

Wald, J., & Losen, D. J. (2003). Defining and redirecting a school-to-prison pipeline. *New Directions for Youth Development, 2003*(99), 9-15.

Walton, G. M., & Wilson, T. D. (2018). Wise interventions: Psychological remedies for social and personal problems. *Psychological Review, 125*(5), 617-655.

Zvoch, K., & Robertson, M. C. (2017). Multivariate summer school effects. *Studies in Educational Evaluation, 55*, 145-152.

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

Tables

Table 1. Descriptive Statistics

	Analytic Sample			
	Two-Cohort		One-Cohort	
Variable	Aim High (n=4,024)	Never Aim High (n=53,535)	Aim High (n=906)	Never Aim High (n=13,947)
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)	0.04 (0.87)	0.06 (0.99)	-0.03 (0.91)	0.02 (1.00)
Math Score (standardized)	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)
African American	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)

Note: 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 ITT sample with full information and who are enrolled for 175 to 180 days in a given year from grades 1 through 8; N=57,559 student-year observations, which includes 7,908 unique total students, 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 intent-to-treat 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 California Assessment of Student Performance and Progress (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=14,853 student-year observations, which includes 4,322 unique students, 248 of whom were ever in Aim High. Eighty percent of students show up three or more grades in the sample. See Appendix B for more information on attrition from the samples. ^Students in the two-cohort sample only have California Assessment of Student Performance and Progress (CAASPP) ELA and math test scores in grades five through eight.

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

Note: The intent-to-treat 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 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 to 180 days in a given year from grades 1 through 8; N=57,559 student-year observations, which includes 7,908 unique total students, 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 Appendix Table C2. Balanced panel data results are very similar to those shown above and can be found in Table D2. Standard errors, clustered at the student level, are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

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

Note: The intent-to-treat 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 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 to 180 days in a given year from grades 1 through 8; N=57,559 student-year observations, which includes 7,908 unique total students, 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 (LPMs). 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 Tables C2. Balanced panel data results are very similar to those shown above and can be found in Table D3. Standard errors, clustered at the student level, are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

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

Note: The intent-to-treat 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 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 to 180 days in a given year from grades 1 through 8; N=57,559 student-year observations, which includes 7,908 unique total students, 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 Tables C2. Balanced panel data results are very similar to those shown above and can be found in Table D4. Standard errors, clustered at the student level, are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

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

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

Note: The intent-to-treat sample consists of students enrolled in fifth grade during the 2015-16 school year (see 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=14,853 student-year observations which includes 4,322 unique students, 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 Table C3. Balanced panel data results are very similar to those shown above and can be found in Table D6. Standard errors, clustered at the student level, are in parentheses *** p<0.01, ** p<0.05, * p<0.10.

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

		Race and Ethnicity^			Gender	
Dependent Variables	Full Sample	African American	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*

Note: The intent-to-treat sample for absence and suspension outcomes consists of two cohorts of students enrolled in fifth grade for 175 to 180 days during the 2011-12 through 2014-15 school years. The intent-to-treat sample for California Assessment of Student Performance and Progress (CAASPP) English language arts 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 ^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).

Appendix A. Aim High Issues and Choices Curriculum Learning Goals

Emerging Sixth Grade Learning Goals

6-1 Time Management: Students will be able to apply time management strategies for the transition to middle school.
6-2 Advocate Against Bullying: Students will develop the skills/tools to become not only self-advocates but for others as well. They will also determine and demonstrate multiple ways one can be assertive in the face of bullying and name calling
6-3 Building Awareness and Relationships: Students will be able to understand the dangers of making assumptions of others and how each person's unique trait can be viewed as a contribution. Students will also be able to participate in discussions using the SAFE guidelines.
6-4 Establishing A Safe Space and Transitioning to Middle School: Students will be able to anticipate changes that will occur as they transition into middle school. They will also identify positive actions and people that I can turn to for advice as they face upcoming changes.
6-5 How to Recognize and Prevent Bullying: Students will be able to give a clear and concise definition of bullying. They will also be able to recognize four main categories of bullying. Students will be able to recognize bullying around them as well as practice tools to effectively deal with bullying situations.
6-6 Support Network: Who can help me to be successful in middle school?: Students will be able to generate lists of people and resources that will serve as their support networks as they participate in various activities related to career-planning, career placement, and career success. They will have the tools to self-advocate.
6-7 Goal Setting: What are my goals for sixth grade?: Students will be able to set goals, timelines and create action plans. They will connect their personal strengths, skills and resources they will utilize to their goals.
6-8 A Day In The Life of a Middle School Student: Students will create an artistic representation (skit, video, poster, PSA, etc.) of a healthy response to challenges that may come up when they begin middle school (time management, how to be an ally when a student is being bullied, etc.).

Emerging 7th Grade Learning Goals

7-1 Evaluating Career Goals: Students will engage in research to learn the process for determining what learning strengths lend themselves well to a particular career, the pros and cons of a career goal, and the experience and educational preparation recommended for reaching a career goal.
7-2 My Career Interests: Students will explore their individual interests and see how those interests connect to a career or job. Students will understand the array of careers available, the training necessary for those careers, and create educational goals related to their career aspirations.
7-3 Career Pathway Fair: Students will create a brochure/poster and presentation of the educational pathway to a career they researched.
7-4 Stress and our brains: Students will recognize and name things in their lives that make them feel stressed. They will also identify ways to help them feel relaxed. Students will also learn how stress impacts their decision-making.
7-5 Mindfulness as a way to relieve stress: Students will understand what mindfulness is and how it can be used to deal with stress. Students will practice mindfulness through: checking in with how they are feeling mentally, emotionally, and physically; and mindful
breathing.
7-6 Explore Identity: Students will explore and identify their personal characteristics, interests, & talents. Students will also be able to describe & examine multiple components of their own identity (culture, beliefs, race, etc.). They will understand how their families and/or backgrounds may have influenced their Identity.
7-7 Media and Gender: Students will analyze media examples to decode societal messages of masculinity and femininity. Students will also increase their awareness of his/her physical self and media influences on his/her self-image and behavior.

Emerging Eighth Grade Learning Goals

8-1 Growth Mindset: Students will explore the concept of growth mindset and how it having a growth mindset can help them in high school. Students will learn how to respond to challenges with a growth mindset.
8-2 Empowering Students to Challenge Stereotypes: Students will write their own statements on whiteboards to challenge
stereotypes that they face.
8-3 High school choice: Students will learn about the high school options available to them and
will engage in the high school selection process. Students will be able to conduct research on local high schools and present their results.
8-4 High School Admissions/Enrollment: Students will be aware of the timeline of the high school enrollment/admissions process for public, independent, and parochial schools in their region.
8-5 A-G Requirements: Students will be aware of the important high school classes in which they
need to thrive in order to be admitted to a California university (or in SFUSD, graduate).
8-6 Introduction to Mindfulness: Students will understand what mindfulness is. They will practice mindfulness. Students will understand the importance of kindness and non-judgment (of self and others). They will practice showing kindness and non-judgment.
8-7 Assumptions and Stereotypes: Students will be able to define "stereotype" and "assumptions". Students will reflect on assumptions that other may have about them and how they may have made incorrect assumptions about others. Students will explore ways to avoid making assumptions about others as well as how to respond when others make assumptions about them. (8th)
8-8 The Impact of Stereotypes: Students will explore and understand the different kinds of stereotypes placed on different social groups, and will think critically about the various levels of consequences that come from the stereotypes. Students will understand how stereotypes impact communities differently.
8-9 Empathy for Different Experiences: Students will understand and practice empathy. Students will connect empathy with the concepts of support and solidarity.
8-10 How does being more mindful help create a stronger community?: Students will understand and practice how mindfulness and kindness can lead to a better community.
8-11 Empowering Students to Challenge Stereotypes: Students will write their own statements on whiteboards to challenge stereotypes that they face.
8-10 How does being more mindful help create a stronger community?: Students will understand and practice how mindfulness and kindness can lead to a better community.
8-12 What kind of high school do I want to go to?: Students will develop clarity around how they learn best and determine the factors that can contribute to their success in high school, considering physical space, kinds of classes, activities, philosophies, and teacher/learning styles.
8-13 How will I make a successful transition to high school?: Students will begin thinking about goals they have for themselves once they are in high school. Students will become familiar with the college and career readiness options available to them in high school

Appendix B. Sample Missingness

Table B1. Observed and Missing Groups in the Two-Cohort Sample

Variables	Observed (n=7,908)	Missing (n=179)	Test	*p*-value
Mean (SD) Absence Rate	2.2 (4.2)	4.5 (7.8)	-2.29	0.02
Mean (SD) State ELA test score^	2513 (97)	2503 (88)	1.10	0.27
Mean (SD) State math test score^	2513 (93)	2501 (86)	1.64	0.10
Suspended this year	0.3 (20)	0.0 (0)	0.45	0.50
Female	48.2 (3,811)	45.8 (82)	0.40	0.53
Special education student	12.0 (951)	19.0 (34)	7.95	<0.01
English learner	22.2 (1,759)	15.6 (28)	4.43	0.04
White	13.3 (1,050)	23.5 (42)	15.55	<0.01
African American	8.0 (630)	16.2 (29)	15.86	<0.01
Latinx	25.3 (1,998)	32.4 (58)	4.70	0.03
Asian	39.5 (3,124)	14.0 (25)	48.0	<0.01
Multiracial/Other ethnicity	7.6 (598)	8.38 (15)	0.17	0.68
Missing race/ethnicity	6.4 (508)	5.6 (10)	0.20	0.65
Parent education				
Not a high school graduate	14.0 (1,107)	8.4 (15)	4.62	0.03
High school graduate	18.5 (1,459)	17.3 (31)	0.15	0.70
Some college	13.9 (1,101)	15.6 (28)	0.43	0.51
College graduate or higher	28.0 (2,210)	36.9 (66)	6.89	<0.01
Not reported	25.7 (2,031)	21.8 (39)	1.39	0.24

Note: The full two-cohort sample includes all students enrolled for 175 to 180 days in fifth grade from 2013-14 and 2014-15 (n=8,087). Students in the two-cohort full sample are tracked and assigned a "1" for all missing values in grades 3 through 8; N=59,447 student-year observations. The 'Observed' group retains fifth grade students from the full sample who we observe at least once in SFUSD's administrative data in grades 3 through 8, over the 2009-10 through 2017-18 school years. Ninety-six percent of students in the 'Observed' group show up in five or more grades. Students must be enrolled between 175 and 180 days each year to be included in the observed sample. The 'Missing' group includes all fifth grade students from the full sample who are not included in the Observed group. ^ Fifth grade students in the two-cohort full sample may be missing either math or English language arts fifth-grade California Assessment of Student Performance and Progress (CAASPP) test scores.

Table B2. Observed and Missing Groups in the One-Cohort Sample

Variables	Observed (n=4,322)	Missing (n=232)	Test	*p*-value
Mean (SD) Absence Rate^	3.4 (4.1)	10.1 (17.1)	-6.39	<0.01
Mean (SD) State ELA test score	2509 (103)	2428 (106)	3.99	<0.01
Mean (SD) State math test score	2507 (96)	2383 (98)	7.27	<0.01
Suspended this year^	0.6 (25)	1.5 (4)	3.10	0.08
Female	50.0 (2,027)	44.7 (225)	5.04	0.03
Special education student	11.7 (472)	25.3 (127)	72.4	<0.01
English learner	26.7 (1,083)	32.3 (84)	3.84	0.05
White	14.1 (569)	11.9 (60)	1.69	0.19
African American	7.6 (307)	14.1 (71)	21.12	<0.01
Latinx	25.9 (1,051)	42.7 (215)	62.91	<0.01
Asian	38.5 (1,051)	15.1 (76)	106.04	<0.01
Multiracial/ Other ethnicity	10.9 (441)	9.5 (48)	0.84	0.36
Missing race/ethnicity	3.1 (125)	6.6 (33)	16.13	<0.01
Parent education				
Not a high school graduate	12.4 (501)	8.4 (42)	6.88	<0.01
High school graduate	14.2 (575)	9.2 (46)	9.69	<0.01
Some college	13.1 (532)	10.5 (53)	2.69	0.10
College graduate or higher	26.2 (1,061)	12.1 (61)	47.7	<0.01
Not reported	34.1 (1,382)	60.0 (301)	127.10	<0.01

Note: The full one-cohort sample includes n=4,554 fifth grade students enrolled during the 2015-16 school year. The 'Observed' group retains students from the sample who were observed at least once in SFUSD's longitudinal data in grades 4 through 7, over the 2014-15 through 2017-18 school years, the only years for which California Assessment of Student Performance and Progress (CAASPP) data are available in the district for these grade levels. Eighty percent of students in the 'Observed' group show up three or more grades in the sample. Estimates are derived from ordinary least squares multiple regression models. The 'Missing' group includes all students from the full sample who are not included in the Observed group. ^ Students in the one-cohort full sample may be missing complete attendance and suspension data.

Table B3. Contribution of Aim High Participation to Probability of Missingness from the Samples

	Sample					
	Two-Cohort			One-Cohort		
Independent Variables	Static (1)	Semi-Dynamic (2)	Dosage (3)	Static (4)	Semi-Dynamic (5)	Dosage (6)
After Aim High participation	-0.005	--	--	0.019	--	--
	(0.006)			(0.020)		
First participated one summer prior	--	-0.006	--	--	0.015	--
		(0.006)			(0.018)	
First participated two summers prior	--	-0.005	--	--	0.037	--
		(0.007)			(0.034)	
First participated three summers prior	--	-0.001	--	--	--	--
		(0.015)				
After one summer of participation	--	--	-0.005	--	--	0.013
			(0.006)			(0.018)
After two summers of participation	--	--	-0.003	--	--	0.026
			(0.008)			(0.035)
After three summers of participation	--	--	0.009	--	--	--
			(0.019)			
p value (H_0: $\delta_0=\delta_{-1}=\delta_{-2}$)	--	0.87	0.72	--	0.39	0.64

Note: The full two-cohort sample includes all students enrolled in fifth grade from 2014-14 and 2014-15 (n=8,087). The analytic sample retains those with full information who were enrolled for 175 to 180 days of the school year. Students in the two-cohort full sample are tracked and assigned a "1" for all missing values in grades 1 through 8; N=59,447 student-year observations. The full one-cohort sample includes all students enrolled in fifth grade in 2015-16 (n=4,554). Students in the one-cohort full sample are tracked and assigned a "1" for all missing values in grades 4 through 7; N=16,855 student-year observations. The dependent variable is a binary indicator of whether a student is missing in a given school year (i.e., "Missing"=1; "Not Missing"=0). All models include student, school year and grade level fixed effects. Standard errors, clustered at the student level, are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Appendix C. Supplementary Models

Table C1. Lagged Passive Effects Versus Dosage Effects of Aim High Participation – Engagement Outcomes

	Dependent Variables			
Independent Variable	Total Absence Rate (1)	Unexcused Absence Rate (2)	Probability of Chronic Absence (3)	Probability of Suspension (4)
First summer of participation	0.16	-0.06	0.000	-0.003
	(0.28)	(0.21)	(0.008)	(0.006)
Only one summer of participation, 1-year lag^	-0.43	-0.13	0.017	0.027
	(0.54)	(0.46)	(0.035)	(0.028)
Second summer of participation	-0.81***	-0.65***	-0.035***	-0.026***
	(0.12)	(0.08)	(0.005)	(0.003)
Only one summer of participation, 2-year lag^	-0.67	-0.41	-0.026***	-0.031**
	(0.44)	(0.35)	(0.006)	(0.012)
Third summer of participation	-1.07***	-0.79***	-0.048***	-0.017
	(0.23)	(0.14)	(0.009)	(0.014)
p value (H_0: 1 dose, 1-year lag = 2 doses)	0.48	0.25	0.14	0.05
p value (H_0: 1 dose, 2-year lag = 3 doses)	0.41	0.30	0.04	0.45

Note: The intent to treat two-cohort sample includes all students enrolled in fifth grade from 2014-14 and 2014-15 (see Appendix B for more information on attrition from the sample). This analytic sample is an unbalanced panel of all students from that sample with full information and who are enrolled for 175 to 180 days in a given year from grades 1 through 8; N=57,559 student-year observations, which includes 7,908 unique total students, 520 of whom were ever in Aim High. This model differentiates the effects of additional summers of participation in Aim High from effects among those who participate only one summer but we observe in later years. Absence rate estimates are derived from ordinary least squares multiple regression models. Chronic absence and suspension estimates are derived from ordinary least squares linear probability models. 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. Standard errors, clustered at the student level, are in parentheses. *** p<0.01, ** p<0.05, * p<0.10. ^ We observe 67 students one year after participating in only one summer of Aim High and 13 students two years after participating in only one summer of Aim High. No students appeared in the sample a year after participating in only two summers of Aim High.

Table C2. Two-Cohort Event Study Model Results – Engagement Outcomes

	Dependent Variables			
Independent Variable	Total Absence Rate (1)	Unexcused Absence Rate (2)	Probability of Chronic Absence (3)	Probability of Suspension (4)
4-year lead	0.02	-0.08	0.005	0.002
	(0.10)	(0.06)	(0.007)	(0.004)
3-year lead	0.04	-0.02	0.006	0.005
	(0.10)	(0.07)	(0.007)	(0.003)
2-year lead	0.08	-0.07	0.014*	0.002
	(0.11)	(0.07)	(0.007)	(0.003)
1-year lead	-0.05	-0.19	0.005	-0.003
	(0.14)	(0.10)	(0.008)	(0.005)
First participated one summer prior	0.17	-0.13	0.006	-0.003
	(0.30)	(0.23)	(0.009)	(0.007)
First participated two summers prior	-0.75***	-0.64***	-0.022**	-0.018***
	(0.16)	(0.11)	(0.009)	(0.005)
First participated three summers prior	-0.97***	-0.80***	-0.033**	-0.019*
	(0.22)	(0.15)	(0.014)	(0.011)
p value (H_0: $\delta_4 = \delta_3 = \delta_2 = \delta_1 = 0$)	0.90	0.19	0.44	0.56
p value (H_0: $\delta_0 = \delta_{-1} = \delta_{-2}$)	<0.01	0.01	<0.01	0.12

Note: The intent-to-treat two-cohort sample includes all students enrolled in fifth grade from 2014-14 and 2014-15 (see Appendix B for more information on attrition from the sample). This analytic sample is an unbalanced panel of all students from that sample with full information and who are enrolled for 175 to 180 days in a given year from grades 1 through 8; N=57,559 student-year observations, which includes 7,908 unique total students, 520 of whom were ever in Aim High. Alternate analyses also retaining students enrolled for fewer than 175 days in any school year yield very similar results. Absence rate estimates are derived from ordinary least squares multiple regression models. Chronic absence and suspension estimates are derived from ordinary least squares linear probability models. 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. The reference category is 5 or 6 years prior to first Aim High participation or never participating in Aim High. Balanced panel results can be found in Table D5. Standard errors, clustered at the student level, are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table C3. One-Cohort Event Study Model Results – Academic Outcomes

	Dependent Variables	
	State ELA Test	State Math Test
Independent Variable	(1)	(2)
2-year lead	0.05	-0.02
	(0.04)	(0.04)
1-year lead	0.05	0.05
	(0.04)	(0.04)
First participated one summer prior	0.11**	0.01
	(0.05)	(0.04)
First participated two summers prior	0.06	-0.02
	(0.07)	(0.07)
p value (H_0: $\delta_2=\delta_1=0$)	0.49	0.06
p value (H_0: $\delta_0=\delta_{-1}$)	0.44	0.62

Note: The intent-to-treat sample consists of students enrolled in fifth grade during the 2015-16 school year (see Appendix B for more information on attrition from the sample). This analytic sample is an unbalanced panel of all students from the sample with full information and who are enrolled in a given year in grades 4 through 7, for school years 2014-15 through 2017-18; N=14,853 student-year observations which includes 4,322 unique students, 248 of whom were ever in Aim High. Eighty percent of students show up three or more grades in the sample. The dependent variables are z-scored 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. The reference category is 3 or 4 years prior to first Aim High participation or never participating in Aim High. Balanced panel results can be found in Table D7. Standard errors, clustered at the student level, are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Appendix D. Robustness Checks

Table D1. The Estimated Static Effect of Aim High on Suspension and Absenteeism (Negative Binomial and Logistic Regression Static DD Models)

Dependent Variable	Estimate	Method
Excused Absences (Incidence rate ratio)	0.86***	Negative Binomial Regression
	(0.79 - 0.93)	
Unexcused Absences (Incidence rate ratio)	0.78***	Negative Binomial Regression
	(0.69 - 0.87)	
Chronic Absenteeism (Odds ratio)	0.43***	Logistic Regression
	(0.26 - 0.71)	
Suspension (Odds ratio)	0.50**	Logistic Regression
	(0.27 - 0.93)	

Note: Note: The intent-to-treat two-cohort sample includes all students enrolled in fifth grade from 2014-14 and 2014-15 (see Appendix B for more information on attrition from the sample). This analytic sample is an unbalanced panel of all students from that sample with full information and who are enrolled for 175 to 180 days in a given year from grades 1 through 8; N=57,559 student-year observations, which includes 7,908 unique total students, 520 of whom were ever in Aim High. All models are static difference-in-differences 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, and foster care status. 95% confidence intervals are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table D2. The Estimated Effects of Aim High Participation on Total, Excused and Unexcused Absence Rates (Balanced Panel)

	Dependent Variables								
	Overall Absence Rate			Excused Absence Rate			Unexcused Absence Rate		
Independent Variables	Static (1)	Semi-Dynamic (2)	Dosage (3)	Static (4)	Semi-Dynamic (5)	Dosage (6)	Static (7)	Semi-Dynamic (8)	Dosage (9)
After Aim High participation	-0.25	--	--	0.03	--	--	-0.28*	--	--
	(0.19)			(0.13)			(0.14)		
First participated one summer prior	--	0.24	--	--	0.24	--	--	-0.01	--
		(0.31)			(0.20)			(0.23)	
First participated two summers prior	--	-0.71***	--	--	-0.17*	--	--	-0.54***	--
		(0.13)			(0.09)			(0.09)	
First participated three summers prior	--	-0.88***	--	--	-0.24	--	--	-0.64***	--
		(0.21)			(0.15)			(0.14)	
After one summer of participation	--	--	0.26	--	--	0.25	--	--	0.01
			(0.31)			(0.20)			(0.23)
After two summers of participation	--	--	-0.69***	--	--	-0.15	--	--	-0.55***
			(0.12)			(0.08)			(0.08)
After three summers of participation	--	--	-0.94***	--	--	-0.26	--	--	-0.68***
			(0.23)			(0.18)			(0.14)
p value (H_0: $\delta_0=\delta_{-1}=\delta_{-2}$)	--	<0.01	<0.01	--	0.07	0.07	--	0.03	0.02

Note: The intent-to-treat 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 Appendix B for more information on attrition from the sample). This analytic sample is a balanced panel of all students from that ITT sample with full information and who are enrolled for 175 to 180 days every year from grades 1 through 8; N=45,344 student-year observations, which includes 5,668 unique students, 463 of whom were ever in Aim High. Estimates are derived from ordinary least squares multiple regression models. Alternate analyses that also retain students enrolled for fewer than 175 days in any school year yield very similar results. 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 Tables D5. Standard errors, clustered at the student level, are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table D3. The Estimated Effects of Aim High on Probability of Chronic Absenteeism (Balanced Panel)

Independent Variables	Static (1)	Semi-Dynamic (2)	Dosage (3)
After Aim High participation	-0.015**	--	--
	(0.007)		
First participated one summer prior	--	-0.001	--
		(0.008)	
First participated two summers prior	--	-0.026***	--
		(0.007)	
First participated three summers prior	--	-0.036***	--
		(0.013)	
After one summer of participation	--	--	-0.001
			(0.008)
After two summers of participation	--	--	-0.033***
			(0.005)
After three summers of participation	--	--	-0.044***
			(0.011)
p value (H_0: $\delta_0=\delta_{-1}=\delta_{-2}$)	--	<0.001	<0.001

Note: The intent-to-treat 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 Appendix B for more information on attrition from the sample). This analytic sample is a balanced panel of all students from that ITT sample with full information and who are enrolled for 175 to 180 days every year from grades 1 through 8; N=45,344 student-year observations, which includes 5,668 unique students, 463 of whom were ever in Aim High. Estimates derived from linear probability models (LPMs). Alternate analyses 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 Table D5. Standard errors, clustered at the student level, are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table D4. The Estimated Effects of Aim High Participation on Probability of Suspension from School (Balanced Panel)

Independent Variables	Static (1)	Semi-Dynamic (2)	Dosage (3)
After Aim High participation	-0.011***	--	--
	(0.004)		
First participated one summer prior	--	-0.004	--
		(0.006)	
First participated two summers prior	--	-0.018***	--
		(0.005)	
First participated three summers prior	--	-0.019*	--
		(0.011)	
After one summer of participation	--	--	-0.004
			(0.006)
After two summers of participation	--	--	-0.027***
			(0.002)
After three summers of participation	--	--	-0.017
			(0.014)
p value (H_0: $\delta_0=\delta_{-1}=\delta_{-2}$)	--	0.175	0.001

Note: The intent-to-treat 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 Appendix B for more information on attrition from the sample). This analytic sample is a balanced panel of all students from that ITT sample with full information and who are enrolled for 175 to 180 days every year from grades 1 through 8; N=45,344 student-year observations, which includes 5,668 unique students, 463 of whom were ever in Aim High. Estimates are derived from ordinary least squares linear probability models. Alternate analyses using logistic regression yield very similar results. 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 Table D5. Standard errors, clustered at the student level, are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table D5. Two-Cohort Event Study Model Results – Engagement Outcomes (Balanced Panel)

	Dependent Variables			
Independent Variable	Total Absence Rate (1)	Unexcused Absence Rate (2)	Probability of Chronic Absence (3)	Probability of Suspension (4)
4-year lead	0.07	-0.02	0.006	0.003
	(0.10)	(0.06)	(0.007)	(0.004)
3-year lead	0.10	0.01	0.005	0.006
	(0.11)	(0.07)	(0.007)	(0.003)
2-year lead	0.19	0.01	0.013*	0.003
	(0.11)	(0.07)	(0.007)	(0.003)
1-year lead	0.09	-0.07	0.005	-0.003
	(0.14)	(0.10)	(0.008)	(0.004)
First participated one summer prior	0.31	-0.02	0.004	-0.002
	(0.33)	(0.24)	(0.009)	(0.007)
First participated two summers prior	-0.63***	-0.55***	-0.021**	-0.016***
	(0.16)	(0.10)	(0.009)	(0.005)
First participated three summers prior	-0.79***	-0.65***	-0.030**	-0.017*
	(0.22)	(0.15)	(0.014)	(0.011)
p value (H_0: $\delta_4 = \delta_3 = \delta_2 = \delta_1 = 0$)	0.57	0.86	0.54	0.44
p value (H_0: $\delta_0 = \delta_{-1} = \delta_{-2}$)	<0.01	0.03	<0.01	0.18

Note: The intent-to-treat 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 Appendix B for more information on attrition from the sample). This analytic sample is a balanced panel of all students from that ITT sample with full information every year from grades 1 through 8; N=45,344 student-year observations, which includes 5,668 unique students, 463 of whom were ever in Aim High. Absence rate estimates are derived from ordinary least squares multiple regression models. Chronic absence and suspension estimates are derived from ordinary least squares linear probability models. 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. The reference category is 5 or 6 years prior to first Aim High participation or never participating in Aim High. Standard errors, clustered at the student level, are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table D6. The Estimated Effects of Aim High on Standardized Test Scores (Balanced Panel)

	Dependent Variables					
	English Language Arts			Mathematics		
Independent Variables	Static (1)	Semi-Dynamic (2)	Dosage (3)	Static (4)	Semi-Dynamic (5)	Dosage (6)
After Aim High participation	0.05*	--	--	0.00	--	--
	(0.03)			(0.03)		
First participated one summer prior		0.06*	--	--	0.00	--
		(0.03)			(0.03)	
First participated two summers prior	--	0.02	--	--	-0.02	--
		(0.06)			(0.06)	
After one summer of participation	--	--	0.06*	--	--	0.00
			(0.03)			(0.03)
After two summers of participation	--	--	0.02	--	--	-0.02
			(0.06)			(0.06)
p value (H_0: $\delta_0=\delta_{-1}$)	--	0.53	0.51	--	0.74	0.80

Note: The intent-to-treat sample consists of students enrolled in fifth grade during the 2015-16 school year. This analytic sample is a balanced panel of all students from that ITT sample with full information every year from grades 4 through 7, for school years 2014-15 through 2017-18; N=13,247 student-year observations, which includes 3,538 total unique students, 208 of whom were ever in Aim High. Estimates are derived from ordinary least squares multiple regression models. The dependent variables are z-scored 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 Table D7. Standard errors, clustered at the student level, are in parentheses *** p<0.01, ** p<0.05, * p<0.10.

Table D7. One-Cohort Event Study Model Results – Academic Outcomes (Balanced Panel)

	Dependent Variables	
	State ELA Test	State Math Test
Independent Variable	(1)	(2)
2-year lead	0.05	-0.04
	(0.04)	(0.04)
1-year lead	0.05	0.04
	(0.05)	(0.04)
First participated one summer prior	0.10**	0.00
	(0.05)	(0.04)
First participated two summers prior	0.07	-0.02
	(0.07)	(0.07)
p value (H_0: $\delta_2=\delta_1=0$)	0.46	0.07
p value (H_0: $\delta_0=\delta_{-1}$)	0.62	0.73

Note: The intent-to-treat sample consists of students enrolled in fifth grade during the 2015-16 school year. This analytic sample is a balanced panel of all students from that ITT sample with full information every year from grades 4 through 7, for school years 2014-15 through 2017-18; N=13,247 student-year observations, which includes 3,538 total unique students, 208 of whom were ever in Aim High. Estimates are derived from ordinary least squares multiple regression models. The dependent variables are z-scored 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. The reference category is 3 or 4 years prior to first Aim High participation or never participating in Aim High. Standard errors, clustered at the student level, are in parentheses *** p<0.01, ** p<0.05, * p<0.10.

Table D8. The Estimated Effect of Aim High on State English Language Arts Test Scores among Those First Participating in Aim High in the Summer Before Seventh Grade.

Independent Variable	Static (1)	Event Study (2)
After Aim High participation	0.08**	--
	(0.04)	
2-year lead	--	0.04
		(0.04)
1-year lead	--	0.05
		(0.04)
First participated one summer prior	--	0.11**
		(0.05)
p value (H_0: $\delta_2=\delta_1=0$)	--	0.55

Note: The intent-to-treat sample consists of students enrolled in fifth grade during the 2015-16 school year. This analytic sample is an unbalanced panel of all students from that ITT sample with full information in a given year, excluding the 64 students who begin participating in Aim High in the summer prior to sixth grade; N=14,597 student-year observations; N=4,257 unique students. The dependent variable is a z-scored continuous measure of English Language Arts score on the California Assessment of Student Performance and Progress (CAASPP) exam each year. Both models include student, school year grade level fixed effects and the following time-varying student-year controls: Special education, parent's highest education level, and foster care status. For the event study model, the reference category is 3 years prior to first Aim High participation or never participating in Aim High. Standard errors, clustered at the student level, are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.