

Accelerating Opportunity: The Effects of Instructionally Supported Detracking

Thomas S. Dee 

Graduate School of Education, Stanford University

Elizabeth Huffaker 

College of Education, University of Florida

The pivotal role of algebra in the educational trajectories of U.S. students continues to motivate high-profile policies focused on when students access the course, their peers, and how it is taught. This random-assignment partnership study examined an innovative district-level reform—the Algebra I Initiative—that placed ninth grade students with prior math scores below grade level into Algebra I classes coupled with teacher training instead of a remedial pre-algebra class. We found that this reform significantly increased grade 11 math achievement (extreme spread = 0.2 SD) without lowering the achievement of classroom peers. This initiative also increased attendance and district retention. These results suggest that higher expectations for the lowest-performing students coupled with aligned teacher supports is a promising model for realizing students’ mathematical potential.

KEYWORDS: ability grouping, acceleration, mathematics education, professional development

Introduction

High school mathematics attainment has significant consequences for postsecondary and labor-market outcomes (Altonji, 1995; Goodman, 2019; Kim, 2018; Long et al., 2012). Notably, Goodman (2019) estimated that an additional year of high school math completion improves future income by 10% for Black (but not White) students.¹ Improved access to lucrative science,

THOMAS S. DEE is the Barnett Family Professor in the Graduate School of Education at Stanford University (tdee@stanford.edu). His research focuses on the use of the quantitative methods to inform contemporary issues of public policy and practice.

ELIZABETH HUFFAKER is an assistant professor of educational leadership and policy at the University of Florida’s College of Education (ehuffaker@ufl.edu). She uses quantitative methods to study the efficacy of instructional interventions and the influence of K–12 math policies on students’ educational opportunities, engagement, and achievement.

technology, engineering, and mathematics (STEM) careers drives this effect and implies that greater development of latent human capital in minoritized youth through high school math coursework is a lever to reduce economic inequality. Yet, enrollment in advanced courses (i.e., those requiring completion of Algebra II) is starkly stratified by race, class, and ethnicity (Ayalon & Gamoran, 2000; Schiller & Hunt, 2011). Black, Hispanic, and poor students complete fewer college preparatory math classes than their White, Asian, and affluent peers (Conger et al., 2009). Fragmented math enrollment patterns also undermine school-level equity goals by exacerbating within-school ethnoracial segregation (Clotfelter et al., 2021; Dalane & Marcotte, 2020). Because stratification is driven primarily by within-school assignment practices (Antonovics et al., 2022; Betts, 2011; Clotfelter et al., 2021), *tracking*—the practice of sorting students on the basis of prior achievement or perceived ability—has long been critiqued as an inherently unequal method for distributing educational opportunities (e.g., Oakes, 2005).

Policy efforts to reduce these disparities have centered on the foundational high school math course: Algebra I. The well-established correlations between the accelerated take-up of algebra and improved outcomes such as stronger high school math test scores and course progressions (Gamoran & Hannigan, 2000; Stein et al., 2011) motivate this focus. Toward the end of the 20th century, the Algebra-for-All movement criticized Algebra I assignment practices for creating a bottleneck in student entry to rigorous math classes. Early Algebra I access was recognized as a concern grounded in fundamental fairness and educational civil rights (Moses & Cobb, 2001). Conversely, restricting early take-up of algebra was seen as gatekeeping (Gamoran & Hannigan, 2000; Oakes et al., 1990).

The Algebra-for-All movement dramatically increased the prevalence of acceleration into early Algebra I in the United States. For example, Chicago Public Schools eliminated prealgebra in ninth grade so that all freshmen would enroll in Algebra I or higher (Allensworth et al., 2009). In California, the share of eighth graders enrolled in Algebra I grew from 16% in 1999 to 65% in 2013 (McEachin et al., 2020). Many districts opted to accelerate middle schoolers into Algebra I—widely and even universally—to reduce academic stratification across secondary math courses. That is, *acceleration* became the predominant strategy to reduce the role of tracking in stratifying student access to advanced math courses. More recently, some districts have instead attempted to preempt the emergence of stratified math tracks through controversial policies that uniformly *delay* Algebra I access until high school (e.g., Huffaker et al., 2025). The Algebra I (A1) Initiative studied here is in the tradition of *detracking* interventions that accelerate access, but, critically, it does so with the addition of novel instructional supports.

A substantial body of quasi-experimental evidence indicates that acceleration into Algebra I academically benefits well-prepared students but carries negative consequences for lower-performing students (Allensworth et al.,

2009; Clotfelter et al., 2015; Dougherty et al., 2015; Heppen et al., 2011; Lafortune, 2018; McEachin et al., 2020). These results suggest some advantages to tracking students into different math courses using baseline achievement measures. Indeed, evaluations of *targeted* acceleration strategies (i.e., using test scores to assign or encourage students to take advanced coursework—a form of tracking) identify positive effects on educational attainment (Austin et al., 2024; Card & Giuliano, 2024; Dougherty et al., 2017). Complementary causal analyses have found that grouping students homogeneously by proficiency can carry benefits *across* the achievement distribution (Card & Giuliano, 2016; Cohodes, 2020; Collins & Gan, 2013; Cortes & Goodman, 2014; Duflo et al., 2011; Figlio & Page, 2002). A dominant explanation for these findings is that tracking facilitates efficiency in instructional targeting (e.g., Duflo et al., 2011). That is, achievement-based grouping lessens the technical burden of differentiation for classroom teachers by reducing within-classroom variation in student preparedness. In a tracked class, teaching to the middle approximates teaching at the speed of learning for a larger share of the class (Good et al., 1978). Detracking, by contrast, introduces greater diversity in within-classroom student needs and can exacerbate pedagogic challenges (Rosenbaum, 1999).

However, there are equity-oriented concerns about tracking students into homogeneous classrooms according to baseline achievement. For example, because prior educational opportunity is correlated with ethnoracial and socioeconomic status, tracking necessarily increases within-school segregation (Clotfelter et al., 2021; Conger, 2005). Where such stratification reduces the prevalence of diverse and inclusive classroom environments, it impedes the democratic and social aspirations for schooling (Brighthouse et al., 2018; Labaree, 1988). This dynamic also can create powerfully self-reinforcing and inequitable patterns in longer-run student engagement. When students lack role models with shared identities in rigorous courses, they are in turn less inclined to pursue advanced math classes (Francis & Darity, 2021). Similarly, Legette and Kurtz-Costes (2021) noted that even after controlling for baseline achievement, students' sense of belonging and motivation in school are negatively associated with assignment to a low-status track. Tracking also can amplify inequity by influencing teacher expectations and effectiveness. A body of qualitative evidence has indicated that teachers perceive students in lower-level tracks as having limited capacity for growth and accordingly reduce the rigor and richness of their pedagogy. Students in lower-level classes are also more likely to be taught by novice teachers (Kalogrides & Loeb, 2013).

In sum, the literatures on math acceleration and tracking surface vexing tensions for any effort that simultaneously seeks to support both mathematical excellence and broad opportunity. An effective solution would need to avoid the negative consequences of tracking students by prior achievement but still harness the academic benefits of appropriately differentiated instruction for

math learners. For example, evidence from Chicago has suggested that the targeted provision of additional instructional time (i.e., double-dose math), although expensive, protects low-performing students in algebra classes from the academic harm of acceleration (Nomi & Allensworth, 2009). This partnership study provides evidence on the student-level impacts of an innovative approach—the A1 Initiative—that bundles the acceleration of lower-performing students into algebra with a different, classroom-focused strategy: capacity building for high-quality, differentiated instruction.

The A1 Initiative presents a unique opportunity to study the impact of an intensive effort to improve instructional quality paired with targeted detracking. Specifically, the A1 Initiative reduced the prevalence of academic tracking within a diverse, medium-sized school district by reducing the number of math tracks for ninth graders deemed at or below grade level (i.e., two thirds of the entry cohort) from three pathways to one.² In its inaugural year, eligible ninth grade students were randomized into either the control condition (i.e., their conventional assignment to a remedial prealgebra class or to algebra, based on prior achievement) or to the treatment condition. The A1 Initiative classrooms therefore featured both heterogeneously grouped students and teachers who received unique professional development (e.g., on strategies for instructional differentiation) and additional resources (e.g., more planning time). Although we cannot disentangle the separate effects of each component, the bundle of interventions that comprises this novel program was piloted under random assignment. Specifically, we address the following:

1. **Preregistered confirmatory research question:** For each baseline achievement level, what is the impact of assignment to the A1 Initiative on a standardized test score?
2. **Exploratory research question:** For each baseline achievement level, what is the impact of assignment to the A1 Initiative on attendance, math course enrollment and attainment, and student enrollment status?

This study presents an opportunity for researchers and practitioners to better understand the promise of instructional improvement as a strategy for promoting both high expectations and inclusivity in math pathways.

The A1 Initiative: Program Features and Related Research

The A1 Initiative was a response to persistent math achievement disparities in an ethnoracially and socioeconomically diverse suburban school district in California's Bay Area (i.e., the Sequoia Union High School District). The district serves students across four comprehensive high school and three alternative high schools. At the time of the intervention, roughly 40% of its students were socioeconomically disadvantaged, 13.5% were classified as English learners, and 43% identified as Hispanic, 8% as Asian, and 39% as White.

2018-19 Math Course	9 th Grade Placement Based on Highest Assessment Score			2019-20 Math Placement	
8 th Grade Math Course	Smarter Balanced (SBAC)	Math Diagnostic Testing Project (MDTP)	Let's Go Learn Diagnostic Online Math Assessment (DOMA)	9 th Grade Math Placement & Support	
8 th Grade Common Core	Grade 7 SBAC Level/Scale Score	High School Math Readiness	¹ Pre-Algebra	Math Course	Support
	Met / Exceeded 2567 to 2778	18 to 45	10 to 14	Algebra I	No
	Nearly Met 2484 to 2566	13 to 17	5 to 9.5	Algebra I	Yes
	Not Met 2250 to 2483	0 to 12	0 to 4.5	³ Algebra Readiness 1 st Sem	Block of 2 periods
Algebra I	Grade 7 SBAC	Geometry Readiness	² Algebra I	Math Course	Support
	Met / Exceeded 2567 to 2778	24 to 45	7.5 to 11	Geometry	No
	Nearly Met 2484 to 2566	14 to 23	5.5 to 7	Algebra I	No
	Not Met 2250 to 2483	0 to 13	0 to 5	Algebra I	Yes
				³ Algebra Topics 2 nd Sem	Yes

Figure 1. Placement chart for rising ninth graders who took Common Core Math-8 in eighth grade (Sequoia Union High School District, 2018).

Although the district performed above state averages on English language arts and mathematics assessments in years preceding the A1 Initiative, scores were highly stratified by race, ethnicity, and residential ZIP Code. Specifically, more than three quarters of White and Asian students in the graduating classes of 2017 through 2020 met University of California admissions criteria in math compared with fewer than half of Black and Hispanic students.

The district identified ninth grade math assignment practices as a potential driver of these disparities. This inference is consistent with the evidence on the pivotal role ninth grade plays in shaping student trajectories (e.g, Krone Phillips, 2019). Prior to the A1 Initiative, the district implemented the California Math Placement Act of 2015 using the placement scheme depicted in Figure 1. The act required districts to use transparent and nondiscretionary methods for math course assignment, reducing the reliance on teacher or counselor recommendations, which may be subject to implicit bias (Grissom & Redding, 2016). Specifically, the district sorted students across freshman math courses using middle school transcripts and test scores. Because incoming Black and Hispanic freshmen were disproportionately likely to be classified as having below-grade-level proficiency, this practice resulted in ethnoracially stratified math enrollment. The A1 Initiative aimed

to redress disparities without reducing achievement for any student group using two key strategies. It combined all treatment-assigned students entering high school at or below grade level into Algebra I classes (i.e., acceleration). Critically, the district also provided targeted professional support and development for teachers of these classes to implement new pedagogic approaches.

Detracking and Heterogeneous Classrooms

By detracking students at or below grade level, the A1 Initiative created math classes with greater heterogeneity in baseline proficiency. Nearly two thirds of the freshman class were eligible to be placed in an A1 Initiative section.³ The remaining third, who entered high school above grade level,⁴ were excluded from the randomization and, therefore, our analytic sample. Instead, they were placed in geometry or higher per the business-as-usual practices summarized in Figure 1. The A1 Initiative thus can be accurately characterized as a *partial* detracking policy. For those randomly assigned to treatment, the A1 Initiative collapsed the number of ninth grade math pathways from three to one. Students who otherwise would have been assigned, based on prior achievement, to “Algebra Readiness” (i.e., remedial prealgebra), “Algebra I with Support” (i.e., a double-dose option), or “Algebra I” (i.e., the standard grade-level track) were combined into A1 Initiative sections. As a result, assignment to an A1 Initiative class exposed students to a different mix of peers than assignment to a business-as-usual section.

These issues suggest that the per se impact of A1 Initiative classes could, in theory, be confounded by other distinctive classroom traits (e.g., teachers and student peers). Discussions with our district partners indicated that there was not a consistently purposive pattern of teacher assignment to A1 Initiative classrooms. However, we explore these issues empirically in Table 1 and Appendix Table A7 in the online version of the journal by reporting the results of auxiliary regressions that examine the impact of the intent to treat (ITT)—random assignment to an A1 Initiative classroom—on measures of students’ classrooms.⁴ These results indicate that the treatment teachers were no more experienced than control-group teachers and that they also were generally similar with regard to the probability of having an advanced degree or national board certification.⁵ For below-grade-level students, being assigned to the A1 Initiative rather than Algebra Readiness does increase the likelihood that students have a Hispanic or Asian or Pacific Islander teacher but decreases the probability they have a White teacher or a same-race teacher (see Supplementary Table A9 in the online version of the journal).⁶ Given the positive relationship observed between student–teacher racial congruence and academic achievement in prior studies (Egalite et al., 2015; Redding, 2019), if anything, this would downwardly bias achievement for students assigned to treatment. Otherwise, teacher demographic traits—including age, gender, and gender match—are broadly equivalent across conditions.

Table 1
Intent-to-treat (ITT) effects of the A1 Initiative on peer and teacher traits

	Peer characteristics							Classroom characteristics			
	% Female (1)	% FRPL (2)	% English learner (3)	% Hispanic (4)	% White (5)	% Asian/Pacific Islander (6)	% Black (7)	Baseline math achievement (8)	Teacher has advanced certification (9)	Years of teacher experience (10)	Class size (11)
Achievement stratum											
ITT × at grade level	-2.5126 (2.0200)	-7.4453** (3.1600)	2.3541 (2.0513)	-1.3714 (3.1055)	3.3942 (3.1711)	-1.8813 (2.5362)	-0.5955 (0.8037)	0.0568 (0.0865)	-0.2644** (0.1196)	-2.6740 (2.1168)	0.3865 (1.0994)
ITT × nearly at grade level	-3.1418 (1.9085)	13.5829*** (2.7789)	-2.5070 (3.0760)	-5.7349* (3.0393)	6.2620** (2.6978)	0.7096 (1.9856)	1.7834* (1.0654)	0.2989*** (0.0851)	-0.0416 (0.0399)	0.9208 (1.5760)	0.2590 (0.7865)
ITT × below grade level	-1.0962 (4.5047)	29.8969*** (3.2444)	34.2722*** (3.6244)	16.3469*** (3.3883)	22.2065*** (2.6553)	-2.6447 (2.1867)	-3.9925 (2.4196)	1.2913*** (0.0669)	0.0322 (0.0499)	0.3685 (1.2256)	6.3366*** (1.7647)
Nearly at grade level	1.0339 (4.4525)	15.1343*** (2.8475)	32.3297*** (4.4692)	-9.7896*** (3.3666)	14.7169*** (2.6548)	-2.5017 (2.0874)	-2.7541 (2.4420)	1.0173*** (0.0825)	0.0644 (0.0398)	-0.5274 (1.5007)	6.6709*** (1.7591)
Below grade level	0.8577 (4.5549)	21.6892*** (3.8452)	38.5134*** (3.7370)	15.4317*** (3.2765)	19.1021*** (2.7929)	-0.1905 (2.6709)	-3.7252 (2.5354)	1.3085*** (0.0991)	0.2303** (0.0974)	2.9618 (2.0187)	6.8426*** (1.8657)
p Value ($H_0: \beta_1 = \beta_2 = \beta_3$)	.4453	.0000	.0000	.0001	.0000	.2130	.2232	.0000	.1207	.1257	.0080
Control mean if below grade level	49.5004	78.5114	50.5284	72.6950	7.9259	10.0512	9.1790	-1.2214	0.9375	9.5312	19.7041
Observations	1,039	1,039	1,039	1,039	1,039	1,039	1,039	1,039	1,035	1,035	1,039

FRPL, free or reduced-price lunch

Notes: This table summarizes the effect of assignment to the A1 Initiative on peer and teacher characteristics. Columns (1) through (8) present the percentage point difference in classroom concentration of baseline student traits between treatment and control conditions for each achievement stratum. Classroom-level demographic shares use data from a total of 1,347 students from the ITT sample and their classroom peers not in the ITT sample. This includes 94 students in grades 10–12. The measure of math achievement in column (8) is constructed by standardizing high school math readiness assessment (Mathematics Diagnostic Testing Project) scores across the 1,269 students for whom it was available. Columns (9) presents the extent to which assignment to the A1 Initiative changes a student's likelihood of receiving a teacher with a graduate degree and/or national board certification, whereas column (10) summarizes the effect of A1 Initiative assignment on average maximum years of teacher experience. The conditional control mean presented was estimated across control observations in the below-grade-level stratum. All models control for ninth grade campus membership. Standard errors are clustered at the ninth grade classroom level.

* $p < .1$; ** $p < .05$; *** $p < .01$; **** $p < .001$.

We also found that the ITT was unrelated to class size except for the students who otherwise would have been in remedial Algebra Readiness sections (see Table 1). Because these remedial classes tended to have fewer students, random assignment to an A1 Initiative section implied a class size that was larger by six students. Prior research has suggested that an increase in class size of this magnitude reduces student engagement (Dee & West, 2011) and test scores (Krueger, 2003). As such, our findings may *understate* the impacts of the A1 Initiative for this key group of students relative to what we would observe if class sizes were held constant.

Assignment to an A1 Initiative classroom had more substantive effects on the traits of classroom peers, as one would expect with a detracking reform. Specifically, lower-achieving (below-grade-level) students saw a substantial (i.e., 1.29 SD) increase in the baseline math achievement of their peers. Furthermore, the share of their peers classified as poor or English learners declined by 30 and 34 percentage points, respectively (see Table 1). Nearly-at-grade-level students, those assigned to the A1 Initiative, experienced modest declines in peer economic disadvantage (i.e., -10 percentage points) and increases in peer achievement (i.e., 0.3 SD). Composition changes for higher-achieving students—the majority share of each class—were not statistically significant.

We note that these changes in peer attributes, although substantial, likely contributed little to the academic gains we found among below-grade-level students. The measured impacts of peer achievement on student academic outcomes are generally small (Angrist, 2014; Lefgren, 2004; Sacerdote, 2014). Feld and Zölitz (2017) estimated that a 1 SD increase in peer grade-point average translates, on average, to a 0.0126 improvement in student grades. This implies that assignment to the A1 Initiative would boost below-grade-level academic achievement by 0.016 SD solely through a peer-effects channel, an impact that is small relative to the main confirmatory findings we report. Furthermore, this is likely an upper bound on the benefits of sharing a classroom with increasing shares of high-skilled peers. Both Feld and Zölitz (2017) and Burke and Sass (2013) uncovered peer effect heterogeneity, suggesting that less skilled students may be harmed by a very low position in the class distribution. Peer effects thus are directionally ambiguous for accelerated students and likely depend on teacher capacity to promote positive peer interactions and growth mindsets.

The impact of being a relatively high- versus low-achieving student relative to one's peers is also subject to countervailing theorized effects on self-perception. On the one hand, the “big fish, little pond” effect may induce greater engagement through more positive self-perception among students with high relative classroom positions (Malamud et al., 2025). For example, a student who just barely fails to qualify for Algebra I and is instead assigned to Algebra Readiness in the control condition may simultaneously feel more positively about themselves as a math learner when they observe the lower

proficiency of their peers relative to a student who barely qualifies for algebra. On the other hand, the self-perception of a marginal Algebra Readiness attendee is reduced if simply being in a remedial pathway introduces a negative contrast with Algebra I students. International comparative work has suggested that the big fish, small pond effect dominates when tracking is more strict, but stigma by course contrast is more influential in the course-by-course assignment system of the United States (Chmielewski et al., 2013).

Altogether, these studies support Nomi and Allensworth's (2013) observation that a change in achievement within a classroom can alter assigned course grade to the detriment of lower achievers. Furthermore, the salience of peer achievement depends on interaction quality between students at different levels of achievement, and students tend to sort themselves by proficiency within classrooms (Feld & Zölitz, 2017; Kang, 2007; Murata, 2013). These dynamics suggest a major role for pedagogic practice (i.e., to facilitate collaborative learning across skill levels) in unlocking the benefits of having high-achieving peers for less prepared students.

A1 Initiative Pedagogy and Resources

The A1 Initiative pedagogy aimed to cultivate broadly effective Algebra I instruction within the specific context of these heterogeneous classrooms. This primarily requires that teachers engage in a continuous process of differentiated evaluation, reflection, and adaptive instruction (Valiande & Koutselini, 2009). From interviews with district administrators, we identified three main ways the A1 Initiative promoted mastery of this approach. First, teachers were trained in specific instructional strategies such as math language routines to foster academic conversation and permit frequent assessments of student comprehension (Zwiers et al., 2017). Instructional leaders emphasized the importance of hearing and seeing student reasoning in A1 Initiative classes. Second, to facilitate responsive pacing, teachers were afforded flexibility in executing the Algebra I curriculum (Rosenbaum, 1999). The A1 Initiative cohort collaborated on unit planning and assessments. Teachers received optional lesson-planning resources but otherwise retained autonomy over day-to-day activities.⁷ By contrast, Algebra I control condition teachers were guided by structured district-defined pacing recommendations tied to a single textbook. Third, the A1 Initiative cohort strongly emphasized high expectations and growth mindsets for students and teachers. This encouraged all students to proceed on the grade-level track to geometry.

To facilitate this approach, teachers received significant resources coordinated by the district as well as support from a math education consultancy. In addition to ~15 full days of professional development, they received an additional class-section release (i.e., equivalent to five additional planning periods per week), four coaching days per site per semester, a districtwide professional learning community, and a partner teacher at their campus.⁸ They

also participated in lesson studies to share and learn promising practices within the cohort. The A1 Initiative introduced considerable resources and training to promote high-quality pedagogy in heterogeneous classrooms.

Effective instruction in heterogeneous classrooms entails complex teacher practices (Prast et al., 2015). In this context, teachers must adapt instruction (i.e., iteratively assess and adjust teaching to each student's zone of proximal development; Vygotsky, 2011) across classmates with very different academic needs. The extent to which professional training can develop teachers' ability to meet this substantial pedagogic challenge is uncertain. In particular, the literature on this topic is limited in two main ways. First, the conceptualization of instructional differentiation is often imprecise. Second, reviews of the literature acknowledge that the empirical base is limited—few high-quality causal studies have measured the direct impacts of differentiation training on student outcomes (Graham et al., 2021).

On the first limitation, some studies define differentiation simply as ability grouping within classes, whereas others include the use of technology to tailor instruction and responsiveness to interim data. Prast et al. (2015) proposed a more general cycle of differentiation that professional development ought to promote. It occurs in five stages: identifying educational needs, setting differentiated goals, providing differentiated instruction, implementing differentiated practice, and evaluating progress. Conversations with our district partners revealed that A1 Initiative professional development emphasized many of these stages explicitly, including training teachers on strategies to frequently assess and respond adaptively to student needs. Vogt and Rogalla (2009) found that an intervention designed to help teachers improve adaptive planning competency aligned with this cycle and resulted in small gains in mathematics achievement (i.e., ~ 0.04 SD). This would explain slightly under one quarter of the effect of the A1 Initiative for below-grade-level students.

Turning to related research that measured the effects of training on teacher practice (although not on student outcomes), Dixon et al. (2014) found that more hours of differentiated instruction per day are positively associated with teacher self-efficacy, which, in turn, is associated with greater levels of instructional differentiation. Using an experimental design, VanTassel-Baska et al. (2008) estimated that 4 days of training in differentiated teaching strategies led to statistically significant increases in the use of these strategies and improvements in overall teacher effectiveness ratings.

Beyond singular training interventions, the results of an observational study by Goddard et al. (2019) suggested that overall strong school-level instructional leadership promotes instructional differentiation and, as a result, student achievement. Specifically, using multilevel modeling techniques, they found that a 1 SD increase in survey-assessed instructional leadership quality was associated with 0.07 SD higher math scores through a differentiated instruction pathway. The A1 Initiative is a test case of this finding because leadership worked closely with teachers and external consultants to

Table 2
Summary of treatment–control contrasts by baseline achievement group

Student group by baseline achievement	Control	Treatment
At grade level	Regular Algebra I (n = 304)	Algebra I with A1 Initiative pedagogy and mixed-achievement peers (n = 366)
Nearly at grade level	Regular Algebra I + support class (n = 121)	Algebra I with A1 Initiative pedagogy and mixed-achievement peers (n = 83)
Below grade level	Algebra Readiness (n = 98)	Algebra I with A1 Initiative pedagogy and mixed-achievement peers (n = 67)

Note: Baseline achievement groups are determined according to Figure 1 using middle school test scores and transcripts.

coordinate activities aimed to develop instructional skill in heterogeneous classrooms.

Features of the Control Conditions

We also note that because students were sorted across three groups in the business-as-usual control condition, the treatment–control contrasts we study (summarized in Table 2) are differentiated by baseline achievement.⁹ Specifically, students *not* in the A1 Initiative were assigned to ninth grade math courses conditional on their eighth grade class (i.e., Common Core Math 8) and their best score from three assessments: the seventh grade Smarter Balanced Assessment Consortium (SBAC) state test in California and two separate diagnostic tests¹⁰ (Figure 1). Teachers and parents/guardians had influence to “level up” student assignment from the objective placement (e.g., parents of a student whose test scores made them eligible for Algebra I with support could opt them out of that support class). Fall 2019 transcript data show that 90% of control group students enrolled in the ninth grade math course indicated by their middle school course taking and test scores.

Below-grade-level students *not* assigned to the A1 Initiative were enrolled in a prealgebra remedial course called *Algebra Readiness*. This course is slower paced and contains less advanced material than Algebra I. It also does not count toward either district or University of California–California State University (UC-CSU) math requirements. These students cannot enroll in Algebra I until after ninth grade. Assignment to the A1 Initiative allowed earlier access to high-school-level coursework among this group, which meaningfully changes the structure of future math opportunities for these students relative to the control condition. Control group students nearly at

grade-level proficiency enrolled in Algebra I and a second (also *not* UC-CSU aligned) block of math instruction called *Support*. Students in this category therefore received a double dose of math. Although double-dose math has had demonstrably positive impacts on student achievement in other contexts (Cortes & Goodman, 2014; Nomi & Allensworth, 2009), it is costly to staff. Administrators also worried that the course crowded students out from taking electives. The treatment–control contrast is smallest among at-grade-level students, for whom instructional time and potential course credits are identical across conditions.

Data and Sample

We used administrative data from the district to examine the effects of the A1 Initiative on an array of student outcomes. We observe enrollment, transcripts, test scores, and attendance for students in the pilot cohort from ninth grade (academic year [AY] 2019–20) through twelfth grade (AY 2022–23). Twelfth-grade data also include an exit indicator for on-time graduation status.

Our analytic sample included 1,039 students from a cohort of 2,124 ninth graders who entered any of the district’s four comprehensive high schools in the fall of 2019. We excluded students enrolled in the district’s small alternative schools because these campuses did not use random assignment for freshmen course placement. We did, however, follow the high school trajectory of students who initially enrolled at a comprehensive high school and then transferred to an alternative campus.

From this cohort, we refined the sample until it was composed only of students included in the randomization. This encompassed nearly all students eligible for either Algebra I or Algebra Readiness (Figure 1). Our ITT population excluded ninth graders enrolled in math classes beyond Algebra I (e.g., Geometry and Algebra II; $n = 760$), students with Individualized Education Programs enrolled in either no math course or in a basic math skills class ($n = 31$), and those enrolled in Algebra I courses designated for newcomer English learners ($n = 77$). Incumbent English learners were enrolled in standard math classes and were retained in our sample. Students who took Algebra I in eighth grade but had to repeat it in ninth grade also were excluded ($n = 29$). Two further freshmen were dropped because they enrolled late in the fall semester, after randomization had occurred. Finally, we eliminated observations for students whose baseline proficiency under the business-as-usual placement scheme (Figure 1) could not be verified. These are students missing baseline data because they did not attend a feeder district in AY 2018–19 ($n = 145$) or whose eighth grade math courses were not included in the assignment table ($n = 41$).

Measures

We examined both academic (i.e., assessment and course progression) and nonacademic (i.e., attendance) student outcomes. Math proficiency, our focal outcome, was measured using multiple sources of assessment data. Of these, we privileged the SBAC assessment administered under the California Assessment of Student Performance and Progress system in spring 2021, during our pilot cohort's eleventh grade year. Scale scores were standardized using the published statewide mean and standard deviation (California Department of Education, 2023). Secondary test metrics included the fall 2020 and fall 2021 results of an interim comprehensive assessment (ICA) created and administered by the district.¹¹ We apply a hybrid item response theory model (see the Supplementary Technical and Data Appendix in the online version of the journal for technical details) to item-level response data to create a standardized measure of achievement across the analytic sample for each ICA (StataCorp, 2023).

Exploratory outcomes including course enrollment and attainment were parsed from transcript data that denoted, for each enrollment, the course name, a teacher identification number, credit type (e.g., math, English language arts, or elective), credits attempted (one semester of a core course is equivalent to 5 credits), and any credits or letter grade received. Due to the disruption of the COVID-19 pandemic, grading was on a credit or no-credit basis in the spring of 2020. Summative measures were constructed from longitudinal ninth through twelfth grade transcript data. These variables included binary indicators of whether a student ever enrolled in and/or earned credit in a course (e.g., Algebra II) and whether a student achieved math subject eligibility for entry to the University of California system.¹² Additional details on course classification are in the Supplementary Technical and Data Appendix in the online version of the journal. We also calculated total credit attainment—by subject and overall—through twelfth grade. Finally, grade-by-grade indicators of course enrollment and completion (i.e., earning at least 10 units, equivalent to two semesters of credit) were used to compare the *pace* of course progression across the treatment and control groups.

We used annual indicators of district enrollment and attendance to measure student engagement. Absence rates were constructed by dividing days reported absent by total enrollment days in each year. We also coded a binary indicator for chronic absenteeism that took on a value of one if a student's absent rate exceeded 10%. Chronic absenteeism is less sensitive to distortion from outliers than absence rate and is used by California as a primary indicator of student connectedness. However, impacts on the continuous absence measure can be more precisely estimated. We therefore present exploratory results using both measures. Enrollment files described student status at the beginning and end of each academic year and denoted timing plus explanatory codes for entry or exit events (see the Supplementary Technical and Data

Appendix in the online version of the journal for more detail on treatment of student exits). These data allowed us to explore student persistence in the district. District administrators expressed confidence in the fidelity of attendance and enrollment data, citing care taken during the COVID-19 pandemic to maintain accurate records.

Sample Description

Table 3 reports summary statistics for the ITT sample of 1,039 students on ninth grade entry. These administrative categories are not necessarily reflective of the full diversity of student identities. All students are listed under a single racial/ethnic category, which does not include a multiracial indicator. In the fall of 2019, about half the students in our sample were eligible for free or reduced-price lunch (FRPL), and 16% were classified as being English learners (ELs). A slight majority was identified as Hispanic, nearly a third as White, 12% as Asian or Pacific Islander, 4% as Black, and 1% as Native American or Alaska Native. Eligible students were dispersed relatively evenly across the four comprehensive high schools in the district, and exactly half were randomly assigned to the A1 Initiative. Following Figure 1, 64% of the sample entered ninth grade at grade level for math, whereas 20% were nearly at grade level and 16% were below grade level. District enrollment declined by 3–4 percentage points within the sample year, from 98% retention through ninth grade to 87% through twelfth grade. We examined the effect of the ITT on student retention because it is both an internal-validity concern and a substantively relevant outcome.

Turning to the key dependent variables, students in the analytic sample scored 0.12 SD above the statewide mean of the eleventh grade math SBAC assessment. Although most passed their ninth grade math class and received credit for two Algebra I semesters by the end of tenth grade, only 62% of students completed a full year of geometry by the end of AY 2020–21. By the end of eleventh grade, 79% of students had received two semesters of geometry credit, and just over half the sample had completed Algebra II. This increased to just >60% after first-time twelfth grade attendance. Similarly, among the 4-year sample including dropouts, half the remaining students completed at least one advanced math course, and 62% met the UC-CSU “C” admission requirement for math. Students earned an average of 33 credits (i.e., they completed just over three courses) during high school. Of the 4-year sample, 90% graduated from high school within that time span. Finally, we observed a decline in attendance-related outcomes over the high school years. Chronic absenteeism rates were 14% in AY 2020–21, 25% in AY 2021–22, and 29% in AY 2022–23 compared with only 4% in AY 2019–20, a pattern of postpandemic growth consistent with state and national data (Dee, 2024). The underlying absence rate increased from 3–5 to nearly 10% as students progressed from ninth through twelfth grade.

Table 3
Descriptive statistics

Variable	Mean	SD	Minimum	Maximum	N
Student characteristics at high school entry					
Female	0.51	0.50	0	1	1,039
Free and reduced-price meals	0.49	0.50	0	1	1,039
English learner	0.16	0.36	0	1	1,039
White	0.30	0.46	0	1	1,039
Hispanic	0.54	0.50	0	1	1,039
Asian or Pacific Islander	0.12	0.32	0	1	1,039
Black	0.04	0.19	0	1	1,039
Native American or Alaska Native	0.01	0.09	0	1	1,039
Attended campus 1 in grade 9	0.28	0.45	0	1	1,039
Attended campus 2 in grade 9	0.24	0.43	0	1	1,039
Attended campus 3 in grade 9	0.25	0.43	0	1	1,039
Attended campus 4 in grade 9	0.22	0.42	0	1	1,039
Assigned placement					
At grade level	0.64	0.48	0	1	1,039
Nearly at grade level	0.16	0.37	0	1	1,039
Not at grade level	0.20	0.40	0	1	1,039
Math test scores (standardized)					
Baseline readiness score	22.00	9.84	0	45	1,039
Tenth grade interim comprehensive assessment	0.00	0.85	-1.80	2.22	850
Eleventh grade state assessment (Smarter Balanced Assessment Consortium test)	0.12	0.84	-1.91	2.32	805
Transcript outcomes					
Completed Algebra I by end of grade 10	0.99	0.09	0	1	908
Completed geometry by end of grade 11	0.63	0.48	0	1	985
Completed geometry by end of grade 11	0.79	0.41	0	1	936
Completed Algebra II by end of grade 11	0.55	0.50	0	1	940
Completed Algebra II by end of grade 12	0.60	0.49	0	1	920
Completed any advanced math by end of grade 12	0.49	0.50	0	1	920
Met math requirements for University of California–California State University	0.61	0.49	0	1	920
Total math credits	32.58	10.57	0	60	920
Graduated on time	0.91	0.29	0	1	920
Attendance and enrollment outcomes					
Chronically absent in grade 9	0.04	0.21	0	1	1,024
Chronically absent in grade 10	0.14	0.34	0	1	984
Chronically absent in grade 11	0.25	0.43	0	1	936
Chronically absent in grade 12	0.29	0.45	0	1	916
Absence rate in grade 9 (%)	2.87	4.27	0	50.28	1,024
Absence rate in grade 10 (%)	5.48	12.36	0	85.56	984
Absence rate in grade 11 (%)	8.72	11.35	0	90.00	936
Absence rate in grade 12 (%)	9.42	11.62	0	90.50	916
Completed grade 9 in district	0.98	0.13	0	1	1,039
Completed grade 10 in district	0.94	0.23	0	1	1,039
Completed grade 11 in district	0.90	0.30	0	1	1,039
Completed grade 12 in district	0.87	0.33	0	1	1,039

Notes: Our analytic sample included 1,039 students who were enrolled in an algebra class for the first semester of ninth grade in one of the four comprehensive high schools in the partner district and who attended a feeder elementary district so that they could be matched with their middle school academic records. All data were sourced from the partner districts or one of the feeder elementary districts.

Estimation Strategy

Our analysis estimated the causal effects of the A1 Initiative by leveraging the random assignment of eligible students into treatment. The strong internal validity of our estimates reflects both institutional information and empirical

evidence consistent with successful randomization. From interviews, we knew that randomization had been conducted by campus-level administrators during the course-assignment process preceding the fall 2019 academic term. Students at three of the four campuses were individually randomized between the A1 Initiative and the business-as-usual condition. They took varied approaches to apportion the mix of students from each placement-level stratum. For example, one school stopped assigning students with below-grade-level achievement to A1 Initiative sections after reaching a 15% threshold. At other campuses, the baseline achievement composition of A1 Initiative classes matched the ITT student distribution. At the only campus that did not randomize at the student level, an administrator used a two-stage process. First, business-as-usual Algebra I and Algebra Readiness sections were created following the district's procedure to balance classes by demographic traits. Second, a randomly selected half of these sections were dissolved and recombined as A1 Initiative classes. We provide empirical evidence on this randomization through inspection of the balance in pretreatment student traits across conditions conditional on campus membership and baseline achievement group. Table 4 presents the results of these tests for individual characteristics and jointly using seemingly unrelated regression. There was no evidence of systematic imbalance.

We estimated ITT effects by baseline achievement group. We did this because the treatment–control contrast and probability of randomization into treatment were both dependent on students' baseline achievement group (see Figure 1). Our preferred specification was

$$y_{ics} = \beta_1(ITT \times atgradelevel)_i + \beta_2(ITT \times nearly)_i + \beta_3(ITT \times below)_i + \beta_4nearly_i + \beta_5below_i + \alpha_s + \varepsilon_{ics} \quad (1)$$

where y_{ics} is an outcome for student i who, in ninth grade, was in class c at school s . Following Institute of Education Sciences (IES) guidance presented in the What Works Clearinghouse (2022) *Procedures and Standards Handbook*, we distinguished between primary, preregistered outcome, for which this research is confirmatory (i.e., standardized assessment scores), and exploratory measures (e.g., credit accumulation and attendance), which provide only suggestive explanations for observed effects. The parameters β_1 , β_2 , and β_3 estimate the impact of assignment to the A1 Initiative for each achievement group. To guide interpretation of the estimand, we note that while β_1 , β_2 , and β_3 technically represent the *intent-to-treat* effects of assignment to the A1 Initiative, compliance with this assignment was high. By the end of the fall 2019 semester, 99% remained in their assigned condition, and 97% remained through the spring of 2020. All regressions included indicator variables for ninth grade campus membership and baseline placement level, with below-grade-level students in the control condition being the reference category. Standard errors are clustered by class section in our preferred model.

Table 4
Balance in student characteristics by intent-to-treat (ITT) × placement group

	Student characteristics							
	Female	FRPL	English learner	Baseline math achievement (standardized)	Hispanic	White	Asian	Black
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Achievement stratum								
ITT × at grade level	-0.0050 (0.0437)	-0.0732 (0.0379)	-0.0231 (0.0276)	0.0540 (0.0503)	-0.0229 (0.0382)	0.0293 (0.0371)	-0.0111 (0.0270)	-0.0047 (0.0159)
ITT × nearly at grade level	-0.0832 (0.0713)	-0.0146 (0.0618)	0.1019* (0.0450)	0.0200 (0.0820)	0.0241 (0.0623)	0.0423 (0.0606)	-0.0185 (0.0441)	-0.0545* (0.0260)
ITT × below grade level	-0.0705 (0.0790)	-0.0260 (0.0685)	-0.1150* (0.0498)	0.1131 (0.0908)	-0.0209 (0.0690)	0.0388 (0.0671)	-0.0453 (0.0488)	0.0388 (0.0288)
At grade level	-0.0818 (0.0620)	-0.3730*** (0.0538)	-0.4824*** (0.0391)	1.9299*** (0.0713)	-0.2352*** (0.0542)	0.2913*** (0.0527)	-0.0023 (0.0383)	-0.0418 (0.0226)
Nearly at grade level	-0.0412 (0.0677)	-0.0874 (0.0587)	-0.3124*** (0.0427)	0.5243*** (0.0779)	0.0284 (0.0592)	0.0442 (0.0575)	-0.0495 (0.0419)	-0.0190 (0.0247)
Control mean if below grade level	0.5612	0.8367	0.5510	-1.3923	0.7449	0.0510	0.1122	0.0816

FRPL, free or reduced-price lunch
Notes: This table shows the balance of characteristics of students in the randomized sample. Each column presents coefficients from a regression of a baseline characteristic on interactions between assignment to treatment and placement level at baseline, controlling for both campus membership and baseline achievement group. The *p* value of a joint-significance test for all 24 coefficients of interest is .6553, indicating no systematic imbalance. **p* < .05; ***p* < 0.01; ****p* < .001 (N = 1,039).

This error selection attends to evolving guidance from recent conceptual and econometric observations on the specification of standard errors for causal inference. A recent study by Abadie et al. (2023) encouraged researchers to consider a design-based rationale for error selection. In the case of clustered sampling and randomization, heterogeneity-robust standard errors can be too small, but in other cases (e.g., when random sampling and assignment occur predominantly at the unit level), clustering is overly conservative. In this case, randomization was largely the unit level in a procedural sense, but treatment status was almost perfectly correlated within classrooms. Therefore, we clustered standard errors by the ninth grade classroom membership in our main analysis.

In additional analysis, we investigated the sensitivity of our results to alternative empirical specifications and across teacher- and student-defined subgroups. This included exploring the implications of alternative standard error specifications on our main test score outcome in several ways. First, we present less conservative heterogeneity-robust Eicker–Huber–White standard errors. Second, we adjust standard errors for the relatively small number of class sections in our sample, following recommendations from Pustejovsky and Tipton (2018) for $N < 50$ clusters. For each of β_1 , β_2 , and β_3 , effective counts of clusters ranged from 38 to 47.

Third, we followed the encouragement of econometricians (e.g., as in Abadie et al., 2020) to consider an alternative to frequentist standard errors. As with many randomized, controlled trials, we used a convenience sample and observed all relevant units. Under the conventional framework, standard errors conferred uncertainty in our estimated parameters relative to the true values that would exist in a hypothetical “superpopulation.” However, the source of uncertainty in our estimates came not from random drawing but from the random assignment of treatment itself. Standard errors then were derived by simulating permutations of the treatment indicator. We used randomization inference testing (*ritest*; Hefß, 2017) to execute the randomization inference procedure.

Sensitivity and Robustness Checks

We also report ITT effects on our main outcomes after the addition of controls for student- and teacher-level characteristics. In Equation (2), \mathbf{X}_{ics} is a vector of student traits that includes race/ethnicity, gender, EL, FRPL status, and sometimes a middle school test score. \mathbf{T}_{cs} includes indicators of two observable teacher characteristics: years of experience and advanced certification status (e.g., a postgraduate degree or national board certification). Equation (1) results are privileged in our main analysis over Equation (2) estimates due to superior performance under Akaike information criterion cross-validation for most outcomes.

$$y_{ics} = \beta_1(ITT \times atgradelevel)_i + \beta_2(ITT \times nearly)_i + \beta_3(ITT \times below)_i + \beta_4nearly_i + \beta_5below_i + \gamma\mathbf{X}_{ics} + \delta\mathbf{T}_{cs} + \alpha_s + \varepsilon_{ics} \quad (2)$$

The inclusion of \mathbf{T}_{cs} in some specifications is one component of our strategy to address an implementation detail that warrants further discussion. That is, because *most* of the A1 Initiative teachers volunteered to join the pilot, it is possible that our analysis is detecting the efficacy of the eight A1 Initiative teachers rather than the impact of the program itself. It is possible that these teachers are particularly enthusiastic and committed instructors. As discussed earlier, we controlled for observable teacher traits that correlate with teacher quality (e.g., Harris & Sass, 2011) in the specification of Equation (2). Additionally, we tested the sensitivity of our findings to a leave-one-out exercise. Under this test, we repeated our main analysis eight times, in turn limiting the data to exclude students enrolled in each teacher's ninth grade class. This check indicated that the results presented below are not due to the influence of any one teacher leading the A1 Initiative classes (see Supplementary Appendix Table A4 in the online version of the journal).

Heterogeneity Analysis

We also produced exploratory estimates of our main effect by gender and economic disadvantage, as measured by FRPL status. Our interest in the presence of heterogeneous impacts by gender was motivated by a notable finding from multiple high-quality studies from the targeted acceleration literature that attainment effects are limited to only female students (Card & Giuliano, 2024; Dougherty et al., 2017). We were constrained from estimating results by ethnic or racial subgroup due to sample size constraints. However, we note that supportive acceleration studies (e.g., of double-dose Algebra I) that report impacts by subgroup (e.g., Cortes et al., 2015) have broadly failed to identify differing impacts by race and ethnicity.

Cost Analysis

Finally, where program impacts were detected, we used a program cost analysis to construct cost-effectiveness ratios. We favored this approach to characterizing the balance of program inputs and returns for three primary reasons. First, benefit-cost analyses (e.g., return-on-investment calculations) impose stronger assumptions because they rely on conjectural estimates on how to monetize the value of educational gains throughout the life cycle and how to place these longitudinal gains on a present-value footing. Second, cost-effectiveness comparisons better approximate the decision faced by many educational practitioners (i.e., how to best use available resources). Third, they facilitate ready comparisons at a policy level because cost-effectiveness ratios have clear and prominent benchmarks (e.g., Jackson & Mackevicius, 2024) in the broader literature on educational interventions.

Procedurally, we followed guidance from the evaluation literature (e.g., Levin et al., 2018) by compiling and pricing all program inputs. In collaboration with our district partners, we identified costs in three broad categories—consulting, teacher labor, and supplementary instruction (e.g., substitute teaching)—and leveraged historical administrative data to estimate their value in 2019 dollars.

Results

Tables 5–7 present our preferred estimates from Equation (1) across a rich collection of exploratory (e.g., attendance and course taking) and confirmatory (e.g., test scores) outcomes.

Exploratory Analysis of A1 Initiative Mechanisms

Attendance is an important determinant of student achievement (e.g., Lamdin, 1996; Liu & Loeb, 2021). Table 5 summarizes our findings on the impacts of assignment to the A1 Initiative on measures of student school engagement. Across all years, assignment to the A1 Initiative reduced absenteeism for students who entered high school below grade level. Their absence rate was between 2 and 7 percentage points lower than for control group students assigned to Algebra Readiness. This translated to a 5 to 9 percentage point reduction in chronic absenteeism, although this was only measured precisely in ninth grade. Attendance was largely unaffected for higher-achieving students. These targeted gains in students' behavioral engagement with school were consistent with—although not direct evidence for—the hypothesis that detracking, coupled with instructional supports, enhanced students' academic self-perception and motivation.

Additionally, for the highest- and lowest-achieving students at baseline, the A1 Initiative improved their chances of remaining in the district rather than transferring to a different school.¹³ For the below-grade-level group, assignment to the A1 Initiative improved their likelihood of remaining in the district for all 4 years of high school by 13 percentage points from a control group base rate of 77%. We also examined within-district school switching, and these results are presented in Supplementary Appendix Table A6, column (7), in the online version of the journal. For the focal group, this intradistrict transfer estimate was not statistically significant but was directionally the same as our estimated interdistrict exit effect. Taken together, these findings indicate an *overall* lower school switching rate, including both intra- and inter-district transfers, among this group of students.

This is a substantive outcome consistent with the A1 Initiative generating heightened levels of engagement for an academically vulnerable population. Although full explication of the mechanism underlying this result is beyond our study's scope, it stands to reason that parents may be less inclined to remove a child from their current school if the student is content and

Table 5
Intent-to-treat effects (ITT) of the A1 Initiative on measures of student engagement

Achievement stratum	9th grade			10th grade			11th grade			12th grade		
	Enrolled through academic year (1)	Chronically absent (2)	Absence rate (%) (3)	Enrolled through academic year (4)	Chronically absent (5)	Absence rate (%) (6)	Enrolled through academic year (7)	Chronically absent (8)	Absence rate (%) (9)	Enrolled through academic year (10)	Chronically absent (11)	Absence rate (%) (12)
ITT × at grade level	0.0002 (0.0078)	-0.0034 (0.0121)	0.0919 (0.2718)	0.0002 (0.0283)	0.0002 (0.0231)	-0.2906 (0.7048)	0.0578** (0.0267)	-0.0067 (0.0266)	-0.0169 (0.6692)	0.0638** (0.0280)	-0.0306 (0.0467)	-0.1830 (0.9597)
ITT × nearly at grade level	-0.0021 (0.0239)	0.0351 (0.0415)	0.8471 (0.8444)	0.0374 (0.0338)	0.0739 (0.0493)	1.2713 (1.5447)	-0.0202 (0.0516)	-0.0130 (0.0645)	-1.0298 (2.1515)	0.0398 (0.0538)	0.0993 (0.0862)	3.0712 (2.0829)
ITT × below grade level	0.0337 (0.0243)	-0.0747** (0.0349)	-1.9784** (0.8480)	0.0103 (0.0329)	-0.0863 (0.0629)	-7.4185*** (1.8294)	0.0728 (0.0483)	-0.0881 (0.0888)	-5.9688** (2.5776)	0.1268* (0.0480)	-0.0495 (0.0845)	-4.3176 (2.8075)
Nearly at grade level	0.0402 + (0.0229)	-0.0892*** (0.0244)	2.6571*** (0.7796)	0.0146 (0.0319)	-0.2431*** (0.0523)	11.0243*** (1.6130)	0.0761* (0.0438)	-0.3044*** (0.0567)	9.3166*** (2.3614)	0.1122*** (0.0408)	-0.2582*** (0.0763)	7.4890*** (2.7399)
Below grade level	0.0152 (0.0240)	-0.0685** (0.0323)	-1.7787** (0.8514)	-0.0214 (0.0311)	-0.2100*** (0.0541)	-8.6745*** (1.7562)	0.0570 (0.0416)	-0.1402** (0.0607)	-3.9867 (2.6608)	0.0659 (0.0459)	-0.1820** (0.0792)	5.4663* (2.9752)
<i>p</i> Value ($H_0: \beta_1 = \beta_2 = \beta_3$)	0.5239	0.1652	0.1212	0.7221	0.3176	0.0018	0.8668	0.7762	0.2193	0.0219	0.5937	0.2403
Control mean if below grade level	0.9490	0.1183	5.0258	0.9286	0.3626	15.0035	0.8061	0.5063	16.6826	0.7347	0.5139	15.4438
Observations	1,039	1,022	1,022	1,039	980	980	1,039	935	935	1,039	907	907

Notes: This table presents impacts of assignment to the A1 Initiative on measures of student engagement and connection to school. For each grade level indicated, this table shows effects on enrollment through the end of that academic year, on the likelihood of chronic absenteeism (i.e., absent >10% of all enrolled days), and on the continuous percent rate of days absent over total days enrolled. The conditional control mean presented was estimated across control observations in the below-grade-level stratum. All models controlled for ninth grade campus membership. Standard errors are clustered at the ninth grade classroom level.

* $p < .1$; ** $p < .05$; *** $p < .01$; **** $p < .001$.

succeeding in their educational environment. Given the negative consequences of reactive moves (e.g., Welsh, 2017), this is another academically protective feature of the A1 Initiative. Interestingly, we see in column (10) of Supplementary Appendix Table A6 that students who enter the district at grade level are 6 percentage points more likely to remain enrolled through twelfth grade if assigned to the A1 Initiative. A potential explanation, drawing from research on the “big fish, small pond” effect (e.g., Malamud et al., 2025), is that their improved *relative* classroom position in A1 Initiative classes boosted their self-perception and satisfaction. However, this study cannot formally decompose aggregate detected effects by each lever through which detracking and/or acceleration could influence student engagement and motivation. Thus, we propose this rationale provisionally.

Empirically, these results carry implications for interpreting the effects on dependent variables measured in later grades. Specifically, downstream outcomes may be biased by the differential rate of district exit (i.e., attrition). We explored the likely direction of this bias conceptually and empirically. First, we hypothesized that A1 Initiative students who would be marginal out-transfers in the unobserved counterfactual were likely to have lower levels of expected achievement. This is because mobility is negatively correlated with achievement (e.g., Mehana & Reynolds, 2004). The disproportionate persistence of these students in our sample’s treatment arm exerted a downward pressure on the average outcomes of below-grade-level A1 Initiative student relative to the observed control group. This reasoning was supported by the observed baseline achievement of attriters—the average control group attriter had lower middle school math scores than the average treatment group attriter (table available on request). The differential loss of higher-performing students from the treatment condition was consistent with the view that although there’s no clear evidence of attrition bias on the eleventh grade SBAC test (see Supplementary Appendix Table A1 in the online version of the journal), if it existed, it would involve negative selection into treatment, implying a downward bias in the study’s main estimates. We also did not observe pretreatment differences in A1 Initiative and non-A1 Initiative attriters across demographic dimensions. Therefore, the estimates we report for the impact of A1 Initiative assignment on downstream outcomes (i.e., graduation and twelfth grade course taking) likely reflect a *lower* bound on the true ITT effects. However, we note that the ITT effect on attrition was not statistically significant in eleventh grade.

In Figure 2 and Table 6, we summarize ITT effects on student course progression, attainment, and on-time graduation. A Sankey graph (Figure 2) illustrates the role that assignment to the A1 Initiative plays in shaping math trajectories for students achieving below grade level at baseline.¹⁴ Considering the ninth to tenth grade transition depicted in Figure 2, it is evident that placement in Algebra I rather than Algebra Readiness was challenging for many of these students. Ninth grade math failure rates were greater in

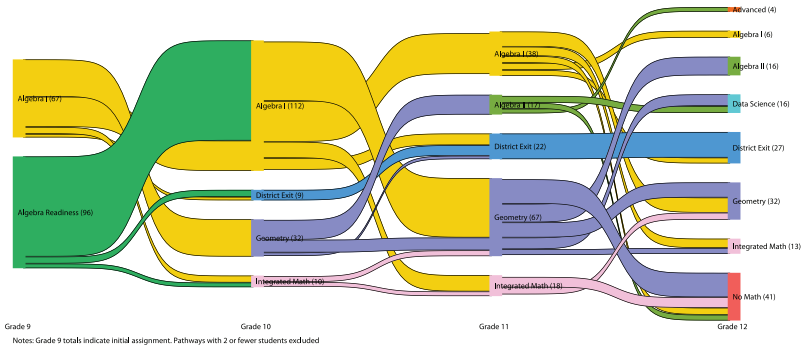


Figure 2. Students entering the district below grade level: A1 Initiative vs readiness pathways.

this stratum for A1 Initiative students. Approximately half had to retake Algebra I or enroll in a pregeometry bridge course as sophomores. However, the other half continued the college preparatory grade-level pathway to geometry in tenth grade, whereas only one Algebra Readiness student did so. And by the end of twelfth grade, academically underprepared A1 Initiative students were more likely to pass Algebra II than their control group peers (Table 6, column 5). Thus, although the pipeline observed in Figure 2 is leaky, students who were held back in their progression after ninth grade did no worse than students whose progression was delayed *before* ninth grade. In other words, track stability—the propensity for sustained placement on a particular pathway over time (Domina et al., 2019)—was very high for students on the remedial track. Once students were assigned a ninth grade remedial course, we observed that acceleration onto the standard pathway was very rare.

Table 6 supplements these visual patterns with estimates from Equation (1) for a series of transcript outcomes. Specifically, it shows that by the end of tenth grade, below-grade-level A1 Initiative–assigned students were no less likely to have passed two semesters of Algebra I when compared with the control group. Furthermore, they were 22 percentage points more likely to have passed two semesters of geometry. This accelerated progression continues through eleventh grade, when below-grade-level students assigned to the A1 Initiative were 14 percentage points more likely to have earned Algebra II credit. The 11 percentage point magnitude of the analogous twelfth grade estimate reflects a *more than doubling* of the likelihood of completing Algebra II. Twenty-one percent of below-grade-level students assigned treatment earned Algebra II course credit compared with only 9% of comparable control students.

Table 6
Intent-to-treat (ITT) effects of the A1 Initiative on student course access and attainment

	Through grade 10			Through grade 11			Through grade 12		
	Passed Algebra I (1)	Passed geometry (2)	Passed geometry (3)	Passed Algebra II (4)	Passed Algebra II (5)	Passed any advanced math (6)	Met UC-CSU math admission requirement (7)	Total math credits (8)	Graduated (9)
Achievement stratum									
ITT × at grade level	0.0192 (0.0148)	0.0458 (0.0424)	0.0277 (0.0363)	0.0409 (0.0408)	0.0524 (0.0377)	-0.0338 (0.0513)	0.0107 (0.0437)	0.7642 (0.8995)	0.0024 (0.0269)
ITT × nearly at grade level	0.0152 (0.0447)	-0.1298** (0.0580)	-0.0410 (0.0736)	-0.0258 (0.0734)	-0.0181 (0.0795)	0.0274 (0.0673)	-0.0163 (0.0777)	0.3732 (1.6619)	-0.0807 + (0.0458)
ITT × below grade level	0.0784 (0.0819)	0.2187*** (0.0596)	0.1204 (0.0786)	0.1434*** (0.0520)	0.1138* (0.0644)	0.0594 (0.0607)	0.0041 (0.0618)	2.7074* (1.3518)	0.0302 (0.0602)
Nearly at grade level	0.3174*** (0.0451)	0.6990*** (0.0534)	0.4949*** (0.0622)	0.6338*** (0.0557)	0.5998*** (0.0545)	0.5480*** (0.0669)	0.6037*** (0.0645)	15.0331*** (1.2274)	0.1150** (0.0474)
Below grade level	0.2211*** (0.0551)	0.3719*** (0.0510)	0.2310*** (0.0678)	0.2574*** (0.0500)	0.2176*** (0.0601)	0.1302** (0.0549)	0.1704** (0.0649)	6.0274*** (1.3073)	0.0736 (0.0479)
p Value ($H_0: \beta_1 = \beta_2 = \beta_3$)	.5317	.0006	.3639	.0218	.1024	.6593	.9909	.1232	.3519
Control mean if below grade level	0.6413	0.0109	0.3250	0.0125	0.0909	0.0779	0.1429	20.1461	0.8052
Observations	991	991	940	940	923	922	922	923	923

UC-CSU, University of California–California State University

Notes: This table presents estimates of the impact of assignment to the A1 Initiative on key academic outcomes through the indicated grade level. All models controlled for ninth grade campus membership. The conditional control mean presented was estimated across control observations in the below-grade-level stratum. Standard errors are clustered at the ninth grade classroom level.

* $p < .1$; ** $p < .05$; *** $p < .01$; **** $p < .001$.

Treated students in this group also obtained more math credits (column 8) but were no more likely than their Algebra Readiness–assigned peers to complete an optional fourth (i.e., post–Algebra II) math course. Because Algebra Readiness confers math credit rather than elective credit, this result is expected. However, although this district requires only 10 math credits to graduate, it is possible that in a district with higher expectations for math completion, the fact that math credit is withheld from control group students until tenth grade could be consequential. Additionally, the results of course taking through twelfth grade should be understood in the context of some selection bias. We found that the reform increased the likelihood that educationally vulnerable students remained in the district into twelfth grade (see Table 5). While this indicates a positive effect of the reform, it creates negative selection into treatment for twelfth grade outcomes (e.g., advanced math taking). This implies that the estimated effects of the reform on such outcomes are a lower bound on the true effect. It remains notable that our results indicated that the reform led to a weakly significant increase in the probability of completing Algebra II. We highlight this result in particular because of the changing and contested nature of “advanced math” in California (Fensterwald, 2024). Specifically, “Introduction to Data Science,” available to the students in this study, is no longer considered advanced or college ready, whereas completion of Algebra II is emphasized. This collection of results underscores the importance of students’ ninth grade placement in the hierarchical math sequence in structuring opportunity for the rest of their high school career.

Turning to the other baseline-achievement groups, those at grade level achieve comparably regardless of their treatment status. We observed a slight divergence in medium-term course progression and graduation within the nearly-at-grade-level stratum. Specifically, these students were less likely to have completed two semesters of geometry by the end of tenth grade if assigned to the A1 Initiative despite passing Algebra I at equivalent rates to control group students. We ruled out baseline imbalance or differential attrition as explanations based on the results in Tables 4 and 5. Notably, control group students from this stratum qualified for double-dose math, so treated students received less math instructional time. Given the sturdy evidence base in support of double-dose math for borderline-proficient students (e.g., Allensworth et al., 2009), its removal may explain these results. However, by the end of eleventh and twelfth grades, there was no difference by treatment assignment in cumulative math attainment for students with nearly-at-grade-level middle school achievement.

Confirmatory Analysis of Test Scores

Our main, preregistered confirmatory analysis aimed to measure the effect of A1 Initiative assignment on student math achievement as directly as possible. Specifically, we used assessment outcomes as the main metric

for student proficiency (Table 7). Otherwise, course outcomes alone were imprecise proxies for cognitive achievement. This is because course content may vary across similarly titled classes. And, depending on the grading autonomy and strategies afforded to teachers, credit attainment may not indicate content mastery. That is, absent other measures of proficiency, the superior course-taking outcomes we observe *could* reflect social promotion rather than learning.

Despite this concern, for students entering high school with below-grade-level math proficiency, assignment to the A1 Initiative yielded large and positive test score effects. On the eleventh grade SBAC test—the psychometrically validated exam for high school math proficiency in California—we detected a substantial and precisely estimated +0.19 SD (Table 7, column 3) to +0.20 SD (Table 7, column 4) ITT impact for these students using state-normed scores. We detected null eleventh grade test score ITT impacts among the nearly-at-grade-level and at-grade-level student groups. This absence of effects strongly suggests that A1 Initiative teachers were *not* uniquely high-quality math instructors and supports the internal validity of the estimated program benefits for below-grade-level students.

Because this study was originally intended to run through the pilot cohort's tenth grade year, our preregistration plan focused on an 11-item district-constructed assessment (i.e., the ICA) taken in the fall following the pilot year with the intent that a longer-term follow-up study would have a new preregistration. However, project delays related to the COVID-19 pandemic implied that the superior eleventh grade assessments became available to us. For the same group of below-grade-level students, the relevant ITT estimate using the 11-item tenth grade assessment was nearly as large as the SBAC-based effect (i.e., +0.14 SD), although not statistically significant. Following the spirit of our preregistration, in Table 7, we also present Romano–Wolf p values that implement a multiple-comparison correction for the three confirmatory estimates on the test score outcomes. After this correction, the ITT SBAC effect for the below-grade-level group takes on a p value of .0589 conditional on student traits, within the threshold for marginal statistical significance. In sum, the results presented in Table 7 provide strong evidence that the A1 Initiative improved math learning for low-proficiency students as well as—and possibly because of—superior course attainment and school attendance.

The sizable magnitudes of these estimates can be contextualized in several ways. Kraft (2020) deemed 0.2 SD to be a large effect in the distribution of estimates across 700 education randomized, controlled trials. Given a normal distribution of scores, this implies a 7 percentage point effect, placing it in the most effective third of educational interventions (von Hippel, 2025). A math-specific meta-analysis of randomized interventions by Williams et al. (2022) found an average effect size of 0.24 SD for algebra-related interventions but noted that estimates were largely based on short-term outcomes.

Table 7

Intent-to-treat (ITT) effects of the A1 Initiative on standardized math test scores

	10th grade district assessment (ICA)		11th grade state assessment (SBAC)	
Achievement stratum	(1)	(2)	(3)	(4)
ITT × at grade level	0.0270 (0.0626)	0.0036 (0.0613)	0.0745 (0.0739)	0.0603 (0.0690)
Romano–Wolf adjusted <i>p</i> value	[.7163]	[.9520]	[.2857]	[.3626]
ITT × nearly at grade level	0.0309 (0.0856)	0.0208 (0.0890)	0.1332 (0.0857)	0.1377 (0.0881)
Romano–Wolf adjusted <i>p</i> value	[.7682]	[.8551]	[.1698]	[.1518]
ITT × below grade level	0.1433 (0.1108)	0.1613 (0.1051)	0.1904* (0.0863)	0.2027** (0.0768)
Romano–Wolf adjusted <i>p</i> value	[.2358]	[.1768]	[.0729]	0.0589]
At grade level	1.0631*** (0.0817)	0.8944*** (0.0889)	1.2677*** (0.1037)	1.0657*** (0.1117)
Nearly at grade level	0.2812** (0.0876)	0.1950* (0.0934)	0.3299*** (0.0937)	0.2641* (0.0992)
Includes controls for student traits	No	Yes	No	Yes
<i>p</i> Value ($H_0: \beta_1 = \beta_2 = \beta_3$)	.5375	.4949	.0498	.0209
Control mean if below grade level	-0.8397	-0.8397	-0.8817	-0.8817
Observations	850	850	805	805

ICA, interim comprehensive assessment; SBAC, Smarter Balanced Assessment Consortium test

Notes: This table presents estimates of the impact of assignment to the A1 Initiative on standardized test score outcomes. Columns (1) and (2) present impacts on a tenth grade district-level ICA, whereas columns (3) and (4) summarize A1 Initiative impacts on the state SBAC assessment taken in eleventh grade. ICA scores are standardized over the sample, whereas SBAC scores are standardized using the mean and standard deviation of the statewide distribution. All models controlled for campus membership. Columns (2) and (4) also feature controls for baseline student race/ethnicity, gender, English learner status, and free or reduced-price meals status. The conditional control mean presented was estimated across control observations in the below-grade-level stratum. Standard errors are clustered at the ninth grade classroom level. Multiple comparison adjustments are indicated with Romano–Wolf *p* values (resample $n = 10,000$).

* $p < .05$; ** $p < .01$; *** $p < .001$.

Thus, the persistence of an A1 Initiative effect through an eleventh grade assessment is notable. This impact is also large relative to learning trajectories at this age. Bloom et al. (2008) found that a 0.19 SD effect across grades 9 to 11 roughly translated into an entire additional year of math learning. Finally, we reiterate that because of the asymmetric rate of district exit observed across the treatment and control conditions, our estimate is plausibly a *lower* bound on the true A1 Initiative impact on math learning.

Cost-Effectiveness

We constructed cost-effectiveness ratios of the A1 Initiative by scaling the main effect by the per-student cost of the A1 Initiative. Because we only observed statistically significant effects on primary outcome for one of the three treatment–control comparisons, we focused on effectiveness for only this group but also recognized the A1 Initiative’s total costs (i.e., spreading total costs over targeted gains). We computed the A1 Initiative’s total costs using district data on the consultant fees for professional development and on-site coaching and state data on teacher costs in the districts (see the Supplementary Technical and Data Appendix in the online version of the journal for details).

Altogether, we estimated the A1 Initiative cost at ~\$264,000. Given the comparatively small number of below-grade-level students assigned to A1 Initiative classes (i.e., the only students with clear academic benefits), the implied cost per relevant student then would be ~\$3,950. The estimated 0.19 SD test score gain among these students implies that the A1 Initiative generated 0.048 SD in test score gains per \$1,000 spent. Notably, this is roughly six times the return on general increases in school spending. Specifically, Jackson and Mackevicius (2024) concluded that a \$1,000 increase in annual spending per pupil repeated over 4 years increases test scores by only 0.032 SD.

Effect of Heterogeneity and Spillovers

Supplementary Appendix Table A5 in the online version of the journal presents results for exploratory analyses of A1 Initiative assignment effects by student gender and socioeconomic status (i.e., as proxied by FRPL). Given the small sample sizes—which precluded further subsetting across racial and ethnic categories—and the type I error risk associated with estimating so many effects, we consider these results to be merely suggestive of directions for future research. Still, we note a few intriguing patterns. First, estimates of the positive influence of the A1 Initiative on attendance and district retention for below-grade-level students are consistently larger when the sample is subset to only include girls versus when effects are estimated among boys. Prior research *has* identified girls as more sensitive to negative behavioral peer effects than boys (Imberman et al., 2012). This result is consistent with a symmetric propensity for positive peer influence. Additionally, as in

prior studies of targeted acceleration (e.g., Dougherty et al., 2017), the attainment effects of A1 Initiative assignment for below-grade-level students—as measured by the likelihood of ever passing Algebra II—are larger and only statistically significant for female students and non-FRPL students. Conversely, and perhaps surprisingly, the test score benefits of the A1 Initiative for below-grade-level students are concentrated among boys and FRPL-eligible students. It is possible that acceleration may benefit male versus female students or students of differing economic backgrounds through different mechanisms, and this heterogeneity warrants future investigation.

We found mixed evidence of spillover effects into performance in other subjects such as English, science, and electives courses. At the 10% level, we detected negative impacts from assignment to the A1 Initiative on credit accumulation for nearly-at-grade-level students in science and elective courses as well as overall (see Supplementary Appendix Table A6, columns 4 and 5, in the online version of the journal). These findings comport with the marginally significant but negative 8 percentage point on-time graduation effect. As discussed earlier, the treatment–control contrast for this group is a complicated one. Specifically, for this student group, assignment to the A1 Initiative also removed a double-dose support class. Therefore, control group students within this stratum had more math instructional time in ninth grade than treated students. Although the short-term negative effect on math attainment (Table 6, column 2) did not persist, we do observe weakly significant impacts on academic achievement in other subjects (see Supplementary Appendix Table A6 in the online version of the journal). We also observed a negative impact from A1 Initiative assignment on the English language arts eleventh grade SBAC test for the nearly-at-grade-level group.

The observed deficit in elective credit (see Supplementary Appendix Table A6, column 5, in the online version of the journal) is partly mechanical due to these students' nonenrollment in the elective-credit-bearing support course. Similarly, the negative impact on elective credit attainment for below-grade-level students can be explained by their acceleration *out* of the prealgebra Algebra Readiness elective credit course. On the other apparent negative spillover results, we can only speculate from ninth grade credit-taking data (see Supplementary Appendix Table A8 in the online version of the journal). Notably, nearly-at-grade-level students in the treatment condition attempted 2.5 fewer total credits than control group students (see Supplementary Appendix Table A8, panel A, column 1, in the online version of the journal), implying that they may just be taking a less academic course-load in the absence of the support class rather than replacing it with equivalent time in other subjects. Because these results fall outside our preregistered analysis, and due to the presence of multiple comparisons, this should motivate caution in interpreting results from Supplementary Appendix Tables A6 and A8.

We did not detect effects on standardized English language arts scores for the other baseline achievement strata. This result is consistent with our topline confirmatory result for below-grade-level students of math-specific learning gains due to the A1 Initiative program (i.e., a + 0.19 SD ITT impact). It also suggests that acceleration to the more rigorous A1 Initiative pathway did not crowd out effort in other classes among below-grade-level students.

Sensitivity and Robustness Checks

Our main results, especially those we highlighted for the below-grade-level proficiency group, are robust to a variety of alternative specifications. Supplementary Appendix Table A3 in the online version of the journal presents the results of Equation (2) where student (panels B and C) and teacher (panels D and E) characteristics are controlled for. As we would expect given successful randomization, the magnitudes of the ITT estimates did not differ substantially across specifications. In panel C, we find that our topline results are robust to the addition of a linear control for baseline math achievement that uses, due to missingness, a combination of middle school test scores (i.e., $ES \approx 0.17$ vs. 0.19 SD). Supplementary Appendix Table A4 in the online version of the journal similarly shows that our findings are unlikely to be driven by a “superstar” teacher (or its opposite) because dropping each A1 Initiative teacher from our sample in turn left our main findings broadly unaltered. Additionally, although we did not report full results by campus due to small cell sizes and anonymity concerns, we found that our topline test score effects were no smaller at the campus that was deemed “low buy-in” and did not have volunteer teachers leading A1 Initiative classes. This suggests that our findings are indeed robust to lower levels of teacher enthusiasm.

Finally, in Supplementary Appendix Table A2 in the online version of the journal, the standardized test score effect for our focal student group remained large and at least marginally statistically significant using alternate standard errors, except for when randomization inference was combined with our nonpreferred strategy of including a vector of baseline student characteristics.

Discussion

In contrast to much of the existing research on the acceleration of very low-proficiency students (e.g., Clotfelter et al., 2015; Domina et al., 2019; Lafortune, 2018; McEachin et al., 2020), this study identified academic and nonacademic benefits from a program that randomly assigned academically underprepared ninth graders into Algebra I rather than to a remedial prealgebra course. Given established links between high school math attainment and post-secondary school and labor market attainment (e.g., Altonji, 1995; Goodman, 2019), these findings carry significant implications for these students’ long-run outcomes. Our results indicate that the A1 Initiative positively

shaped the trajectories of students who were the lowest achievers in middle school.

On the state math assessment, we detected a substantial 0.19 effect size. The A1 Initiative also improved math course attainment, as measured by Algebra II completion as well as student attendance throughout high school. The limited catch-up in Algebra II completion suggests that the test score measured in the eleventh grade is likely to persist. Additionally, given college testing timelines, earlier mastery of key content such as Algebra II has meaningful implications for student readiness and therefore long-term access to math learning opportunities. And because district retention of treatment-assigned students was 13 percentage points higher than of control students, our estimates likely reflect a lower bound on the true impacts of the A1 Initiative. We found no evidence that students entering high school at grade level were negatively influenced by assignment to the A1 Initiative. Given this strikingly positive collection of results for students deemed to have low levels of proficiency, several implementation and policy details of the A1 Initiative deserve particular attention.

Implications for Policy and Practice

First, pedagogic quality should be a central implementation concern for policies that expand access to rigorous content for less prepared students and/or increase within-classroom variation in baseline achievement. The A1 Initiative provided support for the development (i.e., dedicated professional learning) and execution (i.e., additional planning time) of an appropriate instructional approach for a mixed-achievement classroom environment. For example, teachers were provided with flexibility to responsively pace their courses, strategies to help surface student misconceptions (e.g., math language routines), and community support including a partner teacher and on-site coaching. A1 Initiative trainings also strongly emphasized that teachers hold high expectations for *all* students to continue progressing through a college-preparatory math sequence. Our findings in this study are consistent with a protective dynamic wherein supportive practices for high-quality instructional differentiation offset the academic risks of acceleration.

Second, the superior attendance and school retention outcomes among accelerated A1 Initiative students relative to their control group peers suggests that a supported-acceleration model may improve school engagement for underprepared students relative to traditional remediation. The conceptual literature has defined three broad forms of student engagement: behavioral, emotional, and cognitive (Fredricks et al., 2004). Thus, our attendance, retention, and course-taking results are consistent with a positive behavioral engagement effect, although merely suggestive of other dimensions of engagement. Prior tracking research has suggested that such effects may operate through social-emotional changes such as improved motivation and

decreased stigma (e.g., Chmielewski et al., 2013; Gamoran, 1992). It is also possible that the improved engagement may instead be a result of positive behavioral peer effects (e.g., from exposure to more engaged classmates) or a direct result of high expectations and a growth mindset promoted in the intervention (Good et al., 2018; Imberman et al., 2012). Although our results are consistent for the role targeted acceleration has in shaping belongingness or motivation, it is beyond the scope of this study to confirm or untangle any of these potential pathways. We look to future research to more directly explore the social-emotional dynamics of placing students on accelerated versus remedial pathways.

Third, the persistence of impacts in the eleventh and twelfth grades underscore the status of ninth grade as a “make or break year” for the rest of high school (Krone Phillips, 2019). Most empirical studies of math acceleration and tracking have focused on middle school practices (e.g., eighth grade Algebra I). However, the patterns we observed in this high school-based study suggest that ninth grade acceleration or remediation plays a significant role in how math educational opportunities are structured in a time-limited setting (i.e., generally within 4 years to complete a hierarchical sequence). This study demonstrates that high school is not too late to positively transform students’ academic trajectories.

Implications for Classroom Diversity

Furthermore, although our analysis focused on student-level outcomes, the A1 Initiative did advance the district’s broader equity and inclusion goals. Ethnoracial and socioeconomic diversity was greater in A1 Initiative sections relative to remedial sections, where poor and minoritized students were disproportionately concentrated (see Table 2). We can quantify the differences in A1 Initiative versus control classroom-level segregation using a dissimilarity index. This is a measure of the evenness in distribution of students and can be interpreted as the share of students within each treatment arm that would have to switch classes for all sections to include a balanced proportion of the indicated student groups. A higher ratio reflects more severe between-class segregation. The dissimilarity index between poor (i.e., qualifying for FRPL) and nonpoor students among A1 Initiative sections was 0.212 compared with 0.479 for control sections. The same measures comparing segregation of EL and non-EL students were 0.243 and 0.580. Ethnoracial segregation also was lower in the detracked condition, with the distribution of White to non-White, White to Hispanic, and Hispanic to non-Hispanic students being 1.7 to 1.9 times more uneven across control classes versus across A1 Initiative classes.

Cost-Effectiveness and Policy Alternatives

Even though only the lowest-achieving students benefited from the A1 Initiative, we still found that it was substantially more cost-effective than

most education spending, generating 0.048 SD in test score gains per \$1,000 spent, compared with estimates by Jackson and Mackevicius (2024) that 4 sequential years of spending \$1,000 only boosted test scores by 0.032 SD. Furthermore, most districts could fund a comparable intervention at much lower cost than our partner district. The average teacher was paid just under \$110,000 in the A1 Initiative district for AY 2019–20, nearly double the nationwide mean salary of \$63,645 (National Education Association, 2021). However, this calculation likely overstates the actual cost of the A1 Initiative because these classes were somewhat larger and required less supplementary staff (i.e., coteaching and staffing for double-dose sections) than business-as-usual classes.¹⁴ In general, these results suggest that the A1 Initiative, which generated quite large gains targeted among a uniquely important subgroup of students, is comparatively cost-effective to untargeted educational spending.

We also can specifically contrast this cost-effectiveness with returns from the primary alternative strategies districts used to support lower-achieving students in on-level classes: tutoring and double-dose math. Guryan et al. (2023) estimated 0.16–0.37 SD end-of-year test score effects from a high-impact tutoring program that cost ~\$3,800 in 2013. Adjusting to 2019 dollars, this yields a 0.038–0.089 range for ratio of effect per \$1,000 of spending. The cost-effectiveness ratio of the A1 Initiative lies squarely within this range. Notably, the low-end estimate compared tutoring with double-dose math, which aligns with estimated double-dose effects of ~0.25 SD (Nomi & Allensworth, 2009).

Cortes et al. (2015) provided double-dose impact estimates ranging from 0.14 to 0.22 SD for eleventh grade test scores of students who were lower-achieving students in the ninth grade, an apt comparison point for our eleventh grade test score effects estimates for below-grade-level students. The A1 Initiative again produced results in the range of empirically supported alternatives. Double dose is, however, more cost-effective than the A1 Initiative in its initial form. The back-of-the-envelope estimated cost of staffing a double-dose class section in our high-salary partner district in 2019 (i.e., again calculating one fifth of an average teacher salary) was ~\$22,000, or about \$1,000 per student (i.e., generating a cost-effectiveness ratio of 0.14–0.22).

As a practical matter, these cost estimates should be considered through the lens of opportunity cost—what are districts giving up by providing these interventions? Double-dose costs, for instance, may be covered by reducing staffing for other subjects. Our analyses also did not account for the effort costs *to the student* of spending twice the instructional time in double-dose math to students in the A1 Initiative. Similarly, the A1 Initiative’s professional development could occur without additional financial cost if it replaced other forms of in-service training. We note that when the district later scaled the A1 Initiative to cover all students, it reduced costs by providing only some teachers (i.e., those who also took on instructional leadership roles) with a course release and cutting most consultant-provided supplemental professional

development and coaching. It is plausible that eliminating these supportive features reduced program efficacy along with expenses.

As initially implemented, however, the A1 Initiative was comparably effective at improving achievement for very low-proficiency students as other supportive acceleration practices and similarly cost-effective per student per hour of instruction. Given the mixed results we observed for nearly-at-grade-level students, it is also plausible that tutoring and double dose could be effective complements to an A1 Initiative-type program, resources permitting.

External Validity and Feasibility

The scalability of A1 Initiative-like programs is not constrained merely by pecuniary considerations. Although the A1 Initiative is a novel and promising proof point, the perennial challenges of replicating promising pilots implies a healthy agnosticism about realizing these effects elsewhere. We can, however, draw some optimistic observations from research on related interventions. Double dose was successfully implemented in a very different context, the large urban Chicago Public Schools district (Nomi & Raudenbush, 2016). Academic trajectory improvements also were identified in a randomized evaluation (Arshavsky et al., 2025) of an early college high school intervention in North Carolina that shared similar features with the A1 Initiative (e.g., acceleration of low-proficiency students into ninth grade Algebra I, high expectations, and supportive practices for students and teachers). There is nothing to indicate that the A1 Initiative program success reflects singular features of this district, which is a medium-sized, diverse, suburban, and like numerous others in its basic instructional capacity (i.e., districtwide instructional administrators and coaches).

Still, potential challenges to replicating the promising academic, nonacademic, and inclusion benefits we observed merit emphasis. For example, policies that promote detracking often face pushback from some parents who oppose the removal of selective tracks (Kariya & Rosenbaum, 1999; Tucker, 2023; Wells & Serna, 2010). In this respect, the A1 Initiative was a *moderate* policy because the highest-achieving third of the student body was unaffected by the reform. Additionally, we note that our analysis did not bear out concerns that the addition of below-grade-level students to Algebra I classrooms would induce negative peer effects. A partial detracking model typified by the A1 Initiative may promote greater buy-in than broader detracking reforms.

Our findings also may not generalize to populations with different distributions of baseline achievement. In the district, positive effects were concentrated among a small—and uniquely high-need—group of students. Similar policies *could* produce even larger average gains in districts where a larger share of students would otherwise be placed in remediation. Conversely, it is possible that in classes with large shares of the lowest-proficiency students (i.e., a reverse of the composition observed in Table 2), high achievers would

experience negative peer impacts, and low achievers would not see benefits from having more proficient peers. In sum, our findings did not necessarily generalize to policies that would *fully* detrack ninth grade math, such as by removing middle school acceleration pathways.¹⁵ Such an expansion may risk the collective effects identified by Penner et al. (2015): The broader the scope of a detracking program, the greater would be the risk created for unintended consequences such as political backlash, negative peer effects, and burdensome staffing and pedagogic demands. We also cannot speak directly to the benefits of similar instructional supports in a *more* tracked system, but it is plausible that A1 Initiative pedagogies and supports could promote student learning in homogeneous classrooms as well.

A final category of threats to replication and scalability is the feasibility of high-fidelity implementation in other contexts; that is, whether other districts can support and nurture reforms in the manner of districts that pioneered those reforms is uncertain. Interventions such as the A1 Initiative that rely on substantial shifts in within-classroom practice are challenging to promote successfully at scale (Elmore, 2010). Additionally, although our results are robust to multiple checks for teacher-selection effects (i.e., conditioning on observable traits, a “leave one out” exercise), it is possible the A1 Initiative would be less effective if teachers were mandated to adopt it against their preferences. Optimistically, however, the program improved outcomes for low-proficiency students *despite* the COVID-19 pandemic interrupting a planned program of teacher support in the spring of 2020. The implementation realized in this trying context still generated outcomes above and beyond those of the status quo.

Our findings did highlight two specific areas for improvement of the A1 Initiative program. First, in twelfth grade, we observed a partial fadeout of the A1 Initiative’s relative influence on Algebra II completion—and null impacts for coursework beyond Algebra II. This may merely reflect differential attrition. However, it also raises a concern that students who sit out senior math lose the subject continuity that may contribute to a successful college transition (Wainstein et al., 2023). Thus, there could be a role for proactively guiding students to enroll in a twelfth grade math class above and beyond graduation requirements. In our sample, 10% of the students, including 27% of below-grade-level students, who completed Algebra II by the end of eleventh grade took no math course in twelfth grade. In states that do not require a fourth math course for high school graduation, such as California, these results suggest that merely raising the floor in earlier grades will not translate to higher rates of college-level math completion if expectations and counseling practices do not exert upward pressure toward the end of students’ high school experiences. Second, evidence of small negative impacts from assignment to the A1 Initiative on academic attainment for students who would have received a “support” class in the control group suggests that the A1 Initiative could be more effective for some students if paired with additional

instructional time. However, staffing demands for co-implementation of the A1 Initiative *and* double-dose math could be prohibitive in some districts.

Limitations

In addition to the constraints just noted on this study's ability to speak conclusively about the external validity of our findings, three other key limitations are worth highlighting as areas for future work. First, we acknowledge that this study design cannot disentangle each mechanism through which the A1 Initiative specifically—and targeted detracking policies broadly—mediates student outcomes. Supported acceleration practices alter not only the content and rigor of coursework that high-need students receive but also their peer group, their relative rank, and, often, the supplemental services they receive. Teacher perception and quality also may change. While we emphasize that the results of this field experiment reflect the “real world” nature of this class of policies as inherently composed of “bundled” interventions, we are limited to reasoned speculation rather than confirmatory analysis in parsing the drivers of A1 Initiative benefits on student learning engagement. There is much to be learned from evaluation of iterative programs—such as what happens if double dose is not removed from the nearly-at-grade-level students? Or does instructionally supported acceleration still generate similar gains to double dose and tutoring if teachers do not receive additional planning time and/or individualize coaching?

Second, while we have laid out several reasons why secular teacher quality is unlikely to be driving these results, we cannot entirely rule out its influence in the absence of randomized teacher assignment. Third, a full elucidation of the social-emotional dynamics of the A1 Initiative and its impact on student engagement, belonging, and self-perception is beyond the scope of this study. Although the revealed preferences implied by improved student attendance and retention are compelling, supplementing administrative data with student survey or qualitative study could more explicitly identify the role of supported acceleration in shaping student self-perception and sense of belonging.

Conclusion

In sum, this study shows that raising the floor of academic expectations for educationally vulnerable students can be a successful strategy to link equity goals with improved achievement. Critically, the A1 Initiative paired reforms to course assignment with aligned supports for students and teachers. The positive impacts of the A1 Initiative on student attendance imply that any academic benefits of tracking likely come at the cost of student engagement. This finding should motivate more empirical study of the theorized roles motivation, belongingness, and stigma play in the educational experiences of remedial-track students. Furthermore, even the *partial* detracking induced by this reform decreased within-school segregation by race, ethnicity, and

class. The A1 Initiative presents a provocative proof point for high school classes in which students with disparate levels of prior math achievement excel together.

Acknowledgments

We thank Victoria Dye, C. J. Carlson, Diana Wilmot, and Bonnie Hansen at the Sequoia Union High School District for their crucial contributions to this study.

Funding

We are also grateful for the support of the Stanford Sequoia K–12 Research Collaborative and Diana Mercado-Garcia’s facilitation of this researcher–practitioner partnership.

ORCID iDs

Thomas S. Dee  <https://orcid.org/0000-0001-7524-768X>

Elizabeth Huffaker  <https://orcid.org/0000-0002-8050-4844>

Notes

¹Goodman (2019) estimated noisy nulls with caveats for Hispanic students due to reporting issues.

²Roughly one third of entering ninth graders were above grade level (i.e., had passed Algebra I in middle school) and assigned to geometry (or higher). They were ineligible for randomization.

³A few students with Individualized Education Programs (i.e., those placed in basic math courses absent from the conventional placement policy; Figure 2) and some newcomer English learners (i.e., those placed in Algebra I sections specifically for students deemed to have limited English proficiency) also were excluded from our intent-to-treat sample because they were ineligible for randomization.

⁴We describe the data construction, variables, and this regression specification in detail below.

⁵For students at grade level, treatment reduces incidence of teachers with higher-level credentials. This implies that our results may *understate* the impact of a detracked classroom for these students.

⁶Most primary teachers in the sample ($n = 26$) identify as White ($n = 14$), seven as Asian/Pacific Islander and five as non-White Hispanic. An equal number of teachers are male and female.

⁷Initiative teachers were encouraged to use Illustrative Math, a curriculum and professional leaning open resource to retrieve or inspire their lesson materials. The problem-solving emphasis of Illustrative Math broadly supports the IES–suggested practice of using multiple strategies to solve algebra problems (Institute for Education Sciences, 2014), but the efficacy Illustrative Math has itself not yet been independently evaluated.

⁸Second-semester on-site coaching and site visit plans were interrupted by the COVID-19 pandemic. The final professional-development session was held on March 11.

⁹In the fall of 2019, only 79% of control students eligible for Algebra I with a support class (nearly at grade level) took both of those courses. Control group compliances with assignments to Algebra I (at grade level) and Algebra Readiness (below grade level) were 99% and 93%, respectively.

¹⁰For example, referring to Figure 1, a student who scored a 2490 on the SBAC test, a 6 on the Diagnostic Online Math Assessment test, but a 20 on the Mathematics Diagnostic Testing Project test should have been assigned to Algebra I *without* support, even though they only met the qualifying threshold on one of the three tests (the Mathematics Diagnostic Testing Project test in this case).

¹¹We preregistered the tenth grade ICA score when we anticipated following the cohort for 2 years. It included two performance tasks and only 11 closed-response items (e.g., multiple choice), which constrains construct validity. We prioritized SBAC test scores for this reason, in addition to the test's validated psychometric properties and known statewide distribution (California Department of Education, 2023).

¹²In the simple case, this means that a student has passed Algebra I, geometry, and Algebra II with at least a C– course grade, although our variable construction accounts for the full range of alternative course validations sanctioned by the University of California system.

¹³We did not detect evidence of systematic differences in dropout rates across treatment conditions.

¹⁴Similar figures for other strata are not included because of the null impacts observed in Table 6.

¹⁵Recent backlash to a broad range of equity-motivated policies, such as detracking in English language arts, suggest a more comprehensive detracking A1 Initiative would have drawn greater community pushback.

References

- Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. M. (2020). Sampling-based versus design-based uncertainty in regression analysis. *Econometrica*, 88(1), 265–296. <https://doi.org/10.3982/ECTA12675>
- Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. M. (2023). When should you adjust standard errors for clustering? *The Quarterly Journal of Economics*, 138(1), 1–35. <https://doi.org/10.1093/qje/qjac038>
- Allensworth, E., Nomi, T., Montgomery, N., & Lee, V. E. (2009). College preparatory curriculum for all: Academic consequences of requiring algebra and English I for ninth graders in Chicago. *Educational Evaluation and Policy Analysis*, 31(4), 367–391. <https://doi.org/10.3102/0162373709343471>
- Altonji, J. G. (1995). The effects of high school curriculum on education and labor market outcomes. *The Journal of Human Resources*, 30(3), 409–438. <https://doi.org/10.2307/146029>
- Angrist, J. D. (2014). The perils of peer effects. *Labour Economics*, 30, 98–108. <http://dx.doi.org/10.1016/j.labeco.2014.05.008>
- Antonovics, K., Black, S. E., Cullen, J. B., & Meiselman, A. Y. (2022). *Patterns, determinants, and consequences of ability tracking: Evidence from Texas public schools* (NBER Working Paper 30370). National Bureau of Economic Research. <https://doi.org/10.3386/w30370>
- Arshavsky, N., Edmunds, J. A., Unlu, F., & Fesler, L. (2025). *Improving college readiness in mathematics in the context of a comprehensive high school reform*. (EdWorkingPaper: 25–1131). Retrieved from Annenberg Institute at Brown University: <https://doi.org/10.26300/bd45-mf87>
- Austin, M., Backes, B., Goldhaber, D., Li, D., & Streich, F. (2024). Leveling up: An academic acceleration policy to increase equity in advanced high school course taking. *American Educational Research Journal*, 62(1), 136–179. <https://doi.org/10.3102/00028312241282793>

- Ayalon, H., & Gamoran, A. (2000). Stratification in academic secondary programs and educational inequality in Israel and the United States. *Comparative Education Review*, 44(1), 54–80. <https://doi.org/10.1086/447591>
- Betts, J. R. (2011). The economics of tracking in education. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the economics of education* (Vol. 3, pp. 341–381). Elsevier.
- Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness*, 1(4), 289–328. <https://doi.org/10.1080/19345740802400072>
- Brighouse, H., Ladd, H. F., Loeb, S., & Swift, A. (2018). *Educational goods: Values, evidence, and decision-making*. University of Chicago Press.
- Burke, M. A., & Sass, T. R. (2013). Classroom peer effects and student achievement. *Journal of Labor Economics*, 31(1), 51–82. <https://doi.org/10.1086/666653>
- California Department of Education. (2023, May 31). *California assessment of student performance and progress smarter balanced summative assessment 2021–22 Technical Report*. <https://www.cde.ca.gov/ta/tg/ca/documents/sbcaasppptechrpt22.docx>
- Card, D., & Giuliano, L. (2024). *Targeted acceleration in middle school math: Impacts on college entry, degree completion, and STEM*. <https://events.bse.eu/live/files/5327-gempdf>
- Card, D., & Giuliano, L. (2016). Universal screening increases the representation of low-income and minority students in gifted education. *Proceedings of the National Academy of Sciences*, 113(48), 13678–13683. <https://doi.org/10.1073/pnas.1605043113>
- Chmielewski, A. K., Dumont, H., & Trautwein, U. (2013). Tracking effects depend on tracking type. *American Educational Research Journal*, 50(5), 925–957. <https://doi.org/10.3102/0002831213489843>
- Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2015). The aftermath of accelerating algebra evidence from district policy initiatives. *Journal of Human Resources*, 50(1), 159–188. <https://doi.org/10.3368/jhr.50.1.159>
- Clotfelter, C. T., Ladd, H. F., Clifton, C. R., & Turaeva, M. R. (2021). School segregation at the classroom level in a southern “new destination” state. *Race and Social Problems*, 13(2), 131–160. <https://doi.org/10.1007/s12552-020-09309-w>
- Cohodes, S. R. (2020). The long-run impacts of specialized programming for high-achieving students. *American Economic Journal: Economic Policy*, 12(1), 127–166. <https://doi.org/10.1257/pol.20180315>
- Collins, C. A., & Gan, L. (2013). *Does sorting students improve scores? An analysis of class composition* (NBER Working Paper No. 18848). National Bureau of Economic Research. <https://doi.org/10.3368/10.3386/w18848>
- Conger, D. (2005). Within-school segregation in an urban school district. *Educational Evaluation and Policy Analysis*, 27(3), 225–244. <https://doi.org/10.3102/01623737027003225>
- Conger, D., Long, M. C., & Iatarola, P. (2009). Explaining race, poverty, and gender disparities in advanced course-taking. *Journal of Policy Analysis and Management*, 28(4), 555–576. <https://doi.org/10.1002/pam.20455>
- Cortes, K. E., & Goodman, J. S. (2014). Ability-tracking, instructional time, and better pedagogy: The effect of double-dose algebra on student achievement. *American Economic Review*, 104(5), 400–405. <https://doi.org/10.1257/aer.104.5.400>
- Cortes, K. E., Goodman, J. S., & Nomi, T. (2015). Intensive math instruction and educational attainment long-run impacts of double-dose algebra. *Journal of Human Resources*, 50(1), 108–158. <https://doi.org/10.2139/ssrn.2248171>

- Dalane, K., & Marcotte, D. E. (2020). *The segregation of students by income in public schools* (EdWorkingPapers.com). Annenberg Institute at Brown University.
- Dee, T. S. (2024). Higher chronic absenteeism threatens academic recovery from the COVID-19 pandemic. *Proceedings of the National Academy of Sciences*, 121(3), e2312249121. <https://doi.org/10.1073/pnas.2312249121>
- Dee, T. S., & West, M. R. (2011). The non-cognitive returns to class size. *Educational Evaluation and Policy Analysis*, 33(1), 23–46. <https://doi.org/10.3102/0162373710392370>
- Dixon, F. A., Yssel, N., McConnell, J. M., & Hardin, T. (2014). Differentiated instruction, professional development, and teacher efficacy. *Journal for the Education of the Gifted*, 37(2), 111–127. <https://doi.org/10.1177/0162353214529042>
- Domina, T., McEachin, A., Hanselman, P., Agarwal, P., Hwang, N., & Lewis, R. W. (2019). Beyond tracking and detracking: The dimensions of organizational differentiation in schools. *Sociology of Education*, 92(3), 293–322. <https://doi.org/10.1177/0038040719851879>
- Dougherty, S. M., Goodman, J. S., Hill, D. V., Litke, E. G., & Page, L. C. (2017). Objective course placement and college readiness: Evidence from targeted middle school math acceleration. *Economics of Education Review*, 58, 141–161. <https://doi.org/10.1016/j.econedurev.2017.04.002>
- Dougherty, S. M., Goodman, J. S., Hill, D. V., Litke, E. G., & Page, L. C. (2015). Middle school math acceleration and equitable access to eighth-grade algebra: Evidence from the Wake County Public School System. *Educational Evaluation and Policy Analysis*, 37(1), 80S–101S. <https://doi.org/10.3102/0162373715576076>
- Duflo, E., Dupas, P., & Kremer, M. (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *American Economic Review*, 101(5), 1739–1774. <https://doi.org/10.1257/aer.101.5.1739>
- Egalite, A. J., Kisida, B., & Winters, M. A. (2015). Representation in the classroom: The effect of own-race teachers on student achievement. *Economics of Education Review*, 45, 44–52. <https://doi.org/10.1016/j.econedurev.2015.01.007>
- Elmore, R. (2010). Getting to scale with good educational practice. *Harvard Educational Review*, 66(1), 1–27. <https://doi.org/10.17763/haer.66.1.g73266758j348t33>
- Feld, J., & Zölitz, U. (2017). Understanding peer effects: On the nature, estimation, and channels of peer effects. *Journal of Labor Economics*, 35(2), 387–428. <https://doi.org/10.1086/689472>
- Fensterwald, J. (2024, March 5). *UC professors' math problem: How does data science fit in?* EdSource. <https://edsources.org/2024/uc-confirms-data-science-cant-sub-for-algebra-ii-unresolved-what-can-it-qualify-for/707043>
- Figlio, D. N., & Page, M. E. (2002). School choice and the distributional effects of ability tracking: Does separation increase inequality? *Journal of Urban Economics*, 51(3), 497–514. <https://doi.org/10.1006/juec.2001.2255>
- Francis, D. V., & Darity, W. A. (2021). Separate and unequal under one roof: How the legacy of racialized tracking perpetuates within-school segregation. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, 7(1), 187–202. <https://doi.org/10.7758/RSF.2021.7.1.11>
- 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. <https://doi.org/10.3102/00346543074001059>
- Gamoran, A. (1992). The variable effects of high school tracking. *American Sociological Review*, 57(6), 812–828. <https://doi.org/10.2307/2096125>
- Gamoran, A., & Hannigan, E. C. (2000). Algebra for everyone? Benefits of college-preparatory mathematics for students with diverse abilities in early secondary

- school. *Educational Evaluation and Policy Analysis*, 22(3), 241–254. <https://doi.org/10.3102/01623737022003241>
- Goddard, Y. L., Goddard, R. D., Bailes, L. P., & Nichols, R. (2019). From school leadership to differentiated instruction: A pathway to student learning in schools. *The Elementary School Journal*, 120(2), 197–219. <https://doi.org/10.1086/705827>
- Good, T. L., Grouws, D. A., & Beckerman, T. M. (1978). Curriculum pacing: Some empirical data in mathematics. *Journal of Curriculum Studies*, 10(1), 75–83. <https://doi.org/10.1080/0022027780100106>
- Good, T. L., Sterzinger, N., & Lavigne, A. (2018). Expectation effects: Pygmalion and the initial 20 years of research. *Educational Research and Evaluation*, 24(3–5), 99–123. <https://doi.org/10.1080/13803611.2018.1548817>
- Goodman, J. (2019). The labor of division: Returns to compulsory high school math coursework. *Journal of Labor Economics*, 37(4), 1141–1182. <https://doi.org/10.1086/703135>
- Graham, L. J., de Bruin, K., Lassig, C., & Spandagou, I. (2021). A scoping review of 20 years of research on differentiation: Investigating conceptualisation, characteristics, and methods used. *Review of Education*, 9(1), 161–198. <https://doi.org/10.1002/rev3.3238>
- Grissom, J. A., & Redding, C. (2016). Discretion and disproportionality: Explaining the underrepresentation of high-achieving students of color in gifted programs. *AERA Open*, 2(1). <https://doi.org/10.1177/233285841562217>
- Guryan, J., Ludwig, J., Bhatt, M. P., Cook, P. J., Davis, J. M. V., Dodge, K., Farkas, G., Fryer, R. G., Jr., Mayer, S., Pollack, H., Steinberg, L., & Stoddard, G. (2023). Not too late: Improving academic outcomes among adolescents. *American Economic Review*, 113(3), 738–765. <https://doi.org/10.1257/aer.20210434>
- Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality and student achievement. *Journal of Public Economics*, 95(7), 798–812. <https://doi.org/10.1016/j.jpubeco.2010.11.009>
- Heppen, J. B., Walters, K., Clements, M., Faria, A. -M., Tobey, C., Sorensen, N., & Culp, K. (2011). *Access to Algebra I: The effects of online mathematics for grade 8 students* (NCEE 2012-4021). <https://files.eric.ed.gov/fulltext/ED527394.pdf>
- Heß, S. (2017). Randomization inference with Stata: A guide and software. *The Stata Journal: Promoting Communications on Statistics and Stata*, 17(3), 630–651. <https://doi.org/10.1177/1536867X1701700306>
- Huffaker, E., Novicoff, S., & Dee, T. S. (2025). Ahead of the game? Course-taking patterns under a math pathways reform. *Educational Researcher*, 54(2), 91–102. <https://doi.org/10.3102/0013189X241309642>
- Imberman, S. A., Kugler, A. D., & Sacerdote, B. I. (2012). Katrina's children: Evidence on the structure of peer effects from hurricane evacuees. *American Economic Review*, 102(5), 2048–82. <https://doi.org/10.1257/aer.102.5.2048>
- Institute for Education Sciences. (2014). *Teaching strategies for improving algebra knowledge in middle and high school students*. <https://ies.ed.gov/ncee/wwc/PracticeGuide/20>
- Jackson, C. K., & Mackevicius, C. L. (2024). What impacts can we expect from school spending policy? Evidence from evaluations in the United States. *American Economic Journal: Applied Economics*, 16(1), 412–446. <https://doi.org/10.1257/app.20220279>
- Kalogrides, D., & Loeb, S. (2013). Different teachers, different peers: The magnitude of student sorting within schools. *Educational Researcher*, 42(6), 304–316. <https://doi.org/10.3102/0013189X13495087>

- Kang, C. (2007). Classroom peer effects and academic achievement: Quasi-randomization evidence from South Korea. *Journal of Urban Economics*, 61(3), 458–495. <https://doi.org/10.1016/j.jue.2006.07.006>
- Kariya, T., & Rosenbaum, J. E. (1999). Bright flight: Unintended consequences of detracking policy in Japan. *American Journal of Education*, 107(3), 210–230.
- Kim, S. (2018). *Return to Algebra II: The effect of mandatory math coursework on post-secondary attainment* (SSRN Scholarly Paper ID 3351235). Social Science Research Network. <http://dx.doi.org/10.2139/ssrn.3351235>
- Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, 49(4), 241–253. <https://doi.org/10.3102/0013189X20912798>
- Krone Phillips, E. (2019). *The make-or-break year*. New Press.
- Krueger, A. (2003). Economic considerations and class size. *The Economic Journal*, 113(485), F34–F63. <https://doi.org/10.1111/1468-0297.00098>
- Labaree, D. F. (1988). *The making of an American high school: The credentials market and the central high school of Philadelphia, 1838–1939*. Yale University Press.
- Lafortune, J. (2018). *Essays on the distribution and effectiveness of educational resources*. University of California Press.
- Lamdin, D. J. (1996). Evidence of student attendance as an independent variable in education production functions. *The Journal of Educational Research*, 89(3), 155–162. <https://doi.org/10.1080/00220671.1996.994132>
- Lefgren, L. (2004). Educational peer effects and the Chicago Public Schools. *Journal of Urban Economics*, 56(2), 169–191. <https://doi.org/10.1016/j.jue.2004.03.010>
- Legette, K. B., & Kurtz-Costes, B. (2021). Curricular tracking, students' academic identity, and school belonging. *The Journal of Early Adolescence*, 41(7), 961–981. <https://doi.org/10.1177/0272431620977659>
- Levin, H. M., McEwan, P. J., Belfield, C., Bowden, A. B., & Shand, R. (2018). *Economic evaluation in education: Cost-effectiveness and benefit-cost analysis*. Sage. <https://doi.org/10.4135/9781483396514>
- Liu, J., & Loeb, S. (2021). Engaging teachers: Measuring the impact of teachers on student attendance in secondary school. *Journal of Human Resources*, 56(2), 343–379.
- Long, M. C., Conger, D., & Iatarola, P. (2012). Effects of high school course-taking on secondary and postsecondary success. *American Educational Research Journal*, 49(2), 285–322. <https://doi.org/10.3102/0002831211431952>
- Malamud, O., Mitrut, A., Pop-Eleches, C., & Urquiola, M. (2025). Self-, peer-, and teacher perceptions under school tracking. *The Review of Economics and Statistics*, 1–41. https://doi.org/10.1162/rest_a_01555
- McEachin, A., Domina, T., & Penner, A. (2020). Heterogeneous effects of early Algebra across California middle schools. *Journal of Policy Analysis and Management*, 39(3), 772–800. <https://doi.org/10.1002/pam.22202>
- Mehana, M., & Reynolds, A. J. (2004). School mobility and achievement: A meta-analysis. *Children and Youth Services Review*, 26(1), 93–119. <https://doi.org/10.1016/j.childyouth.2003.11.004>
- Moses, R. P., & Cobb, C., Jr. (2001). Organizing Algebra: The need to voice a demand. *Social Policy*, 31(4), 4–12.
- Murata, A. (2013). Diversity and high academic expectations without tracking: Inclusively responsive instruction. *Journal of the Learning Sciences*, 22(2), 312–335. <https://doi.org/10.1080/10508406.2012.682188>
- National Education Association (2021). *Estimates of school statistics, 1960 through 2021; and unpublished tabulations*. U.S. Department of Commerce, Bureau of Economic Analysis, National Income and Product Accounts, table 6.6D.

- Nomi, T., & Allensworth, E. (2009). "Double-dose" algebra as an alternative strategy to remediation: Effects on students' academic outcomes. *Journal of Research on Educational Effectiveness*, 2(2), 111–148. <https://doi.org/10.1080/19345740802676739>
- Nomi, T., & Allensworth, E. (2013). Sorting and supporting: Why double-dose algebra led to better test scores but more course failures. *American Educational Research Journal*, 50(4), 756–788. <https://doi.org/10.3102/0002831212469997>
- Nomi, T., & Raudenbush, S. W. (2016). Making a success of "Algebra for All": The impact of extended instructional time and classroom peer skill in Chicago. *Educational Evaluation and Policy Analysis*, 38(2), 431–451. <https://doi.org/10.3102/0162373716643756>
- Oakes, J. (2005). *Keeping track: How schools structure inequality*. Yale University Press.
- Oakes, J., Ormseth, T., Bell, R. M., & Camp, P. (1990). *Multiplying inequalities: The effects of race, social class, and tracking on opportunities to learn mathematics and science*. RAND Corporation.
- Penner, A. M., Domina, T., Penner, E. K., & Conley, A. (2015). Curricular policy as a collective effects problem: A distributional approach. *Social Science Research*, 52, 627–641. <https://doi.org/10.1016/j.ssresearch.2015.03.008>
- Prast, E. J., Weijer-Bergsma, E. V. de, Kroesbergen, E. H., & Luit, J. E. H. V. (2015). Readiness-based differentiation in primary school mathematics: Expert recommendations and teacher self-assessment. *Frontline Learning Research*, 3(2), article 2. <http://journals.sfu.ca/flr/index.php/journal/article/view/163>
- Pustejovsky, J. E., & Tipton, E. (2018). Small-sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models. *Journal of Business & Economic Statistics*, 36(4), 672–683. <https://doi.org/10.1080/07350015.2016.1247004>
- Redding, C. (2019). A teacher like me: A review of the effect of student–teacher racial/ethnic matching on teacher perceptions of students and student academic and behavioral outcomes. *Review of Educational Research*, 89(4), 499–535. <https://doi.org/10.3102/0034654319853545>
- Rosenbaum, J. E. (1999). If tracking is bad, is detracking better? *American Educator*, 23(4), 24.
- Sacerdote, B. (2014). Experimental and quasi-experimental analysis of peer effects: Two steps forward? *Annual Review of Economics*, 6(1), 253–272. <https://doi.org/10.1146/annurev-economics-071813-104217>
- Schiller, K. S., & Hunt, D. J. (2011). Secondary mathematics course trajectories: Understanding accumulated disadvantages in mathematics in grades 9–12. *Journal of School Leadership*, 21(1), 87–118. <https://doi.org/10.1177/105268461102100105>
- Sequoia Union High School District. (2018). *9th Grade Math Placement Guidelines*. https://www.carlmonths.org/documents/18-19%20Math%20Placement%20Chart%20to%20Board_022018.pdf
- StataCorp. (2023). *Stata 18 item response theory reference manual*. Stata Press.
- Stein, M. K., Kaufman, J. H., Sherman, M., & Hillen, A. F. (2011). Algebra: A challenge at the crossroads of policy and practice. *Review of Educational Research*, 81(4), 453–492. <https://doi.org/10.3102/0034654311423025>
- Tucker, J. (2023, May 10). *S.F. could bring back algebra in eighth grade. Here's what you need to know about city's math wars*. San Francisco Chronicle.
- Valiande, S., & Koutselini, D. M. I. (2009). *Application and evaluation of differentiation instruction in mixed ability classrooms*. University of Cyprus. <https://scispace.com/pdf/application-and-evaluation-of-differentiation-instruction-in-2s8gkugy8s.pdf>

- VanTassel-Baska, J., Xuemei Feng, A., Brown, E., Bracken, B., Stambaugh, T., French, H., McGowan, S., Worley, B., Quek, C., & Bai, W. (2008). A study of differentiated instructional change over 3 years. *Gifted Child Quarterly*, 52(4), 297–312. <https://doi.org/10.1177/0016986208321809>
- Vogt, F., & Rogalla, M. (2009). Developing adaptive teaching competency through coaching. *Teaching and Teacher Education*, 25(8), 1051–1060. <https://doi.org/10.1016/j.tate.2009.04.002>
- von Hippel, P. T. (2025). Multiply by 37 (or Divide by 0.027): A Surprisingly Accurate Rule of Thumb for Converting Effect Sizes From Standard Deviations to Percentile Points. *Educational Evaluation and Policy Analysis*, 47(3), 960–969. <https://doi.org/10.3102/01623737241239677>
- Vygotsky, L. S. (2011). The dynamics of the schoolchild's mental development in relation to teaching and learning. *Journal of Cognitive Education and Psychology*, 10(2), 198–211.
- Wainstein, L., College, R., Miller, C. E., Phillips, M., Yamashiro, K., & Melguizo, T. (2023). *Twelfth grade math and college success* (LAERI Research Report). <https://laeri.luskin.ucla.edu/wp-content/uploads/sites/22/2023/09/LAERITwelfthGradeMathandCollegeSuccessReport.pdf>
- Wells, A. S., & Serna, I. (2010). The politics of culture: Understanding local political resistance to detracking in racially mixed schools. *Harvard Educational Review*, 66(1), 93–119. <https://doi.org/10.17763/haer.66.1.274848214743t373>
- Welsh, R. O. (2017). School hopscotch: A comprehensive review of K–12 student mobility in the United States. *Review of Educational Research*, 87(3), 475–511. <https://doi.org/10.3102/0034654316672068>
- What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook, version 5.0*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. https://ies.ed.gov/ncee/wwc/Docs/referenceresources/Final_WWC-HandbookVer5_0-0-508.pdf
- Williams, R., Citkowicz, M., Miller, D. I., Lindsay, J., & Walters, K. (2022). Heterogeneity in mathematics intervention effects: Evidence from a meta-analysis of 191 randomized experiments. *Journal of Research on Educational Effectiveness*, 15(3), 584–634. <https://doi.org/10.1080/19345747.2021.2009072>
- Zwiers, J., Dieckmann, J., Rutherford-Quach, S., Daro, V., Skarin, R., Weiss, S., & Malamut, J. (2017). *Principles for the design of mathematics curricula: Promoting language and content development*. Understanding Language. <https://ul.stanford.edu/resource/supporting-ells-mathematics>

Manuscript received October 16, 2024

Final revision received October 20, 2025

Accepted November 4, 2025