Effects of double-dose algebra on college persistence and degree attainment

Takako Nomi[a], Stephen W. Raudenbush[b,c,1], and Jake J. Smith[c]

[a]School of Education, Saint Louis University, St. Louis, MO 63103; [b]Department of Sociology, The University of Chicago, Chicago, IL 60637; and [c]Harris School of Public Policy, The University of Chicago, Chicago, IL 60637

Contributed by Stephen W. Raudenbush, April 6, 2021 (sent for review September 9, 2020; reviewed by Adam Gamoran and Florencia Torche)

In 2003, Chicago Public Schools introduced double-dose algebra, requiring two periods of math—one period of algebra and one of algebra support—for incoming ninth graders with eighth-grade math scores below the national median. Using a regression discontinuity design, earlier studies showed promising results from the program: For median-skill students, double-dose algebra improved algebra test scores, pass rates, high school graduation rates, and college enrollment. This study follows the same students 12 y later. Our findings show that, for median-skill students in the 2003 cohort, double-dose significantly increased semesters of college attended and college degree attainment. These results were not replicated for the 2004 cohort. Importantly, the impact of the policy on median-skill students depended largely on how classes were organized. In 2003, the impacts on college persistence and degree attainment were large in schools that strongly adhered to the cut-score-based course assignment, but without grouping median-skill students with lower-skill peers. Few schools implemented the policy in such a way in 2004.

STEM education | double-dose algebra | inequality | regression discontinuity | multilevel analysis

In the United States, low-income and minority students are completing college at low rates compared to higher-income and majority peers. Even as overall college graduation rates have risen in the United States in recent decades, social and racial inequalities in 4-y degree attainment rates have persisted, and in some cases risen (1). Among 25- to 29-y-olds, the gap in bachelor's degree completion between Blacks and Whites has held steady for nearly two decades, such that by 2017 42% of Whites but only 19% of Blacks had a bachelor's degree. For Hispanics, 16% had such a degree. Although racial gaps in associate's degree completion have narrowed slightly, the Black–White gap currently stands at 16 percentage points, while the Hispanic–White gap stands at 25 percentage points (1). The disparity by income level is stark as well: In 2017, 58% of students from the top quartile of family income, but only 11% from the bottom quartile of family income quartile, completed a bachelor's degree (2).

For scholars and policymakers concerned with social and ethnic inequality these persistent patterns are troubling as unequal college degree attainment is a key driver of income inequality (3, 4). The wage premium for college completion has grown steadily since the 1960s and is currently higher in the United States than in many other developed countries (5–7). As a result, persistent disparities in college completion among different social groups contribute substantially to income inequality (8).

Academic readiness, especially in math, is a key factor that accounts for disparities in college completion (9–11). In particular, success in high school algebra—the conventional focus of ninth-grade mathematics education in the United States—has been found to be a "gateway" to further math courses (12, 13). Indeed, passing high-school algebra predicts success in subsequent academic math courses such as geometry, trigonometry, and calculus; success in these courses, in turn, predicts college enrollment, as well as completing high school (14, 15). In college, foundational math skills, including algebra, are closely related to college completion: Among those who attend college, failing to pass the minimum requirements in math is arguably the primary obstacle to completing a degree (16–18). However, remedial education as a strategy to support students in college has shown disappointing results (19–22). Alternatively, improving math preparation in high school, and thus alleviating the need for remediation in college, may increase college attendance, persistence, and degree attainment, but research evidence supporting this proposition, reviewed below, is limited.

Given the outsized importance of math preparation for college access and success, one promising intervention for increasing college completion rates is double-dose algebra. Chicago Public Schools—the third-largest district in the nation, which predominantly serves low-income and minority students—implemented double-dose algebra for 2 y in the early 2000s. For entering ninth-grade students whose eighth-grade math scores were below the national median, this intervention doubled instructional time by assigning two periods of math: the required algebra class, plus a second support class designed to bolster requisite skills for algebra. Prior studies have found that double-dose algebra in Chicago produced positive effects on median students' test scores, high school graduation, and college entry. However, the impact of double-dose on college persistence and completion has not been studied.

In this paper, we provide analysis of the long-term impacts of double-dose algebra on college outcomes, tracking student outcomes in the 12 y following the intervention. We use a regression discontinuity design (RDD) to estimate the impact

Significance

To obtain high-paying jobs, students from low-income backgrounds need to complete college at higher rates. Could remedial math coursework in high school help them do so? Some have argued that adolescence is too late to intervene, or that schools cannot overcome systemic social problems that hinder student success. Yet, evidence from Chicago suggests otherwise. We find that ninth graders assigned to "double-dose algebra"—an algebra class plus a second class designed to bolster algebra skills—were more likely to stay in, and complete, college compared to similar students who did not take double-dose algebra. However, when students were placed in double-dose classes with much-lower-skilled peers the program had no effect.

Author contributions: T.N. and S.W.R. designed research; T.N. and S.W.R. performed research; S.W.R. contributed new reagents/analytic tools; T.N. and S.W.R. analyzed data; and T.N., S.W.R., and J.J.S. wrote the paper.

Reviewers: A.G., W. T. Grant Foundation; and F.T., Stanford University.

The authors declare no competing interest.

See online for related content such as Commentaries.

[1]To whom correspondence may be addressed. Email: sraudenb@uchicago.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2019030118/-/DCSupplemental.

Published June 28, 2021.

of this intervention on college persistence and degree attainment. The RDD approximates a randomized trial for students whose eighth-grade math scores were in the neighborhood of the median score on a math pretest. These students attended 67 neighborhood high schools in Chicago. Such schools serve the majority of high school students in Chicago. Students in our sample whose eighth-grade scores were within 25 percentiles of the median were 67% African American and 27% Hispanic, and 94% were low-income as indicated by eligibility for free or reduced lunch. Their eighth-grade math scores were, on average, very close to the national median on the Iowa Test of Basic Skills. In a year before the policy was introduced, however, only 6.4% of such students had received a 4-y college degree 9 y after entering high school in the absence of double-dose.

We report three main findings. First, for the median-skill students entering high school in 2003, taking double-dose algebra, on average, produced positive impacts on semesters of college attendance and college degree attainment. These effects were substantial: For those induced to take double-dose by scoring below the cut point in 2003, the percentage of students obtaining a college degree increased from about 12 to 17% and the percentage of students obtaining a 4-y degree increased approximately from 6 to 10%.

Second, the effects of the intervention in 2003 varied substantially across schools. The degree of compliance with the policy and the process by which students were assigned to classrooms appear decisive in producing these favorable results. In particular, in schools where median-skill students taking double-dose were assigned to algebra classes composed of students of similar skill, the effects were large. However, in schools where students taking double-dose were assigned to classes composed of low-skill students, double-dose had no measurable effect. Based on interviews with key policy makers, we consider barriers to sustained implementation that must be overcome if the policy is to realize its potential to increase the odds that disadvantaged students will succeed in college, and, consequently, have better economic opportunities down the road.

Third, for students entering high school 1 y later, in 2004, positive effects of double-dose did not appear. We propose and empirically test an explanation with supporting evidence for why positive effects were not replicated. We show that, for the 2004 cohort, compliance with the cut-score-based assignment declined substantially. Moreover, most schools with high compliance placed median-skill students in classes with low-skill peers. Once we account for the implementation of the policy, results from 2003 and 2004 were consistent. This finding is corroborated by an analysis of schools that substantially changed policy implementation between 2003 and 2004.

We begin by considering the importance of increasing college attainment for low-income students, not only for reducing inequality but also for increasing economic productivity of the nation. We then review what is known about educational interventions that aim to increase college attainment. After presenting our empirical strategy and results, we conclude that double-dose algebra holds promise as a potentially simple and replicable strategy for achieving this goal. However, how the policy is implemented is crucial to the success of this reform. We identify open research questions concerning the optimal assignment of students to math classes and the impact of double-dose algebra on students whose initial skills are significantly below the median.

The Social and Economic Benefits of College Completion

Increasing Equality of Economic Opportunity. As mentioned above, low-income urban students are less likely than their more advantaged peers to obtain a 4-y degree (1, 2) and there is a large premium for college attendance and completion in terms of employment and wages (23). This premium persists across subgroups defined by gender and race (24). The college premium is not limited to 4-y degrees: There is evidence that 2-y college degrees and related occupational certifications also deliver wage returns (25). Thus, increasing college attendance, persistence, and completion among students from low-income backgrounds would likely enhance those students' success in a competitive labor market.

Increasing Economic Growth. Although rates of college completion are growing, evidence suggests that the supply of highly skilled labor in the United States is nonetheless still failing to keep up with demand. From 1980 to 2005, real earnings of individuals with a college degree rose much faster than real earnings of those without a college degree, while the fraction of the population receiving such degrees stagnated (4). Since 2005, college attendance and degree attainment have grown, yet the gap in wage returns between college- and non-college-educated adults has remained very large, suggesting that the demand for skills gained in college persistently outpaces the supply. (5, 26) It is therefore reasonable to conclude that any barriers that keep large segments of the population, including low-income and minority segments, from completing college are inhibiting the supply of skilled labor, which inhibits economic progress writ large.

Reducing Economic Inequality. Earlier economic research interpreted the large wage premium for college degree attainment as a key driver of income inequality (4, 26). More recent analyses have argued that the growth in the college wage premium reflects, in part, the fact that medium- and low-skilled workers have been losing their high-wage jobs due to automation and offshoring. In this view, there has been a "wage polarization" wherein technological innovation has simultaneously reduced the pay of less-skilled workers while increasing the pay of more-skilled workers (27). Increasing college completion of the broader population—and therefore increasing the supply of college-educated workers—is likely to put downward pressure on the college premium, by ensuring that supply more closely matches demand; this will arguably mitigate economic inequality and also reduce the number of low- and medium-skilled workers for the jobs that are increasingly disappearing. However, this view is not universally shared, as some analyses have found that increasing access to college education would only moderately reduce overall heterogeneity in earnings (28, 29).

Positive Effects of More Education on Nonacademic Outcomes. In addition to wage returns, additional education has also been shown to improve outcomes in other important areas of life that are largely stratified by social and ethnic background. Several studies have assessed the benefits of additional years of schooling on nonpecuniary outcomes by exploiting exogeneous changes in compulsory schooling laws; these studies have found favorable results in such outcomes as mortality, (30) blood pressure, (31) civic participation (30, 32), criminal activity (30, 33), divorce (30), teen pregnancy (30, 34), and overall life satisfaction (30).

Can Better K–12 Education Improve College Persistence and Completion?

Improving primary and secondary schooling for disadvantaged youth may seem like an obviously plausible strategy for increasing access to college. Yet, over the last half century, scholars have debated the extent to which school improvement can overcome inequality in skills and educational attainment. In 1965, Coleman et al. conducted a highly influential survey which concluded that family and household attributes were strong predictors of racial and socioeconomic disparities in academic outcomes, and that differences in schools' resources did not predict students' cognitive skills (35). Many subsequent surveys largely replicated that result, finding significant impacts of such nonschool factors as neighborhood deprivation (36–38) and parent job loss (39, 40) on students' educational attainment. However, recent experimental and quasi-experimental studies have shown that a range of policies, including expanding preschool education (41), improving elementary school curriculum (42, 43), using assessments to tailor instruction to children of varied skill (44–46), and reducing class size (47, 48) can all increase achievement and reduce social and ethnic disparities in educational outcomes.

Many of these interventions have been focused on preschool and elementary school. Indeed, some scholars have argued that interventions targeting adolescents come too late to shape the skills linked to academic success and success in the labor market (49). However, recent research has called this argument into question. For instance, certain innovative secondary charter schools have been found to improve academic performance among low-income, predominately minority youth (50, 51). Small high schools of choice (52), intensive tutoring of high-schoolers paired with cognitive behavioral therapy (53), and assigning high school students to effective teachers (50) have all been shown to improve the likelihood of graduating high school. Most relevant to our present study, earlier studies of ninth-grade double-dose algebra in Chicago showed increased test scores for median-skill students as well as higher rates of high school graduation and college enrollment (54, 55).

Yet, comparatively few of these studies have produced evidence that changes in schooling can increase college completion. In part this reflects the challenge of following the same students long enough to obtain data on completed schooling. Moreover, the few studies that have followed students long-term have often focused on small samples of deeply disadvantaged students for whom college attendance was too rare to allow for statistically powerful evaluation of impacts on college degree attainment (56, 57).

There are, however, important exceptions: Krueger and Whitmore found that class size reduction in elementary school increased application to college (58). Chetty et al. provided compelling evidence that assignment to effective teachers in kindergarten and elementary school predicted college persistence and earnings, as well as reduced teenage pregnancy (59, 60). Deming et al. found that students who won a random lottery to attend their most-preferred high school achieved higher rates of college enrollment and degree attainment (61). In the Early College High School Project, where high schools were randomly assigned to offer college courses in partnership with local colleges and increase social and academic support to help students succeed, these high schools saw their students enroll in and complete college at higher rates (62). Jackson found that offering high-school students and teachers monetary incentives, paired with teacher training, improved students' college persistence and wages (63).

Many of these interventions, however, are likely to be costly, producing large new expenses for schools, and/or unintended consequences for students. Reducing class size requires schools to invest in additional personnel, which may also disrupt local educator labor markets in unanticipated ways. For instance, California's class size reduction policy in 1996 inadvertently intensified statewide competition for a limited supply of highly qualified teachers, with some evidence that low-income districts suffered the most from this competition, and the results of the policy were disappointing (64). Policies that introduce new incentives could similarly induce competition for scarce resources. Furthermore, opening new innovative charter schools, recalibrating school-choice policies, and modifying teacher recruitment and evaluation are system-wide changes that require substantial investment and/or community buy-in. Individualized tutoring also incurs substantial costs if implemented on a broad scale. By contrast, double-dose algebra is a comparatively simple curricular change that a high school may adopt without broader structural change in governance, costs, or incentives. The main constraint is that the policy, while not increasing the number of teachers overall, does require more math teachers.

The present study adds to the slim body of direct empirical evidence that school interventions, even those instituted as late as high school, can, in fact, increase college attainment. Moreover, it considers how a double-dose approach—a relatively low-cost and simple intervention—may lead to greater rates of 4-y degree completion.

Classroom Composition, "Algebra for All," and Double-Dose Algebra

For most of the 20th century, US high schools used "tracking" to organize instruction and accommodate the heterogeneity of the student body. In the typical comprehensive high school, a vocational track composed mostly of boys focused instruction on skilled trades (e.g., carpentry, electricity, and printing), while a general track, disproportionately composed of girls, emphasized home economics (e.g., cooking and sewing) and clerical skills (65, 66). A much smaller academic track provided college-preparatory instruction in math and science as well as history, literature, and foreign languages. Academic math focused on algebra, geometry, and trigonometry, while math in the nonacademic tracks emphasized simple arithmetic skills needed for the most basic financial literacy, limiting the postsecondary schooling options for these "low-track" students (66).

Detracking and Classroom Composition. During the latter half of the century, tracking was widely criticized for its role in educational stratification. Students from low-income families and minority students were overrepresented in vocational and general tracks, and the tracking system was frequently regarded as an engine for reproducing social and racial inequality (65). As these criticisms grew, so did public concern about the overall low performance of US students, especially in math and sciences, and particularly in light of a job market that increasingly valued skills in math, science, and engineering. In response, policymakers called for raising academic rigor by standardizing curricula. The National Governor's Association recommended toughening high school graduation requirements, and by the early 2000s "College-Prep Curriculum for All" was in place in nearly 20 states (67). In math, an increasing number of districts and schools, particularly in urban areas, began to require algebra for all students in ninth grade or earlier.

The movement to abolish rigid tracking and to replace it with academically challenging instruction for all had profound implications not only for the content of instruction but also for the composition of the high school classroom. With all students now expected to study algebra, geometry, and more advanced subjects, teachers, policymakers, and scholars continued to debate whether high school classes should be segregated by students' prior achievement, which tended to be correlated with students' socioeconomic and racial–ethnic background.

Some prior research provides evidence that assigning higher- and lower-skill students to the same high-school classroom can be successful. In one New York district, students at all skill levels saw math score gains after accelerated mathematics curriculum was brought to mixed-skill classrooms (68). However, to understand its success, we highlight several important aspects of this reform. First, this reform took place in a very-high-achieving school (only 6% of students scored below the 40th percentile in the national distribution). Thus, heterogeneous classes there look vastly different from those in an urban district like Chicago, where many students are far behind. Second, this reform also offered new math workshops for students with weak math skills. Third, the school was in an affluent district serving predominantly middle-class students. Instructional resources needed to counter the challenge of teaching students of heterogeneous in skill are more likely to be available in affluent districts than in low-income districts.

An alternative view is that, while all students should experience high-level academic instruction, it would be more efficient to assign students to classes based on prior skill. Qualitative and survey evidence show that teaching in mixed-skill classes is often difficult, as teachers struggle to present course material at a level appropriate for students of all skill levels (69–71). Several randomized and quasi-experimental studies have found that heterogeneous classroom peer composition reduced the quality of instruction (72, 73), and increasing instructional time in elementary school produced the greatest gains when no one was placed with substantially lower-skilled peers (74, 75). Similarly, when schools experienced a large influx of low-skill students, the achievement of incumbent students has been found to decline (76).

Algebra for All. Chicago was at the forefront of the detracking movement, introducing "Algebra-for-All" in 1997 as part of the "College-Prep-for-All Curriculum" policy to address low academic achievement. Under this reform, low-skill students who would formerly have been assigned to remedial math were instead assigned to standard algebra classes. This reform shifted the composition of the typical ninth-grade algebra class: Math skills of incoming students were, on average, both lower and more heterogeneous than in the prepolicy year. A rigorous evaluation of this reform found that lower-skill students did not benefit from algebra instruction; moreover, the reform reduced the test scores of high-performing students, as they experienced declines in the skill level of their peers (77). Given that the student population in Chicago was predominantly from low-income, minority backgrounds, this reform likely suppressed the math achievement of high-performing students from these backgrounds, which likely intensified achievement disparities between them and other high-achieving students in middle-class, suburban high schools.

Double-Dose Algebra. The failure of "Algebra for All" highlighted an underlying problem: Many students lacked sufficient mathematical background to benefit from algebra instruction. To remedy this problem, the district introduced double-dose algebra in 2003. This new policy required all incoming ninth-grade students with eighth-grade math scores below the national median to take algebra, plus an additional period of instruction designed to build foundational prealgebra skills. However, schools varied dramatically in the degree to which they complied with cut-score-based course assignment, and in the extent to which they segregated double-dose classrooms by students' prior skills. Double-dose thereby created a natural experiment that made it possible to study the impact of two key aspects of math instruction: instructional time and classroom peer composition.

The design of double-dose facilitates valid causal inference regarding the effect of the policy. Specifically, for students with eighth-grade math scores near the national median, course assignment was effectively random. Thus, a comparison of outcomes between those who scored just above and just below the cut point approximates a randomized experiment. However, the key limitation is that any generalization drawn from such a comparison applies only to those we call "median-skill" students.

In many schools, such median-skill students who were assigned to double-dose tended to take classes composed primarily of low-skill peers. This peer composition reflected how schools organized algebra classes. Specifically, the district recommended that the algebra and remedial classes be taught "back-to-back," that these two classes be taught by the same teachers, and that students in the remedial class also take the algebra class together. To follow these recommendations, some schools tended to separate algebra classes based on the double-dose eligibility cut point. For double-dose students, this meant that their algebra classes were primarily composed of low-skill peers. However, in other schools, median-skill double-dose students attended algebra class with classmates whose skills were similar to their own. Because double-dose was implemented in each of 67 neighborhood high schools, we can regard this policy as launching 67 natural experiments and estimate the effect of scoring below the cut point on median-skill students in each school; we can then investigate how this effect depends on the extent to which the cut-score-based assignment affected the classroom's peer composition.

The current study builds on earlier studies of this policy, which showed promising results. Using RDDs, prior studies showed assignment to double-dose improved 10th-grade algebra test scores and pass rates as well as subsequent outcomes, including 11th-grade math achievement, high school graduation, and college enrollment (29, 30). Importantly, however, the policy impact varied widely across schools and depended substantially on how schools organized algebra classes. Specifically, the impact on algebra scores was largest in schools where the students just below the national median took algebra alongside peers with relatively similar skills. In comparison, no discernable impact was found in schools where the median-skill students took the course with low-skill peers (78).

Research Questions. By following the same students 9 y after entering grade 9, we pursue answers to three main questions:

1) For median-skill students, did assignment to double-dose affect semesters of college attendance and degree attainment?
2) Not all schools followed the cut-score-based course assignment rule; therefore, what were the effects of taking double-dose on college outcomes on those students induced to do so by having scored below the cut point?
3) To what extent did the effects of the policy on college outcomes depend on the skill composition of the algebra class? In particular, did median-skill students benefit more from double-dose if their class was composed of similarly skilled peers, rather than less-skilled peers?

To answer these questions, we study the first 2 y of policy implementation, 2003 and 2004.

Research Design and Methods

Sample. Our analytic sample consists of first-time ninth-grade students entering regular Chicago high schools in 2003 and 2004. We exclude selective enrollment schools because most of their students had eighth-grade test scores above the eligibility cut point; thus, the policy did not apply to these schools. The resulting sample size is 33,033 students attending 67 regular high schools. Students were largely non-White and on the free/reduced lunch program. Nearly 20% were special education students. In the average high school, a small majority of students scored below the national median on the pretest (double-dose–eligible).

Data Sources. The National Student Clearinghouse provides college enrollment and graduation records, and we linked these records to precollege data, allowing us to follow ninth-grade students for 9 y after the end of grade 9. The outcomes of interest are semesters of college attendance, attainment of any college degree (including a 2-y degree or certificate or a 4-y degree), and attainment of a 4-y college degree. We count college semesters for students who enrolled in college by the third semester after graduating from high school (thereby allowing one gap year), and if they took no more than one semester off after starting college.

Causal Identification. We use an RDD, taking advantage of the cut-score-based double-dose course assignment. Our first set of analyses consider the "intention to treat" (ITT) effect—that is, the impact of being assigned to double-dose by virtue of scoring below the cut point on the pretest. The ITT effect, though of interest to policymakers, will underestimate the impact of actually taking double-dose because not all students followed the policy assignment: Some students scoring below the 50th percentile cut point did not take double-dose algebra, and some students scoring above the cut-point did take double-dose. Thus, our second analysis considers the impact of taking double-dose. This is applicable to those induced to do so by scoring below the cut point. We use the scoring below the cut-point as an instrumental variable to estimate this effect, known as the "complier average causal effect" (CACE) (79).

Using an RDD design, we model the association between the eighth-grade math scores and each outcome, and a discontinuity in this association at the cut score is regarded as evidence of the impact. We adopt three strategies to detect evidence of such discontinuities: 1) graphical evidence, 2) evidence from a school-specific parametric analysis that uses a piecewise quadratic functional form using all data as well as students within the 20th percentile points of the cut point, and 3) nonparametric analyses, using a local-linear fit, that ignore school membership but restrict our attention to those students whose eighth-grade test score was within 10 or 20 percentile points of the cut point. The graphical results enable us to visualize impacts and assess the validity of our functional form. The parametric analysis has the advantage of fully accounting for school-specific effects of the policy and is comparatively precise statistically but relies on functional form assumptions. The nonparametric analyses relax those assumptions at some cost to statistical precision and do not incorporate school-specific effects. Detailed descriptions of these models can be found in *SI Appendix*, sections A and B. Our results are largely convergent across these analyses.

Results

Policy Implementation. Fig. 1*A* shows the relationship between double-dose algebra enrollment rates on the vertical axis and math percentile scores on the horizontal axis for the 2003 and 2004 cohorts, where the math score of zero indicates the cut score (the 50th percentile). There is a large discontinuity at the cut point such that scoring below the cut point is associated with a large increase in the likelihood of taking double-dose. This increase was much smaller in 2004 than 2003. Scoring below the cut point also affected the skills of one's peers in algebra classes, as shown in Fig. 1*B*, with classroom mean peer skills on the vertical axis. A large average drop in mean peer skills as a function of scoring below the cut point was observed in both years.

The Average Impact of Being Assigned to Double-Dose Algebra (ITT). Results were reasonably similar across specifications. For median-skill students entering high school in 2003, our findings suggest that being assigned to double-dose algebra increased semesters of college attendance and the probability of degree attainment. However, the data do not support such conclusions for students entering in 2004 (Table 1).

Semesters of college attended. For median-skill students in 2003, we find consistent evidence of positive ITT impacts on total semesters: Estimated effects for the global, 20th percentile bandwidth, and 10th percentile bandwidth models are, respectively, 0.468 semesters (SE = 0.138), 0.427 semesters (SE = 0.201), and 0.540 semesters (SE = 0.258). In comparison, no discernable impacts were found for 2004; the estimated effects of the three models are −0.055 semesters (SE = 0.147), 0.043 semesters (SE = 0.224), and −0.156 semesters (SE = 0.261). See Fig. 2.

Degree attainment. Among those median-skill students not assigned to double-dose, the estimated probability of receiving a college degree (including a certificate, a 2-y degree, or a 4-y degree) was 0.146. We estimate assignment to double-dose increased this probability by 0.034 (SE = 0.014) under the global model, 0.031 (SE = 0.017) under the 20% bandwidth model, and 0.032 (SE = 0.021) under the 10% bandwidth model. For 2004, we see little evidence of policy impacts; estimated effects from the three models are 0.017 (SE = 0.013), 0.001 (SE = 0.017), and −0.010 (SE = 0.020). See Fig. 3.

Four-year degree attainment. Among those median-skill students not assigned to double-dose, the estimated probability of receiving a 4-y college degree was 0.070. We estimate assignment to double-dose increased this probability by 0.033 (SE = 0.012) under the global model, 0.025 (SE = 0.013) under the 20% bandwidth model, and 0.026 (SE = 0.015) under the 10% bandwidth model. For 2004, we see little evidence of policy impacts; estimated effects from the three models are 0.017 (SE = 0.013), 0.001(SE = 0.017), and −0.010 (SE = 0.020). See Fig. 4.

The Average Impact of Taking Double-Dose Algebra (CACE). Not all students scoring below the cut point took the assigned support course, and some students scoring above the cut point did take the support course. Our second analysis examines the impact of taking double-dose algebra on those induced to do so by having scored below the cut point. Fig. 2*A* portrays our theoretical model. The average impact of scoring below the cut point on taking double-dose—that is, the probability of taking double-dose if scoring below the cut point minus the probability of taking double-dose if scoring above the cut point—is denoted by γ. The average impact of taking double-dose on the outcome for the compliers—those induced to

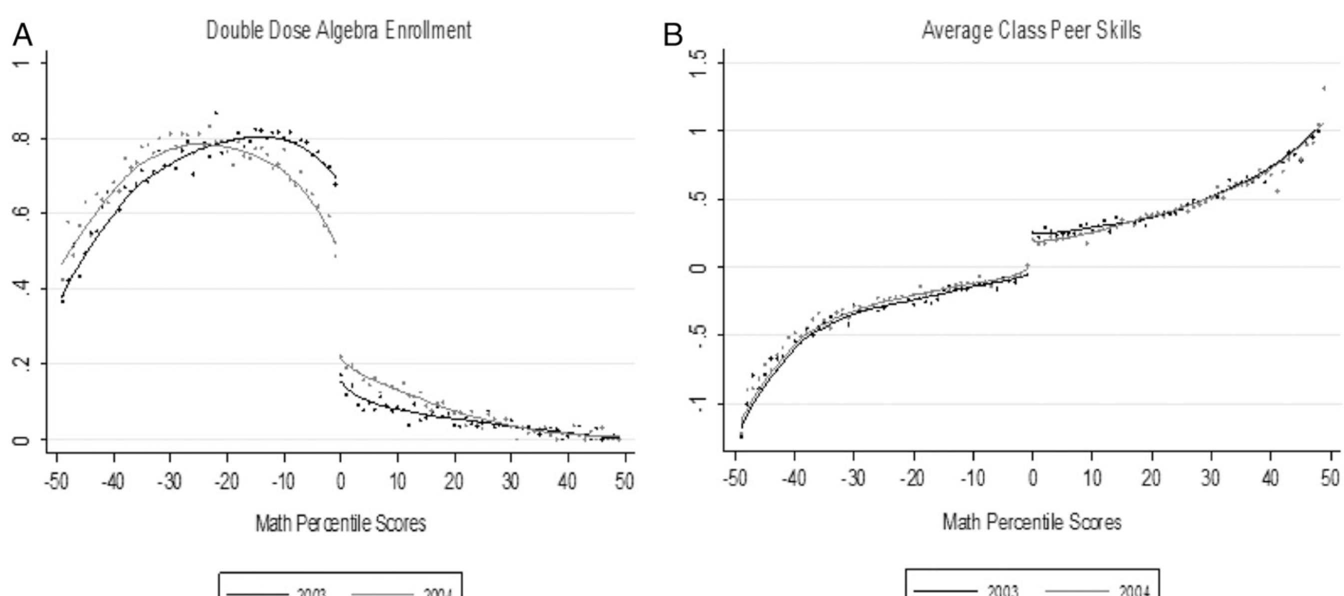

Fig. 1. (*A*) Proportion of students enrolled in double-dose as a function of eighth-grade pretest. (*B*) Average pretest of classroom peers (in SD units) as a function of eighth-grade pretest. Note: The dark line is a LOESS (locally estimated scatterplot smoothing) curve for the 2003 cohort and the light line is a LOESS curve for the 2004 cohort. Reprinted from ref. 54 by permission of Informa UK Limited, trading as Taylor & Francis Group, https://www.tandfonline.com/.

https://doi.org/10.1073/pnas.2019030118

Table 1. The average impact of being assigned to double-dose algebra

Outcomes	2003 Estimate	2003 SE	2004 Estimate	2004 SE
Total semesters				
Control mean	9.082***	0.220	9.496***	0.371
Treatment effects: all students	0.468***	0.138	−0.055	0.147
20% bandwidth	0.427**	0.201	0.043	0.224
10% bandwidth	0.540**	0.258	−0.156	0.261
Any degree/certificate				
Control mean	0.146***	0.018	0.141***	0.028
Treatment effects: all students	0.034**	0.014	0.017	0.013
20% bandwidth	0.031*	0.017	0.001	0.017
10% bandwidth	0.032	0.021	−0.010	0.020
4-y degree				
Control mean	0.070***	0.013	0.084***	0.023
Treatment effects: all students	0.033***	0.012	0.007	0.009
20% bandwidth	0.025*	0.013	−0.021	0.012
10% bandwidth	0.026*	0.015	−0.023	0.015
No. of students				
All students	19,800		20,491	
20% bandwidth	10,286		10,418	
10% bandwidth	5,368		5,408	
No. of schools	62		66	

***$P < 0.01$, **$P < 0.05$, *$P < 0.1$.

do so by scoring below the cut point—is δ. The average impact of scoring below the cut point on the outcome (ITT impact) is therefore

$$\beta = \gamma\delta \quad [1]$$

under several key assumptions: 1) Among students whose skill is near the median, scoring below the cut point is effectively random (exogeneity); 2) scoring below the cut point cannot reduce the probability of taking double-dose (monotonicity); and 3) scoring below the cut point can have no impact on the outcome other than by affecting participation in double-dose (exclusion restriction). We must also assume $\gamma > 0$, which is clearly met here. Under these assumptions, we can identify the impact of taking double-dose algebra on the compliers ("CACE") as

$$\delta = \beta/\gamma. \quad [2]$$

The pattern of results for CACE is similar to the ITT effects but a bit larger in magnitude. Though the results vary slightly depending on the model specification, the evidence consistently suggests those induced to take double-dose stayed in school longer: Estimates from the global, 20th, and 10th percentile bandwidth models are, respectively, 0.713 semesters (SE = 0.183), 0.621 semesters (SE = 0.250), and 1.014 semesters (SE = 0.488). Students were also more likely to obtain a degree with estimates of 0.053 (SE = 0.016), 0.047 (SE = 0.026), and 0.060 (SE = 0.039) from the three models. They were more likely to obtain a 4-y degree (estimates of 0.048 [SE = 0.014], 0.036 [SE = 0.019], and 0.048 [SE = 0.029]). In contrast, the estimated impacts for 2004 were not consistent and none of them was statistically significant (Table 2).

Accounting for Differences between Cohorts. In sum, for the 2003 cohort we see clear evidence that, on average, double-dose algebra increased college persistence and degree attainment. However, we see no such evidence for the 2004 cohort. We find that substantial differences in how schools implemented the policy largely account for this discrepancy. First, as mentioned previously, overall compliance with the policy declined substantially in 2004: Fewer schools followed the cut-score-based double-dose algebra assignment in 2004 than in 2003. Second, most schools that did strongly comply in 2004 did so by placing their median-skill double-dose students in low-skill algebra classrooms. We will describe the empirical evidence supporting our conclusion that these two

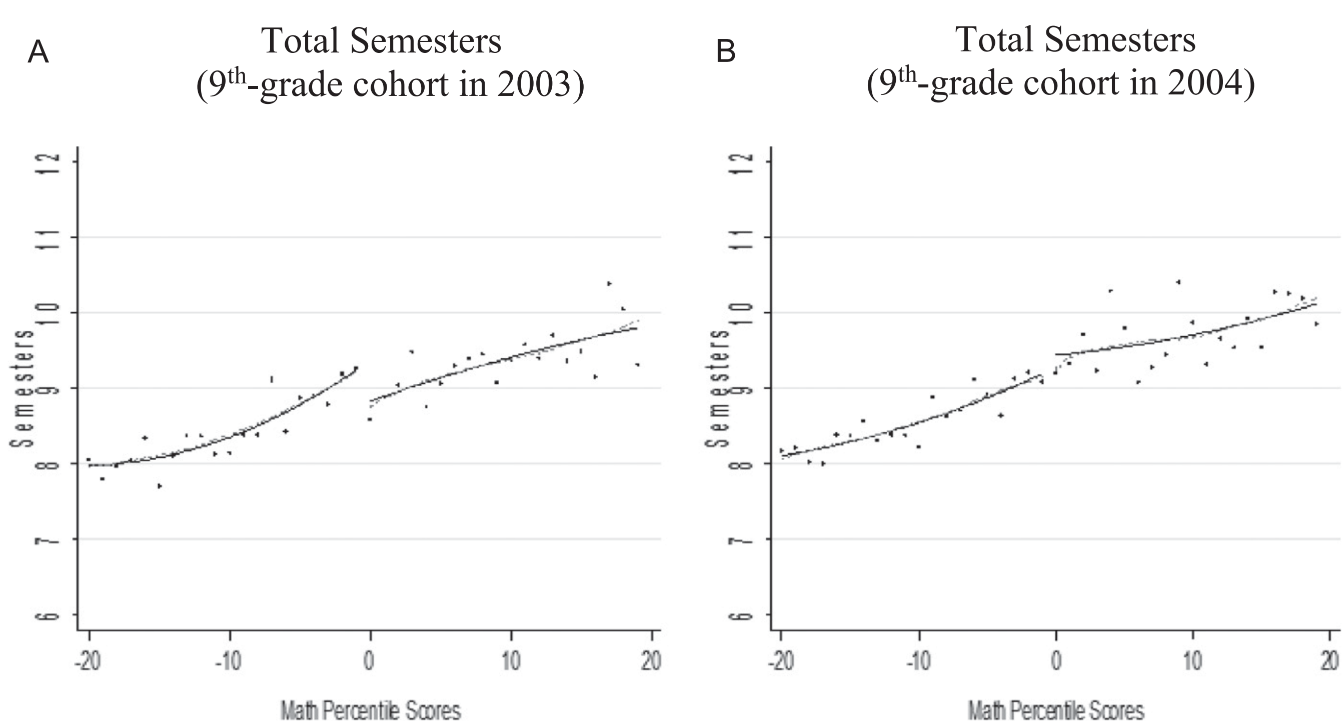

Fig. 2. Average total semesters as a function of math pretest. (*A*) 2003 cohort. (*B*) 2004 cohort. Note: The dashed line is a LOESS line and the solid line is a quadratic fit.

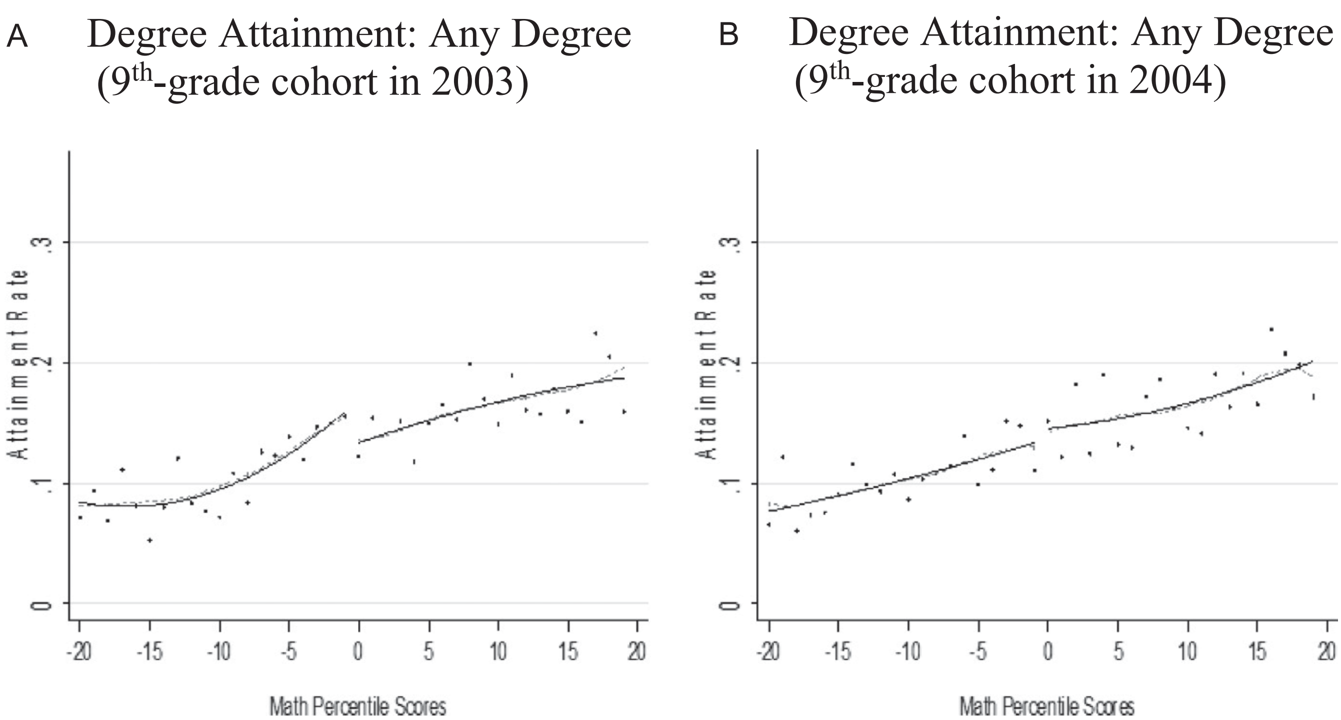

Fig. 3. Proportion who receive college degree as function of math pretest. (*A*) 2003 cohort. (*B*) 2004 cohort. Note: The dashed line is a LOESS line and the solid line is a quadratic fit.

factors, in combination, largely explain the difference between our findings for the two cohorts.

Compliance with the policy. We denote the compliance in each school j as γ_j, which measures the school-specific impact of scoring below the cut point on the probability of taking double-dose. We estimate the average value of $\gamma_j$to be 0.54 in 2004 and 0.31 in 2004, as illustrated graphically in Fig. 1*A*.

Classroom peer skill. Implementing the policy affected not only instructional time but also the average math skills of algebra classmates. As we saw in Fig. 1*B*, for students with median skills the

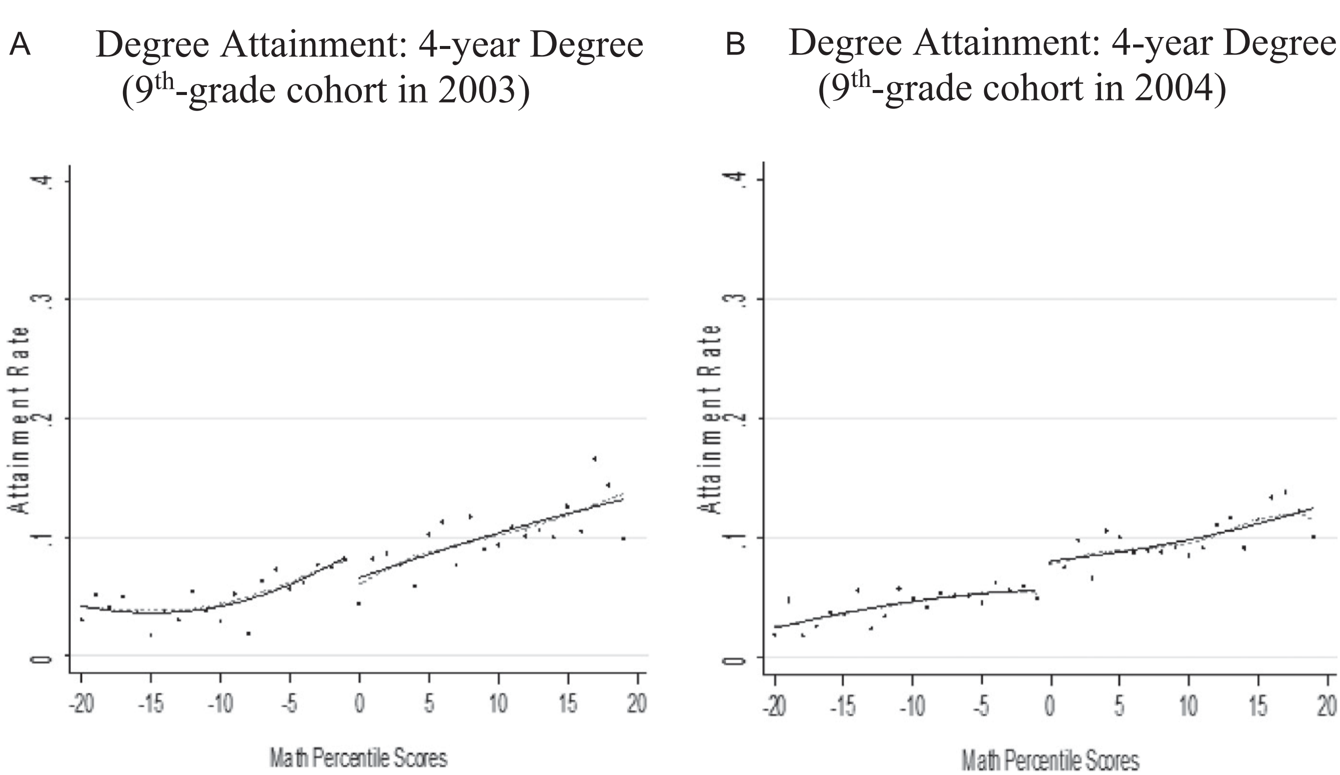

Fig. 4. Proportion who receive 4-y degree as function of math pretest. (*A*) 2003 cohort. (*B*) 2004 cohort. Note: The dashed line is a LOESS line and the solid line is a quadratic fit.

Table 2. The average impact of taking double-dose algebra

	2003		2004	
Outcomes	Estimate	Robust SE	Estimate	Robust SE
Total semesters				
Complier control mean	8.626***	0.211	9.477***	0.333
Treatment effects: all students	0.713***	0.183	0.181	0.252
20% bandwidth	0.621**	0.250	−0.114	0.301
10% bandwidth	1.014**	0.488	−0.548	0.913
Any degree/certificate				
Complier control mean	0.124***	0.017	0.153***	0.026
Treatment effects: all students	0.053***	0.016	0.028	0.019
20% bandwidth	0.047*	0.026	0.007	0.025
10% bandwidth	0.060	0.039	−0.034	0.071
4-y degree				
Complier control mean	0.055***	0.012	0.096***	0.020
Treatment effects: all students	0.048***	0.014	0.013	0.015
20% bandwidth	0.036*	0.019	−0.022	0.020
10% bandwidth	0.048*	0.029	−0.081	0.052
No. of students				
All students	19,800		20,491	
20% bandwidth	10,286		10,418	
10% bandwidth	5,368		5,408	
No. of schools	62		66	

***$P < 0.01$, **$P < 0.05$, *$P < 0.1$.

classroom-average peer skill was, on average, lower if they scored just below the cut point than had they scored just above the cut point. We denote the impact of scoring below the cut point on classroom peer skills in school j as θ_j.

Policy impact as the combined effect of instructional time and classroom composition. To account for the discrepancies between the two cohorts, we elaborate our theoretical model as displayed in Fig. 5, which indicates that taking double-dose algebra can affect the outcome variable through two channels: increasing instructional time and changing classroom composition such that one takes algebra with lower-skill peers. The dynamics of this process can vary from school to school. Under this model, the impact of scoring below the cut point on classroom peer skill in school j is

$$\theta_j = \gamma_j \lambda_j, \qquad [3]$$

where $\lambda_j$is the average impact of taking double-dose on classroom peer skill among the compliers. The effect of scoring below the cut point on the outcome is then

$$\beta_j = \gamma_j\left(\delta_{1j} + \lambda_j \delta_{2j}\right) = \gamma_j \delta_{1j} + \theta_j \delta_{2j}, \qquad [4]$$

where δ_{1j} is the school-specific impact on compliers of doubling instructional time, holding classroom peer skill constant, and δ_{2j} is the school-specific impact on compliers of classroom peer skill, holding constant instructional time. Model **4** could be elaborated to include a statistical interaction between γ_j and θ_j, but because we observe ITT impacts only when γ_j is high this additional complexity is unnecessary.

Models of this type were first estimated by Kling et al. (80). Reardon and Raudenbush (81) and Raudenbush et al. (82) revealed key additional assumptions required for valid estimation, in addition to those mentioned in *The Average Impact of Taking Double-Dose Algebra (CACE)*. Specifically, they are 1) scoring below the cut point cannot affect peer composition unless doing so induces one to take double-dose (exclusion restriction with respect to classroom peer skill); 2) at least some values of $\gamma_j$and θ_j are nonzero; 3) $\gamma_j$and θ_j are not collinear; and 4) θ_j cannot be associated with the causal effects δ_{1j} or δ_{2j} (independence of compliance and effect). Assumption 1 is standard and plausible. Assumptions 2 and 3 are clearly supported by our data (see Figs. 1 and 6). However, the fourth assumption is quite strong and not directly testable, and we shall take special pains to check its credibility in *Within-School Comparison of Policy Impact*.

Obtaining precise estimates of the causal effects δ_{1j} and $\delta_{2j}$for each school is not possible, but we can gain insight about our theoretical model by stratifying the schools as shown in Fig. 6. The figure displays a scatter plot of schools with the estimate $\hat{\gamma}_j$ on the horizontal axis and $\hat{\theta}_j$ on the vertical axis for each year. The lines on the horizontal and vertical axis represent the average values of these estimates across the two policy years ($\hat{\gamma}. = .494$, $\hat{\theta}. = -.212$). We see a negative correlation between these two estimates: Schools with large compliance (high $\hat{\gamma}_j$) had larger declines in classroom peer skills (low $\hat{\theta}_j$) for students with the median skills.

The following implementation differences between 2003 and 2004 appear key to understanding why the policy produced positive

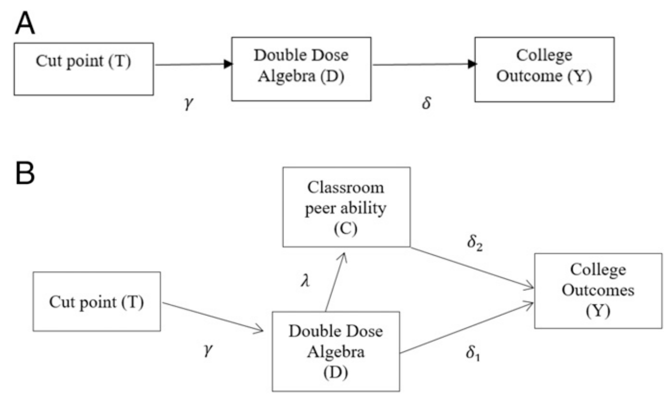

Fig. 5. Causal model. (*A*) Effect of scoring below the cut point on outcome through double-dose enrollment. (*B*) Effect of scoring below the cut point on outcome through double-dose enrollment and classroom peer skills. Reprinted with permission from ref. 78.

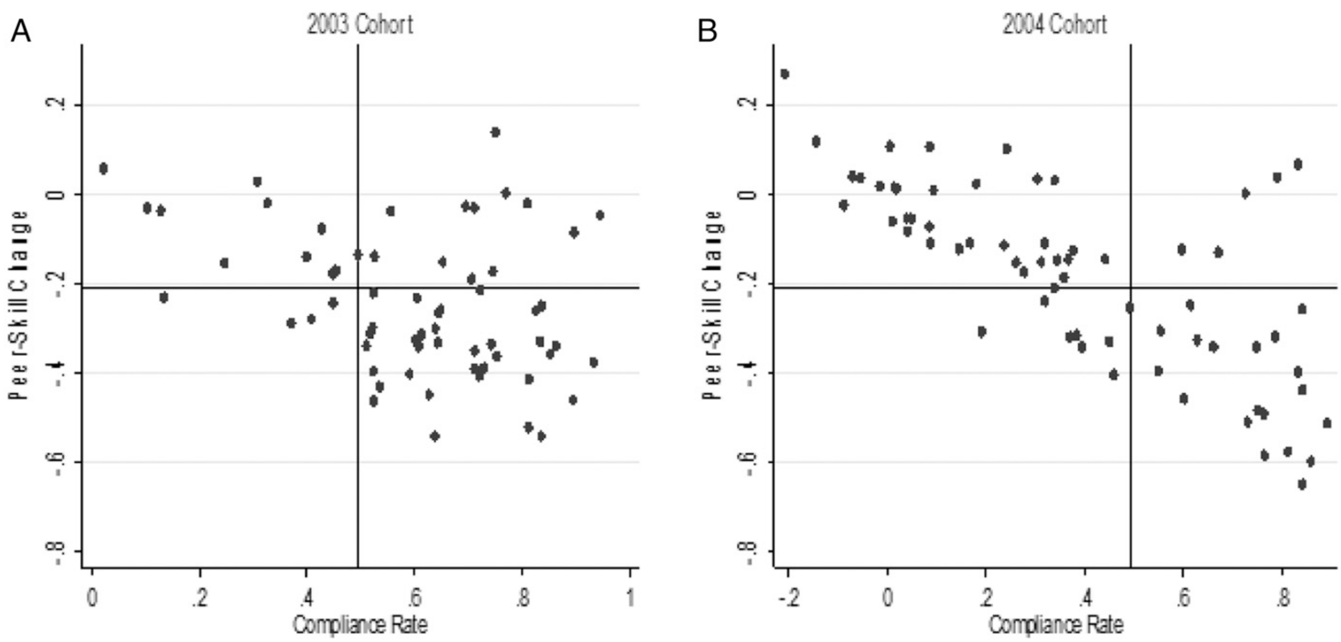

Fig. 6. School-average impact of scoring below the cut point on taking double-dose (horizontal axis) classroom-peer skill (vertical axis). (*A*) 2003 cohort. (*B*) 2004 cohort.

impacts in 2003 but not in 2004. First, overall compliance declined substantially in 2004. Second, the negative correlation between $\gamma_j$and θ_j was much stronger, $r = -0.74$ in 2004 as opposed to -0.37 in 2003. The low correlation in 2003 implies the presence of schools with relatively high compliance (high $\hat{\gamma}_j$) but where assignment to double-dose had only a small effect on classroom peer skills (less negative $\hat{\theta}_j$). These schools are in the upper-right quadrant of Fig. 6, and 13 such schools were found in 2003. In comparison, only five schools fall in this category in 2004, as indicated by high correlation between $\hat{\gamma}_j$ and $\hat{\theta}_j$.

Additional analysis of the 2003 and 2004 cohorts showed the largest ITT impacts occurred in schools in this upper-right quadrant (Table 3), although the large positive ITT impact estimates for the 2004 cohort are imprecise based on the small sample size (five schools). In contrast, there was no evidence of impact in schools in the lower-right quadrant for either cohort, which is composed of schools with high course compliance but large declines in peer skills for double-dose–eligible students at the cut point. There were no significant impacts in schools with very low levels of compliance for either cohort. This pattern of results is consistent with the results of an earlier study on 10th-grade algebra test scores: The effectiveness of the double-dose policy was undermined by the exposure of median skill students to low-skill peers (78).

Taken together, our results suggest that assignment to double-dose algebra had countervailing effects. High levels of compliance enhanced the effects of scoring below the cut point, but assigning median-skill students to classrooms composed of low-skill peers reduced these effects. We consider policy implications of this finding in *Discussion*.

Within-School Comparison of Policy Impact among Schools Differing in Classroom Peer Skills over Time with the Same Level of High Course Compliance. The validity of the differential policy impact as a function of peer skills relies on the assumption that θ_j is not correlated with δ_{1j} or δ_{2j} in Eq. **1** (assumption 4 above) (80, 81). This would be violated, for example, if schools that were particularly ineffective in teaching double-dose algebra were the very schools that assigned median-skill students to low-skill algebra classes, or if schools that can provide effective instruction regardless of classroom academic compositions are the schools that did not create low-skill classrooms for median-skill students. Such a violation would bias our assessment. To check the validity of this assumption, we investigated the change in policy effectiveness within the same schools among a subset of schools that had high course compliance in both years but that differed in policy implementation with regard to peer skills. Because the analysis compares ITT effects within schools, the assumption of no correlation between θ_j and δ_{1j} or δ_{2j} is no longer relevant. Furthermore, given that the overall instructional capacity of schools changes slowly, we reasoned that, if we were to find

Table 3. ITT impacts by peer skills: Schools with above-average course compliance in 2003

	Small declines (13 schools, 2,090 students)		Large declines (35 schools, 5,922 students)	
Outcomes	Estimate	SE	Estimate	SE
2003				
Total semesters	0.877*	0.489	0.190	0.247
Any degree/certificate	0.090**	0.040	0.014	0.020
4-y degree	0.060**	0.030	0.017	0.017
	(5 schools, 273 students)		(19 schools, 2,765 students)	
	Estimate	SE	Estimate	SE
2004				
Total semesters	2.456	1.531	0.255	0.419
Any degree/certificate	0.092	0.125	0.026	0.034
4-y degree	0.160	0.100	0.008	0.026

**$P < 0.05$, *$P < 0.1$. Note: This analysis uses 20th percentile bandwidth because the parametric model using all students is sensitive to outliners. The result using the 10th percentile is similar. Also, the analysis by pooling the two cohorts produced similar results.

strong associations between the change in classroom peer academic composition and policy effectiveness over a 1-y period in schools where compliance stayed the same, this would provide additional evidence of how peer skill composition shapes the overall policy impact. This within-school analysis is based on 11 high schools which had high course compliance in both years but where the impact of assignment to double-dose led to large shifts in peer skill one year (large negative θ_j) and small changes in the other year (null or small negative θ_j). Nine of them had small peer skill declines in 2003 and large declines in 2004; two schools had the opposite pattern. These schools were pooled for this analysis. Our statistical model is again a piecewise quadratic model that estimates the impact of scoring below the cut point on our outcomes but allows this impact to differ between 2003 and 2004 as a function of whether the decline in peer skills associated with double-dose assignment was small or large (see *SI Appendix*, section D for the statistical model).

The results for within-school comparison replicate those of between-school comparison (Table 4): No discernable ITT impacts were found when students at the cut point experienced large declines in classroom peer skills. However, ITT impacts within the same school were larger when such students experienced only a small decline in algebra-class peer skills.

Discussion

Our results reveal the potential power of high school curricular reform to improve the life chances of low-income, predominately minority students. Our results are generalizable to students whose pre-high school math skill is near the national median. Nearly 95% of these students in Chicago were minority students from low-income families. We find that offering an extra course to support algebra learning can significantly increase college persistence and degree attainment of these students. However, considering the implementation of this intervention appears key to understanding its effectiveness. Specifically, our RDD analysis revealed substantially positive results for schools that 1) strongly complied with the policy by following the cut-score-based double-dose course assignment and 2) placed median-skill students in algebra classes composed of peers with similar skills. In 2004, compliance declined, and in most schools that did comply strongly median-skill students tended to have low-skill peers.

These findings pose key unanswered questions: 1) Why was classroom peer-skill composition critical to the effectiveness of the policy for median-skill students; 2) what were the consequences of double-dose for low-skill students; and 3) why was effective implementation far less prevalent in 2004 than in 2003?

Why Does Class Peer-Skill Composition Matter? The findings of this study, as well as those of an earlier study of double-dose algebra (78) and other peer-effects studies (73–77), suggest that exposure to low-skill peers undermines learning of higher-skill students. A body of research on this topic supplies several potential explanations for this finding. The most plausible explanation, we believe, is that teachers have been found to tailor the conceptual level and pace of instruction to average skill level of their students (83). In a typical US high school, instructional differentiation occurs between classrooms; teachers then use "whole-class" instruction wherein the entire class moves at the same pace. In this setting, the progress of median-skill students would be delayed if instruction becomes slower-paced and less conceptually demanding in classes primarily composed to lower-skill peers. (65, 84)

Other explanations are also possible. School leaders may assign high-skill teachers to classrooms with high-skill students, leaving lower-skill students with less effective teachers (85). It could also be the case that lower-skill students have become discouraged, and that their low motivation may affect median-skill kids as well (86).

What Is the Effect of Double-Dose Algebra on Low-Skill Students? If we optimize for median-skill students by assigning them to algebra classes composed of students with similar skill, what would be the consequences for students of low skill? Would they suffer by not having higher-skill classmates? Would they benefit from taking an additional math course? How does one then design an optimal policy for instructional time and assignment of students to classes? Unfortunately, a study using the RDD cannot answer these important questions. However, an earlier study comparing observably similar students pre- and postpolicy provides some evidence that low-skill students benefited little from this policy (54). Students with pretest scores in the bottom quartile of the national distribution showed little improvement postpolicy compared to prepolicy. In comparison, students between the 25th and 50th percentile made significant gains, as we would expect from the results of the RDD analysis.

These results suggest that for students with very low incoming math skills additional instructional time alone was not enough to prepare them for algebra. We estimate that ~30% percent of the student population in Chicago fell into the "very-low-skill" category (based on a model predicting the likelihood of a student's scoring below the national median, using as predictors students' testing history, race, and indicators for socioeconomic status). The math skills of the average student in this very-low-skill population ranked at the 20th national percentile; nearly 60% of this group was identified as needing special education services. Of these students, 37% failed ninth-grade algebra, and fewer than half graduated from high school. The lack of test score improvement for these students post-double-dose suggests that this approach failed to meet their developmental needs. We speculate that a math intervention far more intensive than double-dose is essential to improve their high school and postsecondary outcomes.

Why Was Effective Implementation Far Less Prevalent in 2004 than in 2003? While the results of double-dose in 2003 are very encouraging, the decline in estimated impacts in 2004 is troubling. Why did implementation change in 2004? What are the obstacles to optimally implementing the policy and sustaining the positive effects? To answer these questions, we interviewed key architects of the policy. We learned that the district recommended 1) cut-score-based assignment, 2) the two math courses being scheduled "back to back," and 3) that both courses be taught by the same teacher and attended by the same students. Our interviews suggested that, while these guidelines were intended to provide coherent instruction, implementation may have imposed burdens on the schools, potentially disrupting the standard procedures for scheduling and teacher assignment. These challenges in the first year of the policy may have undermined the motivation to continue following the implementation guidelines in 2004. Importantly, however, any such change did not result in abandonment of the double-dose program. Even though compliance to the cut-score-based assignment declined in 2004, our data show

Table 4. Within-school comparison of ITT impacts by peer skills changes (high course compliance)

	Total semesters		Any degree		4-y degree	
	Estimate	SE	Estimate	SE	Estimate	SE
Intercept	9.109***	0.312	0.138***	0.021	0.081***	0.016
Below	−0.485	0.421	0.008	0.034	0.017	0.027
Below*small	1.046***	0.287	0.053**	0.023	0.033*	0.019
Small	−0.844***	0.204	−0.023	0.017	−0.004	0.013

Number of students, 3,502; number of schools, 11. ***$P < 0.01$, **$P < 0.05$, *$P < 0.1$.

that the overall student participation rates in double-dose in 2003 and 2004 were virtually identical.

Thus, it appears that what changed in 2004 was not a willingness to implement double-dose but rather that school personnel were more likely in 2004 than in 2003 to use their own judgement about who should take double-dose. Such a change would not necessarily reduce the impact of double-dose—that is, the policy might still have led students' outcomes to improve more than they would have if the district had not implemented the policy. However, because the RDD relies on adherence to the cut-score-based assignment to identify the program's impact, low compliance with the assignment rule reduces the capability of RDD to detect that impact.

Finally, an important difference between 2003 and 2004 was that the average incoming skills of all ninth-grade students declined in 2004. This might have put additional constraints on the schools' decisions about how to implement the program, including who should take the course, scheduling, and academic composition of algebra classes. Moreover, our data also showed that schools were more likely to enroll special education students in double-dose in 2004 than in 2003. This, in part, explains why, in those schools with high compliance in both years, median-skill students tended to attend classrooms with much-lower-skill peers in 2004.

Taken together, the evidence encourages us to speculate that an optimal policy would assign median-skill students to algebra classes with similarly skilled peers, while providing much more intensive and targeted instruction to students with very low skills. More research is required to test this speculation. In addition, despite the large effects we have described, the 4-y graduation rate for double-dose students with the national median skill remains low (10%). This suggests that additional, more intensive interventions could further raise 4-y degree completion among the urban low-income students that these Chicago students represent.

Conclusion

Double-dose algebra shows real promise for helping median-skill students in Chicago neighborhood schools to gain math skills, take more college-preparatory math classes, achieve more years of schooling, and earn college degrees. The success of the double-dose approach is consistent with evidence of a similar successful intervention in college—offering a remedial course along with a core math course (e.g., statistics) to those who do not meet the minimum math requirement (87). However, the longevity of the effects we observe for the 2003 cohort is unusual among studies of schooling interventions (88, 89). The double-dose strategy is plausibly cost-effective relative to other types of interventions with a comparable long-term impact, such as reducing class size or providing tutoring, which require substantial labor costs. More generally, our results bear on large questions concerning prospects for secondary-school reform. Scholars have long questioned what role, if any, high school improvement can play in improving workforce outcomes for low-income children, and whether high school may be too late for any meaningful intervention. Our results provide some cause for optimism.

A key caveat is that the effectiveness of this policy depends on assigning moderately skilled students to algebra classes with similarly (or higher-) skilled peers. This finding raises key questions for subsequent research. In particular, we need to know how to design optimal policies for lower-skill students and how to ensure that an effective double-dose policy can be sustained over time within the constraints of the organizational design of urban neighborhood high schools.

SOCIAL SCIENCES

Data Availability. Data for re-analysis can be obtained with permission of the Chicago Public Schools (CPS). Applications should be submitted to the Research Review Board of CPS.

ACKNOWLEDGMENTS. This study was supported by Grant R305A170602 from the Institute of Education Sciences entitled "Doubling Up? Understanding the Long-Term Effects of Ninth-Grade Algebra Reform on College Persistence and Graduation."

1. US Department of Education, "The condition of education 2018" (NCES 2018-144, US Department of Education, National Center for Education Statistics, 2018).
2. M. Cahalan, L. W. Perna, M. Yamashita, J. Wright, S. Santillan, "2018 indicators of higher education equity in the United States: Historical trend report" (The Pell Institute for the Study of Opportunity in Higher Education, Council for Opportunity in Education (COE), and Alliance for Higher Education and Democracy of the University of Pennsylvania, 2018).
3. J. Rothwell, "The hidden STEM economy" (Metropolitan Policy Program at Brookings, 2013).
4. C. Goldin, L. F. Katz, *The Race Between Education and Technology* (Harvard University Press, 2008).
5. D. H. Autor, Skills, education, and the rise of earnings inequality among the "other 99 percent". *Science* **344**, 843–851 (2014).
6. E. A. Hanushek, G. Schwerdt, S. Wiederhold, L. Woessmann, Returns to skills around the world: Evidence from PIAAC. *Eur. Econ. Rev.* **73**, 103–130 (2015).
7. J. James, "The college wage premium" (Federal Reserve Bank of Cleveland, 2012).
8. C. Funk, K. Parker, "Women and men in STEM often at odds over workplace equity" (Pew Research Center, 2018).
9. G. J. Duncan *et al*., School readiness and later achievement. *Dev. Psychol.* **43**, 1428–1446 (2007).
10. N. C. Jordan, D. Kaplan, C. Ramineni, M. N. Locuniak, Early math matters: Kindergarten number competence and later mathematics outcomes. *Dev. Psychol.* **45**, 850–867 (2009).
11. L. C. Landivar, "The relationship between science and engineering education and employment in STEM occupations" (US Department of Commerce, Economics and Statistics Administration, US Census Bureau, 2013).
12. J. Kaput, "Transforming algebra from an engine of inequity to an engine of mathematical power by "algebrafying" the K-12 curriculum" in *The Nature and Role of Algebra in the K-14 Curriculum: Proceedings of a National Symposium* (National Academies Press, 1998), pp. 25–26.
13. National Mathematics Advisory Panel, "The final report of the national mathematics advisory panel" (US Department of Education, 2008).
14. S. Y. Byun, M. J. Irvin, B. A. Bell, Advanced math course taking: Effects on math achievement and college enrollment. *J. Exp. Educ.* **83**, 439–468 (2015).
15. A. J. Bowers, R. Sprott, S. A. Taff, Do we know who will drop out? A review of the predictors of dropping out of high school: Precision, sensitivity, and specificity. *High Sch. J.* **96**, 77–100 (2012).
16. L. E. Richland, J. W. Stigler, K. J. Holyoak, Teaching the conceptual structure of mathematics. *Educ. Psychol.* **47**, 189–203 (2012).
17. K. B. Givvin, J. W. Stigler, B. Thompson, What community college developmental mathematics students understand about mathematics, part 2: The interviews. *MathAMATYC Educ.* **2**, 4–8 (2011).
18. P. Martorell, I. McFarlin, Jr, Help or hindrance? The effects of college remediation on academic and labor market outcomes. *Rev. Econ. Stat.* **93**, 436–454 (2011).
19. J. C. Calcagno, B. T. Long, "The impact of postsecondary remediation using a regression discontinuity approach: Addressing endogenous sorting and noncompliance" (NBER Working Paper 14194, National Bureau of Economic Research, 2008).
20. J. Scott-Clayton, O. Rodriguez, Development, discouragement, or diversion? New evidence on the effects of college remediation policy. *Educ. Finance Policy* **10**, 4–45 (2015).
21. G. J. Ngo, H. Kosiewicz, How extending time in developmental math impacts student persistence and success: Evidence from a regression discontinuity in community colleges. *Rev. Higher Educ.* **40**, 267–306 (2017).
22. D. Xu, M. Dadgar, How effective are community college remedial math courses for students with the lowest math skills? *Community Coll. Rev.* **46**, 62–81 (2018).
23. T. R. Bailey, C. R. Belfield, The false dichotomy between academic learning & occupational skills. *Daedalus* **148**, 164–178 (2019).
24. S. T. Cooper, E. Cohn, Internal rates of return to college education in the United States by sex and race. *J. Educ. Finance* **23**, 101–133 (1997).
25. J. E. Rosenbaum, C. E. Ahearn, J. E. Rosenbaum, *Bridging the Gaps: College Pathways to Career Success* (Russell Sage Foundation, 2001).
26. D. Autor, C. Goldin, L. F. Katz, "Extending the race between education and technology" (NBER Working Paper 26705, National Bureau of Economic Research, 2020).
27. M. Goos, A. Manning, A. Salomons, Explaining job polarization: Routine-biased technological change and offshoring. *Am. Econ. Rev.* **104**, 2509–2526 (2014).
28. R. Breen, I. Chung, Income inequality and education. *Sociol. Sci.* **2**, 454–477 (2015).
29. C. Jencks *et al*., *Inequality: A Reassessment of the Effect of Family and Schooling in America* (Basic Books, 1972).
30. P. Oreopoulos, K. G. Salvanes, Priceless: The nonpecuniary benefits of schooling. *J. Econ. Perspect.* **25**, 159–184 (2011).
31. N. Powdthavee, Does education reduce the risk of hypertension? Estimating the biomarker effect of compulsory schooling in England. *J. Hum. Cap.* **4**, (2010).
32. K. Milligan, E. Moretti, P. Oreopoulos, Does education improve citizenship? Evidence from the U.S. And the U.K. *J. Public Econ.* **88**, 1667–1695 (2004).

33. L. Lochner, E. Moretti, The effect of education on crime: Evidence from prison inmates, arrests, and self-reports. *Am. Econ. Rev.* **94**, 155–189 (2004).
34. M. Fort, "Just a matter of time: Empirical evidence on the causal effect of education on fertility in Italy" (European University Institute Working Paper No. 2007/22, 2007).
35. J. S. Coleman *et al*., *Equality of Educational Opportunity* (National Center for Educational Statistics, 1966).
36. C. L. Garner, S. W. Raudenbush, Neighborhood effects on educational attainment: A multilevel analysis. *Sociol. Educ.* **64**, 251–262 (1991).
37. R. Chetty, N. Hendren, L. F. Katz, The effects of exposure to better neighborhoods on children: New evidence from the Moving to Opportunity project. *Am. Econ. Rev.* **106**, 855–902 (2016).
38. J. Burdick-Will *et al*., "Converging evidence for neighborhood effects on children's test scores: An experimental, quasi-experimental, and observational comparison" in *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*, G. J. Duncan, R. J. Murnane, Eds. (Russell Sage Foundation, New York, 2011), pp. 255–276.
39. A. Kalil, P. Wightman, Parental job loss and children's educational attainment in black and white middle-class families. *Soc. Sci. Q.* **92**, 57–78 (2011).
40. M. B. Coelli, Parental job loss and the education enrollment of youth. *Labour Econ.* **18**, 25–35 (2011).
41. S. W. Raudenbush, R. D. Eschmann, Does schooling increase or reduce social inequality? *Annu. Rev. Sociol.* **41**, 443–470 (2015).
42. G. D. Borman *et al*., Final reading outcomes of the national randomized field trial of success for all. *Am. Educ. Res. J.* **44**, 701–731 (2007).
43. G. D. Borman, N. M. Dowling, C. Schneck, A multisite cluster randomized field trial of open court reading. *Educ. Eval. Policy Anal.* **30**, 389–407 (2008).
44. E. M. Hassrick, S. W. Raudenbush, L. Rosen, *The Ambitious Elementary School: Its Conception, Design, and Implications for Educational Equality* (University of Chicago Press, 2017).
45. C. M. Connor *et al*., Individualizing student instruction precisely: Effects of child x instruction interactions on first graders' literacy development. *Child Dev.* **80**, 77–100 (2009).
46. S. W. Raudenbush *et al*., Longitudinally adaptive assessment and instruction increase numerical skills of preschool children. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 27945–27953 (2020).
47. J. D. Finn, C. M. Achilles, Answers and questions about class size: A statewide experiment. *Am. Educ. Res. J.* **27**, 557–577 (1990).
48. J. D. Finn, Academic and non-cognitive effects of small classes. *Int. J. Educ. Res.* **96**, 125–135 (2019).
49. P. Carneiro, J. Heckman, "Human capital policy" (Institute for the Study of Labor [IZA] Discussion Paper 821, IZA, 2003).
50. D. Neal, *Information, Incentives, and Educational Policy* (Harvard University Press, 2018).
51. J. A. Angrist, P. A. Pathak, C. R. Walters, Explaining charter school effectiveness. *Am. Econ. J. Appl. Econ.* **5**, 1–27 (2013).
52. H. S. Bloom, R. Unterman, Can small high schools of choice improve educational prospects for disadvantaged students? *J. Policy Anal. Manage.* **33**, 290–319 (2014).
53. P. J. Cook *et al*., "The (surprising) efficacy of academic and behavioral intervention with disadvantaged youth: Results from a randomized experiment in Chicago" (NBER Working Paper 19862, National Bureau of Economic Research, 2014).
54. T. Nomi, E. Allensworth, Double-dose algebra as an alternative strategy to remediation: Effects on students' academic outcomes. *J. Res. Educ. Eff.* **2**, 111–148 (2009).
55. K. E. Cortes, J. S. Goodman, T. Nomi, Intensive math instruction and educational attainment: Long-run impacts of double-dose algebra. *J. Hum. Resour.* **50**, 108–158 (2015).
56. J. J. Heckman, S. H. Moon, R. Pinto, P. A. Savelyev, A. Yavitz, The rate of return to the high/scope perry preschool program. *J. Public Econ.* **94**, 114–128 (2010).
57. W. S. Barnett, L. N. Masse, Comparative benefit–cost analysis of the Abecedarian program and its policy implications. *Econ. Educ. Rev.* **26**, 113–125 (2007).
58. A. B. Krueger, D. M. Whitmore, The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from project STAR. *Econ. J. (Lond.)* **111**, 1–28 (2001).
59. R. Chetty, J. N. Friedman, J. E. Rockoff, Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *Am. Econ. Rev.* **104**, 2633–2679 (2014).
60. R. Chetty *et al*., "How does your kindergarten classroom affect your earnings? Evidence from Project STAR" (NBER Working Paper 16381, National Bureau of Economic Research, 2010).
61. D. J. Deming, J. S. Hastings, T. J. Kane, D. O. Staiger, School choice, school quality and postsecondary attainment. *Am. Econ. Rev.* **104**, 991–1013 (2014).
62. M. Song, K. L. Zeiser, "Early college, continued success: Longer-term impact of early college high schools" (American Institutes for Research, 2019).
63. C. Jackson, Do college-preparatory programs improve long-term outcomes? *Econ. Inq.* **52**, 72–99 (2014).
64. C. Jepsen, S. Rivkin, *Class Size Reduction, Teacher Quality, and Academic Achievement in California Public Elementary Schools* (Public Policy Institute of California, 2002).
65. S. Bowles, H. Gintis, *Schooling in Capitalist America* (Basic Books, New York, 1976).
66. J. Oakes, *Keeping Track: How Schools Structure Inequality* (Yale University Press, 2005).
67. Achieve, "Closing the Expectations Gap Report" (Achieve, American Diploma Project Network, 2009).
68. C. C. Burris, J. P. Heubert, H. M. Levin, Accelerating mathematics achievement using heterogeneous grouping. *Am. Educ. Res. J.* **43**, 137–154 (2006).
69. J. E. Rosenbaum, If tracking is bad, is detracking better? *Am. Educ.* **23**, 24–29, 47 (1999).
70. A. Gamoran, "Tracking and inequality: New directions for research and practice" in *The Routledge International Handbook of the Sociology of Education*, M. Apple, S. J. Ball, L. A. Gandin, Eds. (Routledge, London, 2010), pp. 213–228.
71. T. Loveless, "Tracking and detracking: High achievers in Massachusetts middle schools" (The Thomas B. Fordham Institute, 2009).
72. C. R. Gibbs, "Experimental evidence on early intervention: The impact of full-day kindergarten" (working paper, University of Virginia, Charlottesville, VA, 2014).
73. E. Duflo, P. Dupas, M. Kremer, Peer effects and the impacts of tracking: Evidence from a randomized evaluation in Kenya. *Am. Econ. Rev.* **101**, 1739–1774 (2011).
74. G. Hong, Y. Hong, Reading instruction time and homogeneous grouping in kindergarten: An application of marginal mean weighting through stratification. *Educ. Eval. Policy Anal.* **31**, 54–81 (2009).
75. C. M. Connor *et al*., A longitudinal cluster-randomized controlled study on the accumulating effects of individualized literacy instruction on students' reading from first through third grade. *Psychol. Sci.* **24**, 1408–1419 (2013).
76. S. Imberman, A. D. Kugler, B. Sacerdote, Katrina's children: Evidence on the structure of peer effects from hurricane evacuees. *Am. Econ. Rev.* **102**, 2048–2082 (2012).
77. T. Nomi, The unintended consequences of an algebra-for-all policy on high-skill students: Effects on instructional organization and students' academic outcomes. *Educ. Eval. Policy Anal.* **34**, 489–505 (2012).
78. T. Nomi, S. W. Raudenbush, Making a success of "algebra for all": The impact of extended instructional time and classroom peer skill in Chicago. *Educ. Eval. Policy Anal.* **38**, 431–451 (2016).
79. J. D. Angrist, G. W. Imbens, D. B. Rubin, Identification of causal effects using instrumental variables. *J. Am. Stat. Assoc.* **91**, 444–455 (1996).
80. J. R. Kling, J. B. Liebman, L. F. Katz, Experimental analysis of neighborhood effects. *Econometrica* **75**, 83–119 (2007).
81. S. F. Reardon, S. W. Raudenbush, Under what assumptions do site-by-treatment instruments identify average causal effects? *Sociol. Methods Res.* **42**, 143–163 (2013).
82. S. W. Raudenbush, S. F. Reardon, T. Nomi, Statistical analysis for multisite trials using instrumental variables with random coefficients. *J. Res. Educ. Eff.* **5**, 303–332 (2012).
83. R. Barr, R. Dreeben, *How Schools Work* (University of Chicago Press, Chicago, 1983).
84. A. Gamoran, M. Nystrand, M. Berends, P. C. LePore, An organizational analysis of the effects of ability grouping. *Am. Educ. Res. J.* **32**, 678–715 (1995).
85. S. Kelly, Are teachers tracked? On what basis and with what consequences. *Soc. Psychol. Educ.* **7**, 55–72 (2004).
86. R. N. Page, *Lower Track Classroom: A Curricular and Cultural Perspective* (Teachers College Press, New York, 1991).
87. A. W. Logue, D. Douglas, M. Watanabe-Rose, Corequisite mathematics remediation: Results over time and in different contexts. *Educ. Eval. Policy Anal.* **41**, 294–315 (2019).
88. T. R. Bailey, S. S. Jaggars, "When college students start behind: College completion series: Part five" (The Century Foundation, 2016).
89. E. Taylor, Spending more of the school day in math class: Evidence from a regression discontinuity in middle school. *J. Public Econ.* **117**, 162–181 (2014).