EdWorkingPaper No. 23-825

Interscholastic Policy Debate Promotes Critical Thinking and College-going: Evidence From Boston Public Schools

Beth E. Schueler
University of Virginia

Katherine E. Larned
Harvard University

Few interventions reduce inequality in reading achievement, let alone higher order thinking skills, among adolescents. We study "policy debate"—an extracurricular activity focused on improving middle and high schoolers' critical thinking, argumentation, and policy analysis skills—in Boston schools serving large concentrations of economically-disadvantaged students of color. Student fixed effects estimates show debate had positive impacts on ELA test scores of 0.13 SD, equivalent to 68% of a full year of average 9th grade learning. Gains were concentrated on analytical more than rote subskills. We find no harm to math, attendance, or disciplinary records, and evidence of positive effects on high school graduation and postsecondary enrollment. Impacts were largest among students who were lowest achieving prior to joining debate.

VERSION: August 2023

Suggested citation: Schueler, Beth E., and Katherine E. Larned. (2023). Interscholastic Policy Debate Promotes Critical Thinking and College-going: Evidence From Boston Public Schools. (EdWorkingPaper: 23-825). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/e8at-8836

INTERSCHOLASTIC POLICY DEBATE PROMOTES CRITICAL THINKING AND COLLEGE-GOING: EVIDENCE FROM BOSTON PUBLIC SCHOOLS

Beth E. Schueler, University of Virginia
Katherine E. Larned, Harvard University

Abstract: Few interventions reduce inequality in reading achievement, let alone higher order thinking skills, among adolescents. We study "policy debate"—an extracurricular activity focused on improving middle and high schoolers' critical thinking, argumentation, and policy analysis skills—in Boston schools serving large concentrations of economically-disadvantaged students of color. Student fixed effects estimates show debate had positive impacts on ELA test scores of 0.13 SD, equivalent to 68% of a full year of average 9th grade learning. Gains were concentrated on analytical more than rote subskills. We find no harm to math, attendance, or disciplinary records, and evidence of positive effects on high school graduation and postsecondary enrollment. Impacts were largest among students who were lowest achieving prior to joining debate.

Author's note: This research was generously funded by the Smith Richardson Foundation. We are grateful to a long list of people who helped this project come together, including Kimberly Willingham, Michael Wasserman, Steve Stein, Roger Nix, Marisa Suescun, Sarah Mayper, Mark Steinmeyer, Robert Litan, Apryl Clarkson, Elena Lanin, Judith Marcella, Briana Mezuk, and Daniel Shackelford. Isabelle Saillard provided research assistance and Shannon Kontalonis provided administrative support. We received valuable feedback from Joshua Goodman, Brendan Bartanen and Laura Bellows (along with other University of Virginia EdPolicyWorks affiliates), Matthew Steinberg, David Houston, Association of Education Finance and Policy annual conference participants, and our *Educational Evaluation and Policy Analysis* editor and anonymous reviewers. Finally, we are grateful to all the debaters, coaches, judges, administrators, parents, volunteers, and fellow pen-twirlers who made this research possible. Please address correspondence to Beth E. Schueler, Ridley Hall 268, University of Virginia, P.O. Box 400277, Charlottesville, VA 22904; beth_schueler@virginia.edu.

WHY STUDY DEBATE?

Educational inequality on the basis of race and social class in the U.S. was already unacceptably high prior to the pandemic (Duncan & Murnane, 2011; Reardon, 2011; Hanushek et al., 2020; Hashim et al., 2020). This was true both for K-12 indicators such as test scores (Reardon, 2011) as well as longer-term outcomes including years of schooling (Duncan & Murnane, 2011) and college going (Bailey & Dynarski, 2011). Unfortunately, all signs suggest educational disparities have only widened in the aftermath of COVID-19 due to differential pandemic-era experiences (Fahle et al., 2023; West & Lake, 2021). On the first post-pandemic, national civics exam, which measures 8th grade students' abilities in "evaluating, taking, and defending positions" (among other things), only 22 percent of students demonstrated proficiency. The exam further revealed inequality in civics performance based on economic and racial/ethnic dimensions (NAEP, 2023). These are skills critical for success in college and beyond (National Assessment Governing Board, 2018). Gaps in postsecondary attainment are particularly concerning given access to higher education is one of the few available pathways to social mobility for low-income Americans (Chetty, Friedman, Saez, Turner & Yagan, 2017) and affects lifetime well-being on a range of dimensions (Oreopoulos & Salvanes, 2011). For K-12 school systems to play a greater role in mitigating educational inequality, they must find avenues to more effectively serve low-income learners, as well as Black, Hispanic, and Indigenous students.

Much of the prior attention devoted to understanding how schools can accomplish this goal has focused on improving traditional classroom instruction in early elementary grades. This has been especially true when it comes to reading achievement where researchers emphasize the third grade turning point. Indeed, if students have not learned to read by this time, they are at higher risk of challenges later in life (Feister, 2013). In contrast, a review of experimental evaluations illustrates the rarity of research-based interventions that improve reading outcomes once students are in secondary school (Kraft, 2020). Relatedly, Carniero and Heckman (2003) argue that benefit-cost ratios of educational interventions decline across the life course, justifying greater investments for young children. Many observers conclude that intervening in the secondary school years may be "too late" or not worth the costs (Cook et al., 2014).

Additionally, the literature has focused more heavily on reforms designed to catch students up to proficiency benchmarks by mastering basic skills rather than the higher order analytic, critical thinking, and argumentation skills needed for long-term academic success as well as effective civic participation. This is likely, at least in part, due to the federal No Child Left Behind (NCLB) Act's target of universal proficiency in math and reading which encouraged the adoption and evaluation of policies designed to meet that goal. In more recent years, the promotion and adoption of Common Core State Standards (CCSS) was intended to refocus school systems on improving the higher order skills for college and career readiness. Though several states stepped back from the CCSS testing consortia, many have retained CCSS-aligned standards rebranded with a state-specific identity (Jochim, 2016). Despite widespread concern about student competencies in these areas, limited evidence exists on activities designed to strengthen higher order skills, particularly among economically-disadvantaged students of color.

Policy debate represents just such an activity intended to train secondary school students in literacy, critical thinking, argumentation, and policy analysis through preparation for and participation in interscholastic competition. Typically, extracurricular debate programs are disproportionately found in private and high-income public schools (Jacques, Basch, Fera & Basch, 2022). However, not-for-profit organizations have emerged over the past four decades to expand access in school systems that serve large concentrations of low-income students and students of color. We study debate in the Boston Public Schools, home to a relatively well-established league. We capitalize on districtwide longitudinal data to address the following research question: What is the effect of participating in policy debate on student academic achievement and attainment? We use student fixed effects methods to estimate impacts on K-12 outcomes, comparing students to themselves over time in years when they do and do not participate in debate. We find debate had positive effects on English Language Arts (ELA) test scores and the magnitude of these impacts is meaningful for an intervention targeting secondary school students. Gains were largest among the students who were lowest performing at baseline and impacts were concentrated on the reading standards that represent higher order subskills. We find no evidence of slippage on math achievement, school attendance or disciplinary infractions.

We also examine effects on high school graduation and postsecondary enrollment by comparing debaters to observably similar students who never participated in debate, using multiple comparison groups and matching methods with a rich set of covariates to minimize selection bias. We find evidence of positive impacts on graduation and postsecondary enrollment, driven by increased enrollment in 4-year and private schools. Overall, results provide policymakers a rare promising program for reducing inequality in reading achievement, analytical thinking skills, and educational attainment among secondary school students.

WHAT IS POLICY DEBATE?

Policy debate is an interscholastic, competitive, extracurricular activity for which teams of students engage in structured argumentation about public policy issues. While “speech and debate” and “forensics” refer to a diverse set of activities from impromptu speaking to mock trial, our study focuses on a particular public speaking activity known as “policy debate” or “cross-examination debate.” In policy debate, participants focus on a single resolution (i.e., topic) for the entire academic year, requiring them to learn about one policy area in depth (e.g., criminal justice, immigration, arms sales). In contrast, for impromptu speaking, students receive the topic minutes before they speak. As a result, the focus of many speech activities is more on developing a polished speaking style than on the substance of arguments or policies. Policy debate students, on the other hand, must rely on their substantial knowledge of the evidence related to their policy topic and their effective use of evidence to back up an argument to succeed, in addition to their ability to speak persuasively about the issues and think on their feet.

Policy debaters work in two-person teams (usually within a school-wide team) to research the topic and then craft and defend their arguments at interscholastic tournaments where they alternate between affirming and negating the resolution. Typically, the affirmative team proposes a specific “plan”—a policy proposal consistent with the resolution—and the negative team raises various disadvantages to that proposal, counter-proposals, and philosophical critiques. The debate consists of eight speeches plus four “cross-examination” periods for questioning as well as a pre-set amount of preparation time debaters can use at any time during the debate. This allows for several rounds of back-and-forth between the teams refuting initial arguments and then refuting counter-arguments to their responses such that a high level of clash and analysis is possible, beyond a superficial presentation of the main arguments on two sides of an issue. Students have a limited time to lay out their initial positions and refute opponents’ arguments—each speech and cross-examination period are timed and students must stop speaking when the timer rings. At the end of the debate round, the judge—typically a coach, college debater, teacher, or parent volunteer—who has taken notes on the debate, provides a written and verbal decision about which team won and why. A debate round includes a total of

69 minutes of speech, cross-examination, and prep time. Including the judge's deliberation and decision delivery, the whole debate typically lasts between 75 and 90 minutes.

Why might we expect all of this to pay off academically? First, the central skill that ensures success in policy debate is the ability to construct and deliver a compelling argument that is well-supported by both reasoning and evidence (Mitchell, 1998). This skill is assessed by mandatory ELA exams—particularly those aligned with Common Core standards—and is also required for success in writing effective papers and participating verbally in class for advanced coursework in high school as well as college. While research supports the benefits of extended learning time (e.g., Kidron & Lindsay, 2014) and extracurricular activities for student achievement (Lipscomb, 2007; Crispin, 2017; Stevenson, 2010; Cuffe et al., 2017), debate may be of higher value for academic achievement than sports or clubs given its alignment with the academic language and writing skills schools seek to develop through curricula.

Second, the research that debate students conduct in advance of tournaments requires them to develop skills in reading and interpreting advanced non-fiction texts, often including social science research, to find evidence that supports their positions. During debate rounds themselves, debaters are often confronted with new advanced non-fiction texts that their opponents submit as evidence. They must then read and refute those texts within a limited amount of time and under pressure. Debaters can distinguish themselves by explicitly contrasting the strengths of various pieces of text-based evidence. They are trained to consider not only the content of the texts but also the relative credibility and objectivity of the sources. These skills should also pay off on standardized exams, as well as in coursework.

Third, debate may provide a mechanism for motivating academic engagement. In many ways, debate flips the roles found in a traditional classroom. Rather than passively listening to an adult deliver a lecture, in a debate, students are at the front of the room, taking ownership of their learning, and teaching the adult judge (a teacher, debate coach, or community volunteer) as well as peers on the opposing team about their own policy proposals. Students have opportunities to be more actively engaged in the learning experience than in a typical classroom. The topics are directly related to current events and involve direct questioning of authority as the topics ask debaters to consider what policies the federal government should adopt and which it should abandon or amend. Debate may therefore provide some students with a reason to come to school.

Another goal of the debate leagues is to change school culture such that a debate team's success at competitions energizes the school population as a whole, much like interscholastic sports, providing further motivation for debater effort and engagement. The time pressure and competitive nature of the activity adds additional excitement, consistent with sociologist James Coleman's (1959; 1961) seminal studies documenting the power of interscholastic competition. Debate has been identified by scholars of "deeper learning" as a promising avenue for developing student engagement with rigorous academic content. Mehta & Fine (2019) argue that the most powerful learning experiences successfully, "integrate seemingly opposing virtues: mastery, identity, and creativity" (p. 42). Debate does this by asking students to master a unique skill, to become experts in a particular policy area, and to identify with a group and culture. It also encourages significant creativity given the nearly infinite numbers of arguments and proposals that can be marshalled to win any given debate and the strategic advantage that comes with surprising an opponent with an unexpected argument or novel piece of evidence.

Fifth, debate may help students develop a host of 21st century learning skills that promote success in school. Given the regimented structure of a debate round, with timed speeches and a limited number of minutes of "prep time" students can use throughout the debate, debaters are incentivized to learn time management and the critical thinking skills needed to prioritize the arguments that will result in maximum impact within the limited time they are allotted. Many debaters at the highest levels of competition deliver their speeches at a rapid pace that would be unintelligible to a newcomer for the purpose of getting out as much content as possible, making it difficult for the other team to respond to all arguments (typically, if a team fails to respond to an argument, the judge considers that argument conceded). For these reasons, policy debate may have more in common to a strategic thinking game than a public speaking activity.

Debate also requires independent organization and preparation—skills also needed for academic success—in advance of debate rounds in which students will be required to make decisions on their own without the guidance of a teacher or coach. Students must develop the ability to effectively listen to their opponents' arguments, to respectfully cross-examine (question) their opponents, and to work as a team toward a shared goal, making the theory of change consistent with a substantial research literature demonstrating the benefits of learning activities that incorporate cooperation with competition (Johnson, Johnson & Stanne, 2000; Slavin, 1983; Johnson, Maruyama, Johnson, Nelson & Skon, 1981; Slavin, 1980).

Finally, debate may encourage academic success through exposure to a college-going culture. The judges who watch debate rounds at tournaments, determine winners and losers, and provide feedback after each debate, are often current or former college-level debaters, and become mentors for the middle and high school debaters. Tournaments are sometimes held on college campuses and there is a strong culture of attending debate summer camps, also often held on college campuses, to prepare for the upcoming season. This exposure may encourage college-going, consistent with experimental research showing that even college visits can positively impact engagement, self-management, and college-going self-efficacy (Swanson et al., 2019).

Debate advocates point to a long list of well-known high-achieving adults—particularly those in public policy-related positions—including U.S. presidents, Supreme Court justices, legislators, media personalities, academics, lawyers, and policy analysts and advisors, who have participated in debate and credit much of their professional success to the activity (Litan, 2018).

THE BOSTON DEBATE LEAGUE

Over one million students have been members of the National Speech and Debate Association since it was founded in 1925 (NSDA, 2019). Historically, opportunities to participate in these activities have been concentrated in schools serving limited numbers of low-income students of color, mirroring broader patterns of unequal access to high-quality extracurricular programs (Pederson, 2005; Quinn, 1999; Fredricks & Simpkins, 2012). To address this opportunity gap, the first "urban debate league" (as they were original described) was established in the mid-1980s as a partnership between Emory University (where one of the top college debate programs at the time was housed) and the Atlanta Public Schools (Winkler, 2011). In 2002, the National Association of Urban Debate Leagues (NAUDL) was founded with the goal of extending debate access to majority-low-income, urban school districts nationwide. NAUDL currently serves over 11,000 students in 679 schools located in 22 cities across the country and aims to nearly double its reach to serve 20,000 debaters by 2023 (NAUDL, 2019).

Our research focuses on the Boston Debate League (BDL)—one of the 22 leagues associated with NAUDL—which partners with the Boston Public Schools (BPS) "to integrate argumentation and competitive debate into public schools in Boston to develop critical thinkers ready for college, career, and engagement with the world around them… with a commitment to serving students of color and other students who have been denied these educational opportunities" (Boston Debate League, 2022). Although it operates several initiatives related to argumentation training, the BDL's flagship program is "After-School Debate" for which BDL "partners with schools and teacher-coaches to launch and grow debate teams, and hosts city-wide debate tournaments, a summer camp, and additional programs to spark students' critical thinking, and engagement with the world around them" (Boston Debate League, 2022).

BDL was founded in 2005 with seven partner schools and today includes 40 school-based teams in public middle and high schools in Boston, Chelsea, and Somerville, Massachusetts. Our research focuses on Boston public schools and the period from 2007-08 to 2016-17 for which we were able to link debate participation data with student-level administrative data provided by BPS. As one of the more established NAUDL member leagues, Boston is a ripe context to study the impact of policy debate. While debate is most often offered at the high school level, the BDL also offers middle school debate, and runs a small but growing Spanish language league. The greater Boston area is home to several nationally competitive debate teams from suburban districts that typically do not participate in BDL competitions. However, experienced BDL debaters sometimes participate in regional or national tournaments with these students.

THE LIMITS OF EARLIER DEBATE RESEARCH

Despite the spread of debate programs, and the considerable anecdotal evidence of the notable successes of former debate participants, limited rigorous research has been devoted to understanding the causal impacts of debate participation on student academic success. There is a small existing literature on competitive policy debate, but previous studies have been limited by the outcomes under study, their access to narrow comparison groups selected by the districts providing the data prior to the analysis phase, and their use of matching methods without accompanying quasi-experimental identification strategies that could more persuasively isolate debate impacts. This is particularly important given that debate is an opt-in activity. It is therefore likely that debaters differ on both observable and unobservable dimensions from those who do not elect to join debate. As a result, a key challenge for researchers is to isolate the causal effect of debate from these pre-existing differences between debaters and non-debaters that could influence their academic achievement regardless of debate participation.

Despite their limitations, previous studies show that after participating in debate, students experience higher levels of academic achievement than non-participants. Studying high school debate in Chicago Public Schools, Mezuk (2009) as well as Anderson and Mezuk (2012) compare debate participants to non-participants attending the same schools and find debaters were three times less likely to drop out of high school and 70% more likely to score at or above ACT college readiness benchmarks in reading, even after accounting for 8th grade test scores and GPA. When analyzing a similar sample of students utilizing propensity score matching to account for observed differences between students who do and do not opt to participate, Mezuk, Bondarenko, Smith & Tucker (2011) find debaters were 19% more likely to graduate and 15% more likely to score at or above ACT college readiness benchmarks in English and reading.

Another report from the Minneapolis Public Schools (2015), again using OLS methods with covariate adjustments, finds participants had higher test scores and attendance than non-debaters. Specifically, debaters grew approximately 14% more than non-debaters on state standardized exams and attended 6% more days of school. Shackelford (2019) studied middle school debate in Baltimore Public Schools using propensity score matching techniques. He finds effects on the order of 0.21 SD units on 8th grade reading test scores and 0.13 in math. Furthermore, these middle school debaters had a 10% lower probability of being chronically absent. A more recent study similarly finds high school debaters in Houston had higher GPA and SAT math and reading/writing scores than a comparison group matched on 8th grade baseline achievement and other observable sociodemographic characteristics using propensity score methods (Ko & Mezuk, 2021). In the Boston context, there is one unpublished study that examined Boston debate's effects on student achievement from 2008 to 2012, again using propensity score matching. This study shows participants had greater growth on standardized ELA exams and higher GPAs than non-participants (Winkler & Fortner, 2014).

Our study makes several contributions. First, while previous studies have relied on matching methods alone, for our analysis of K-12 outcomes, we capitalize on quasi-experimental methods to more credibly isolate the causal impact of debate from pre-existing unobservable differences between students who do versus do not join debate. Specifically, we use difference-in-differences methods to compare the change in outcomes for debaters after joining debate to the change in outcomes for observably similar students who did not join debate over the same period. Additionally, we rely on student fixed effects methods that allow us to compare students to themselves over time in years when they did versus did not debate, addressing potential bias introduced not only due to observed differences but also all fixed unobserved differences between students who do and do not opt into debate when making within-student comparisons.

Second, previous studies have often relied on limited comparison groups, such as those pre-selected by data providers. We instead capitalize on districtwide data from BPS which allows us to test the sensitivity of results to different comparison group choices. These earlier studies have often focused only on comparing students within debate schools who did and did not participate in debate. However, these comparisons may inflate estimates of debate impacts in a biased manner given students who choose to participate in debate are likely quite different on observable and unobservable dimensions from those who do not participate. For example, perhaps debaters are more academically motivated, more interested in public policy, more argumentative, or more willing to challenge authority than their peers who do not elect to participate, and it is these differences that drive the higher outcomes rather than debate participation. We overcome this by exploring comparisons between debaters and observably similar students in non-debate schools who we suspect would have been likely to participate in debate had they been given the opportunity but who were not able to debate simply because debate was not offered at their school in the years they were enrolled.

Third, we examine the impact of debate on novel outcomes. For ELA achievement, we examine debate effects on subskills—language and reading—that allow us to say something about the types of competencies debate promotes and the mechanisms through which it improves achievement. Additionally, we examine outcomes beyond short-term test scores, including high school graduation, college going, and the type of postsecondary institution attended. While we are not able to deploy student fixed effects methods to estimate effects on these postsecondary outcomes (because they are not measured repeatedly over time), we provide suggestive evidence that our estimates are not particularly biased.

DATA

Debate Participation Data. We make use of student-level data provided by the Boston Debate League (BDL) that tracks which students participated in debate in which years from the 2007-08 to the 2016-17 school year. Going forward, we refer to each school year by its spring year (e.g., we call the 2016-17 school year "2017"). These participation data also include information on the number of tournaments attended, the division (novice, junior varsity, and varsity) in which a student competed at each tournament, and whether the student participated in the Spanish-speaking league by year. Participation data were merged with districtwide, longitudinal, student-level administrative data covering the same time-period and provided by the Boston Public Schools (BPS). Linking was accomplished using unique student ID numbers or names to locate ID numbers when numbers were unavailable in the participation data. The match rate was 94%. Ultimately, we include 3,515 unique students who participate in debate at some point in our panel. Administrative data contain annual information on student enrollment, school, and grade, as well as demographic characteristics, and academic performance.

K-12 Outcome Measures. The K-12 outcome measures include scores on the statewide ELA and mathematics exams given in 3rd to 8th and 10th grades. In most years, this was the Massachusetts Comprehensive Assessment System (MCAS). However, in 2015 and 2016 a subset of students took the Common Core-aligned Partnership for Assessment of Readiness for College and Careers (PARCC) exam. In 2016 and 2017, some students took online versions of the test. We standardize scores within year, grade, subject, exam, and mode (paper vs. online) using the full districtwide sample to have a mean of zero and SD of one. For a subset of years, 2012 to 2017, we have access to item-level test data in ELA and can map these items onto the two major categories of ELA sub-standards that exist across years—language and reading—to generate two separate scores and assess the relative impact of debate on these subskills. We also examine two non-test outcomes measured annually: school attendance and disciplinary infractions. More specifically, we use percentage of total school days present and days suspended. We have a larger sample with which to estimate the impacts on the non-test outcomes than test-based outcomes because 9th graders do not take statewide exams and most debate participants are high school students, as we describe in more detail below.

Postsecondary Outcome Measures. To examine debate effects on postsecondary outcomes, we use data from the National Student Clearinghouse (NSC) that provides information on students' high school graduation status and date, postsecondary enrollment and date, and the type of postsecondary institution in which a student enrolled (2-year or 4-year and public or private). We construct six binary outcomes, measuring whether a student ever graduated from high school (according to either the BPS or NSC records), whether a student ever enrolled in a postsecondary institution within two years of their expected high school graduation date based on when they enrolled in 9th grade for the first time, and whether they enrolled in a 2-year, 4-year, public, or private institution. NSC has a good coverage rate of higher education institutions for Massachusetts, specifically 98.2% averaging across all years in our data (NSC, 2022).

SAMPLE

There is a two-step selection process through which a student participates in debate. First, the student's school must join the debate league run by BDL. As we will show below, the schools that join the league are not a random sample of BPS schools, however, we provide evidence that the precise timing of their entry into the league appears to be exogenous. In Appendix Figure A1, we show the number of new schools that joined the debate league in each year. This ranges from seven schools in the first year to only one new school in the final year of our panel. In total, we observe 40 BPS schools that were ever part of the debate league. Once the school is a part of the league, the student must opt to join the team to participate in debate.

In Table 1, we describe the sample of students separated into three categories. The first column includes all students who attended schools that were part of the debate league at some point, regardless of whether they themselves participated in debate, and the second column represents students in schools that never joined the debate league. A comparison between the first two columns reveals that schools that joined the debate league were relatively disadvantaged compared to schools that did not join. Specifically, debate schools had average baseline test scores that were more than a quarter of a SD lower in both ELA and math than schools that were never part of the league (e.g., -0.13 vs. 0.14 SD in ELA). Debate schools also served a greater concentration of students classified as economically disadvantaged (82%) relative to non-debate schools (68%), had a higher proportion of students classified as English learners (36 vs. 26%), and a higher percentage of students with an Individualized Educational Plan (IEP) (23 vs. 20%). Debate schools also served student populations with a greater concentration of Hispanic students and a lower share of White students. These differences are statistically significant.

A different story emerges when looking within debate schools. Columns three and four are limited to students in schools that were a part of the debate league in years when that school offered debate. Column three includes debate participants ("ever debaters") while column four includes students at debate schools who opted not to join debate ("never debaters"). Students who participated in debate were not representative of the broader student population at their schools. Debaters were substantially higher achieving at baseline, on average, than those who did not debate. For example, they have average ELA scores of 0.21 SD units versus -0.18 for those who have the option but never debate. They also have a slightly higher attendance rate (by one percentage point), 0.04 fewer days of suspension per year, on average, and are less likely to have an IEP (14 vs. 24%) than those who did not join debate. Interestingly, debaters are more likely female (56 vs. 47%), more likely classified as economically disadvantaged (87 vs. 84%), and more likely identified as Black (42 vs. 36%) than students in debate schools who do not debate.

Given these non-trivial observable baseline differences between students who do and do not choose to participate in debate within debate schools, and the known process through which students must individually opt in to debate participation, it is likely that debaters are also different on unobservable dimensions from those students who do not choose to debate. Therefore, students in debate schools who never participated in debate may not be the best comparison group for isolating the impact of debate. Instead, we view the students in schools that never joined the debate league as a more useful comparison group because a subset of these comparison students are those who were likely to have participated in debate had they been given the opportunity but who did not participate simply because debate was not offered at their school in the years when they attended. These students are also observationally more similar to the treated group than students in debate schools who did not debate in terms of their achievement measures when they were not yet old enough to participate in debate. Comparing columns two and three, students in non-debate schools have more similar baseline test scores to the debate group (e.g., 0.18 SD in math vs. 0.17 SD for debaters, and 0.14 SD in ELA vs. 0.21 SD). Therefore, we ultimately exclude from our analytic sample students who were in debate schools in years when the school offered debate but who opted not to participate, focusing our comparison on students in non-debate schools. (However, we also confirm our results are similar when using the full sample and making within-school comparisons.)

In Table 1, we provide additional detail on debate participation. Specifically, the average debater participated for 1.39 years. A large majority of debaters only participated for a single academic year. Debaters competed at an average of three tournaments per year with a maximum of 10. Regional tournament participation was quite rare, meaning debaters in our sample mostly participated in tournaments limited to BDL schools rather than competitions that included neighboring suburban districts or national teams. A small subset (6%) participated in BDL's Spanish-speaking debate league (established in 2014). Only 16% of debaters participated in varsity level competition, at least among the 70% of debaters for which we know their highest division. Among those, 25% made it to junior varsity, while the largest share remained in the novice division for the duration of their participation. The average debater began debating in 9th

grade and only 28% ever participated in middle school. This has important implications given that Massachusetts high school students only sit for statewide ELA or math tests in their 10th grade year. Therefore, our test score estimates of the impact of a year of debate participation rely on a subsample of debaters which excludes a non-trivial number of debaters who began debating in 9th grade and only debated for a single year.

In Table 2, we describe the sample of debaters that contribute to our postsecondary analysis. This is different than the sample used to explore K-12 outcomes because we limit the postsecondary sample to the 2,269 debaters who were in cohorts for which we can observe outcomes five years after their first 9th grade year (first-time 9th graders in 2008 to 2013). Postsecondary estimates therefore represent effects for cohorts participating in the earlier years of BDL's operations. As was the case for the K-12 analysis sample, debaters contributing to postsecondary estimates were more likely than non-debaters to be Black, female, economically disadvantaged, and high achieving at baseline, and less likely to be Hispanic or have an IEP.

ANALYTIC METHODS

Pooled Estimates of K-12 Outcomes. To assess the effect of debate participation on academic achievement outcomes observed during students' K-12 schooling, we present results from two difference-in-differences models, one using a rich set of covariates and the other using student fixed effects. The first model tests whether the changes in outcome measures for debaters after they participated in debate were different from changes in outcomes for observationally similar students who did not debate. Again, we exclude students who attended debate schools in years when that school was part of the league but who themselves did not debate (students who we know do not take up debate when it is available). This focuses the comparison on debaters versus students in non-debate schools who never had the opportunity to participate in debate, some of whom would have participated if given the chance. We begin by estimating the following model pooling across all cohorts of debaters:

$$Y_{ist} = \beta_0 + \beta_1 Debate_{ist} + \delta_t + X'_{is} + \partial'_{s} + \varepsilon_{ist} \qquad (1)$$

Here, Y_{ist} represents an outcome, such as a test score, for student *i* in school *s* and year *t*. On the right-hand side of the equation, we include a dummy indicator ($Debate_{ist}$) equal to one if a student ever participated in debate and if the student was participating in debate in a given year. Therefore, β_1 is the difference-in-differences estimate of the debate effect. This indicator turns "on" and "off" in cases where a student, for example, began debating in ninth grade but quit in eleventh grade. We consider this a conservative approach that may underestimate the impact of debate if a single year of participation has long-lasting effects after a student quits. We also include year fixed effects (δ_t), as well as a host of student- and school-level covariates to address known differences between debaters and comparison group students. At the student level (X'_{is}), we control for grade, gender, race, age, economic disadvantage, English learner status, first language, whether the student ever had an IEP, and baseline measures of math and ELA achievement, attendance, and suspensions. We also include a vector of school-level covariates (∂'_{s}) for the same list of characteristics in the student-level controls. Finally, we control at the student level for whether a student took the PARCC exam (vs. MCAS) and whether the student took an online exam (vs. paper) in a given year. We cluster standard errors at the school level.

Although this rich set of controls addresses bias stemming from observed differences between debaters and comparison students, there still may be differences between these groups on unobserved dimensions that could influence outcomes regardless of debate participation. To address this possibility, we estimate a second model using student fixed effects methods to test whether, within individual students, performance is different in years when students did versus did not participate in debate. Student fixed effects allow us to control for all observed and unobserved student characteristics that are fixed over time for comparisons made within-students. For example, if debaters are simply more academically motivated than non-debaters, these models would separate those factors out from the resulting estimate of the debate effect. This is our student fixed effects model when pooling across all cohorts:

$$Y_{ist} = \beta_0 + \beta_1 Debate_{ist} + \delta_t + \theta_i + \varepsilon_{ist} \qquad (2)$$

There are a few differences between this model and model (1). First, we add student fixed effects (θ_i) and omit covariates that are constant within student. We retain controls for exam (PARCC vs. MCAS) and test mode (online vs. paper) as these vary within student over time.

Event Study Estimates of K-12 Outcomes. The above models pool all debate years across cohorts and across the year of debate participation (i.e., year 1, year 2, year 3, etc.) to calculate the average effect of debate. We pair these results with findings from event-study analyses for two reasons. First, it allows us to formally test the parallel trends assumption by examining estimates in the pre-debate period. Second, the methodological literature on difference-in-differences has proliferated in recent years and suggests bias can arise when researchers rely on models estimating only two time periods ("pre" and "post") in situations, like our own, where there are multiple time periods and is variation in the timing of treatment (i.e., students joining debate in different years) (Callaway & Sant'Anna, 2020; Sun & Abraham, 2020). Pooled estimates can be biased if treatment effects vary over time (Baker et al., 2021; Goodman-Bacon, 2018). Therefore, we assess the extent to which this is true in our context by estimating effects by year of treatment (event studies), overall and separately by cohort. We then examine the extent to which patterns are consistent across cohorts. When estimating effects separately by cohort, we exclude all other ever-treated students from other cohorts, as Goodman-Bacon (2018) documents that bias can also stem from the use of already-treated units as controls. To generate the event-study estimates, we replace the main effect of debate participation in our pooled model with separate indictors for each year leading up to and after debate participation interacted with an indicator for whether a student was debating in that year. The models take the following form:

$$Y_{ist} = \beta_0 + \sum_{r=-5}^{5} \beta_r Debate_{ist} + \delta_t + X'_{is} + \partial'_s + \varepsilon_{ist} \qquad (3)$$

Here, $\sum_{r=-5}^{5} Debate_{ist}$ represents a series of indicators for the number of years of debate participation, centered on and omitting the year prior to debate participation (Year 0) as the reference category. The post-debate annual indicators allow us to assess the effect of debate by the number of years of participation. However, we caution readers against drawing strong conclusions based on the greater years (e.g., Year 5) as continued participation in debate is likely non-random and, as we will show, we have fewer observations on which to estimate the effect of four years of debate participation versus one year, for example. Again, standard errors are clustered at the school level. The sample restrictions we use mirror those we describe above for the pooled estimates. We also run a student fixed effects version of our event study model:

$$Y_{ist} = \beta_0 + \sum_{r=-5}^{5} \beta_r Debate_{ist} + \delta_t + \theta_i + \varepsilon_{ist} \qquad (4)$$

Here again we omit non-time-varying covariates and include student fixed effects but the model is otherwise the same as model (3).

Estimates of Postsecondary Outcomes. We are unable to use student fixed effects methods to estimate the impact of debate on high school graduation and postsecondary enrollment outcomes because, unlike K-12 outcomes that are measured repeatedly within student, these outcomes are terminal and therefore only observed once (e.g., whether a student ever graduated high school). Therefore, to analyze these outcomes, we collapse our data so we have one observation per student, keeping the observation for the first year each student was enrolled in 9th grade. We then compare the probability of each postsecondary outcome for students who ever participated in debate to those who never participated, controlling for a rich set of covariates known to predict debate participation using the following model:

$$Y_{ist} = \beta_0 + \beta_1 Ever\_Debater_{ist} + \delta_t + X_i + \varepsilon_{ist} \qquad (5)$$

Here Y_{ist} represents a postsecondary outcome, such as whether a student ever graduates high school. The $Ever\_Debater_{ist}$ indicator is a binary variable equal to one if a student ever participated in debate, regardless of what grade they were in when they debated or the number of years they participated. The coefficient on this variable is therefore our estimate of the debate effect. Our model also includes year fixed effects (δ_t) to control for temporal trends in graduation and college enrollment rates and a vector of student-level covariates (X_i) including indicators for racial/ethnic categories, gender, economic disadvantage, English Learner status, first language, and receipt of an IEP. We also include measures of baseline math and ELA achievement, school attendance, and discipline. We impute the average for any student missing on a baseline measure and include dummy variables for missingness on each baseline measure. We report estimates from linear regression models for ease of interpretation but confirm our conclusions are robust to logistic models. We cluster standard errors at the school level.

Matching Strategies. We begin by estimating these models with the full sample but given our concerns about selection, as before, we also estimate after excluding students who never debated but who were enrolled in debate league schools in years when the school offered debate. We also present results from two matched comparison groups based on different approaches to matching that both attempt to reduce observable differences between the treatment and comparison groups. For the first approach, we run a logistic regression in which we predict whether a student ever participated in debate based on all the observable characteristics we have available, described as covariates above (Pseudo R2 = 0.12). We then identify a "high propensity to debate" comparison group by limiting the comparison students to those with a predicted probability equal to or higher than the mean probability to debate among students who actually debated (0.09). When estimating models with this sample we also add in the predicted probability of debating as a covariate into our model to ensure that we are comparing debaters to comparison students who were also similarly likely to join the debate team. For the second matching approach, we rely on coarsened exact matching to generate matching strata based on student characteristics that predict debate participation including race, ethnicity, gender, economic disadvantage, English Learner status, IEP, and ventiles of baseline academic performance. This generated 3,790 strata. We have a match rate of 90% and all treatment strata have at least 26 students in the comparison group. When estimating these models, we add matching strata fixed effects. In Table 2 we show that these matched samples are more similar on observable characteristics to debaters than the full sample of non-debaters, there are still some statistically significant differences. As a result, we are careful to control for all observable characteristics when estimating our models.

RESULTS

Debate Impacts on ELA Achievement. We begin by presenting the pooled estimates for ELA in the top row of Table 3, suggesting effects of between 0.126 and 0.137 SD (with and without student fixed effects). In general, our findings are consistent across the two models, although somewhat attenuated when including student fixed effects, suggesting these models may do more to address selection bias. We confirm that none of our results (in ELA or any outcome) are due to differential changes in missingness on the outcome measures between the treatment and comparison groups (Appendix Figure A1).

To assess the parallel trends assumption, or the extent to which debaters were on a different achievement trajectory leading up to joining the debate team than non-debaters, we present event study estimates in Table 4. Among the pre-treatment estimates, only one of the coefficients is statistically significant (true for both models). Joint F-tests of all pre-debate coefficients reveal no evidence that the pre-trends are statistically different among debaters than non-debaters across all models (results in Table 4). Additionally, for the student fixed effects models, none of the three years leading up to treatment are significant (Year -3 through Year -1), allaying potential concerns that findings could be the result of differences between the treatment and comparison groups occurring immediately prior to the intervention which might reflect an "Ashenfelter's Dip" or anticipatory effects. We display these estimates graphically in Figure 1 and do not see significant evidence of differential pre-trends for debaters versus non-debaters.

The event study treatment effects results are consistent with our pooled estimates. Across sample restrictions, we continue to observe statistically significant, positive impacts of debate on ELA scores. The impacts for the first year of debate are on the order of 0.097 SD to 0.125 SD (with and without student fixed effects) and generally increase with additional years of participation (with the exception of Year 3). Estimates for Year 3 and 4 are positive but do not always achieve statistical significance. As a reminder, these are based on smaller samples of debate participants (e.g., 26 treated students contribute to the Year 4 estimates) as most debaters are only on the team for a single year. We provide the sample sizes of debate students used to estimate each coefficient in the first column of Table 4.

We also conduct a formal pre-trend test using a comparative interrupted time series (CITS) model both without and with student fixed effects. For these models, we pool all post-debate years and include a single dummy indicator equal to one if a student participated in debate and if it is a post-debate year ("Debate"). We include a linear time trend ("Year") as well as an interaction between an indicator for whether a student ever participated in debate and the linear time trend ("Ever Debate X Year"). For the first version of these CITS models, we include a main effect for whether a student ever participated in debate ("Ever Debate") as well as the same set of covariates included in model (1). For the second version of the model we include student fixed effects and omit the colinear covariates. We cluster standard errors at the school level. We display results from the CITS model in Appendix Table A1, confirming results are robust to this specification. Additionally, the insignificance and small magnitude of the interaction ("Ever Debate X Year") implies that there is no significant difference in the pre-treatment trend between debaters and non-debaters driving our estimates. We can also use the event study estimates to calculate predicted Year 1 differences based on the pre-treatment coefficients (and resulting slopes and intercepts). Based on our preferred student fixed effects estimates, we find a predicted Year 1 difference of 0.024 SD. When subtracting this from the pooled estimate (0.126), the coefficient is slightly attenuated but still substantial in size and statistically significant (0.102).

Debate Impacts on ELA by Baseline Achievement. Interestingly, despite the fact that debaters are generally higher performing than the rest of the students at their schools at baseline, we find debate had the largest impacts among students who were lowest performing in ELA prior to joining the debate team. In Table 5, we estimate debate impacts separately for four quartiles of students based on their baseline ELA scores, from lowest performing in the top row to highest in the bottom row. While the direction of the effects is always positive, no matter the group or model, the magnitude of the impact is largest among the lowest performing quartile and decreases for each performance quartile. For example, based on student fixed effects estimates (column two), we observed debate impacts of 0.235 SD for those in the lowest quartile and 0.096

SD for the highest performing quartile. In short, the gains in ELA, while not exclusive to the lowest-performing students, are largest among those students who are lower-achieving at baseline (more than twice as large for those in the lowest quartile as the highest quartile).

Debate Impacts on ELA Subskills. We also explore whether ELA gains vary between the two major categories of ELA standards—language and reading—that we are able to track over time for a subset of our analytic window: 2012 to 2017. We confirm that the main effects on ELA achievement replicate within this subsample. We present the results in Table 6 and begin with the estimates based on the pooled difference-in-difference models in the top panel. Although the direction of effects for both subskills is always positive, the magnitudes are larger for the reading subskill. For example, using the student fixed effects model, we observe statistically significant effects of 0.101 SD in reading versus non-significant effects of 0.057 in language. In the bottom panel, we present event study estimates which imply that these effects on reading achievement emerge within the first year of debate participation and do not appear to be due to pre-trend differences on the subskills between debaters and comparison group students.

In Appendix Figure A3, we illustrate the differences between the language and reading standards by displaying the Massachusetts Curriculum Framework for English Language Arts and Literacy for Grade 8 in 2017. For reading, we use the standards for informational text as our example for the purpose of simplicity but the overall pattern of contrasts between language and reading standards remains the same regardless of these choices. In general, language standards focus on language conventions, such as grammar, capitalization, punctuation, and spelling, as well as vocabulary acquisition. In contrast, reading standards focus more heavily on comprehension and analysis, such as "determine a text's central idea(s)." Reading standards also include argumentation-related skills, such as "cite the textual evidence that most strongly supports analysis of what a text states explicitly as well as inferences drawn from the text" and "delineate and evaluate the argument and specific claims in a text, assessing whether the reasoning is sound and the evidence is relevant and sufficient." Additional standards emphasize social perspective taking and multiple viewpoints, such as, "determine an author's point of view or purpose in a text and analyze how the author acknowledges and responds to conflicting evidence or viewpoints." In short, we find evidence that debate improves ELA achievement through the development of analytic and argumentation skills more so than rote memorization or the mastery of basic language conventions.

Debate Impacts on Math Achievement. In the third and fourth columns of Table 3 we display pooled estimates of debate effects on math achievement. We find suggestive evidence of positive spillover effects in math, although smaller than the ELA impacts, ranging from 0.059 to 0.060 SD (with and without student fixed effects). However, these estimates do not always achieve statistical significance. Turning to the event study estimates in Table 4, the estimates of debate impacts are positive and statistically significant for Year 1 (0.065 SD) when using student fixed effects). None of the estimates based the later years achieve statistical significance and the significance of the Year 1 effect is sensitive to our choice of model. We do not observe evidence of pre-trend differences. None of the pre-debate coefficients are statistically different from zero with the student fixed effects model and we do not observe visual evidence of pre-trend when we plot the coefficients over time in Figure 1. We see no evidence that differential pre-trends were driving math impacts when we formally test for pre-trend differences using the CITS model (results displayed in Table A1). In short, we do not observe strong evidence that debate has a positive impact on math achievement, although see no evidence of harm. The much smaller magnitude of the math impacts relative to the ELA effects provides confidence that the ELA impacts are not simply an artifact of selection.

Debate Impacts on Attendance and Discipline. Next, we turn to the effect of debate on non-test measures of academic success, beginning with the percent of school days for which a student was marked present in a given school year. Pooled estimates in Table 3 suggest small positive impacts (a two to three percentage-point increase). The CITS models suggest no evidence of pre-trend differences given the small and statistically insignificant estimate for the interaction between whether a student ever debated and the linear time trend (see Table A1). However, event study estimates in Table 7 complicate our ability to fully rule out the possibility that increases could be due to pre-trend differences. Therefore, we are reluctant to make strong claims about the impact of debate on attendance, though we certainly find no evidence that debate is harmful on this dimension. We find no impacts of debate on the number of suspensions students receive in a year. Pooled estimates in Table 3 are all negative in sign, implying again that debate did not increase suspensions. However, none of these estimates achieve statistical significance. In Table 7, we show that the declines were concentrated in the later years for which we have a more limited sample. We find no strong evidence of pre-trends on this outcome based on the joint F-tests from the event study models (see Table 7) or the CITS model (Table A1).

Additional Robustness Checks for Estimates of K-12 Outcomes. As we previewed in the methods section, additional concerns with difference-in-difference methods in cases where treatment adoption is staggered, are that bias can occur when treatment effects vary across calendar time (over cohorts) or from the use of already-treated units as part of the comparison group. We show results for which we estimate event studies separately by cohort (for each cohort excluding from the sample all other ever-treated students from other cohorts) for test outcomes in

Appendix Table A2 and the non-test outcomes in Table A3. We do so after excluding students from debate schools who did not choose to debate. In general, we find the patterns of results are consistent across cohorts. For ELA, all but one cohort demonstrates positive effects of debate in Year 1 and for only two of the cohorts do we find some evidence of pre-trends based on the joint F-tests. Of course, given the smaller samples, not all of these estimates achieve statistical significance. Similarly in math, all but one of the cohorts demonstrated positive effects in Year 1 of debate participation. Only one of the cohorts appears to have pre-trend issues. For attendance, the results are fairly consistent across cohorts and when compared to the pooled event study estimates. For suspensions, the results are much more variable—in both direction and magnitude—across cohorts which could help to explain why the impacts do not achieve statistical significance when pooling across all cohorts.

We also run a diagnostic developed by Chaisemartin and D'Haultfoeuille (2021) to determine whether our pooled estimates rely on "negative weights" which would indicate potential bias in a pooled difference-in-differences estimate (with variation in treatment timing). A very small fraction (0.6%) of the weights in our model for ELA achievement are negative, obviating the need for updated estimators designed to address such bias. However, even when we use one of these new estimators—the "did_multiplegt" command in Stata (Chaisemartin and D'Haultfoeuille, 2018)—we get similar results (a pooled ELA effect of 0.07 SD). Therefore, we find no evidence impacts are driven by bias due to variation in treatment timing.

Our preferred sample restriction is based on the idea that students at non-debate schools are a better comparison group than students within debate schools who did not elect to participate in debate. However, this strategy requires comparing between, rather than within, schools and relies on the assumption that there are not systematic time-varying differences between schools that do versus do not join the debate league that are driving the results. In particular, if debate schools were also embarking on additional non-debate reforms in the same year in which the school joined the debate league, our estimates could be reflecting these non-debate policy changes. We present a few pieces of evidence that this is not the case. First, recall that debate schools, on average, are lower-achieving at baseline than non-debate schools. If schools were driving the debate impacts, we would expect debate schools to be higher performing than non-debate schools. Second, we formally test whether a school joining the league had an effect on achievement by running our event study model after replacing the indicators for student-level debate participation with indicators for whether the school had joined the debate league in a given year. As we show in Appendix Table A4, we see no evidence that a school joining the debate league has an impact on academic achievement. In particular, we see no statistically significant effects of a school joining the league on any outcome, with the exception of four coefficients that all tend to be in later years of debate league participation.

Another potential concern is that results could be driven by a subset of debaters who stick with the activity for multiple years. This could be particularly problematic if this group is non-randomly sorting into long-term participation. In the second panel of Table 3, we re-estimate effects after dropping students who participate in debate for more than one year. The ELA impacts are slightly attenuated, but our overall conclusions remain the same for all outcomes. Additionally, our event study models further suggest that benefits begin to emerge in the first year of debate participation and, therefore, the impacts are not driven only by the relatively small sample of students for whom we can estimate the effect of three years (n=65), four years (n=26), or five years of debate (n=14). Another possibility is that it could be that students are systematically sorting into schools that offer debate for the purpose of participating in debate. In the bottom panel of Table 3, we exclude all students who first began debating in ninth grade. The idea being that these are the students who most likely selected a high school based on whether it offered debate. We find that our results are also not sensitive to this sample exclusion.

Debate Impacts on Postsecondary Outcomes. We present estimates of the effect of debate on postsecondary outcomes in Table 8, beginning with high school graduation. Using the sample for which we exclude non-debaters from debate schools, we estimate that debate increases the probability of ever graduating from high school within five years by 12 percentage points, from a comparison group mean graduation rate of 68%. The magnitude of the coefficient is comparable across all sample restrictions and matching approaches, although slightly attenuated to 9 percentage points with the high propensity to debate comparison group. Next, we turn to effects on enrollment in a postsecondary institution within two years of expected high school graduation. We find impacts on the order of 12 percentage points that are basically unchanged when varying the comparison group. Among comparison students, the mean college enrollment rate is 41%. In Table 9, we show impacts on college-going are largely driven by increased enrollment in 4-year and private higher education institutions.

We examine whether postsecondary impacts are heterogenous across subgroups and find two notable patterns. First, like the impacts on K-12 outcomes, the impacts on both high school graduation and college-going are largest among those students who were lowest achieving at baseline. In Table 10 we report the impact estimates for four quartiles of baseline ELA exam performance. The graduation impacts are positive and statistically significant for all four quartiles but largest for the lowest performing group, getting progressively smaller as baseline ELA achievement increases. For postsecondary enrollment, again the estimates for all four groups are positive in direction but only statistically significant for the bottom two performance quartiles and again become progressively smaller the higher the baseline performance. When examining effects on type of postsecondary institution, we see positive effects emerge for enrollment in 2-year colleges and public colleges among the lowest performing group. Second, the postsecondary impacts are entirely concentrated among those students who participated in debate during their high school years, as we demonstrate in Appendix Table A5.

Robustness Checks for Estimates of Postsecondary Outcomes. The postsecondary effects do not seem to be driven by those students who debated for multiple years, as we illustrate in the middle panel of Table 8. The graduation estimates are basically unchanged and the postsecondary enrollment estimates are only slightly attenuated when we limit the sample to one of the matched comparison groups. We also see no evidence that students with a greater ability to select into debate schools (those who started debating in 9th grade) are driving the effects. If anything, the estimates are larger for both graduation and postsecondary enrollment when we exclude these students from our sample (see the bottom panel of Table 8).

We also use the analytic approach for estimating impacts on postsecondary outcomes to estimate effects on ELA achievement. This allows us to compare these ELA estimates to the student fixed effects estimates for which we are more confident in our ability to minimize selection bias. As we show in Table 8, the estimates are roughly 0.11 SD higher (almost twice as high) using these methods than student fixed effects. At first, this may seem to suggest that our models used to estimate graduation and college-going effects have not entirely removed selection bias and therefore may be higher than the true impact of debate. Conservatively, if they are inflated by roughly the same amount as the ELA impacts (nearly doubled), they suggest a 5.5 percentage point increase in high school graduation and a 6.5 percentage point increase in postsecondary enrollment, correcting for inflation. However, it is important to remember that the

OLS estimates are not directly comparable to the student fixed effects estimates, as they are generated with a more limited sample of students. When we estimate ELA impacts using the student fixed effects models and the more limited sample of students that are included in the OLS estimates, we get a coefficient of 0.20 SD ($p<0.001$), much closer to the OLS estimates (~0.24 SD) than the student fixed estimates using the full sample (~0.13 SD). This suggests that the differences in estimates are more due to sample differences than modelling differences, providing evidence consistent with the idea that the OLS estimates are not substantially inflated.

DISCUSSION

We examine a unique educational activity designed to train secondary school students in literacy, argumentation, critical thinking, and public policy analysis skills in a context serving large concentrations of economically-disadvantaged students of color. We find notable positive impacts of participation in policy debate on students' English Language Arts achievement, as measured by standardized exams. We find no evidence of harm to math achievement nor to non-test outcomes such as attendance or discipline. We also find evidence of substantial positive effects on the likelihood of high school graduation (by 12 percentage points) and postsecondary enrollment (by 12 percentage points), driven by enrollment in 4-year and public institutions.

The size of the average impact of debate on ELA achievement is large, particularly for middle and high school students, when considered alongside other policy-relevant benchmarks. The impact is comparable to the typical amount of annual growth the average 9th grader in the U.S. makes on nationally normed reading tests (0.19 SD) and represents roughly 20% of the national 8th grade reading achievement gap between students who do and do not qualify for subsidized lunch (-0.66 SD) (Hill, Bloom, Black & Lipsey, 2007). Researchers have uncovered very few interventions that generate impacts of this magnitude for secondary school students, especially on literacy outcomes. In a review of effect size magnitudes from randomized control trials in education, Kraft (2020) finds that the average effect size on reading exams from high school interventions is 0.05 and middle school 0.06 SD (Kraft, 2020), less than half the size of the ELA effects generated by debate in Boston. Therefore, policy debate appears to be a rare strategy for improving literacy skills among middle and high school students and helps to demonstrate that secondary school is not "too late" to support student progress in reading.

Our results are consistent with prior research documenting debate participants' higher post-debate ELA test scores compared to non-debaters (Minneapolis Public Schools, 2015; Ko & Mezuk, 2021; Mezuk, 2009; Anderson and Mezuk, 2012; Winkler & Fortner, 2014). Using propensity score matching, Shackelford (2019) found that debate improved 8^{th} grade reading scores by 0.21 SD units and math scores by 0.13 SD. Our results are smaller in magnitude in reading (0.13) and positive but not always significant in math (0.062), possibly reflecting our ability to better address selection bias than previous studies. Earlier research documents debate's association with higher rates of school attendance (Minneapolis Public Schools, 2015; Shackelford, 2019), consistent with the direction of effects we observe, although our event study methods allow us to say that these positive effects cannot be fully separated out from differences in pre-treatment trends between debaters and comparison students. Our results are also in line with previous observational studies which have found debate is associated with an increased probability of graduating from high school and enrolling in 4-year college (Mezuk, 2009; Anderson and Mezuk, 2012; Mezuk, Bondarenko, Smith & Tucker, 2011). Shackelford, Ratliff & Mezuk, 2019 found debaters were more likely to matriculate to 4-year colleges than comparison students when controlling for demographic characteristics, neighborhood poverty and 8th grade test scores and ACT performance. Similarly, we show that impacts on college doing are largely driven by enrollment in 4-year colleges.

One way we build on previous findings is by providing evidence on the types of students most likely to benefit from debate participation. While the impacts were positive, on average, among all quartiles of students based on their ELA performance prior to joining debate, the greatest benefits were concentrated among those students who were relatively lower performing prior to joining debate. For example, the effects were more than twice as large for students with average baseline ELA scores of -1.355 SD than for students with average baseline ELA scores of 1.138. A similar pattern emerges for high school graduation and postsecondary enrollment. This suggests that policy debate need not be reserved for relatively high-achieving students. In contrast, when provided to students who have historically not been particularly high-performers, policy debate represents a rare program with high potential to help reduce educational inequality on literacy and critical thinking skills among secondary school students.

Our study is not without limitations. While we use student fixed effects to separate out pre-existing unobservable differences between debaters and non-debaters from the causal impact of debate on K-12 outcomes, we are unable to control for possible differences between these two groups that vary over time. Additionally, although ours is one of the earliest studies to estimate impacts of debate participation on postsecondary outcomes—like the authors of previous earlier studies—we are unable to capitalize on a strong source of exogenous variation to do so. We are therefore not able to rule out the possibility that some or all of the estimated effects on postsecondary outcomes are driven by selection bias. Indeed, the magnitude of the impacts on postsecondary outcomes is quite large, potentially suggesting that we may not have successfully eliminated bias and that these estimates could be inflated. Ideally, future researchers would be able to capitalize on a field experiment in which the offer to participate in debate or timing of debate participation was randomly assigned, fully isolating the debate impact.

Another limitation is that we do not have data on what other extracurricular activities beyond debate students in our data participated in, such as sports, theater, drama, clubs, community service, student government, or even public speaking activities other than policy debate. It is possible that non-debaters participated in other extracurricular activities. Therefore, our study compares the marginal benefit of participating in policy debate to the benefits of (potentially) participating in other extracurriculars (not a lack of any extracurricular activity). Future research should examine the relative effectiveness of different extracurricular activities.

Relatedly, the gains we observed in ELA were concentrated on reading subskills that represent analytical thinking competencies more so than on language subskills that represent foundational knowledge and application of English language rules and conventions. This finding should provide readers with confidence that the results reflect the impact of debate participation rather than participation in any extracurricular activity which we would not expect to move the needle more on argumentation-oriented subskills. The finding also suggests that policy debate is an activity with high potential to develop students' higher order, critical thinking skills, another goal for which strategies are currently in short supply. Future research should probe this finding further with better measures of critical thinking, argumentation skills, and other competencies needed for academic and civic participation such as social perspective taking, media literacy, ability to distinguish fact from opinion, and engagement with the policy process.

We see a few policy implications of this work. First, the expansion of debate programming, particularly among lower-achieving students in schools serving low-income students of color appears a promising avenue for reducing inequality in literacy and postsecondary outcomes among secondary school students. The results presented by cohort do not suggest declining returns as the program expanded to serve a larger population. Additionally, the BDL estimates a current program cost of roughly $1,360 per student per year, one-third the expense of the well-regarded MATCH high-dosage tutoring program which costs roughly $4,000 per student (Cook et al., 2014). Kraft (2015) estimates this model produces impacts between 0.15 and 0.25 SD on literacy outcomes for high school students. Therefore, policy debate programs appear to generate twice the impact on ELA test scores per dollar spent than state-of-the-art high-dosage tutoring. These extracurricular programs therefore have potential for scalability.

Second, there are likely implications for teachers working in traditional classrooms with groups of students who may not elect to participate in extracurricular debate. A handful of organizations, including the BDL, have developed and implemented professional development programs designed to help teachers infuse key principles of debate pedagogy into regular classrooms. Often called “debate-centered instruction,” the goal is to give a wider audience an opportunity to benefit from debate-like learning opportunities without intensive out-of-school participation (Litan, 2018). Researchers should explore the effectiveness of these programs to help uncover the extent to which the impacts of debate would generalize to students who do not opt into extracurricular debate. Burgess, Rawal and Taylor’s work (2022) further suggests this could be a promising pathway in that they find a reallocation of class time toward more open discussion and opportunities for work with classmates is associated with improvements in student’s English achievement. In short, while our study demonstrates exciting results for extracurricular debate participants, there may be even greater dividends to incorporating some of these practices into regular classroom-based instruction for a wider population of students.

REFERENCES

Anderson & Mezuk (2012). Participating in a policy debate program and academic achievement among at-risk adolescents in an urban public school district: 1997-2007. *Journal of Adolescence*.

Backes, B. & Cowan, J. (2019). Is the pen mightier than the keyboard? The effect of online testing on measured student achievement. *Economics of Education Review*, 68, 89-103.

Bailey, M. & Dynarski, S. (2011). "Inequality in postsecondary education" in Greg Duncan and Richard Murnane eds., *Whither opportunity: Rising inequality, schools, and children's life chances.* New York, NY: Russell Safe Foundation.

Baker, Larker & Wang (Forthcoming). How much should we trust staggered difference-in-differences estimates? Journal of Financial Economics.

Boston Debate League (2022). What we believe. www.bostondebate.org/what-we-believe/

Burgess, S., Rawal, S. & Taylor, E. (2022). Teachers' use of class time and student achievement. EdWorkingPaper 22-685. Retrieved from Annenberg Institute at Brown University.

Callaway & Sant'Anna (2020). Difference-in-differences with multiple time periods. *Journal of Econometrics*.

Carneiro, P. & Heckman, J. (2003). “Human capital policy” in *Inequality in America: What role for human capital policies?* Edited by James J. Heckman and Alan B. Krueger. Cambridge, MA: MIT Press. Pp 77-240.

Cascio, E. & Staiger, D. (2012). Knowledge, tests, and fadeout in educational interventions. NBER Working Paper No. 18038.

Chaisemartin, C. and D’Haultfoeuille, X. (2018). Fuzzy differences-in-differences. *The Review of Economic Studies*, 85(2), 999–1028.

Chaisemartin, C. & D’Haultfoeuille, X. (2021). Two-way fixed effects estimators with heterogenous treatment effects. *American Economic Review*, 110(9), 2964-96.

Chetty, Friedman, Saez, Turner & Yagan (2017). Mobility report cards: The role of colleges in intergenerational mobility. NBER Working Paper No. 23618.

Center for Research on Education Outcomes (2015). Urban charter school study: Report of 41 regions. http://urbancharters.stanford.edu/summary.php

Coleman, J. S. (1959). Academic achievement and the structure of competition. *Harvard Educational Review*, *29*, 330–351.

Coleman, J. S. (1961). *The adolescent society*. New York: Free Press of Glencoe.

Cook, Dodge, Farkas, Fryer, Guryan, Ludwig, Mayer, Pollack & Steinberg (2014). The (surprising) efficacy of academic and behavioral intervention with disadvantaged youth: Results from a randomized experiment in Chicago. NBER Working Paper No. 19862.

Crispin, L. M. (2017). Extracurricular participation, "at-risk" status, and the high school dropout decision. *Education Finance and Policy*, *12*(2), 166–196.

Cuffe, H., Waddell, G. and Bignell, W. (2017). Can School Sports Reduce Racial Gaps in Truancy and Achievement?" *Economic Inquiry* 55 (4):1966–1985.

Cunningham, S. (2021). Causal Inference: The Mixtape. New Have, CT: Yale University Press.

Deming, D. (2017). The growing importance of social skills in the labor market. *Quarterly Journal of Economics*, 132(4), 1593-1640.

Duncan & Murnane (2011). "Introduction: The American Dream, Then and Now" in Greg Duncan and Richard Murnane eds., *Whither opportunity: Rising inequality, schools, and children's life chances.* New York, NY: Russell Sage Foundation.

Fahle, E., Kane, T., Patterson, T., Reardon, S., Staiger, D. & Stuart, E. (2023). School district and community factors associated with learning loss during the COVID-19 pandemic.

Feister, L. (2013). Early warning confirmed: A research update on third-grade reading. Baltimore, MD: Annie E. Casey Foundation.

Fredricks, J., & Simpkins, S. (2012). Promoting positive youth development through organized after-school activities. *Child Development Perspectives*, *6*(3), 280–287.

Goodman-Bacon (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2), 254-277.

Hanushek, E., Peterson, P., Talpey, L. & Woessmann, L. (2020). Long-run trends in the U.S. SES-achievement gap. NBER Working Paper No. 26764.

Hashim, Kane, Kelley-Kemple, Laski & Staiger (2020). Have income-based achievement gaps widened or narrowed? NBER Working Paper No. 27714.

Hill, C., Bloom, H., Black, A. & Lipsey, M. (2007). Empirical benchmarks for interpreting effect sizes in research. MDRC Working Paper on Research Methodology, p. 1-11.

Hochschild & Scovronick (2004). *The American dream and the public schools*. New York, NY: Oxford University Press.

Jacques, Basch, Fera & Basch (2022). Debate in public versus independent secondary schools in New York City. *Journal of Community Health*, 47, 745-749.

Jochim, A. (2016). The politics of the Common Core assessments. *Education Next*, 16(4).

Johnson, D. W., Johnson, R. T., & Stanne, M. B. (2000). *Cooperative learning methods: A meta-analysis*.

Johnson, D. W., Maruyama, G., Johnson, R., Nelson, D., & Skon, L. (1981). Effects of cooperative, competitive, and individualistic goal structures on achievement: A meta-analysis. *Psychological Bulletin*, *89*, 47–62.

Kidron, Y. & Lindsay, J. (2014). The effects of increased learning time on student academic and nonacademic outcomes: Findings from a meta-analytic review. American Institutes for Research.

Ko, T. & Mezuk, B. (2021). Debate participation and academic achievement among high school students in the Houston Independent School District: 2012 – 2015. *Educational Research and Reviews*, 16(6), 219-225.

Kraft, M. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, 49(4), 241-253.

Kraft, M. (2015). How to make additional time matter: Extending the school day for individualized tutorials. *Education Finance and Policy*, 10(1), 81-116.

Kuhfeld, M., Soland, J. & Lewis, K. (2022). Test score patterns across three Covid-19-impacted school years. Annenberg EdWorkingPaper 22-521.

Lipscomb, S. (2007). Secondary school extracurricular involvement and academic achievement: A fixed effect approach. *Economics of Education Review*, *26*(4), 463–472.

Litan, R. (2018). A counterintuitive proposal for improving education and healing America: Debate-centered instruction. Washington, D.C.: The Brookings Institution.

Mehta, J. & Fine, S. (2019). *In search of deeper learning: The quest to remake the American High School*. Cambridge, MA: Harvard University Press.

Mezuk, Bondarenko, Smith & Tucker (2011). Impact of participating in a policy debate program on academic achievement. *Educational Research and Reviews*, Vol. 6(9), 622-635.

Mezuk, B. (2009). Urban debate and high school educational outcomes for African American males: The case of the Chicago debate league. *The Journal of Negro Education*, 78(3), 290-304.

Mitchell, G. (1998). Pedagogical Possibilities for Argumentative Agency in Academic Debate. Argumentation and Advocacy, 35, 41-60.

Mosteller, F. (1995). The Tennessee study of class size in the early school grades. *The Future of Children*, 5(2), p. 113-126.

National Assessment of Educational Progress (2023). 2022 NAEP Civics Assessment. https://www.nationsreportcard.gov/highlights/civics/2022/

National Association of Urban Debate Leagues (2019). NAUDL Mission. Retrieved from: https://www.urbandebate.org/who-we-are/

National Speech and Debate Association (2019). History. Retrieved from: https://www.speechanddebate.org/history/

National Student Clearinghouse (2022). Working With Our Data. https://nscresearchcenter.org/workingwithourdata/

Oreopoulos and Salvanes (2011). Priceless: The nonpecuniary benefits of schooling. *Journal of Economic Perspectives*, 25(1), 159-184.

Pedersen, S. (2005). Urban adolescents' out-of-school activity profiles: Associations with youth, family, and school transition characteristics. *Applied Developmental Science*, *9*(2), 107–127.

Quinn, J. (1999). Where need meets opportunity. *The Future of Children*, *9*(2), 96–116.

Reardon, S. (2011). "The widening academic achievement gap between the rich and the poor" in Greg Duncan and Richard Murnane eds., *Whither opportunity: Rising inequality, schools, and children's life chances.* New York, NY: Russell Sage Foundation.

Research, Evaluation and Assessment, Minneapolis Public Schools (2015). 2014-15 Urban Debate League MPS Evaluation.

Shackelford (2019). The BUDL effect: Examining academic engagement outcomes of preadolescent Baltimore Urban Debate League participants. *Educational Researcher*, 48(3).

Shackelford, ScottRatliff & Mezuk (2019). Participating in a high school debate program and college matriculation and completion: Evidence from the Chicago Debate League. *Educational Research and Reviews*, 14(11), 397-409.

Slavin, R. E. (1980). Cooperative learning. *Review of Educational Research*, *50*(2), 315–342.

Slavin, R. E. (1983). When does cooperative learning increase student achievement? *Psychological Bulletin*, *94*, 429–445.

Sun & Abraham (Forthcoming). Estimating dynamic treatment effects in event studies with heterogenous treatment effects. Journal of Econometrics.

Stevenson, B. (2010). Beyond the classroom: Using Title IX to measure the return to high school sports. *Review of Economics and Statistics*, *92*(2), 284–301.

Swanson, Kopotic, Zamarro, Mills, Greene & Ritter (2019). An evaluation of the educational impact of college campus visits: A randomized experiment. https://www.sree.org/conferences/2019s/program/download/abstract/2735_0.pdf

West, M. & Lake, R. (2021). How much have students missed academically because of the pandemic? A review of the evidence to date. Center for Reinventing Public Education.

Winkler, C. & Fortner, K. (2014). Boston Debate League Results 2008-2012. Unpublished internal report.

Table 1. Sample Characteristics, K-12 Outcomes Analysis

	All BPS Schools		Debate Schools	
	Debate schools	Non-debate schools	Ever debaters	Never debaters
Demographics				
Asian	8***	10	8	9
Black	37*	36	42***	36
Hispanic	44***	33	39**	44
Native American	0	0	0	0
White	9***	19	9***	9
Other	1***	2	2	1
Female	48**	50	56***	47
Economically disadvantaged	82***	68	87***	84
English learner	36***	26	37***	38
First language English	55***	62	58**	53
Has IEP	23***	20	14***	24
Baseline achievement measures				
Standardized ELA Score (SD)	-0.13***	0.14	0.21***	-0.18
Standardized Math Score (SD)	-0.13***	0.18	0.17**	-0.17
Attendance rate	94*	94	95***	94
Suspensions	0.04	0.05	0.05**	0.09
N of students	56,306	27,471	3,515	39,903

	Ever debaters (n=3,515)			
Describing debate participation	Mean	SD	Min	Max
Years of debate	1.39	(0.75)	1	5
N of tournaments per year	3.11	(1.88)	1	10
N of regional tournaments per year	0.07	(0.43)	0	5
Grade in first year of debate	9.09	(1.88)	6	12
Ever middle school league	28			
Ever Spanish league	6			
Highest division				
Novice	29			
Junior Varsity	25			
Varsity	16			
Missing	30			

Note: *** p<0.001, ** p<0.01, * p<0.05. All values represent percentages unless otherwise specified. Sample is limited to BPS students who are observed in grade 6 or higher. Stars represent differences between the column with stars and the column immediately to its right. Never debaters are those who were enrolled in debate schools in years when the school offered debate while debate schools are students ever in schools that offered debate (which explains why the sum of columns three and four does not equal column one). Baseline achievement measures are from elementary school years when all students were too young to participate in debate, making these pre-treatment measures for all students.

Table 2. Sample Characteristics, Postsecondary Outcomes Analysis

		Never debaters		
	Ever debaters	All never debaters	High propensity to debate comparison	Coarsened exact matched comparison
Demographics				
Asian	9	10*	9	8
Black	44	35***	48***	43
Hispanic	38	40*	36	38
Native American	0	0	0***	0
White	8	13***	4***	9
Other	2	2	2	1*
Female	57	49***	63***	54
Economically disadvantaged	91	82***	96***	85***
English learner	39	33***	40	21
First language English	55	56	57	59
Has IEP	16	23***	7***	14
Baseline achievement measures				
Standardized ELA Score (SD)	0.14	0.02***	0.36***	0.07***
Standardized Math Score (SD)	0.13	0.06***	0.31***	0.08
Attendance rate	0.95	0.95***	0.96***	0.95*
Suspensions	0.03	0.04**	0.01***	0.01
Predicted probability of debate	0.11	0.06***	0.14***	0.04***
N of students	2,269	35,419	7,235	11,249

Note: *** p<0.01, ** p<0.05, * p<0.1. All values represent percentages unless otherwise specified. Stars represent differences between the column with the starts and the first column ("ever debaters"). Baseline achievement measures are from elementary school years when all students were too young to participate in debate, making these pre-treatment measures for all students.

Table 3. Pooled Estimates of the Impact of Debate Participation on Academic Outcomes

	Full Sample							
	ELA		Math		Attendance		Suspensions	
Debate	0.137***	0.126***	0.060	0.059*	0.030***	0.017***	-0.051	-0.008
	(0.032)	(0.032)	(0.038)	(0.028)	(0.005)	(0.003)	(0.041)	(0.026)
N of Observations	93,401	91,255	93,050	90,973	145,426	178,932	145,665	180,072
	Excluding Students Who Debated for Multiple Years							
	ELA		Math		Attendance		Suspensions	
Debate	0.123***	0.115***	0.062	0.059*	0.029***	0.016***	-0.045	-0.001
	(0.034)	(0.032)	(0.040)	(0.029)	(0.005)	(0.003)	(0.045)	(0.026)
N of Observations	92,833	90,633	92,498	90,361	144,323	177,483	144,562	178,622
	Excluding Students Who Started Debate in Ninth Grade							
	ELA		Math		Attendance		Suspensions	
Debate	0.131***	0.120***	0.049	0.047	0.025***	0.012***	-0.030	0.007
	(0.035)	(0.033)	(0.040)	(0.027)	(0.005)	(0.003)	(0.048)	(0.034)
N of Observations	89,528	87,373	89,190	87,094	139,858	172,884	140,090	174,009
Covariates	x		x		x		x	
Grade fixed effects	x		x		x		x	
Student fixed effects		x		x		x		x

Note: *** p<0.001, ** p<0.01, * p<0.05. All samples exclude non-debaters in debate schools in years when debate was offered. All models include year fixed effects, and a "debate" indicator equal to one if a student was participating in debate in a given year. Standard errors are clustered at the school level. Test score outcomes are in standard deviation units. Attendance is in percent of days present and suspensions are number of days suspended.

Table 4. Event Study Estimates of the Impact of Debate Participation on Achievement

	N of treated students	ELA		Math	
Year -5	728	-0.000	-0.069	-0.011	-0.028
		(0.018)	(0.052)	(0.019)	(0.048)
Year -4	1023	-0.026	-0.102*	-0.030*	-0.057
		(0.015)	(0.043)	(0.014)	(0.040)
Year -3	1556	-0.003	-0.056	-0.010	-0.018
		(0.013)	(0.036)	(0.014)	(0.031)
Year -2	2025	-0.002	-0.048	0.019	0.016
		(0.013)	(0.030)	(0.014)	(0.026)
Year -1	2163	0.047**	0.008	0.024	0.013
		(0.016)	(0.021)	(0.022)	(0.020)
Debate Year 1	1766	0.125***	0.097***	0.064	0.065*
		(0.034)	(0.027)	(0.040)	(0.028)
Debate Year 2	513	0.174***	0.141**	0.069	0.053
		(0.050)	(0.046)	(0.061)	(0.044)
Debate Year 3	65	0.125	0.083	-0.091	-0.064
		(0.112)	(0.131)	(0.116)	(0.136)
Debate Year 4	26	0.244**	0.144	0.102	0.091
		(0.092)	(0.111)	(0.069)	(0.084)
Debate Year 5	14	0.402***	0.355***	0.072	0.108
		(0.094)	(0.093)	(0.056)	(0.100)
Observations		93,401	90,996	93,050	90,973
Joint F-test		0.16	2.70	0.05	0.29
		[0.69]	[0.10]	[0.83]	[0.59]
Covariates		x		x	
Grade fixed effects		x		x	
Student fixed effects			x		x

Note: *** p<0.001, ** p<0.01, * p<0.05. All samples exclude non-debaters in debate schools in years when debate was offered. All models include year fixed effects, and a set of "debate" indicators for each pre and post-debate year that equal one if a student ever participated in debate and the given year was a certain number of years before or after the student participated. We omit year zero (the year prior to debate participation). Standard errors are clustered at the school level. Test score outcomes are in standard deviation units. Joint F-tests examine whether the coefficients for all pre-debate years combined are statistically different from zero.

Table 5. Pooled Estimates of the Impact of Debate Participation on ELA Achievement, by Baseline ELA Performance Quartile

	Mean Baseline ELA Score	ELA		Math		Attendance		Suspensions	
		0.232***	0.235***	0.114	0.098	0.046***	0.023***	-0.194*	-0.169**
Lowest Performing	-1.355	(0.057)	(0.056)	(0.061)	(0.054)	(0.008)	(0.005)	(0.092)	(0.057)
		18,702	17,869	18,908	18,072	29,890	29,677	29,953	29,750
		0.180**	0.164**	0.076	0.052	0.037***	0.023***	-0.046	-0.018
Low Performing	-0.290	(0.062)	(0.051)	(0.064)	(0.045)	(0.006)	(0.004)	(0.058)	(0.051)
		21,648	21,169	21,591	21,086	32,988	32,803	33,045	32,866
		0.148***	0.120***	0.053	0.056	0.028***	0.014***	-0.027	0.015
High Performing	0.416	(0.044)	(0.033)	(0.055)	(0.040)	(0.005)	(0.004)	(0.044)	(0.039)
		23,904	23,363	23,761	23,209	36,863	36,581	36,931	36,651
		0.109***	0.096***	0.046	0.061*	0.018***	0.011***	0.014	0.045
Highest Performing	1.138	(0.025)	(0.025)	(0.031)	(0.031)	(0.004)	(0.003)	(0.025)	(0.024)
		29,147	28,652	28,789	28,277	45,685	45,376	45,736	45,429
Covariates		x		x		x		x	
Grade fixed effects		x		x		x		x	
Student fixed effects			x		x		x		x

Note: *** p<0.001, ** p<0.01, * p<0.05. All samples exclude non-debaters in debate schools in years when debate was offered. Each row represents a regression run on a sample limited to a single baseline ELA performance quartile. All models include year fixed effects, and a "debate" indicator equal to one if a student was participating in debate in a given year. Standard errors are clustered at the school level. Test score outcomes are in standard deviation units. Attendance is in percent of days present and suspensions are number of days suspended.

Table 6. Pooled and Event Study Estimates of the Impact of Debate on ELA Subskills

	Language		Reading	
Debate (Pooled)	0.045	0.057	0.107**	0.101**
	(0.032)	(0.035)	(0.033)	(0.033)
Observations	76,063	73,118	76,063	73,118
	Language		Reading	
Year -5	-0.040	-0.127**	0.017	-0.082*
	(0.028)	(0.042)	(0.021)	(0.041)
Year -4	0.010	-0.072	0.002	-0.097**
	(0.031)	(0.040)	(0.019)	(0.036)
Year -3	0.020	-0.021	0.029	-0.038
	(0.020)	(0.032)	(0.018)	(0.035)
Year -2	-0.010	-0.033	0.021	-0.024
	(0.019)	(0.029)	(0.015)	(0.025)
Year -1	0.015	-0.015	0.036	-0.011
	(0.023)	(0.028)	(0.019)	(0.024)
Debate Year 1	0.036	0.027	0.119***	0.089**
	(0.037)	(0.034)	(0.035)	(0.030)
Debate Year 2	0.058	0.050	0.076	0.052
	(0.050)	(0.057)	(0.047)	(0.051)
Debate Year 3	0.185*	0.091	0.135	0.110
	(0.087)	(0.102)	(0.121)	(0.111)
Debate Year 4	0.049	0.050	0.168**	0.089
	(0.064)	(0.068)	(0.056)	(0.075)
Debate Year 5	0.213	0.224	0.038	0.128
	(0.245)	(0.288)	(0.232)	(0.237)
Observations	76,063	73,118	76,063	73,118
Covariates	x		x	
Grade fixed effects	x		x	
Student fixed effects		x		x

Note: *** p<0.001, ** p<0.01, * p<0.05. All samples exclude non-debaters in debate schools in years when debate was offered. Models in the top row include student fixed effects, year fixed effects, and a "debate" indicator equal to one if a student was participating in debate in a given year. All models below the first row include year fixed effects, and a set of "debate" indicators for each pre and post-debate year that equal one if a student ever participated in debate and the given year was a certain number of years before or after the student participated. We omit year zero (the year prior to debate participation). For all models, standard errors are clustered at the school level. Test score outcomes are in standard deviation units.

Table 7. Event Study Estimates of Debate Impacts on K-12 Non-Test Outcomes

	N of treated students	Attendance		Suspension	
Year -5	1309	-0.001	-0.024***	0.033**	0.085
		(0.002)	(0.007)	(0.012)	(0.055)
Year -4	1603	0.000	-0.013*	0.028*	0.075
		(0.001)	(0.006)	(0.013)	(0.053)
Year -3	2033	0.004*	-0.003	-0.003	0.046
		(0.001)	(0.005)	(0.014)	(0.043)
Year -2	2294	0.007***	0.004	-0.025	0.046
		(0.002)	(0.005)	(0.019)	(0.039)
Year -1	2549	0.010***	0.005	-0.067**	0.006
		(0.003)	(0.004)	(0.026)	(0.030)
Debate Year 1	3515	0.030***	0.015***	-0.047	0.022
		(0.005)	(0.003)	(0.045)	(0.020)
Debate Year 2	886	0.031***	0.016***	-0.057	0.016
		(0.006)	(0.004)	(0.047)	(0.034)
Debate Year 3	327	0.029***	0.014**	-0.106*	-0.091*
		(0.007)	(0.005)	(0.044)	(0.038)
Debate Year 4	125	0.044***	0.022**	-0.021	-0.027
		(0.012)	(0.008)	(0.069)	(0.067)
Deabte Year 5	29	0.034**	0.040***	-0.207**	-0.211***
		(0.012)	(0.010)	(0.064)	(0.059)
Observations		145,426	178,932	145,665	180,072
Joint F-test		8.65	1.54	0.33	1.51
		[0.00]	[0.22]	[0.57]	[0.22]
Covariates		x		x	
Grade fixed effects		x		x	
Student fixed effects			x		x

Note: *** p<0.001, ** p<0.01, * p<0.05. All samples exclude non-debaters in debate schools in years when debate was offered. All models include year fixed effects, and a set of "debate" indicators for each pre and post-debate year that equal one if a student ever participated in debate and the given year was a certain number of years before or after the student participated. We omit year zero (the year prior to debate participation). Standard errors are clustered at the school level. Attendance is in percent of days present and suspensions are number of days suspended. Joint F-tests examine whether the coefficients for all pre-debate years combined are statistically different from zero.

Table 8. Pooled Estimates of the Impact of Debate Participation on Postsecondary Outcomes

	Full Sample								
	Graduated High School			Enrolled in Postsecondary Institution			ELA		
Ever Debater	0.117***	0.0887**	0.120***	0.116**	0.124***	0.144***	0.236**	0.197***	0.206***
	(0.033)	(0.026)	(0.029)	(0.042)	(0.032)	(0.030)	(0.082)	(0.039)	(0.044)
Observations	10,596	4,721	12,188	10,596	4,721	12,188	8,020	3,893	9,256
	Excluding Students Who Debated for Multiple Years								
Ever Debater	0.107**	0.085**	0.101***	0.103*	0.113**	0.113***	0.174	0.074	0.142*
	(0.034)	(0.026)	(0.027)	(0.041)	(0.034)	(0.029)	(0.087)	(0.044)	(0.053)
Observations	10,254	4,492	20,898	10,254	4,492	20,898	7,761	3,744	16,935
	Excluding Students Who Started Debate in Ninth Grade								
Ever Debater	0.147***	0.130***	0.142***	0.152***	0.170***	0.163***	0.243**	0.199***	0.203***
	(0.037)	(0.027)	(0.030)	(0.037)	(0.034)	(0.029)	(0.087)	(0.045)	(0.049)
Observations	10,289	4,515	20,934	10,289	4,515	20,934	7,737	3,726	16,913
Excluding Non-Debaters in Debate Schools	x			x			x		
High Propensity to Debate Comparison		x			x			x	
Coarsened Exact Matched Comparison			x			x			x

Note : *** p<0.001, ** p<0.01, * p<0.05. Samples include one observation per student. All models include a binary indicator for whether a student ever participated in debate, year fixed effects, and our full set of covariates. Standard errors are clustered at the school level.

Table 9. Pooled Estimates of the Impact of Debate Participation on Type of Postsecondary Enrollment

	Enrolled in Postsecondary Institution	Enrolled in 2-Year Institution	Enrolled in 4-Year Institution	Enrolled in Public Institution	Enrolled in Private Institution
Ever Debater	0.130***	0.041*	0.116***	0.056	0.094***
	(0.031)	(0.018)	(0.028)	(0.031)	(0.025)
Observations	22,503	22,503	22,503	22,503	22,503

Note : *** p<0.001, ** p<0.01, * p<0.05. Standard errors are clustered at the school level.

Table 10. Pooled Estimates of the Impact of Debate Participation on Postsecondary Outcomes, by Baseline ELA Performance Quartile

	Mean Baseline ELA Score	Graduated High School	Enrolled in Postsecondary Institution	Enrolled in 2-Year Institution	Enrolled in 4-Year Institution	Enrolled in Public Institution	Enrolled in Private Institution
		0.164***	0.205***	0.073*	0.157***	0.127**	0.100**
Lowest Performing	-0.999	(0.025)	(0.043)	(0.033)	(0.044)	(0.039)	(0.033)
		5,633	5,633	5,633	5,633	5,633	5,633
		0.157**	0.180***	0.035	0.149***	0.062	0.128***
Low Performing	0.021	(0.045)	(0.039)	(0.024)	(0.032)	(0.037)	(0.029)
		8,483	8,483	8,483	8,483	8,483	8,483
		0.069*	0.082	0.041	0.061	0.052	0.035
High Performing	0.294	(0.030)	(0.051)	(0.028)	(0.048)	(0.038)	(0.038)
		2,781	2,781	2,781	2,781	2,781	2,781
		0.053*	0.056	-0.014	0.094*	-0.027	0.099*
Highest Performing	1.013	(0.022)	(0.031)	(0.024)	(0.039)	(0.029)	(0.039)
		5,606	5,606	5,606	5,606	5,606	5,606

Note: *** p<0.001, ** p<0.01, * p<0.05. All models include student fixed effects, year fixed effects, and a "debate" indicator equal to one if a student was participating in debate in a given year. Standard errors are clustered at the school level. Test score outcomes are in standard deviation units. Attendance is in percent of days present and suspensions are number of days suspended. Each row represents a regression run on a sample limited to a single baseline ELA performance quartile.

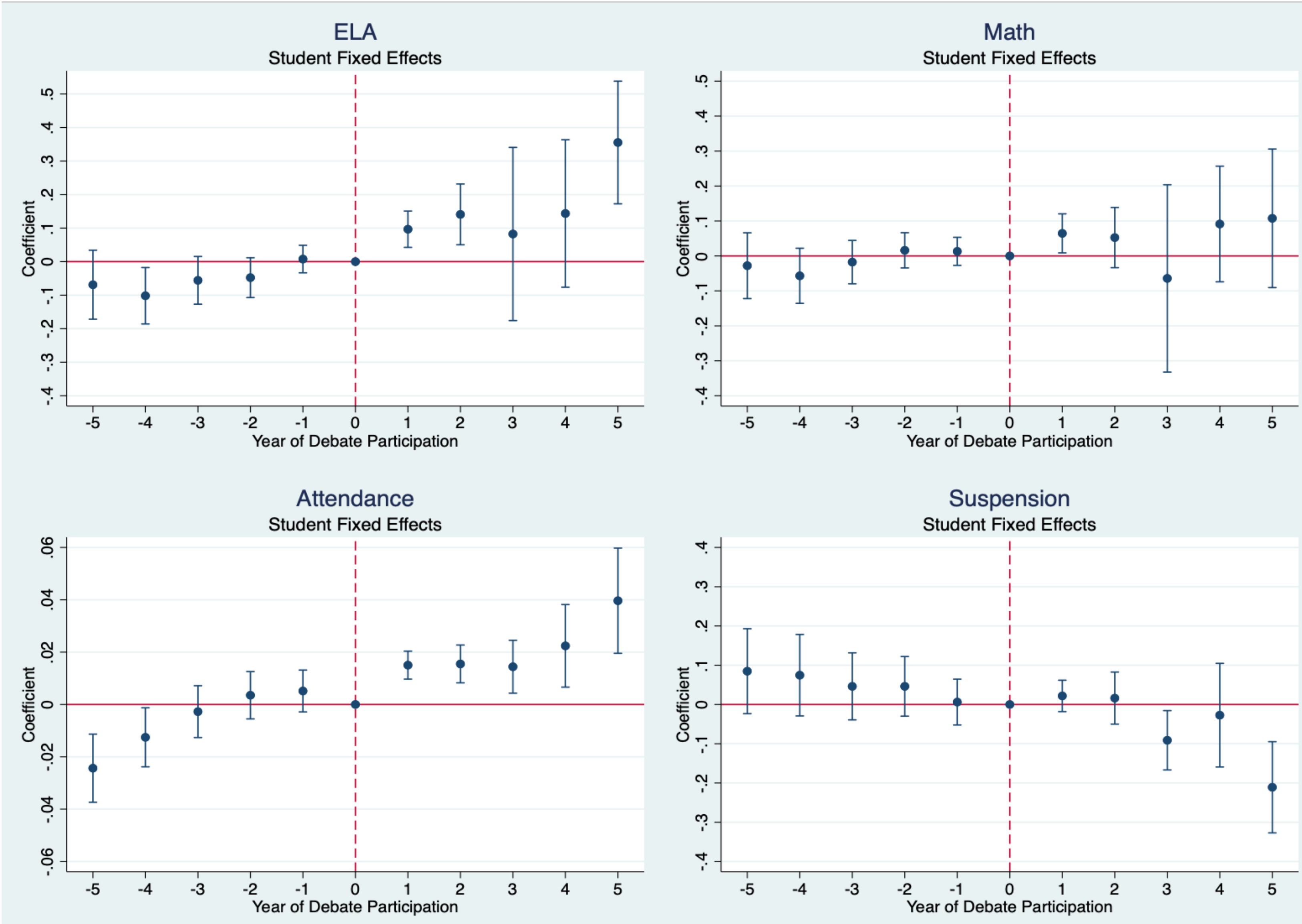


Figure 1. Event Study Estimates of Debate Impacts, Student Fixed Effects Model Excluding Non-Debaters in Debate Schools.

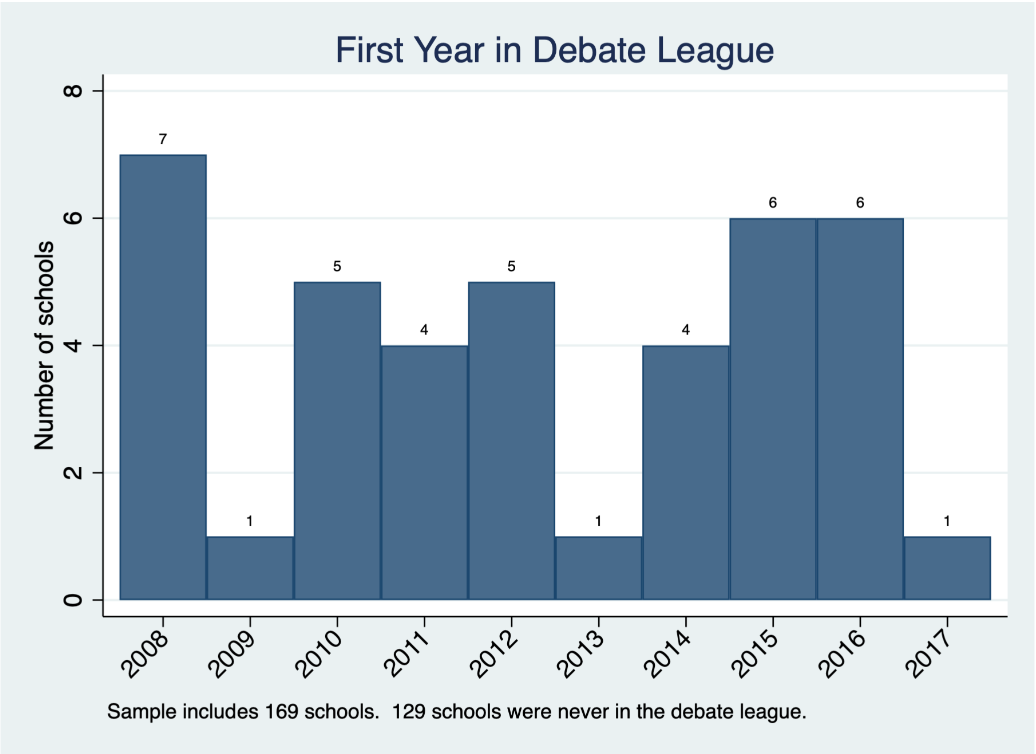
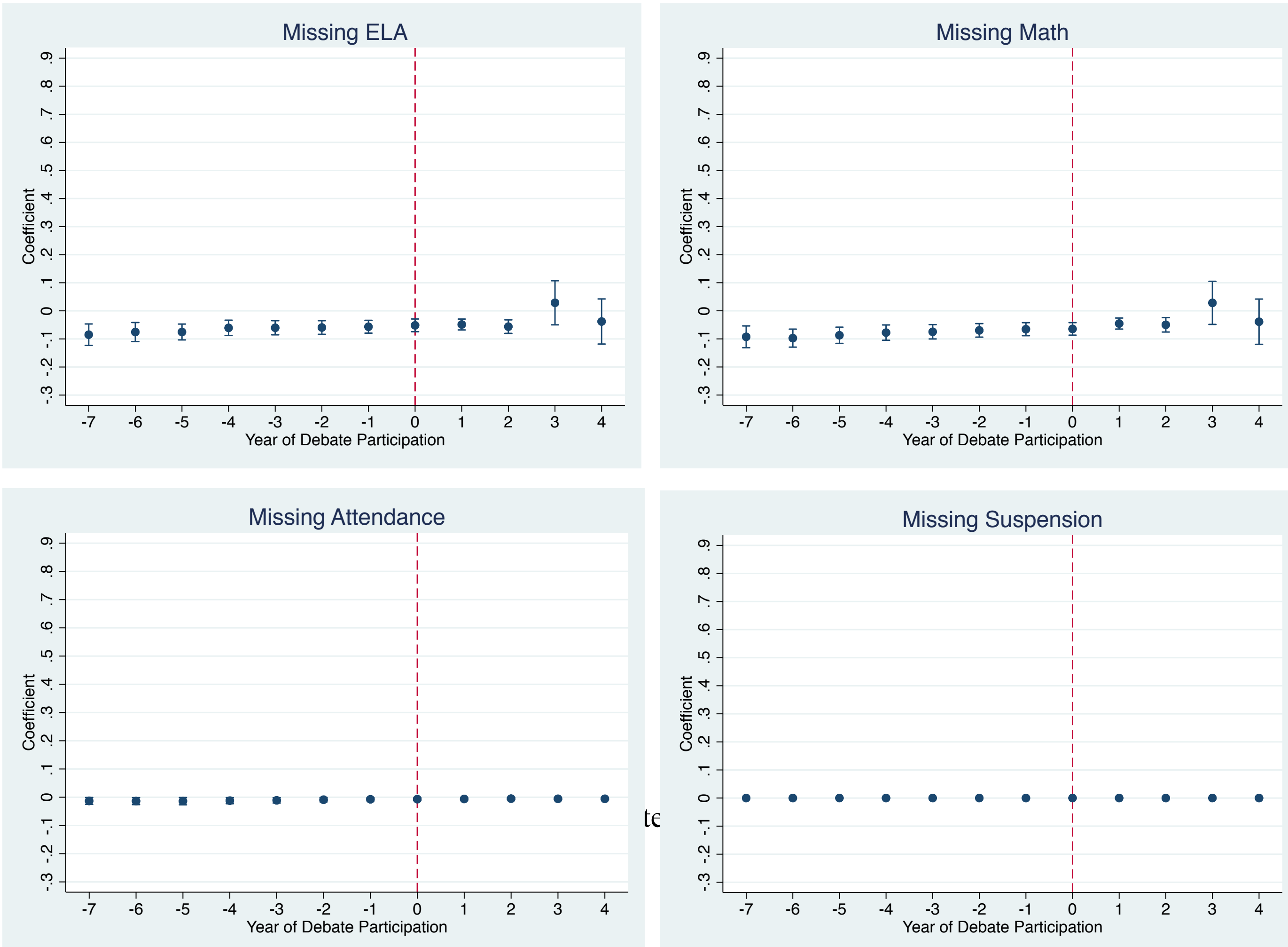


Figure A1. Number of schools joining the Boston Debate League by year.



Appendix Figure A2. Event Study Estimates of Debate Impacts on Outcome Missingness, Unconstrained.

Appendix Figure A3. Comparing Language and Reading Standards, Grade 8 in 2017

Language		Reading	
Conventions of Standard English		*Key Ideas and Details*	
1	Demonstrate command of the conventions of standard English grammar and usage when writing or speaking; retain and further develop language skills learned in previous grades.	1	Cite the textual evidence that most strongly supports analysis of what a text states explicitly as well as inferences drawn from the text, quoting or paraphrasing as appropriate.
2	Demonstrate command of the conventions of standard English capitalization, punctuation, and spelling when writing.	2	Determine a text's central idea(s) and analyze its/their development over the course of the text, including relationships to supporting ideas; provide an objective summary of a text.
		3	Analyze how a text makes connections among and distinctions between individuals, ideas, or events (e.g., through comparisons, analogies, or categories).
Knowledge of Language		*Craft and Structure*	
3	Use knowledge of language and its conventions when writing, speaking, reading, or listening.	4	Determine the meaning of words and phrases as they are used in a text, including figurative, connotative, and technical meanings; analyze the impact of specific word choices on meaning and tone, including analogies or allusions to other texts.
		5	Analyze in detail the structural elements of a text, including the role of specific sentences, paragraphs, and text features in developing and refining a key concept.
		6	Determine an author's point of view or purpose in a text and analyze how the author acknowledges and responds to conflicting evidence or viewpoints.
Vocabulary Acquisition and Use		*Integration of Knowledge and Ideas*	
4	Determine or clarify the meaning of unknown and multiple-meaning words or phrases based on *grade 8 reading and content*, choosing flexibly from a range of strategies.	7	Evaluate the advantages and disadvantages of using different mediums (e.g., print or digital text, video, multimedia) to present a particular topic or idea.
5	Demonstrate understanding of figurative language, word relationships, and nuances in word meanings.	8	Delineate and evaluate the argument and specific claims in a text, assessing whether the reasoning is sound and the evidence is relevant and sufficient; recognize when irrelevant evidence is introduced.
6	Acquire and use accurately grade-appropriate general academic and domain-specific words and phrases; independently research words and gather vocabulary knowledge.	9	Analyze a case in which two or more texts provide conflicting information on the same topic and identify where the texts disagree on matters of fact or interpretation.
		Range of Reading and Level of Text Complexity	
		10	Independently and proficiently read and comprehend literary nonfiction representing a variety of genres, cultures, and perspectives and exhibiting complexity appropriate for at least grade 8.

Note: Reading Standards are taken from standards for informational texts. All standards are drawn from the 2017 Massachusetts Curriculum Framework for English Language Arts and Literacy for Grade 8.

Table A1. Comparative Interrupted Time Series Estimates of Debate Impacts

	ELA		Math		Attendance		Suspensions	
Debate	0.115***	0.104***	0.067*	0.062*	0.023***	0.017***	-0.010	0.012
	(0.027)	(0.025)	(0.031)	(0.026)	(0.004)	(0.003)	(0.022)	(0.020)
Ever Debate X Year	0.004	0.010	-0.004	-0.001	0.001	0.000	-0.010	-0.009
	(0.003)	(0.009)	(0.005)	(0.007)	(0.001)	(0.001)	(0.008)	(0.009)
Year	0.002	0.009	0.002	-0.000	-0.004***	-0.010***	0.017**	0.036***
	(0.002)	(0.009)	(0.003)	(0.007)	(0.001)	(0.001)	(0.006)	(0.008)
Ever Debate	-7.581		7.823		-1.975		19.880	
	(6.699)		(8.736)		(1.592)		(16.200)	
Observations	93,401	91,255	93,050	90,973	145,426	178,932	145,665	180,072
Covariates	x		x		x		x	
Grade fixed effects	x		x		x		x	
Student fixed effects		x		x		x		x

Note: *** p<0.001, ** p<0.01, * p<0.05. All models include a "debate" indicator equal to one if a student was participating in debate in a given year (the interaction between debate status and whether a student was ever treated), an interaction between whether a student ever participated in debate and the linear time trend, and a linear time trend "year." Standard errors are clustered at the school level. Test score outcomes are in standard deviation units. Attendance is in percent of days present and suspensions are number of days suspended.

Table A2. Event Study Estimates of the Impact of Debate Participation on ELA and Math Achievement, by Cohort

	ELA								Math							
	2010 Cohort	2011 Cohort	2012 Cohort	2013 Cohort	2014 Cohort	2015 Cohort	2016 Cohort	2017 Cohort	2010 Cohort	2011 Cohort	2012 Cohort	2013 Cohort	2014 Cohort	2015 Cohort	2016 Cohort	2017 Cohort
Year -5					-0.009	-0.144*	-0.002	-0.111					-0.032	-0.076	0.062	-0.015
					(0.071)	(0.064)	(0.083)	(0.080)					(0.067)	(0.072)	(0.082)	(0.078)
Year -4				-0.013	-0.110	-0.163**	-0.059	-0.050				-0.125	-0.110	-0.073	0.050	0.014
				(0.073)	(0.062)	(0.053)	(0.071)	(0.069)				(0.067)	(0.059)	(0.060)	(0.083)	(0.080)
Year -3			-0.181**	-0.008	-0.061	-0.071	0.052	-0.040			-0.073	-0.084	-0.013	-0.021	0.112	-0.047
			(0.060)	(0.070)	(0.060)	(0.048)	(0.056)	(0.057)			(0.058)	(0.057)	(0.046)	(0.054)	(0.072)	(0.074)
Year -2		-0.012	-0.128*	-0.046	-0.007	-0.067	0.034	-0.082		-0.041	0.029	-0.054	0.054	-0.001	0.102	-0.061
		(0.065)	(0.060)	(0.055)	(0.058)	(0.042)	(0.047)	(0.055)		(0.065)	(0.052)	(0.052)	(0.042)	(0.043)	(0.071)	(0.066)
Year -1	0.125	0.001	-0.072	0.005	0.063	-0.049	0.082	-0.000	0.144*	-0.052	0.025	0.008	0.083	-0.010	0.088	-0.170*
	(0.072)	(0.071)	(0.057)	(0.050)	(0.042)	(0.045)	(0.054)	(0.054)	(0.065)	(0.066)	(0.039)	(0.048)	(0.044)	(0.049)	(0.066)	(0.072)
Debate Year 1	0.142	0.011	0.103	0.110	0.216***	-0.048	0.195**	0.084	0.180	0.002	0.093	0.065	0.149*	-0.064	0.146	0.024
	(0.098)	(0.065)	(0.068)	(0.067)	(0.064)	(0.065)	(0.067)	(0.057)	(0.091)	(0.087)	(0.053)	(0.046)	(0.064)	(0.057)	(0.083)	(0.053)
Debate Year 2	0.282**	0.211*	0.019	0.170*	0.027	0.177**	0.357**		0.052	0.064	0.055	-0.051	-0.002	0.083	0.174	
	(0.098)	(0.086)	(0.111)	(0.072)	(0.089)	(0.062)	(0.112)		(0.142)	(0.060)	(0.112)	(0.101)	(0.082)	(0.078)	(0.112)	
Debate Year 3	0.974***	-0.381***	0.249*	-0.039	0.083	0.085			0.363***	0.001	0.135	-0.417*	0.101	0.163		
	(0.058)	(0.076)	(0.121)	(0.209)	(0.247)	(0.297)			(0.055)	(0.091)	(0.185)	(0.175)	(0.197)	(0.117)		
Debate Year 4		0.728	0.227	-0.077	-0.185					0.355	0.140	0.153	-0.175			
		(0.489)	(0.146)	(0.090)	(0.441)					(0.433)	(0.094)	(0.130)	(0.263)			
Debate Year 5		1.591**	0.290**	0.241**							0.173	0.018				
		(0.515)	(0.111)	(0.086)							(0.126)	(0.110)				
Observations	78,658	78,928	80,228	79,955	81,195	81,223	80,514	80,142	77,315	77,581	78,874	78,590	79,831	79,866	79,160	78,802
N of Treated Students (in post-period)	53	67	266	209	330	452	320	281	52	65	262	208	330	444	319	284
Joint F-test	3.04	0.01	6.36	0.09	0.24	7.38	0.18	1.42	4.92	0.60	0.02	1.94	0.01	0.74	1.70	0.90
	[0.08]	[0.93]	[0.01]	[0.76]	[0.63]	[0.01]	[0.67]	[0.24]	[0.03]	[0.44]	[0.89]	[0.17]	[0.93]	[0.39]	[0.19]	[0.34]

Note: *** p<0.001, ** p<0.01, * p<0.05.

Table A3. Event Study Estimates of the Impact of Debate Participation on Non-Test Outcomes, by Cohort

	Attendance								Suspensions							
	2010 Cohort	2011 Cohort	2012 Cohort	2013 Cohort	2014 Cohort	2015 Cohort	2016 Cohort	2017 Cohort	2010 Cohort	2011 Cohort	2012 Cohort	2013 Cohort	2014 Cohort	2015 Cohort	2016 Cohort	2017 Cohort
Year -5					-0.016**	-0.016**	-0.019***	-0.014**					0.095	0.140	0.128	0.275
					(0.006)	(0.005)	(0.005)	(0.004)					(0.075)	(0.085)	(0.103)	(0.178)
Year -4				-0.007	-0.012*	-0.008	-0.010**	-0.009**				0.000	0.095	0.141	0.169	0.313
				(0.006)	(0.005)	(0.004)	(0.004)	(0.003)				(0.066)	(0.074)	(0.089)	(0.103)	(0.173)
Year -3			-0.010	-0.007	-0.006	-0.002	-0.006*	-0.007*			-0.044	0.010	0.084	0.128	0.128	0.257
			(0.005)	(0.005)	(0.004)	(0.004)	(0.003)	(0.003)			(0.059)	(0.062)	(0.074)	(0.090)	(0.094)	(0.153)
Year -2		0.010	-0.014**	-0.010*	-0.002	-0.002	-0.001	-0.004		-0.179	-0.031	0.035	0.101	0.179*	0.080	0.216
		(0.008)	(0.005)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)		(0.138)	(0.058)	(0.052)	(0.078)	(0.088)	(0.084)	(0.115)
Year -1	0.011	0.015*	-0.006	-0.004	-0.003	-0.002	-0.002	-0.003	0.003	-0.118	-0.092	0.016	0.101	0.125	0.034	0.079
	(0.011)	(0.007)	(0.004)	(0.004)	(0.003)	(0.004)	(0.002)	(0.003)	(0.008)	(0.129)	(0.065)	(0.061)	(0.076)	(0.076)	(0.069)	(0.095)
Debate Year 1	0.010	0.003	0.001	0.004	0.000	0.005	0.002	0.007*	-0.011	-0.131	-0.069	0.157	0.132	0.029	0.002	-0.001
	(0.013)	(0.017)	(0.004)	(0.005)	(0.004)	(0.003)	(0.004)	(0.003)	(0.009)	(0.126)	(0.079)	(0.132)	(0.094)	(0.061)	(0.161)	(0.091)
Debate Year 2	0.012	0.014	0.014*	0.000	0.010**	0.009*	0.010		0.011	-0.202	-0.241**	0.180	-0.042	-0.061	-0.112	
	(0.008)	(0.021)	(0.006)	(0.007)	(0.003)	(0.004)	(0.006)		(0.009)	(0.154)	(0.088)	(0.150)	(0.052)	(0.090)	(0.169)	
Debate Year 3	0.018*	0.030***	-0.004	0.001	0.009	0.006			-0.003	-0.373**	-0.296***	-0.195*	-0.287*	-0.072		
	(0.008)	(0.006)	(0.010)	(0.010)	(0.006)	(0.005)			(0.013)	(0.127)	(0.080)	(0.089)	(0.121)	(0.101)		
Debate Year 4		0.021	0.011	0.016*	-0.005					-0.458**	-0.192*	0.184	2.196*			
		(0.011)	(0.008)	(0.007)	(0.008)					(0.148)	(0.083)	(0.593)	(1.050)			
Debate Year 5		0.054***	0.013	-0.002						-0.499	-0.356**	-0.262*				
		(0.012)	(0.010)	(0.006)						(0.271)	(0.125)	(0.130)				
Observations	78,635	78,905	80,205	79,932	81,171	81,200	80,491	80,119	78,658	78,928	80,228	79,955	81,195	81,223	80,514	80,142
N of Treated Students (in post-period)	122	211	476	373	658	566	490	477	122	211	476	373	658	566	490	477
	1.03	3.43	5.54	2.67	4.23	2.98	10.04	8.88	0.10	1.31	0.95	0.08	1.70	3.38	2.03	2.87
Joint F-test	[0.31]	[0.07]	[0.02]	[0.10]	[0.04]	[0.09]	[0.00]	[0.00]	[0.76]	[0.25]	[0.33]	[0.78]	[0.19]	[0.07]	[0.16]	[0.09]

Note: *** p<0.001, ** p<0.01, * p<0.05.

Table A4. Impact of School Joining Debate League

	ELA	Math	Attendance	Suspension
Year -8	-0.105	-0.281	-0.008	0.073
	(0.145)	(0.174)	(0.021)	(0.103)
Year -7	-0.104	-0.118	0.016	0.021
	(0.080)	(0.082)	(0.013)	(0.050)
Year -6	-0.002	0.046	0.018*	0.002
	(0.090)	(0.112)	(0.011)	(0.046)
Year -5	0.076	0.118	0.026**	-0.017
	(0.055)	(0.081)	(0.011)	(0.041)
Year -4	0.064	0.072	0.028***	-0.016
	(0.049)	(0.076)	(0.010)	(0.034)
Year -3	-0.005	0.005	0.035**	-0.026
	(0.049)	(0.064)	(0.016)	(0.030)
Year -2	-0.046	-0.059	0.007	-0.024
	(0.054)	(0.063)	(0.009)	(0.046)
Year -1	-0.027	-0.055	-0.004	-0.010
	(0.035)	(0.043)	(0.005)	(0.037)
League Year 1	-0.035	-0.004	-0.002	-0.022
	(0.037)	(0.037)	(0.006)	(0.021)
League Year 2	-0.049	-0.013	-0.005	0.085
	(0.049)	(0.046)	(0.006)	(0.064)
League Year 3	-0.010	-0.005	0.007	0.151**
	(0.049)	(0.052)	(0.014)	(0.076)
League Year 4	-0.030	-0.047	0.012	0.044
	(0.045)	(0.050)	(0.016)	(0.059)
League Year 5	-0.033	-0.036	0.001	-0.006
	(0.068)	(0.065)	(0.016)	(0.066)
League Year 6	-0.052	-0.113*	0.003	-0.071
	(0.071)	(0.065)	(0.018)	(0.066)
League Year 7	-0.036	-0.111	0.004	-0.073
	(0.063)	(0.083)	(0.021)	(0.084)
League Year 8	-0.071	-0.055	0.026	-0.162**
	(0.126)	(0.097)	(0.027)	(0.066)
League Year 9	-0.016	0.083	0.022	-0.119*
	(0.092)	(0.106)	(0.026)	(0.069)
League Year 10	0.020	0.054	0.028	-0.124
	(0.108)	(0.131)	(0.031)	(0.076)
Observations	216,480	215,956	407,302	409,060

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A5. Heterogeneity of Debate Impacts on Postsecondary Outcomes, by Student Demographic Characteristics

	Graduated High School (Outcome)										
	Middle School	High School	Male	Black	Hispanic	Econ. Dis-advantaged	First Language English	ELA Baseline	Math Baseline	Attendance Baseline	Suspension Baseline
Ever Debater X Subgroup	-0.175	0.118***	0.053	-0.035	0.073*	-0.048	-0.040	-0.067***	-0.074***	-0.682*	0.136***
	(0.118)	(0.026)	(0.028)	(0.029)	(0.031)	(0.074)	(0.031)	(0.016)	(0.014)	(0.334)	(0.028)
Observations	22,503	22,503	22,503	22,503	22,503	22,503	22,503	22,503	22,503	22,503	22,503
	Enrolled In Postsecondary Institution (Outcome)										
	Middle School	High School	Male	Black	Hispanic	Econ. Dis-advantaged	First Language English	ELA Baseline	Math Baseline	Attendance Baseline	Suspension Baseline
Ever Debater X Subgroup	-0.122	0.133***	-0.013	-0.047	0.060	0.035	-0.110**	-0.094***	-0.097**	0.055	0.004
	(0.127)	(0.032)	(0.035)	(0.039)	(0.043)	(0.054)	(0.039)	(0.025)	(0.028)	(0.361)	(0.059)
Observations	22,503	22,503	22,503	22,503	22,503	22,503	22,503	22,503	22,503	22,503	22,503

Note: *** p<0.001, ** p<0.01, * p<0.05.