Impact of participating in a policy debate program on academic achievement: Evidence from the Chicago Urban Debate League

Article *in* Educational Research and Reviews · January 2011

CITATIONS
9

READS
282

4 authors, including:

Briana Mezuk
University of Michigan
156 PUBLICATIONS **3,443** CITATIONS

SEE PROFILE

Irina Bondarenko
Centers for Disease Control and Prevention
22 PUBLICATIONS **704** CITATIONS

SEE PROFILE

Educational Research and Reviews Vol. 6(9), pp. 622-635, 5 September, 2011
Available online at http://www.academicjournals.org/ERR
ISSN 1990-3839 ©2011 Academic Journals

Full Length Research Paper

Impact of participating in a policy debate program on academic achievement: Evidence from the Chicago Urban Debate League

Briana Mezuk[1]*, Irina Bondarenko[2], Suzanne Smith[3] and Eric Tucker[4,5]

[1]Department of Epidemiology and Community Health, Virginia Commonwealth University, Richmond VA.
[2]Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor MI.
[3]Department of Sociology, University of Chicago, Chicago IL.
[4]Temple University, Philadelphia PA.

Accepted 14 July, 2011

Policy makers have advanced out-of-school time learning as a means to address far-reaching class and racial/ethnic disparities in high school achievement and college readiness, particularly in urban districts. However, limited data have hindered large-scale efforts to evaluate the influence of such activities on student achievement. Recent federal policy has encouraged the development of data systems that track students over the academic life course, and while these datasets hold great opportunity for research they pose inherent methodological challenges. This study applies a novel statistical approach in a comprehensive administrative dataset to evaluate the relationship between participating in a policy debate program and academic achievement in the Chicago Public School (CPS) district from 1997 to 2006 (N = 9145). Using multiple imputation to account for missing data and selective attrition, and propensity score matching to account for self-selection, we find that debaters were more likely to graduate, more likely to meet ACT college-readiness benchmarks, and had greater gains in cumulative grade point average (GPA) over the course of high school relative to comparable peers. This is the largest evaluation study of a debate program on achievement, and these findings suggest that debate programs may offer a means to extend learning time and promote engagement with scholastic materials in a manner that translates into academic performance.

Key words: Education, out-of-school time learning, afterschool activities, achievement gap.

INTRODUCTION

In the United States pronounced disparities across place, social class, and race/ethnicity remain in academic achievement, graduation rates, and college-readiness, despite recent reform efforts. On average, graduation rates are particularly low in urban public school districts (Roscigno et al., 2006; Balfanz and Legters, 2004). For example, the 2005 graduation rate for the Chicago Public School district was 51.0% (an improvement of nine percentage points from a decade earlier), as compared to 70.6% for the US as a whole (Swanson, 2009). Racial and ethnic minorities disproportionately experience even poorer educational attainment within these underperforming urban settings. For example, within Chicago the 2005 graduation rate for White students was 11.4 percentage points higher than for Black students (58.3% versus 46.9%, respectively) (Chicago Public Schools, 2009). Minority and low-income students in urban areas also lag in college readiness (Green and Forster, 2003; Green and Winters, 2005), and those students who do graduate from high school and enroll in college are more likely to be enrolled in remedial classes (Venezia et al., 2004) and are still less likely to complete college (Adelman, 2004). Disparities in graduation rates and coll-

*Corresponding author: Email: bmezuk@vcu.edu. Tel: (804) 628-2511. Fax: (804) 628-9773.

ege readiness contribute to the perpetuation of social inequality, since educational achievement is tightly linked with earning power, job advancement, and social class in adulthood (Blau and Duncan, 1976) as well as health conditions and health behaviors (LaVeist, 2005).

Some of the most pressing social disparities are those that exist across social class, racial/ethnic groups, and their intersection, as a disproportionate number of Black and Latino students come from lower social class families (LaVeist, 2005). The class-based achievement gap emerges early and persists through high school completion and college matriculation, even among high-achieving students (Wyner et al., 2009). This gap appears to grow during the summer months when students are not in school, as evidenced by the growth in achievement disparities between low and high income students each fall (Downey et al., 2004; Alexander et al., 2007). Educational disparities across racial/ethnic groups, often labeled the “Black-White” achievement gap, are also present early and persist throughout schooling, even among students within the same school (Stiefel et al., 2007). For example, in 2004, 17% of White 13-year old students (approximately 8th grade) were at or above the most proficient level on the National Assessment of Educational Progress (NAEP) measure, as compared to 5% of Black and 4% of Latino students (US Department of Education, 2006).

Mirroring the disparities seen in basic literacy, the achievement gaps extend to secondary literacy skills. Students in urban schools, particularly Black and Latino students, have low secondary literacy rates relative to White, suburban students (Snipes and Horowitz, 2008). Most school-based interventions to improve reading skills among adolescents are focused on basic rather than more complex aspects of literacy; interventions tend to be focused at the early childhood and elementary levels, with relatively little focus on secondary literacy (Miller, 2009). Despite investments in basic literacy, students who lack secondary literacy skills are more likely to drop out of school and are less likely to be college-ready (Neild and Balfanz, 2006).

Sources of the achievement gaps

Recent evidence suggests that race and class achievement gaps are substantially the result of unequal access to opportunities. The achievement gap between Whites and disadvantaged minority students (Black and Latino) has been primarily (but not exclusively) attributed to differences in the quality of instructional opportunities provided between and within schools (Condron 2009). Between schools, the “Black-White” achievement gap results from inequities in resources provided to primarily-minority-serving schools; within schools, the achievement gap is driven by inequalities in educational opportunities presented to White students versus their minority schoolmates (Fryer and Levitt 2001; Condron 2007). These disparities in school-based resources suggest school-based solutions to the achievement gap problem; however, intervention efforts address multiple aspects of the gap, including opportunities for academic engagement, scholastic resources, and quality of teaching (Condron, 2009; Rothstein, 2004; Lareau, 2003).

Programs aimed at reducing the achievement gaps

A comprehensive review of research on interventions aimed at reducing these achievement gaps is beyond the scope of this article, except to say that substantial evidence suggests that they are most effective and have the most substantial return on investment (in terms of economic and social outcomes in adolescence and adulthood) when initiated at an early age (Heckman et al., 2009). However, evaluations of the long-term impacts of highly-effective early interventions have produced mixed results (Currie and Thomas, 2000; Schweinhart and Weikart, 1997), particularly in terms of scholastic outcomes in adolescence.

There is limited evidence supporting the effectiveness of programs aimed at improving academic outcomes for low income and minority secondary school students (that is, out-of-school-time programs). In the majority of instances, effects are decidedly modest, and it is unclear whether these gains translate into high school completion or college-readiness (Lauer et al., 2006; Cullen et al., 2006). One of the broadest reaching efforts to improve college-readiness among low-income adolescents is the Gaining Early Awareness and Readiness for Undergraduate Program (GEAR UP), a federal partnership program aimed at socioeconomically-disadvantaged adolescents (ACT, 2007). A recent non-experimental evaluation of two GEAR UP cohorts reported inconsistent evidence of a modest impact on college-readiness. In one cohort, GEAR UP participants were 16% more likely to score at the college-readiness ACT benchmark in English and 27% more likely to score at the benchmark for Reading (ACT, 2007), but not better than comparison students in Science or Mathematics. The second GEAR UP cohort did not perform better than the comparison group on any section of the ACT.

In recent years, federal policy has encouraged states and school systems to establish longitudinal data systems for tracking students and monitoring achievement gains. Though designed primarily for reporting and evaluation, these datasets offer a rich resource for academic research that attempts to understand the mechanisms driving differential achievement. However, because these datasets are designed from an administrative perspective, they present unique methodological challenges for evaluation research. Methodologically, this paper attempts to confront these challenges and employs a novel research design to account for missing data and

self-selection. Empirically, this article builds on recent research that has identified an innovative approach to addressing achievement gaps for urban school students: Urban Debate Leagues.

Debate as an academic activity

Policy debate is an interscholastic and co-curricular activity in which students to face off against each other in a structured exchange centered on pressing policy issues. It is an academic competition centered on communication of evidence-based argument (Breger, 2000; Mitchell, 1998). Each debate round consists of two teams of two students who take opposing (for example, affirmative and negative) positions on a topic. The team members take turns presenting and defending their arguments over the course of a 90 minutes competition, after which a judge assesses the research presented, evaluates student persuasiveness, and renders a decision. The majority of research on debate suggest that intensive participation, at least at the college level, is associated with improved critical thinking (Allen, 1999), although the study design and sample composition preclude generalizing these findings to other settings (Colbert, 2002). Participating in debate expands time spent with teachers and debate peers through after-school practices and weekend tournaments throughout the school year that each consist of three to six debate rounds.

In practical terms, the activity of policy debate is characterized by the training of six academic skills: (1) reading and interpreting complex non-fiction text, (2) developing and writing arguments based on these texts, (3) verbally expressing and defending evidence-based claims, (4) listening to and interpreting opponents' arguments, (5) collaborating with peers, and (6) time-management (Mitchell 1998). Policy debate involves the practice of "secondary literacy" skills including comprehension and interpretation of arguments from non-fiction (informational) texts. According to the US Department of Education, a student who possesses secondary literacy skills can extract and incorporate ideas from a variety text into a broader understanding of a subject (US DOE National Assessment Governing Board, 2008).

However, despite widespread testimonial and anecdotal evidence of the impact of academic debate on critical thinking skills, personal development, and scholastic success (Lee, 1998; Warner and Bruschke, 2001; Collier, 2004), few studies have systematically evaluated this relationship, let alone in urban settings. Initial investi-gations of the influence of urban debate indicate that Black male students who participated in debate were more likely to graduate and had significantly higher scores on the English and Reading, but not Mathematics or Science, sections of the ACT (Mezuk, 2009). However, the authors also found that students who elected to debate in high school differed systematically from their non-debater counterparts in important ways, including better performance on 8th grade standardized assessments of reading and mathematics (although still below state standards for adequate performance, on average). From this initial research it is unclear whether the apparent beneficial influence of debate would translate to other social groups. These initial results also demonstrated the need for statistical methods that account for the differential selection of participants in order to isolate the influence of debate participation from confounding academic (for example, honors curriculum, 8th grade achievement) and non-academic factors (that is, poverty, racial composition of school) that influence achievement.

Present study

Academic debate may provide a compelling practice to address the shortcomings of schools serving low income and minority students. Debate programs may provide an opportunity to address both school-based and family- or household-based resource inequalities that pertain to academic achievement: these programs aim to provide rigorous academic opportunities, increase the amount of time engaged with educational material, and provide scholastic mentorship from coaches and academically-oriented peers (Breger, 2000).

Diverse urban public school systems, such as Chicago Public Schools, provide a unique opportunity to evaluate whether an instructional practice such as policy debate can influence outcomes among high school students. The goal of this study is to examine the influence of participating in a policy debate program on high school achievement, indicated by cumulative grade point average (GPA), ACT standardized test scores, and high school graduation, among students in an urban public school setting. The study hypotheses are: (1) students who elect to participate in a debate program will have better pre-debate achievement as indicated by standardized test scores than those who do not participate; and (2) accounting for differential selection into the debate program, students who elect to participate will have better academic achievement than similar students who do not participate.

METHODS

Sample

Details of the sample have been described previously (Mezuk 2009). Briefly, data come from Chicago Public Schools (CPS), in part-nership with the Consortium on Chicago School Research (CCSR) at the University of Chicago. The CCSR houses CPS enrollment, demographic, attendance, and achievement data dating from 1991 (Consortium on Chicago School Research at the University of Chicago, 2008). The sample is derived from the academic records of CPS and is restricted to students who attended at least one year of high school in CPS from 1997/8 to 2006/7

school years (that is, it does not include students who attended private or charter schools or who left the CPS district prior to 9th grade). CPS currently consists of 116 high schools [39 of which participated in the Chicago Debate League (CDL)] and serves approximately 112,000 students in grades 9 – 12. The district is racially diverse, and is approximately 8% White, 47% Black, 39% Latino, and 3% Asian, with multi-racial students making up the majority of the remainder (Chicago Public Schools, 2009).

Students who participated in the Chicago-based Urban Debate League, the CDL, were identified through tournament registration records. Tournament records spanned from the 1997/8 to 2006/7 school years and indicate whether a particular student attended each tournament (out of five to seven held each year), the number of rounds completed, the level of competition (varsity or JV), win-loss record, and any other awards. These tournament registration records were linked with CPS enrollment data by the CCSR. In a small portion of cases (4.2%), students could not be accurately identified from the tournament registration records (for example, the student was identified by initials on the roster). These cases were removed from analysis.

Enrollment records were used to derive a random sample of students who attended the same CPS schools that participated in the CDL. For each student who participated in the CDL from 1997/8 to 2006/7, CCSR randomly sampled four comparison students who (a) attended the same school and (b) were in the same 9th grade cohort as the debate participant (actual sampling ratio: 1:3.978). This was done to account for institutional-level differences between schools that had debate programs and those that did not. A total of 12,179 CPS students, enrolled at some point during the 1997/8 to 2006/7 school years, were selected for the sample, of which 2,449 (20%) participated in at least one CDL tournament. The final selected sample was well-representative of the CPS student population in terms of race/ethnicity (15% White, 48% Black, and 32% Latino). This analysis is limited to the sample of students who, based on the academic year in which they entered 9th grade, were due to graduate by Spring 2007, and were not in a special education program in high school (N = 9,145).

Independent variable

Students who participated in the CDL were identified through tournament rosters. A student was considered a "debater" if she or he had competed in at least one tournament at some point (regardless of the number of rounds debated). The average number of rounds completed per debater was 21.3 (SD = 19.5), the equivalent of approximately four 5-round tournaments. Records were also used to determine the intensity of debate participation, including the win-loss record, and number of tournaments attended and rounds debated.

Dependent variables

Three indicators – graduation from high school, ACT standardized test scores, and GPA – were used to assess the relationship between debate participation and achievement. These data were derived from CPS administrative records.

The dichotomous graduation outcome indicated the mutually exclusive categories of graduate or drop-out of high school (1 = graduated). For this analysis, we interpreted the graduation outcome as "missing" for students who transferred out of CPS (N = 1,336) and then imputed an ultimate outcome (graduate or drop-out) for these cases, as further described.

The ACT is composed of four sections – Reading, English, Mathematics, and Science – each of which has a total possible score of 36. The total score on the ACT is indicated by the average of a student's scores on these four sections. The ACT has established benchmarks of "college-readiness" in particular subject areas. For example, students who meet the benchmark in Reading (score ≥ 21) have a 50% chance of earning a grade of B or better in relevant college courses, that is, history and other social sciences) (ACT, 2006). The benchmarks for the other three sections are 18 for English, 22 for Mathematics, and 24 for Science. The ACT outcome was assessed as both a continuous outcome (range: 1 – 36 for each of the four sections), indicating average expected difference in ACT performance comparing debaters and non-debaters; and as a dichotomous outcome indicating whether or not the ACT score met or exceeded the standardized benchmark for college-readiness (1 = met or exceeded benchmark).

Cumulative weighted GPA, measured on a continuous 5-point scale, was reported for the fall and spring semesters for each year of high school from CPS administrative data. Because any given cumulative GPA represents the average of all previous semesters, a filter was applied to flag GPA values that were deemed to be likely reporting errors (for example, cumulative GPA of 1.0 in Fall 2000 followed by a GPA of 4.0 in Spring 2001, or a cumulative GPA of 4.0 in Fall 2000 followed by a GPA of 0.00 the following semester). These logical inconsistencies (N = 196 (0.3%) out of 73,160 student-semesters) were removed and then new plausible values were imputed using the strategy described below.

Covariates

Several covariates were identified from the CCSR data, including age in 9th grade, gender, race (categorized as White, Black, Latino, or other), eligibility for the free lunch in the 9th grade (coded yes/no), 9th grade cohort year, average number of absences in 9th grade, and whether the student took any honors coursework in the 9th grade (coded yes/no).

The results of two standardized 8th grade mathematics and reading assessments were used to account for pre-high school achievement (and thus pre-debate participation, as only high school students can participate in the CDL): the Illinois Standards Achievement Test (ISAT, administered from 1998 to 2007) and the Iowa Test of Basic Skills (ITBS, administered until 2005). The state of Illinois uses these state-wide standardized 8th grade tests to establish a baseline of achievement information, facilitate monitoring student development, identify areas of relative strengths and weaknesses, and help determine whether students have the scholastic background and skills necessary to learn at their instructional level (Illinois State Board of Education, 2009). Because the score ranges on the ITBS and ISAT are not identical (ISAT range: 120 to 200 for both math and reading, ITBS range: 1 to 337 for math, 349 for reading), the scores were standardized according to their overall mean, and each student was assigned a z-score for each of the reading and math sections. While the goals and foci of the two tests differ somewhat (Easton et al., 2003), the reading sections of both tests cover several domains, including vocabulary meaning, analyzing idioms, analogies, figurative expressions, and etymologies. The mathematics sections cover problems using whole numbers, percents, proportions, and exponents. These standardized test scores (both reading and math) were highly correlated with 8th grade GPA (r^2 = 0.70), and thus these scores were used to indicate pre-debate achievement rather than 8th GPA because of the large amount of missing data for this variable. Missing data on these 8th grade achievement indicators (N = 4,686) was largely due to students not being part of CPS in that grade (that is, they transferred into the system in high school). These test scores were used to determine whether middle school achievement influenced selection into debate.

Several contextual indicators were also assessed, including the concentration of poverty for the census block group of the student residence (indicated by the percent of adult males employed and

the percent of families with incomes above the poverty line) and the social status for the census block group of the student residence (indicated by the mean level of education of adults and percentage of employed persons in white-collar occupations). Both measures were derived from Census data (1990 for dates prior to fall 1999 and 2000 for fall 1999 onward) and are standardized with a mean of zero. Because many students in Chicago do not attend their local or "neighborhood" school, these indicators were also used to assess the average poverty concentration and social status of the students attending each of the sample's 39 schools. Other indicators of the school environment included the school type (general, vocational, magnet/college preparatory, or alternative) and the school's racial composition (integrated (at least 30% White), mixed (between 15-30% White), predominantly Black (at least 85% Black), predominantly Latino (at least 85% Latino), or predominantly minority (at least 85% Black and Latino, but neither group represents 85% of the total enrollment)).

This study was approved by the Institutional Review Board of the University of Michigan and Chicago Public Schools (Project ID 262).

Imputation of missing data

Due to the administrative nature of the CPS data and the attrition of students out of the sample due to transferring or dropping out, many observations had at least some missing data. Approximately 15% of students were missing data on high school completion, 40% were missing on ACT scores, and 59% were missing at least one value for at least one covariate (that is, free lunch status, semester GPA, honors curriculum indicator, school characteristics, 8th grade standardized test scores) over the 10-year study period. Analyzing only the fully observed portion of the data would cause a decrease in statistical efficiency, and, because the data were not missing completely at random, analysis based on the complete cases might introduce a bias in the estimates of the effect of debate participation. To overcome these issues, we used the Sequential Multiple Impu-tation (SMI) framework to account for missing data (Raghunathan, 2004). Missing values were replaced by several plausible sets of values to yield several completed data sets. This approach generated five independently imputed completed data sets. Each completed data set was analyzed separately and the resulting point estimates and standard errors are combined to yield unbiased results while incorporating uncertainty due to missing values. Imputations were created as draws from the predictive distribution of the missing values conditional on the observed values, and these predictive distributions were based on a sequence of regression models (Raghunathan et al., 2001).

Propensity score balancing

Although the debate team at each school is open to all students (there are no "try-outs"), students were not randomized to participate or not in debate and earlier work suggests self-selection plays an important role into who chooses to participate. We predict the choice to participate may be significantly associated with a number of baseline covariates that are also predictive of high school achieve-ment (that is, selection into debate may correlate with other characteristics that have a causal influence on achievement). One technique for accounting such selection bias is through propensity score matching or balancing (Rosenbaum and Rubin, 1983; D'Agostino 1998). The propensity score (PS) is a scalar summary of several covariates which represents the conditional probability that a student participated in debate, given these covariates (Rosenbaum and Rubin, 1983).

The PS for each student was estimated using multiple logistic regression model with a binary variable indicating whether or not a student participated in debate (yes = 1) as the dependent variable, predicted by a set of independent variables. Independent variables used to estimate the PS to debate were: (1) age at 9th grade, (2) gender, (3) race, (4) 8th grade standardized math test score (ITBS or ISAT), (5) 8th grade standardized reading test score (ITBS or ISAT), (6) eligibility for free lunch in 9th grade, (7) 9th grade cohort, (8) poverty concentration of residence, (9) social status of residence, (10) type of school, (11) racial composition of school, (12) average social status of students attending the school, and (13) average poverty level of students attending the school. We also evaluated and included seven two-way interactions that involved age, race, gender, reading test scores, and social status. The model was selected on the basis of the overall fit and balancing properties.

To balance students on their likelihood of being a debater, we grouped our sample into five equal size strata (quintiles) with respect to the students' PS to debate, a strategy that effectively matches students on the probability of having participated in debate and removes 90% of selection bias attributable to the observed variables (Cochran, 1968). To check the balance we examined differences between debaters and non-debaters on each of the thirteen cova-riates within each quintile for comparability. For each comparison there were no statistically significant differences between debaters and non-debaters. Additionally, we assessed the balancing proper-ties of the PS strata using a two-way analysis of variance model which included main effect for PS quintile and participation in debate (Brookhart et al., 2006). Across the five imputed datasets and prior to adjusting for PS, F-statistics ranged from 0.1 to 351 with mean value at 59. After adjusting for PS these values did not exceed 4.2 with mean values 0.7 for participation and 1.2 for interaction with PS quintile, indicating good fit of the propensity score.

Analysis

The influence of participation on achievement – indicated by ACT score, high school graduation, and cumulative GPA – was assessed using propensity score techniques to account for non-random selection into debate. The relationship between debate participation and the two dichotomous outcomes (meeting the ACT benchmark and high school graduation) was assessed using logistic regression stratified by quintiles of propensity to debate (1st quintile = highest propensity) across each of the five imputed datasets (described above). The average effect estimate was generated by averaging across the quintiles of propensity to debate, across the five datasets. The variance of these estimates was based on the Mantel-Haenzel test.

To assess the effect of debate participation on the continuous ACT score we estimated weighted mean difference between debater and non-debater groups within each PS quintile using linear regression. These estimates were then combined into a single weighted estimate (see Appendix Formula 1).

Cumulative GPA was assessed at each semester over the course of high school, unlike the ACT and graduation outcomes which were assessed at only one time for each student. To determine the longitudinal effect of debate on change in cumulative GPA per semester, over the course of high school, we fit mixed effects linear regression models with both random intercept and slope (see Appendix Formula 2). The random intercept accounts for both the repeated GPA observations (and subsequently correlated errors) within students over time from grade 9 to 12 and inherent variability between students in 9th grade GPA (Hardin and Hilbe, 2007). The random slope accounts for the variation in change in individual student GPA due to participating in debate over the course of high school. We chose mixed effects modeling over other longitudinal modeling approaches [that is, generalized estimating equations (GEE)] in order to account for heterogeneity in change in student achievement (Rabe-Hesketh et al., 2004). As with the ACT

Table 1. Baseline characteristics of debate participants and comparison students in the Chicago Public School district: 1997 – 2006 (N = 9,145).

Baseline characteristics	Non-participants		Debate participants		Chi-square	P-value
	N	%	N	%		
Sex						
Male	3367	46.5	738	38.8	35.4	<0.001
Female	3878	53.5	1162	61.2		
Race						
White	1016	14.0	277	14.6		
Black	3303	45.6	937	49.3	16.2	0.001
Hispanic	2429	33.5	554	29.2		
Other	497	6.9	132	6.9		
Age in 9th grade *(years)*						
<14	459	6.3	165	8.7		
14	5715	78.9	1608	84.6		
15	1002	13.8	120	6.3	94.3	<0.001
16	59	0.8	5	0.3		
>16	10	0.1	2	0.1		
Qualify for free lunch						
No	4245	59.5	1055	56.5	5.5	0.019
Yes	2888	40.5	812	43.5		
Honors courses in 9th grade						
No	4974	70.1	942	50.8	243.2	<0.001
Yes	2122	29.9	911	49.2		
8th grade GPA						
<1.5	30	8.2	6	3.8		
1.5 - 2.49	82	22.5	27	17.2	7.6	0.056
2.5 - 3.49	114	31.3	48	30.6		
>3.5	138	37.9	76	48.4		
Average days absent						
1^{st} to < 2^{nd} quartile (Low absenteeism)	2501	35.2	851	45.9		
2^{nd} to < 4^{th} quartile	1489	21.0	436	23.5	137.8	<0.001
4^{th} to < 7^{th} quartile	1277	18.0	306	16.5		
7^{th} to 8^{th} quartile (High absenteeism)	1829	25.8	260	14.0		
Mathematics standardized test scores (ISAT and ITBS)						
<2 SD below the mean (Low math ability)	11	0.3	0	0		
>1 and ≤ 2 SD below	620	17.8	90	9.2		
Within 1 SD of the mean	2305	66.2	680	69.4	60.1	<0.001
>1 and ≤2 SD above	411	11.8	141	14.4		
>2 SD above the mean (High math ability)	137	3.9	69	7.0		
Reading standardized test scores (ISAT and ITBS)						
<2 SD below the mean (Low reading ability)	60	1.7	3	0.3		
>1 and ≤ 2 SD below	603	17.3	61	6.2		
Within 1 SD of the mean	2310	66.4	680	69.5	116.2	<0.001
>1 and ≤2 SD above	425	12.2	194	19.8		
>2 SD above the mean (High reading ability)	82	2.4	41	4.2		

Table 1. Contd.

Residential concentration of poverty						
<1SD below mean (Low poverty)	1018	14.1	271	14.7		
Within 1 SD of the mean	5087	70.6	1283	69.4	0.9	0.638
>1SD above mean (High poverty)	1105	15.3	294	15.9		
Residential social status						
<1SD below mean (Low social status)	1143	15.9	278	15.0		
Within 1 SD of the mean	5097	70.7	1280	69.3	6.3	0.042
>1SD above mean (High social status)	970	13.5	290	15.7		
School concentration of poverty						
<1SD below mean (Low poverty)	552	8.5	136	8.0		
Within 1 SD of the mean	5026	77.0	1299	76.2	2.1	0.343
>1SD above mean (High poverty)	946	14.5	270	15.8		
School social status						
<1SD below mean (Low poverty)	889	13.6	247	14.5		
Within 1 SD of the mean	4817	73.8	1261	74.0	1.8	0.409
>1SD above mean (High poverty)	818	12.5	197	11.6		
Type of school						
General education	4504	62.2	1199	63.1		
Vocational education	109	1.5	27	1.4		
Magnet/College preparatory school	2521	34.8	646	34.0	0.6	0.962
Charter school/Alternative	65	0.9	16	0.8		
Alternative school	46	0.6	12	0.6		
School enrollment racial composition						
Integrated	1766	25.6	435	24.1		
Mixed	1448	21.0	365	20.2		
Predominantly black	1871	27.1	523	29.0	3.6	0.464
Predominantly latino	781	11.3	208	11.5		
Predominantly minority	1031	14.9	275	15.2		

ITBS: Iowa Test of Basic Skills, ISAT: Illinois Standards Achievement Test; GPA: Grade point average. SD: Standard deviation. Number of observations missing data for each variable prior to imputation: Free lunch (n = 145); Honors courses (n = 196), 8th grade GPA (n = 8624), days absent (n = 196), ITBS or ISAT standardized math test scores (4681), ITBS or ISAT standardized reading test scores (n = 4686), poverty of residence (n = 52), social status of residence (n = 52), poverty of school (n = 916), social status of school (n = 916), and racial composition of school (n = 442).

and graduation outcomes, each linear regression was fit within quintiles of the PS in order to account for selection into debate, and the results from these five models were then merged to generate an overall estimate of the effect of debate participation on GPA change per semester using the combining rules suggested by Rubin and colleagues (Rubin, 1987; Rubin and Schenker, 1986). All analysis was performed using SAS (v9.2) and all p-values refer to two-tailed tests.

RESULTS

As indicated by Table 1, CPS students who participated in debate in high school differed from their peers in many ways prior to applying the propensity score balancing: debaters were younger in 9th grade, more likely to be female (61.2% vs. 53.5%), more likely to qualify for free lunch (43.5% vs. 40.5%), more likely to take honors courses in 9th grade (49.2% vs. 29.2%), and had lower absenteeism in 9th grade and higher 8th grade test scores in mathematics and reading. Debaters and non-debaters did not differ significantly in terms of concentration of poverty of residence, but did differ in terms of social status of residence. As expected given the sampling strategy, debaters and non-debaters were similar on all school-level characteristics. Debaters and non-debaters were similar on these characteristics after propensity score adjustment.

Table 2 illustrates the relationship between participating in debate and graduating high school, adjusted for the propensity to debate. Students who participated in debate were 19% more likely to graduate from high school relative to comparable students who did not debate (95% confidence interval (CI): 1.16 – 1.23). Similarly, students who participated in debate were significantly less likely

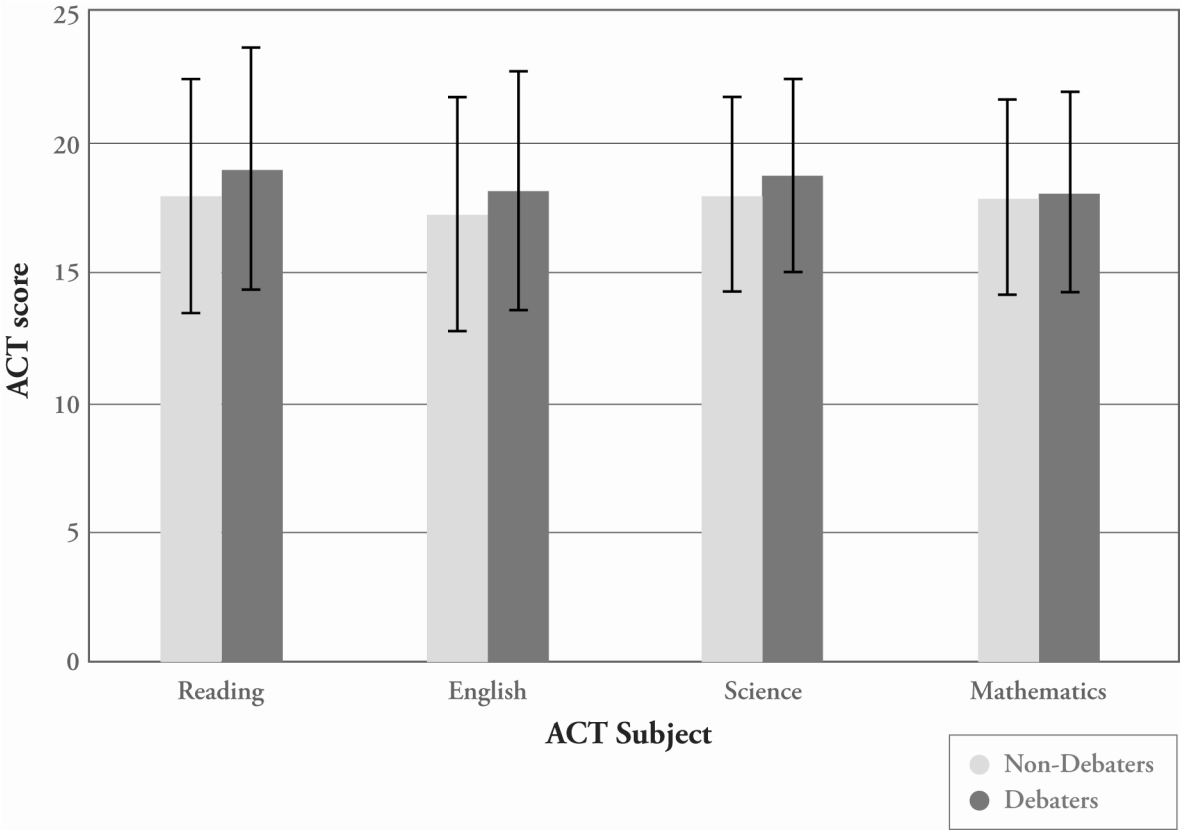

Figure 1. Average adjusted ACT scores for debaters and non-debaters, Chicago Public Schools: 1996 – 2006. Propensity score weighted mean ACT scores and 95% confidence intervals by debater status. Chicago Public Schools 1997 – 2006. N = 9145.

to drop out of high school (RR: 2.5, 95% CI: 2.1 – 2.9) than comparable students who did not debate. This effect was marginally stronger for females than males (RR: 2.7 vs. 2.3) and for Blacks and Latino students relative to Whites (RR: 2.7 and 2.3 vs. 1.9, respectively). Overall, we estimated that 90% of students who debated graduated high school over the study period, as compared to 75% of comparable students who did not debate ($p < 0.01$).

Table 3 and Figure 1 describe the relationship between participating in debate and ACT college-readiness, adjusted for the propensity to debate. The difference between the crude (unadjusted) and PS adjusted estimates indicates that there is substantial confounding by self-selection in the crude estimates. Comparing the unadjusted and propensity-score adjusted β coefficients reveals self-selection's influence on ACT performance relative to the causal effect of participating in debate. For example, based on a crude analysis, 46% of debaters scored at or above the college-ready benchmark in Reading as compared to 30% of comparable non-debaters, whereas corresponding percentages using the PS analysis are equal to 37% and 32%. However, even after accounting for this selection bias, all of the estimates on ACT performance remain statistically significant. As shown by the top portion of the table, participating in debate was associated with approximately a one point increase, on average, on both the Reading and English portions of the ACT relative to non-debaters. Participating in debate was also associated with better performance on the Science ($\beta = 0.88$ points higher) and Mathematics ($\beta = 0.43$ points higher) ACT sections relative to comparable students who did not debate. Overall, accounting for propensity to debate, participation was associated with significantly greater likelihood of scoring at or above the college-readiness benchmarks in Reading (15% more likely), English (15% more likely), Science (27% more likely) and Mathematics (10% more likely) relative to comparable students who did not debate.

In order to better isolate the influence of debate on ACT performance from unobserved factors related to student orientation or proclivity toward achievement, we examined the relationship between debate participation and ACT scores, restricting the sample to those students who graduated from high school. Since the ACT is a college-entrance test, restricting the sample to students who are on-track to graduate provides a more homogenous comparison. The relationship between debate participation and ACT performance was evident even among this

Table 2. Association between debate participation and graduating high school in the Chicago Public School district, 1997 – 2006.

Propensity score adjusted logistic regression model	N	Percent of debate participants	Relative risk	95% confidence interval
Entire sample (unadjusted)	9145		1.25	(1.23, 1.28)
1st PS quintile	1827	37	1.15	(1.10, 1.19)
2nd PS quintile	1830	26	1.14	(1.08, 1.20)
3rd PS quintile	1830	19	1.20	(1.14, 1.27)
4th PS quintile	1830	14	1.23	(1.14, 1.34)
5th PS quintile	1828	8	1.39	(1.24, 1.55)
Average PS weighted effect	9145		1.19	(1.16, 1.21)

Propensity score (PS) weighted effect generated using age at 9th grade, gender, race, 8th grade standardized math test score (ITBS or ISAT), 8th grade standardized reading test score (ITBS or ISAT), eligibility for free lunch in 9th grade, 9th grade cohort, poverty concentration of residence, social status of residence, type of school, racial composition of school, average social status of students attending the school, and average poverty level of students attending the school. The 1st PS quintile describes those students who had the highest propensity to participate in the debate program.

group that successfully graduated: participating in debate was associated with significantly greater scores on all four ACT sections (Reading $\beta = 0.94$, English $\beta = 0.90$, Science $\beta = 0.76$ and Mathematics $\beta = 0.44$) relative to comparable students who did not debate. Notably, the association between debate and ACT scores was strongest on the two sections – Reading and English – most closely associated with the skills practiced in the activity.

Turning to cumulative GPA, students who participated in debate both began high school with greater GPA (expected given the 8th grade test results presented in Table 1), but also gained more in GPA over the course of high school relative to non-debaters (Table 4 and Figure 2). After accounting for both initial GPA and propensity to debate, participating in debate was associated with a gain of 0.02 (95% CI: 0.01 – 0.02) units in cumulative GPA each semester relative to not participating. As shown by Figure 2, participating in debate was associated with improvements in GPA over the course of high school (8 semesters), whereas GPA for non-debaters was substantially unchanged over this period. Among students who successfully graduated from high school, the participating in debate remained associated with significant gains in GPA. High school graduates who participated in debate gained an average of 0.01 points (95% CI: 0.005 – 0.014) in cumulative GPA each semester over equivalent non-debaters. Overall, accounting for propensity to debate, average cumulative GPA at high school graduation students who participated in debate was 3.23 (95% CI: 3.21 – 3.28) for as compared to 2.83 (95% CI: 2.82 – 2.84) for students who did not participate.

DISCUSSION

The primary finding from this study is that even after accounting for the influence of self-selection, students who participated in the CDL were more likely to graduate from high school, performed better on the ACT, and showed greater gains in cumulative GPA relative to similar comparison students. Debate participation was associated with significantly better scores on all four components of the ACT, particularly the Reading and English sections. On average, participation was associated with an additional 1.02 additional points on the Reading and 1.04 additional points on the English sections of the ACT relative to not participation. The ACT suggests that an improvement of 0.50 points is considered “practically important” (ACT, 2006). Students who participated in debate were also more likely to reach the college-readiness benchmarks on all four sections of the test relative to similar students. These results are consistent with the interpretation that participating in debate is associated with statistically significant and substantially meaningful academic performance on the ACT. As expected, there was differential self-selection into debate by better-performing students, demonstrating the need for statistical methods that account for this patterning. This study represents the most comprehensive evaluation of debate in an urban setting to date, utilizing a representative sample of urban public school students, and applying appropriate statistical techniques for accounting for missing data, sample attrition, and identifying predictive relationships.

The relationship between debate participation and ACT performance is particularly relevant in light of the new Common Core Standards, national guidelines that are intended to produce consistent, college-ready high school standards across the fifty states (Council of Chief State School Officials and National Governors’ Association. 2009). The Common Core Standards for English and Language Arts focus on evidence-based argument and informational text mastery as critical language arts skills. As an example, the first Writing standard for grades 9-10 states “Write arguments to support claims in an analysis of substantive topics or texts, using valid reasoning and relevant and sufficient evidence” (National Governors

Table 3. Association between debate participation and ACT score in the Chicago Public School district, 1997 – 2006.

Panel A: Expected difference in ACT score associated with debate participation									
Propensity analysis	**% debaters**	**ACT sections**							
		Reading		**English**		**Science**		**Mathematics**	
		β	**95% CI**	**β**	**95% CI**	**β**	**95% CI**	**β**	**95% CI**
Entire sample (unadjusted)		2.73	(2.42, 3.04)	2.76	(2.44, 3.08)	1.96	(1.67, 2.25)	1.34	(1.02, 1.66)
Adjusted estimates									
1st PS quintile	37	1.21	(0.52, 1.90)	1.28	(0.64, 1.93)	1.08	(0.47, 1.69)	0.69	(0.11, 1.28)
2nd PS quintile	26	0.99	(0.18, 1.80)	1.13	(0.37, 1.89)	0.83	(0.26, 1.40)	0.53	(-0.12, 1.18)
3rd PS quintile	19	1.01	(0.08, 1.94)	0.92	(-0.20, 2.03)	0.98	(0.12, 1.84)	0.37	(-0.69, 1.43)
4th PS quintile	14	0.82	(-0.39, 2.03)	0.83	(-0.22, 1.88)	0.60	(-0.04, 1.24)	0.32	(-0.39, 1.03)
5th PS quintile	8	0.98	(0.08, 1.89)	0.77	(-0.42, 1.97)	0.71	(-0.11, 1.53)	0.06	(-0.99, 1.12)
Average PS weighted effect		1.02	(0.66, 1.39)	1.04	(0.66, 1.41)	0.88	(0.55, 1.21)	0.43	(0.06, 0.80)
Panel B: Association between debate participation and meeting ACT threshold for college readiness									
Propensity analysis	**% debaters**	**ACT sections**							
		Reading		**English**		**Science**		**Mathematics**	
		RR	**95% CI**	**RR**	**95% CI**	**RR**	**95% CI**	**RR**	**95% CI**
Entire sample (unadjusted)		1.53	(1.44, 1.63)	1.45	(1.39, 1.51)	1.75	(1.55, 1.98)	1.40	(1.28, 1.54)
Adjusted estimates									
1st PS quintile	37	1.14	(1.03, 1.26)	1.12	(1.06, 1.19)	1.31	(1.07, 1.61)	1.16	(0.99, 1.35)
2nd PS quintile	26	1.16	(1.00, 1.34)	1.14	(1.04, 1.25)	1.23	(0.91, 1.66)	1.06	(0.85, 1.31)
3rd PS quintile	19	1.16	(0.93, 1.44)	1.16	(1.03, 1.32)	1.08	(0.60, 1.96)	1.09	(0.75, 1.57)
4th PS quintile	14	1.10	(0.79, 1.53)	1.21	(0.99, 1.49)	1.41	(0.81, 2.47)	1.02	(0.69, 1.51)
5th PS quintile	8	1.35	(0.69, 2.63)	1.17	(0.72, 1.91)	1.27	(0.31, 5.14)	0.91	(0.37, 2.19)
Average PS weighted effect		1.15	(1.07, 1.23)	1.15	(1.10, 1.20)	1.27	(1.11, 1.44)	1.10	(0.99, 1.21)

Propensity score (PS) weighted effect generated using age at 9th grade, gender, race, 8th grade standardized math test score (ITBS or ISAT), 8th grade standardized reading test score (ITBS or ISAT), eligibility for free lunch in 9th grade, 9th grade cohort, poverty concentration of residence, social status of residence, type of school, racial composition of school, average social status of students attending the school, and average poverty level of students attending the school. The 1st PS quintile describes those students who had the highest propensity to participate in the debate program. RR: Relative risk. 95% CI: 95% confidence interval.

Association Center for Best Practices, 2010). The results of this study indicate urban debate programs may be an effective means of improving secondary literacy skills among low-income and minority students in urban areas, thereby working toward closing the achievement gaps for these groups. Participating in debate was associated with greater achievement as indicated by GPA as well. This analysis indicates that although students who participate in debate begin high school with higher GPA than non-debaters, after accounting for this initial difference, participating in debate

Table 4. Initial difference and expected change in cumulative GPA over the course of high school by debater status.

Adjusted estimates	Parameter	B (95% CI)		
		Debater	Non-debater	Difference (debater vs. non-debater)
1st PS quintile	Intercept	3.26 (3.17, 3.34)	2.83 (2.76, 2.91)	0.42 (0.32, 0.53)
	Slope	0.02 (0.01, 0.03)	0.01 (-0.01,0.01)	0.02 (0.01, 0.03)
2nd PS quintile	Intercept	2.96 (2.84, 3.08)	2.61 (2.55, 2.67)	0.35 (0.21, 0.49)
	Slope	0.02 (0.01,0.03)	0.01 (0.01, 0.02)	0.01 (0.01. 0.02)
3rd PS quintile	Intercept	2.83 (2.65, 3.00)	2.40 (2.34,2.47)	0.42 (0.26, 0.59)
	Slope	0.02 (0.01, 0.03)	0.01 (-0.01, 0.01)	0.02 (0.01, 0.03)
4th PS quintile	Intercept	2.57 (2.40, 2.75)	2.16 (2.08, 2.23)	0.42 (0.26, 0.58)
	Slope	0.01 (0.01, 0.03)	0.01 (0.01, 0.01)	0.01 (-0.01, 0.02)
5th PS quintile	Intercept	2.13 (1.95, 2.31)	1.74 (1.68, 1.81)	0.39 (0.20, 0.58)
	Slope	0.02 (-0.01, 0.04)	-0.01 (-0.01, 0.00)	0.02 (0.00, 0.04)
Average PS weighted estimate	Intercept	2.92 (2.88, 2.97)	2.28 (2.26, 2.31)	0.40 (0.34, 0.46)
	Slope	0.02 (0.02, 0.02)	0.01 (0.01, 0.01)	0.02 (0.01, 0.02)

Propensity score (PS) weighted effect generated using age at 9th grade, gender, race, 8th grade standardized math test score (ITBS or ISAT), 8th grade standardized reading test score (ITBS or ISAT), eligibility for free lunch in 9th grade, 9th grade cohort, poverty concentration of residence, social status of residence, type of school, racial composition of school, average social status of students attending the school, and average poverty level of students attending the school. The 1st PS quintile describes those students who had the highest propensity to participate in the debate program.

was significantly associated with gains in GPA over the course of high school. Students who debated had an average spring 12th grade GPA of 3.06 as compared to 2.30 for non-debaters students, and among students who graduated debaters had an average GPA of 3.23 as compared to 2.83 for comparable students who did not debate. Previous research has determined that that a cumulative GPA of 3.0 or greater is a key indicator of college-readiness and success in college coursework (Roderick, Nagaoka, and Allensworth 2006). Thus, participating in debate was associated with increased likelihood of meeting college-readiness indicators by two metrics: ACT scores and cumulative GPA.

The precise mechanisms by which debate may influence achievement in this urban setting are unresolved. Debate may influence achievement through indirect pathways, particularly through the provision of mentorship opportunities via interactions with both teachers and fellow students. Previous research indicates that peer group attributes (that is, attitudes towards school) and teachers' expectations of achievement are related to performance on reading and mathematics tests among high school students (Roscigno, 1998). Ethnographic data suggest that peer groups that are competitive, yet supportive of each others' success, are related to academic achievement among Black urban school high school students (Horvat and Lewis, 2003). As described above, participation in policy debate involves connecting with a group of achievement-oriented peers and interacting with supportive teachers as coaches, suggesting additional indirect paths by which participating in debate may influence achievement. Previous work in CPS has found that school absences are a strong predictor of poor achievement, and that students who have positive relationships with teachers and see academic school work as important to their future are more likely to attend class (Allensworth and Easton 2007). Thus debate may improve achievement by fostering a connection to school life in general, including attendance. Further study is necessary to evaluate these hypotheses.

Compared to peers selected from the same 9th grade cohort, students who participated in debate were significantly more likely to graduate from high school. These estimates are not directly comparable with the graduation rates reported by Chicago Public Schools for numerous reasons (for example, the imputation strategy used here assigns students who transferred out of CPS a graduation outcome (complete or drop-out), which is not the approach used by CPS when they report graduation rates, our sample of students was only drawn from that portion of schools in the CPS district that had a debate program, and our graduation estimates are weighted to account for the probability for selecting into the debate program). Despite these qualitative differences in the manner of estimating the likelihood of graduating our results provide evidence that debate is associated with likelihood of completing high school. These results should be interpreted in light of study limitations. Foremost,

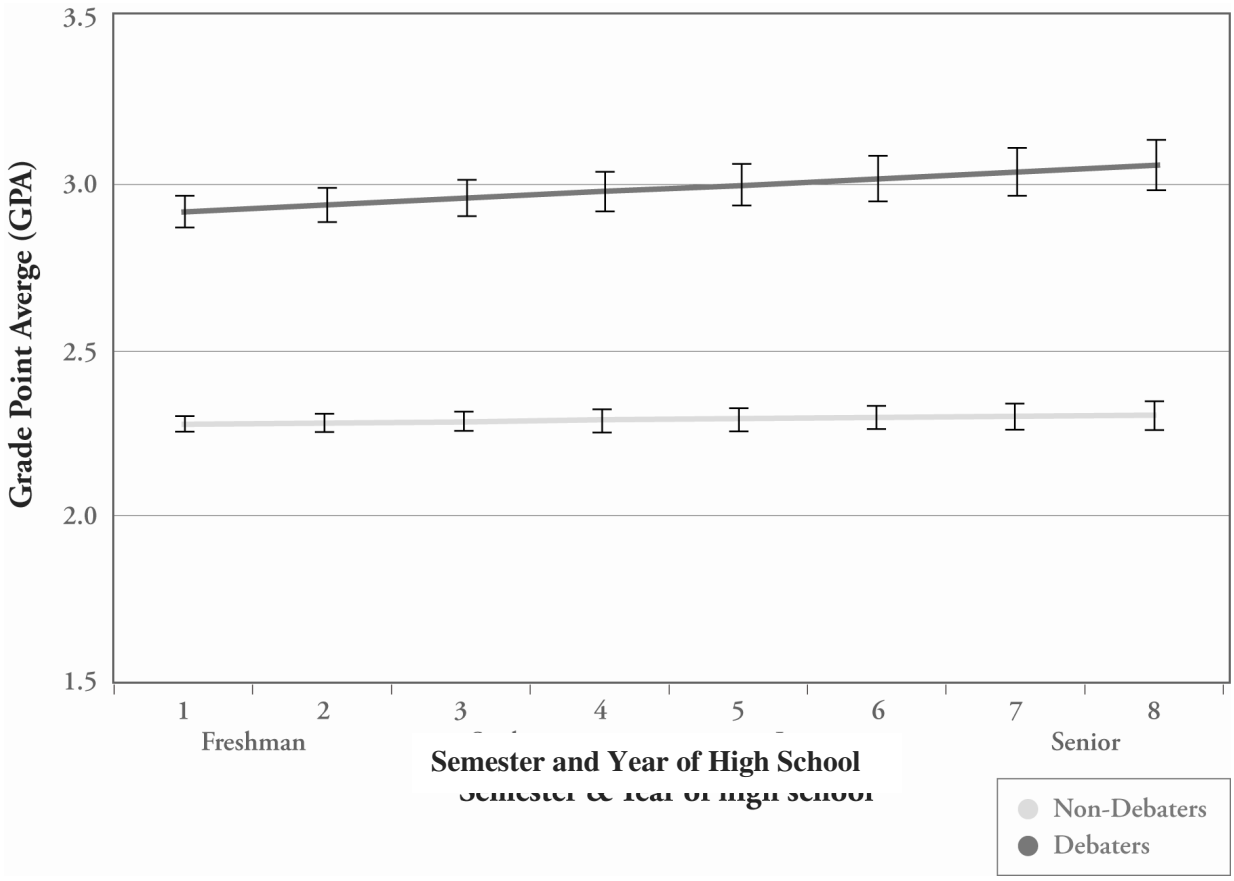

Figure 2. Expected cumulative GPA by debater status over the course of high school. Propensity score weighted mean cumulative GPA over the course of high school by debater status. Bars represent 95% confidence intervals. Chicago Public Schools 1997 – 2006. N = 9145.

although we used propensity score techniques to account for the differential selection into debate, there is still the potential that our estimates of the effect of debate on achievement are biased by other factors that were not taken into account (for example, parent education, attitude toward school) that may have influenced both selection into debate and achievement outcomes (Peikes et al., 2008). Without a randomized, controlled intervenetion for comparison, we cannot quantify the degree of this potential bias. However, provided that unobserved selection factors are correlated with those that we did account for, including 8th grade achievement, our analysis would at least partially account for their effects (Rosenbaum and Rubin, 1983). In order for our findings to be entirely attributable to unobserved selection factors these variables would have to be both very strongly associated with achievement but unrelated to the factors we included in our propensity score; data from other studies in Chicago Public Schools indicates that the factors that matter most for staying on-track and graduating from high school include number of semesters failing a course and school absences, characteristics that we accounted for in this analysis (Allensworth and Easton, 2007). These characteristics were also highly correlated with prior achievement (for example, 8th grade standardized test scores), which we also included. The CCSR also reports that school climate factors (for example, parental support of the school, teacher-parent interactions) did not significantly influence achievement, and that factors such as student-teacher trust only influence achievement insofar as they correlate with improved attendance (Allensworth and Easton, 2007). Therefore, we have believe we have captured the most salient factors that determine high school achievement in our propensity score, and having achieved very good balance on covariates using this approach, we have good confidence that our findings are robust to unobserved selection factors. Further, the finding that debate most strongly influenced performance on the Reading and English portions of the ACT (which most directly relate to the skills practiced as part of debate) compared to the Science and Mathematics sections suggests a degree of specificity consistent with a direct effect of debate on achievement.

A further caveat stems from the fact that these analyses only estimate the average influence of debate

participation, but there is likely heterogeneity in effect that these analyses do not capture (that is, variability in the amount of debate participation and competitive success, as well as variability according to individual and contextual factors), and future research should employ statistical approaches that account for this. Due to the administrative nature of the data, and the fact that the CPS district has a higher drop-out rate than the national average, there was substantial attrition in the sample; this is a common limitation of research on achievement outcomes in urban settings. Our approach to address this limitation was to apply sequential multiple imputation techniques to account for this missing data and selective attrition and the bias it may have introduced; simulation analysis have indicated that in cases such as this with substantial attrition, sequential multiple imputed estimates are less biased and more robust than analysis using only observed data or alternate imputation techniques (Raghunathan, 2004; Allison, 2000).

Conclusion

This study suggests that participating in debate in an urban, underserved setting such as CPS, is associated with practically meaningful achievement in terms of college-readiness and likelihood of graduating high school. Although experimental data are the gold standard for evaluating causal relationships, the quasi-experimental propensity score approach utilized here demonstrated that the relationship between debate and achievement persisted after accounting for numerous factors known to influence academic outcomes. This study also illustrates the value of large administrative datasets for investigating the relationship between out-of-school time activities and achievement over a long period of time. As states and school systems place increasing emphasis on empirical program evaluation and make progress toward developing longitudinal data systems that track students over their academic life-course, researchers must grapple with the methodological challenges of treating such data in order to address pressing education policy needs.

REFERENCES

ACT (2006). Reading between the lines: What the ACT reveals about college readinessin reading. Iowa City IA: ACT.

Adelman C (2004). Principal Indicators of Student Academic Histories in Postsecondary Education, 1972-2000. Washington DC: Department of Education, Institute of Educational Sciences.

Alexander KL, Entwisle DR, Olson LS (2007). Lasting Consequences of the Summer Learning Gap. Am. Sociol. Rev., 72: 167-180.

Allen M, Berkowitz S, Hunt S, Louden A (1999). A meta-analysis of the impact of forensics and communication education on critical thinking. Comm. Educ., 48: 18-30.

Allensworth E, Easton JQ (2007). What matters for staying on-track and graduating in Chicago Public Schools. Report by the Consortium on Chicago School Research at the University of Chicago.

Allison P D (2000). Multiple imputation for missing data: A cautionary tale. Sociol. Methods Res., 28: 301–309.

Balfanz R, Legters N (2004). Locating the Dropout Crisis: Which High Schools Produce the Nation's Dropouts? Where are they Located? Who Attends Them? CRESPAR Report, 70.

Blau P, Duncan OD (1967). The American Occupational Structure. New York: Wiley.

Breger B (2000). Overview of the Urban Debate Program. Rostrum, 75: 14.

Brookhart MA, Schneeweis S, Rothman KJ, Glynn RJ, Avorn J, Sturmer T (2006). Variable selection for propensity score models. Am. J. Epidemiol., 496: 1149-1156.

Chicago Public Schools (2009). Aggregate Citywide Data, 2001–2009. CPS Office of Research, Evaluation, and Accountability. Retrieved January 8, 2010. Available at: http://research.cps.k12.il.us/cps/accountweb/Reports/citywide.html.

Cochran WG (1968). The effectiveness of adjustment by subclassification in removing bias in observational studies. Biometrics, 24: 205-213.

Colbert K (2002). Enhancing critical thinking ability through academic debate. Perspectives in controversy: Selected essays from Contemporary Argumentation and Debate, K. Broda-Bahm, ed. New York, NY: International Debate Education Association, pp. 71-100.

College Board (2008). AP English Language and Composition and English Literature and Composition Course Description. Retrieved January 10, 2010. Available at: http://apcentral.collegeboard.com/apc/public/repository/ap08_englishcoursedesc.pdf.

College Board. 2008 (2010). College Board Standards for College Success: English Language Arts. Retrieved January 10, 2010. Available at: http://www.collegeboard.com/prod_downloads/about/association/academic/english-language-arts_cbscs.pdf.

Collier L (2004). Argument for success: A study of academic debate in urban high schools of Chicago, Kansas City, New York, St. Louis and Seattle. Chicago, IL: National Communication Association Convention Conference Proceedings.

Condron DJ (2009). Social class, school and non-school environments, and Black/White inequalities in children's learning. Am. Sociol. Rev., 74: 683-708.

Condron DJ (2007). Stratification and educational sorting: Explaining ascriptive inequalities in early childhood reading group placement. Soc. Problems, 54: 139-160.

Consortium on Chicago School Research (2008). Data and data requests. Available at: http://ccsr.uchicago.edu/content/page.php?cat=6. Retrieved June 1, 2009.

Council of Chief State School Officials and National Governors' Association (2009). Draft Guidelines: College and Career Readiness Standards for Reading, Writing, and Speaking and Listening. Available at: http://www.corestandards.org/Files/ELAStandardsSources.pdf.

Cullen JB, Jacob BA, Levitt S (2006). The effect of school choice on participants: Evidence from randomized lotteries. Econometrica, 74: 1191-1230.

Currie J, Thomas D (2000). School quality and the longer-term effects of Head Start. J. Hum. Resour., 35: 755-774.

D'Agostino RB (1998). Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Stat. Med., 17: 2265-2281.

Downey DB, Hippel PT, Broh B (2004). Are schools the great equalizer? Cognitive inequality during the summer months and school year. Am. Sociol. Rev., 69: 613-635.

Easton JQ, Correa M, Luppescu S, Park HS, Ponisciak S, Rosenkranz T, Sporte S (2003). How do they compare? ITBS and ISAT reading and mathematics in the Chicago Public Schools, 1999 to 2002. Research data brief: Consortium on Chicago School Research: Chicago IL.

Fryer RG, Levitt SD (2004). Understanding the Black-White test score gap in the first two years of school. Rev. Econ. Stat., 86: 447-464.

Green J, Forster G (2003). Public High School Graduation and College Readiness Rates in the United States. CCI Education Working Paper, No. 3.

Green J, Winters M (2005). Public High School Graduation and College-Readiness Rates: 1991-2002. CCI Education Working Paper, No. 8.

Heckman J, Doyle O, Harmon C, Tremblay R (2009). Investing in early

human development: Timing and economic efficiency. Econ. Hum. Biol., 7: 1-6.
Horvat EM, Lewis KS (2003). Reassessing the "Burden of 'Acting White'": The importance of peer groups in managing academic success. Sociol. Educ., 76: 265-280.
Lauer PA, Akiba M, Wilkerson SB, Apthorp HS, Snow D, Martin-Glenn ML (2006). Out-of-school-time programs: A meta-analysis of effects for at-risk students. Rev. Educ. Res., 76: 275-313.
LaVeist TA (2005). Disentangling race and socioeconomic status: a key to understanding health inequalities. J. Urban Health, 82: 26-34.
Lareau A (2003). Unequal Childhoods. Berkeley, CA: University of California Press.
Lee E (1998). Memoir of a former urban debate league participant. Contemp. Argumentation Debate, 19: 93-96.
Little RJ, Raghunathan T (1999). On summary measures analysis of the linear mixed effects models for repeated measures when data are not missing completely at random. Stat. Med., 18L: 245-2478.
Mezuk B (2009). Urban debate and high school education outcomes for African American males: The case of the Chicago Debate League. J. Negro Educ., 78: 290-304.
Miller M (2009). Seize the Moment: The Need for a Comprehensive Federal Investment in Adolescent Literacy. Washington, DC: Alliance for Excellent Education. (http://www.all4ed.org/files/PolicyBriefSeizeTheMoment.pdf).
Mitchell G (1998). Pedagogical possibilities for argumentative agency in academic debate. Argumentation Advocacy, 35: 41-56.
National Governors Association Center for Best Practices (2010). Common Core State Standards Initiative. Available at: http://www.corestandards.org.
Neild RC, Balfanz R (2006). Unfulfilled promise: The dimensions and characteristic's of Philadelphia's dropout crisis, 2000-2005. Baltimore: Center for Social Organization of Schools, Johns Hopkins University.
Peikes DN, Moreno L, Orzol SM (2008). Propensity score matching: A note of caution for evaluators of social programs. Am. Stat., 62: 222-2311.
Raghunathan TE (2004). What do we do with missing data? Some options for analysis of incomplete data. Ann. Rev. Public Health, 25: 99-117.
Raghunathan TE, Lepkowski JM, VanHoewyk J, Solenberger P (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. Surv. Methodol., 27: 85-95.
Roderick M, Nagaoka J, Allensworth E (2006). From high school to the future: A first look at Chicago Public School graduates' college enrollment, college preparation, and graduation from four-year colleges. Report by the Consortium on Chicago School Research.
Roscigno VJ (1998). Race and the reproduction of educational disadvantage. Soc. Forces, 76: 1033-1061.
Roscigno VJ, Tomaskovic-Devey D, Crowley M (2006). Education and the inequalities of place. Soc. Forces, 84: 2121-2145.
Rosenbaum PR, Rubin DB (1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70: 41-55.
Rothstein R (2004). Class and Schools: Using social, economic, and educational reform to close the Black-White achievement gap. Washington, DC: Economic Policy Institute.
Schafer J (1999). Multiple imputation: A primer. Statistical Methods for Medical Research, 8: 3-15.
Schweinhart LJ, Weikart DP (1997). The High/Scope preschool curriculum comparison study through age 23. Early Childhood Res. Q., 12: 117-143.
Snipes J, Horowitz A (2008). Advancing Adolescent Literacy in Urban Schools. Washington, DC: Council of the Great City Schools. Available at: http://www.cgcs.org/Pubs/ResearchBrief_08.pdf.
Stiefel L, Schwartz AE, Ellen IG (2007). Disentangling the racial test score gap: Probing the evidence in a large urban school district. J. Pol. Anal. Manage., 26: 7-30.
Swanson CB (2009). Cities in Crisis 2009. Bethesda, MD: Editorial Projects in Education, Education Research Center.
U.S. Department of Education (2006). NAEP Trends in Academic Progress. Washington, DC: National Assessment of Educational Progress (NAEP), National Center for Education Statistics; Unpublished tabulations. NAEP Data Explorer. Available at: http://nces.ed.gov/nationsreportcard/nde/.
US Department of Education National Assessment Governing Board (2008). Reading Framework for the National Assessment of Academic Progress. Washington DC: National Assessment Governing Board.
Venezia A, Kirst M, Antonio A (2004). Betraying the College Dream: How Disconnected K-12 and Postsecondary Systems Undermine Student Aspirations. Stanford University Bridge Project Final Report, Available at: http://www.stanford.edu/group/bridgeproject/betrayingthecollegedream.pdf.
Warner E, Bruschke J (2001). 'Gone on Debating': Competitive Academic Debate as a Tool of Empowerment. Contemp. Argumentation Debate, 22: 1-21.
Wyner JS, Bridgeland JM, Diiulio JJ (2009). Achievement Trap: How America is failing millions of high-achieving students from lower-income families. Washington, DC: John Kent Cooke Foundation & Civic Enterprises.