The BUDL Effect: Examining Academic Achievement and Engagement Outcomes of Preadolescent Baltimore Urban Debate League Participants

Daniel Shackelford[1]

This study adds to the limited literature base on extracurricular debate by using doubly robust inverse probability treatment weighting to estimate the average treatment effect for the treated of preadolescent debate participation on a variety of academic and engagement outcomes among a 10-year longitudinal sample of Baltimore City Public School System students. The effect of preadolescent Baltimore Urban Debate League participation for debaters was associated with increases in standardized test scores, a decreased likelihood of chronic absenteeism, and an increased likelihood of attending a selective entrance criteria high school. Although there is a mounting body of research that suggests participation in debate is associated with increases in positive outcomes for high school students, this research constitutes the first quantitative study to examine these relationships among elementary and middle school students. Policy implications for educational interventions that seek to attract low-income students of color in urban areas and influence their trajectories at earlier stages of student development are discussed.

Keywords: achievement; adolescence; policy analysis; program evaluation; quasi-experimental analysis; regression analyses; urban education

Student participation in extracurricular activities has been linked to many positive outcomes (Denault & Poulin, 2009; Eccles & Barber, 1999; Feldman & Matjasko, 2005), including school engagement (Mahoney, Cairns, & Farmer, 2003), academic achievement (Broh, 2002), and overall educational attainment (Gibbs, Erickson, Dufur, & Miles, 2015; McNeal, 1995). Today, extracurricular activities are an important component of students' school lives, and many schools invest substantial resources in support for extracurricular activities (Shulruf, 2010). In fact, more than half of American children between the ages of 6 and 17 participate in an extracurricular activity (U.S. Census Bureau, 2014).

Unfortunately, support for extracurricular activities has not translated into opportunities for participation among all students. Due in large part to resource limitations, youth living in socioeconomically disadvantaged communities are less likely to participate in extracurricular activities than those living in more affluent communities (Pedersen, 2005; Quinn, 1999). Barriers that include transportation, safety conditions, and fees for participation all result in urban youth spending less time engaged in organized activities outside of school compared to wealthier suburban youth (Fredricks & Simpkins, 2012). These statistics on the inequity of extracurricular activity participation are amid the backdrop of education gaps in the United States where pronounced disparities across urbanicity, income, and race remain in achievement outcomes. For example, students in urban schools, particularly Black and Hispanic students, have low literacy rates relative to White suburban students (Snipes & Horowitz, 2008), and only 53% of students graduate high school in urban schools compared to 71% in suburban schools (Kena et al., 2016). This gap is even larger in Baltimore, the site of this study, where only 41% of students graduate from city schools, compared to 81% in the suburbs (Swanson, 2009). Thus, youth who face the greatest difficulties in accessing extracurricular activities are also those who may have the most to gain from participation.

[1]Johns Hopkins University, Baltimore, MD

Enter the National Association of Urban Debate Leagues: an organization whose mission is to extend access to a particular extracurricular activity, competitive debate, to low-income urban school districts. The program currently serves more than 10,000 students from over 600 schools in 23 cities and estimates that nearly 90% of its participants are students of color and 75% are from low-income families (National Association of Urban Debate Leagues, 2016). Research evaluating student outcomes among the organization's participants shows promising results. A 10-year longitudinal study of participants in Chicago shows that high school students who debate have higher 12th-grade grade point averages (GPAs), are more likely to graduate high school, and are more likely to be college ready in reading and English than those who do not participate in debate after adjusting for self-selection into the activity (Mezuk, 2009; Mezuk, Bondarenko, Smith, & Tucker, 2011). Follow-up analysis found that high school debaters have higher social, civic, and school engagement (Anderson & Mezuk, 2015), and are more likely to matriculate to college (Shackelford, Ratliff, & Mezuk, 2018) than non-debaters.

But despite major advances in the research justification for Urban Debate Leagues (UDLs), major gaps remain in the literature. A population largely absent from analysis thus far is elementary and middle school students. This is troublesome because behavioral indicators for dropping out of school become apparent early in a student's educational trajectory. Research indicates that the middle grades are central to students' later academic attainment (Balfanz, Herzog, & Mac Iver, 2007; Kieffer & Marinell, 2012) and that a low commitment to schooling in the late elementary grades are predictors of low academic performance, behavior problems, and poor health in children (Abbot et al., 1999). These findings suggest that the time spent in elementary and middle school are particularly salient periods for altering student trajectories. Because entrenched patterns for students entering high school are extremely difficult to change, the research community has called for significant interventions during the early middle grades in order to prevent most dropout outcomes (Mac Iver, 2010). UDL participation may serve as a constructive intervention during this preadolescent period that improves students' educational attainment.

In order to truly assess the impact participating in extracurricular debate has on student outcomes, UDLs should be studied throughout the stages of student development, which includes the crucial period of the elementary and middle school years. The present study consequently adds to the limited literature base by using doubly robust inverse probability treatment weighting to estimate the average treatment effect for the treated of preadolescent debate participation on a variety of academic and engagement outcomes that include eighth-grade reading and math standardized test scores, attendance rates, and ninth-grade high school destination.

Background

Debate is a competitive extracurricular activity in which teams of students engage in structured argumentation about social policies (Breger, 2000). Students work in two-person teams to craft and defend arguments about a particular topic (called a resolution), which changes annually. Throughout the academic year, debate leagues host tournaments where students participate in switch-side debating (i.e., alternatively debating to affirm or negate a resolution) (Winkler, 2011). As a result, students must become adept at arguing both sides of an issue persuasively. Debates are judged by other coaches, debate alumni, or community volunteers, and students receive individual and team awards at each competitions' conclusion based on their performance. In practical terms, debate is characterized by the training of academic skills such as reading and interpreting complex non-fiction text, developing and writing arguments based on these texts, verbally expressing and defending evidence-based claims, and listening to and interpreting opponents' arguments (Mitchell, 1998). In the mid-1980s, the first UDL began as a partnership between the Atlanta Public School System and Emory University to expand the benefits of debate to underserved populations of impoverished minorities (Winkler, 2011).

The following study on debate participation builds on recent causal evidence found for extracurricular activities as a whole. This evidence is based on models that utilize fixed effect approaches to isolate important self-selection factors (Lipscomb, 2007) as well as exogenous variation from laws and policies that determine participation (Crispin, 2017; Stevenson, 2010). Results from this research show that skills developed through both athletic and club participation are productive in the academic classroom. However, debate is distinct from most extracurricular activities insofar as its content aligns well with many scholastic goals. For example, the first writing standard for Grades 9 and 10 states students should be able to "write arguments to support claims in an analysis of substantive topics or texts using valid reasoning and relevant and sufficient evidence" (National Governors Association, 2010). Furthermore, the English language arts and reading objectives outlined in the Common Core explicitly focus literary education on the analysis of non-fiction texts and oral communication (Porter, McMaken, Hwang, & Yang, 2011). Thus, unlike mentoring programs, sports team, or other extracurricular activities, debate may potentially reinforce the same academic writing and language skills that are the focus of school curricula. Consequently, it is plausible that debate is an extracurricular activity that may influence students' academic achievement more so than extracurricular activities in general.

Debate's competitive nature between groups of students may also influence student learning. Coleman (1961) was one of the first to point to the difference in outcomes if student competition is organized between schools rather than between students. He documents an "adolescent society" in which "interpersonal competition in scholastic matters" between students generates social pressure not to excel, while "interscholastic competition" between schools has the opposite effect. Coleman believed that shifts in the competitive structure of learning environments can change the norms and values of students for the better to encourage academics, and he even cites participation in debate teams as one possible solution to bolster academic competition (Coleman, 1959).

Since Coleman suggested that schools mobilize peer support for effective academic performance through the use of team
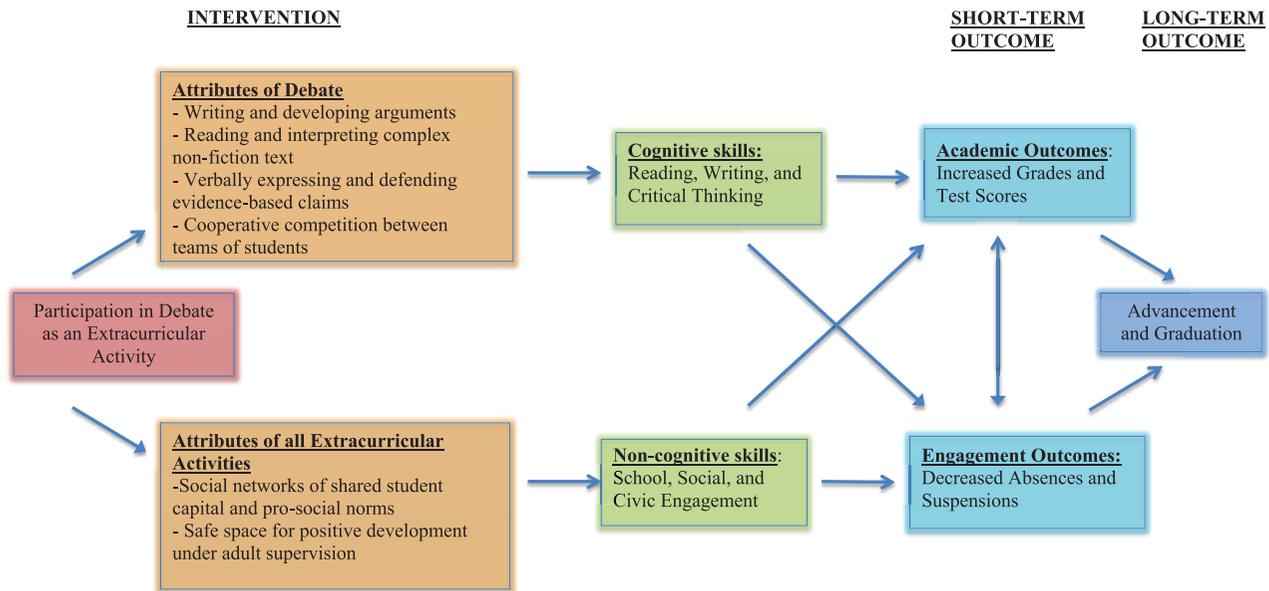
FIGURE 1. *Conceptual model for the theory of change behind participation in debate as an extracurricular activity.*
Note. Figure 1 depicts debate as an extracurricular activity that potentially influences student outcomes from both the developmental structure inherent to all extracurricular activities as well as its unique emphasis on cognitive skills with a cooperative and competitive incentive structure. These attributes are theorized to improve academic achievement and engagement outcomes across a range of indicators.

competition around academic tasks, countless studies have demonstrated efforts that incorporate cooperation with intergroup competition result in higher achievement than interpersonal competition and individualist efforts (Johnson, Johnson, & Stanne, 2000; Slavin, 1983). One meta-analysis of 122 research studies concludes that the overall effects "stand as strong evidence for the superiority of cooperation in promoting achievement and productivity" and that "educators may wish to considerably increase the use of cooperative learning procedures to promote higher student achievement" (Johnson, Maruyama, Johnson, Nelson, & Skon 1981, p. 58). Debate is one such activity that utilizes a cooperative incentive structure in which students are rewarded based on their performance as a team (Slavin, 1980).

Considering debate's unique attributes, the argument for its theorized influence on student outcomes is quite persuasive. The conceptual model for the theory of change behind participation in debate as an extracurricular activity is depicted in Figure 1. The bottom half illustrates the developmental benefits of participation in which researchers theorize that extracurricular activities contribute to academic achievement indirectly by enhancing students' noncognitive skills. The top half conversely shows a direct link between participation and academic outcomes via debate's focus on reading, writing, and verbal communication skills as well as its cooperative competition between teams of students that rewards those skills. This conceptual model, informed by the research literature, illustrates my primary hypothesis that participation in a UDL will be associated with positive academic achievement and engagement outcomes.

Furthermore, if participation in a UDL is protective against declines in school performance, one might expect the strongest benefits to be from students that participate during younger ages, when their trajectories begin to diverge into those on-track

to graduate high school and those who are not, as opposed to participation later in life. Entwisle and Alexander (1992) note that young children are "maximally sensitive to home and school influences" (p. 73), and other research shows that entrenched patterns of students entering the ninth grade are extremely difficult to change (Mac Iver, 2010). Finally, Heckman (2006) documents how early interventions that target disadvantaged children have higher returns than later interventions as an early mastery of a range of cognitive, social, and emotional competencies makes learning at later ages more efficient and therefore easier and more likely to continue.

While the previous research on debate shows positive results for high school participants, few studies investigate the academic benefits of participating in the activity during Grades 4 to 8 or preadolescence, a period largely overlooked both in the research specific to debate as well as in the extracurricular activity research as a whole which primarily focuses on the high school years (Schwartz, Cappella, & Seidman, 2015). Thus, questions remain regarding the direction and strength of the effect when students participate at younger ages. The present study contributes to the literature by providing an understanding of how elementary and middle school participation in a particular UDL from a diverse school district influences student outcomes.

Positive findings will be noteworthy as they may outline to policymakers appropriate and effective means to influence students' overall academic trajectories at earlier stages of development. Consequently, this research may also highlight the need to increase access to extracurricular activities like debate for students of younger ages. The National Institute on Out-of-School Time (2003) estimates that approximately 8 million children between the ages of 5 and 14 are unsupervised after school. As mentioned previously, significant interventions during preadolescence are

required to prevent negative educational trajectories. Participation in a UDL may improve academic achievement and engagement outcomes during this period, thereby preventing students from falling off track.

Data

The present study examines academic achievement and engagement outcomes among a 10-year longitudinal sample of 84,169 Baltimore City Public School System (BCPSS) students who attended a school that participated in the Elementary and Middle School Baltimore Urban Debate League (BUDL) from the 2004–2005 to 2013–2014 school years. Data come from de-identified yearly administrative student-level records from BCPSS in partnership with the Baltimore Education Research Consortium (BERC) that houses the school district's enrollment, demographic, attendance, and achievement data. Students who participated in BUDL were identified through a comprehensive list of tournament registration records. A binary variable was used to signify whether a student experienced "treatment" (i.e., participated in at least one BUDL tournament). A total of 2,263 students in the sample (or 2.69%) participated in the BUDL during preadolescence (Grades 4–8).

The outcome variables of interest in this study include standardized eighth-grade reading and math test scores from the Maryland School Assessment (MSA), average attendance rate in Grades 4–8, and ninth-grade high school destination. The MSA is a test of reading and math achievement given to students in Grades 3–8 that meets the requirements of the federal No Child Left Behind Act. The reading MSA tests a student's general reading processes, informational text comprehension, and literary text comprehension; while the math MSA tests algebra/patterns, geometry/measurement, statistics/probability, number concepts/computation, and processes of mathematics. Between the 2003–2004 to 2013–2014 school years, all students in Maryland (Grades 3–8) were required to take the MSA. For the purposes of this study, MSA reading and math scores from the third grade are used to account for a student's predebate achievement, while scores from the eighth grade are used to measure a student's academic achievement at the end of preadolescence.

BERC enrollment data keep track of days absent and days present for each year of school attended. Because schools vary in the total number of days in session per year, an attendance rate percentage equaling the days present over the sum of days present and days absent was created for each year of school a student attended in the BCPSS. Coding the outcome variable in this way allows for the inclusion of students who transfer schools or leave the district mid-year. An average attendance rate was created for Grades K–3 (in order to account for a student's pre-debate attendance) as well as one for the late elementary and middle schools years (Grades 4–8).

Another way to explore attendance is in terms of chronic absenteeism, a measure that all states are now required to include in their school reports by the Every Student Succeeds Act (Chang, Bauer, & Byrnes, 2018). Chronic absenteeism is defined by missing 10% or more of the school year and the most current national data released by the U.S. Department of Education indicate that nearly 8 million students in the United States were chronically absent in the 2015–2016 school year. Chronic absenteeism has been shown to increase achievement gaps at the elementary, middle, and high school levels (Balfanz & Byrnes, 2012); and in Baltimore, analysis from BERC in 2009 revealed that 9 in 10 BCPSS dropouts were chronically absent (Mac Iver, 2010). Thus, in addition to modeling attendance rate as a continuous variable, a chronically absent binary indicator (created from the average attendance rate from Grades 4–8 using 90% attendance as the cutoff point) is also examined as an outcome.

The final outcome variable of interest in this study is ninth-grade high school destination. Baltimore City's public high school system provides a unique opportunity to study educational attainment insofar as many city students apply for admissions into BCPSS high schools, which can be grouped into five categories. The first two categories (general entrance criteria high schools and career tech entrance criteria high schools) are selective in that they require certain thresholds of middle school performance in order for students to be accepted. These thresholds include high scores on the MSA as well as competitive middle school attendance rates and grades. The last three categories (charter, alternative, and traditional) do not utilize these thresholds when determining admission. Charter schools are externally operated public schools of choice (or lottery admission) and their curricular are often focused on college, career, or specialized career technology programming. Alternative high schools serve students seeking alternative paths to a high school diploma and are specially designed to help students who are overage and severely undercredited earn a diploma. Traditional high schools are the largest and most diverse set of high schools and the majority of BCPSS students attend their local traditional high school. However, it is also possible that some students in the sample are either not promoted to the ninth grade, transfer out of the district after middle school, or drop out before they ever attend a BCPSS high school. Consequently, these three outcomes will be added to the five types of high schools for a total of eight possible outcomes for ninth-grade high school destination.

Each BCPSS high school category has varying graduation and college enrollment rates and thus where a student attends high school can have a significant impact on their later academic attainment. In 2014, for example, the graduation rate of traditional high schools ranged from 50% to 80%, while general entrance criteria high schools had a graduation rate greater than 95% as well as the highest fall college enrollment rates out of any other category (Durham, Stein, & Connolly, 2015). Participating in debate during preadolescence may influence the probability of each outcome of ninth-grade high school destination via middle school performance and consequently aid students in the admissions process for selective entrance criteria BCPSS high schools.

Aside from the aforementioned predebate outcome measures, various demographic and background variables will also be used as covariates during analysis. These include school attended, age in 2017, race-ethnicity (coded as American Indian, Asian, Black, Hispanic, or White), and binary indicators of sex, English Language Learner status, special education services received, and free or reduced-price meals qualification.

Methods

Because this study only includes students who attended a school that participated in the Elementary and Middle School BUDL, any student in one of these schools who wanted could potentially participate. However, the foremost threat to internal validity for research on extracurricular activities stems from the voluntary nature of participation. For example, students motivated to join extracurricular activities are also those who tend to be more positively oriented to school than their peers (Gottfredson, Cross, & Seolé, 2007). Consequently, it is difficult to untangle genuine causal relationships from selection effects between voluntary extracurricular activity participation and student outcomes.

Two types of bias are common in observational data analysis of this type: baseline bias and differential treatment effect bias (Morgan & Winship, 2015). Baseline bias involves the aforementioned condition in which preexisting characteristics are associated with both the treatment and the outcome. In this case, it is possible that debate participation does not directly confer benefits, but that students who are more engaged in school both have better outcomes and are more likely to participate in debate in the first place. Because gathering specific information on when students first participate in extracurricular activities is difficult, especially in nationally representative surveys, adjusting for outcomes prior to participation is not possible. Failure to account for baseline selection bias could therefore artificially inflate estimated effects of participation. Prior empirical research on the Chicago Debate League found that high school debaters had higher average eighth-grade test scores and lower absenteeism in the ninth grade, suggesting that higher-performing students do select into the activity (Mezuk, 2009). High school debaters were also more likely to be female and more likely to qualify for free lunch (Mezuk et al., 2011).

The current study addresses concerns about selection by using inverse probability of treatment weighting to create statistically matched samples for comparison. This type of analysis reduces the potential for confounding factors such as student demographics and improves the confidence in any observed association between debate participation and educational outcomes. Along with a host of demographic characteristics, I will use predebate measures of academic achievement (third-grade reading and math test scores) and engagement measures (average K–3 attendance rates) to attempt to eliminate baseline treatment-effect bias. Analysis will also include a doubly robust method of balancing the data by incorporating covariates into both the propensity score model and the subsequent weighted regression. This supplemental parametric adjustment extends from prior research, which has used propensity score quintiles to examine outcomes for high school debate participants (Mezuk et al., 2011), by providing additional protection against model misspecification and addressing any imbalance that remains after applying weights derived from the propensity scores (Robins & Rotnitsky, 2001).

The second source of selection bias, differential treatment-effect bias, suggests that the associations between experiencing treatment and any observed outcomes may differ across subgroups. For example, there may be differential treatment-effect bias associated with the propensity to participate in debate; sufficient qualitative evidence suggests that some students participate in debate because they expect to gain academic benefits from doing so (Fine, 2001; Winkler, 2011). This self-selection on the individual-level causal effect renders the average treatment effect for those that typically do not participate in debate, as well as the average treatment effect for students in general, unidentified. Consequently, prior studies attempting to estimate the average treatment effect of debate may have upwardly biased estimates if they attempt to infer the size of the overall average treatment effect. Because investigators typically do not have measures for student (or parent) expectations of the benefits they might obtain from participating in debate, the only target parameter that can be estimated with any degree of confidence is the average treatment effect on the treated (ATT). Inverse probability of treatment weighting can be used in a targeted fashion to investigate effects only for the population of students who typically participate in debate. By taking this approach, this study differs from prior research on extracurricular activities by focusing solely on estimating the ATT, which in this case is the effect of preadolescent participation in debate among debaters.

For the initial analysis, I model the treatment selection mechanism. First, descriptive analyses were carried out to examine the extent to which differences exist between debaters and nondebaters on all observed variables. Then, propensity scores were estimated using logistic regression (Rosenbaum & Rubin, 1983). The propensity score is the probability of a student participating in debate during preadolescence, given the student's observed characteristics.

For the second stage of analysis, the following weights were calculated using estimated propensity scores in order to explore the effect of participation depending on the population of students experiencing treatment:

For $d_i = 1$: w_i, ATT $= 1$,

For $d_i = 0$: w_i, ATT $= p_i / 1 - p_i$,

where, for student i, d_i represents whether a student participated in BUDL during preadolescence, p_i represents the estimated propensity score, and w_i,ATT represents the average treatment effect on the treated weight. This weight uses the treatment group, or those students who participated in BUDL during preadolescence, as the target population. Members of the control group with higher propensity scores receive more weight, while members of the control group with low propensity scores receive less weight. The goal is for the weights to effectively align the treatment and control groups, approximating an experimental design were treatment is randomly assigned and unrelated to other characteristics. Balance was assessed between the treatment and control groups by comparing the average standardized mean differences across all covariates as well as the average standardized difference in standard deviations of continuous covariates (Morgan & Todd, 2008; Rubin, 1973).

The final stage of the analysis estimates weighted regressions and assesses causal effects by adopting a counterfactual approach for results from ATT-weighted regressions. In other words, it examines the effect of participating in debate during

Table 1
Summary of Descriptive Statistics for Debaters and Nondebaters

Variables	Preadolescent Debate Sample (n = 2,263)		Nondebate Sample (n = 81,906)	
	Mean	SD	Mean	SD
Covariates				
Age	20.50	3.28	20.75	3.47
Male	0.403		0.510	
Race and Hispanic origin				
American Indian	0.002		0.003	
Asian	0.006		0.010	
Hispanic	0.017		0.030	
Black	0.915		0.864	
White	0.060		0.093	
English language learner	0.019		0.033	
Special education services	0.142		0.235	
Free or reduced-price meals	0.958		0.947	
Predebate measures				
MSA reading Grade 3	406.85	31.35	396.45	31.52
MSA math Grade 3	396.14	39.21	384.07	40.94
Attendance rate (Grades K–3)	94.60	5.36	93.34	6.66
Outcomes				
MSA reading Grade 8	406.30	29.65	392.29	31.09
MSA math Grade 8	404.20	33.81	390.76	35.61
Attendance rate (Grades 4–8)	94.53	6.22	91.47	9.53
Chronic absenteeism (Grades 4–8)	0.117		0.269	
Ninth-grade high school destination				
Attend selective general	0.335		0.153	
Attend selective career tech	0.150		0.121	
Attend charter	0.081		0.062	
Attend alternative	0.008		0.014	
Attend traditional	0.340		0.441	
Dropout before high school	0.007		0.013	
Transfer before high school	0.049		0.152	
Not promoted to high school	0.030		0.044	

Note. N = 84,169. Includes only students that attended a school that participated in the Elementary and Middle School Baltimore Urban Debate League. The average standardized difference in means of covariates = 0.1506, and the average standardized difference in standard deviations of continuous covariates = 0.0804. MSA = Maryland School Assessment.

preadolescence among those students who typically participate. These regressions were restricted to the region of common support, which is the range of the propensity scores for which there are respondents in both the treatment and control groups. This resulted in 619 nondebaters (less than 1% of the sample), present in descriptive analyses, to be excluded from weighted regression analyses.

It is important to note that using inverse probability treatment weighting to estimate average treatment effects on the treated assumes that all variables that predict participation in debate, other than anticipation of the individual-level causal effects, are observed. Furthermore, while weighted regression techniques account for potentially confounding observed variables by balancing treatment and control samples across all observed variables, it cannot illuminate the extent to which these covariates relate to the outcomes of interest. To this end, results

from multiple ordinary least squares regression analyses for continuous outcomes as well as logistic regression analysis for categorical outcomes are presented in the appendix. All analyses included school fixed effects to account for school-level characteristics that may influence both a student's opportunity to participate in the BUDL at his or her school and the outcomes of interest.

Results

Table 1 displays a summary of descriptive statistics of covariates and outcomes of interest for debaters and nondebaters. Differences in outcomes justify further exploration of the effects of preadolescent participation in debate. This table also shows evidence of a baseline bias in many covariates between debaters and nondebaters. For example, the average standardized

Table 2
Logit-Coefficients for a Propensity Score Model of Preadolescent Debate Participation

Variables	Logit-Coefficients
Age	0.0229
	(0.0152)
Male	−0.330*
	(0.0507)
American Indian	0.428
	(0.464)
Asian	−0.0843
	(0.301)
Hispanic	0.00116
	(0.311)
Black	0.566*
	(0.149)
English language learner	−0.154
	(0.187)
Special education services	−0.277*
	(0.0726)
Free or reduced-price meals	0.426*
	(0.169)
Attendance rate (Grades K–3)	0.0286*
	(0.00450)
MSA reading Grade 3	0.00668*
	(0.00139)
MSA math Grade 3	0.00368*
	(0.00112)
Constant	−11.57*
	(0.729)
Model chi-square	408.75
Degrees of freedom	12

Note. N = 84,169. Robust *SE* in parentheses. MSA = Maryland School Assessment.
*$p < .05$ for two-tailed tests with null of 0.

difference in means of covariates between the treatment and control group is 0.1506, while the average standardized difference in standard deviations of continuous covariates is 0.0804. In sum, this table shows that students who end up being debaters differ from nondebaters in important ways prior to participation in the program and consequently highlights the need to adjust for these observed covariates in subsequent analysis.

Table 2 presents the propensity score model predicting likelihood of preadolescent debate participation among covariates and shows statistically significant associations between some covariates and selection into debate. On average, debaters are more likely to be female, more likely to be Black (as opposed to White), less likely to receive special education services, more likely to qualify for free or reduced-priced meals, and more likely to have higher predebate attendance and third-grade achievement as measured by MSA test scores than nondebaters.

Table 3 presents the means and standard deviations of covariates when applying the estimated weights. It demonstrates that the ATT weights constructed from the estimated propensity scores successfully balance the data. More specifically, the

average standardized difference in means of the covariates fell from 0.1506 to 0.0004 and the average standardized difference in standard deviations of the continuous covariates fell from 0.0804 to 0.0490. Furthermore, no statistically significant differences between debaters and non-debaters remain in the weighted sample.

Table 4 summarizes the results from doubly robust inverse probability of treatment weighted regression models predicting the relationship between preadolescent debate participation and the outcomes of interest. For continuous outcomes, estimates of the ATT for preadolescent debate participation are presented as raw coefficients as well as rescaled in standard deviation units. ATT-weighted regression estimates for categorical outcomes are presented as logit-coefficients and average probability differences. For all outcomes, standard errors are presented in parentheses.

Grade 8 MSA reading and math test scores. Before accounting for sample differences, preadolescent debaters scored approximately 14 points higher on average than nondebaters on both the eighth-grade reading and math MSA (see Table 1). This difference is equal to nearly half of a standard deviation. After accounting for potentially confounding covariates through doubly robust inverse probability treatment weighting, the average effect of preadolescent debate participation for debaters was significantly associated with increases in both assessments (reading: $b = 6.35, p < .001$; and math: $b = 4.52, p < .001$). In standard deviation units, the effect of preadolescent debate participation for debaters is an approximate 21% and 13% increase for eighth-grade reading and math MSA test scores, respectively.

Average Grade 4–8 attendance rate and chronic absenteeism indicator. The average Grade 4–8 attendance rate for all students in the sample was 91.6%, with an average attendance rate of 94.5% for preadolescent debaters and 91.5% for nondebaters. This difference of 3% is about one third of a standard deviation unit. The ATT estimate for preadolescent debate participation is 2.03%, or approximately one-fifth of a standard deviation unit ($p < .001$). Converting the attendance rate outcome into a binary indicator of chronic absenteeism with 90% attendance used as the cutoff point, a statistically significant relationship remains ($b = −0.90, p < .001$). Average probability differences demonstrate interpretable effect sizes. For example, the average probability of being chronically absent is 10% lower for debaters than for nondebaters.

Ninth-grade high school destination. Of the 84,169 students who attended a BCPSS school that participated in the elementary and middle school division of BUDL from the 2004–2005 to the 2013–2014 school years, approximately 80% attended a BCPSS high school in the ninth grade. More specifically, approximately 16% attended a selective general entrance criteria high school, 12% attended a selective career tech entrance criteria school, 6% attended a charter or transformation school, 2% attended an alternative school, and 44% attended a traditional high school.

Table 3
Balance Achieved by Weighting

Covariates	Preadolescent Debate Sample (n = 2,263)		Nondebate Sample (n = 81,906)	
	Mean	SD	Mean	SD
Age	20.50	3.28	20.49	3.41
Male	0.403		0.402	
Race and Hispanic origin				
American Indian	0.002		0.002	
Asian	0.006		0.006	
Hispanic	0.017		0.017	
Black	0.915		0.915	
White	0.060		0.060	
English language learner	0.019		0.019	
Special education services	0.142		0.142	
Free or reduced-price meals	0.958		0.958	
Predebate measures				
MSA reading Grade 3	406.85	31.35	406.90	32.17
MSA math Grade 3	396.14	39.21	396.19	41.18
Attendance rate (Grades K–3)	94.60	5.36	94.59	4.93

Note. N = 84,169. Means and standard deviations are weighted by the estimated average treatment effect for the treated (ATT) weight in order to demonstrate achieved balance. The average standardized difference in means of covariates = 0.0004, and the average standardized difference in standard deviations of continuous covariates = 0.0490. MSA = Maryland School Assessment.

The remaining 20% of students in the sample either dropped out before the ninth grade (1%), transferred out of the Baltimore City Public School System (15%), or were held back from attending the ninth grade (4%). With attending a traditional high school used as the base outcome, ATT-weighted multinomial logistic regression was utilized to predict the average treatment effect of preadolescent debate participation for debaters. Statistically significant positive relationships were found for selective general entrance criteria schools ($b = 0.74$, $p < .001$) and selective career tech entrance criteria schools ($b = 0.28$, $p < .001$). The average probability of attending a selective general entrance criteria high school is approximately 12% higher for debaters than nondebaters, while the average probability of attending a selective career tech entrance criteria school is approximately 2% higher for debaters than nondebaters. A statistically significant negative relationship was also found for the odds of transferring out of BCPSS ($b = -0.45$, $p < .001$).

Appendix Tables A1, A2, and A3 present findings from ordinary least squares, logistic, and multinomial logistic regression models, respectively, which can be used to understand the associations that demographic and predebate achievement covariates have with outcomes of interest.

Discussion

The key findings from this study are that preadolescent debate participation in a UDL had statistically significant relationships on many academic achievement and engagement outcomes among debaters. Preadolescent debate participation was associated with a 6.35 point increase in Grade 8 MSA reading scores and a 4.52 point increase in Grade 8 MSA math scores. The

larger association with reading scores is to be expected, as debate is an activity that focuses on informational text comprehension, a concept the MSA reading tests aim to assess. The positive relationship with math scores suggest that debaters may gain skills that aren't explicitly practiced in the activity indirectly through increases in school engagement outcomes (see Figure 1). The positive relationship with debate participation and student attendance rate supports this interpretation.

Increases in attendance during this stage of development may influence a variety of outcomes later in life. As mentioned previously, research has conceptualized eventual educational attainment as part of a long-term process of disengagement from school, with negative developmental pathways that begin during preadolescence. For example, students with greater declines in attendance between Grades 4–8 are less likely to be on track for high school graduation (Kieffer & Marinell, 2012). Furthermore, a majority of students who eventually drop out of high school in Baltimore enter Grade 9 with a pattern of chronic absenteeism that goes back at least several years (Mac Iver, 2010). Thus, the finding that preadolescent debate participation is associated with a 10% decrease in the probability of being chronically absent during this critical period of a student's development is particularly salient for policymakers and practitioners interested in influencing student trajectories.

This is related to the study's last set of findings pertaining to ninth-grade high school destination. Relative to attending a traditional high school, preadolescent debate participation was significantly associated with an increase in the probability of attending a selective general entrance criteria high school or a selective career tech entrance criteria high school. These results may not be surprising considering the aforementioned predicted increases in standardized test scores and attendance rates, two

Table 4
Estimates for the Average Treatment Effect for the Treated (ATT) for Preadolescent Debate Participation on Outcomes of Interest

Continuous Outcomes	Variable-Scaled Coefficients	Effect Size in SD
MSA reading Grade 8	6.35*	0.206
	(0.45)	
MSA math Grade 8	4.52*	0.128
	(0.53)	
Attendance rate (Grades 4–8)	2.03*	0.217
	(0.12)	

Categorical Outcomes	Logit-Coefficients	Effect Size in APD
Chronic absenteeism (Grades 4–8)	−0.90*	−0.100
	(0.08)	
Ninth-grade high school destination (base outcome = attend traditional high school)		
Attend selective general high school	0.74*	0.122
	(0.06)	
Attend selective career tech high school	0.28*	0.015
	(0.07)	
Attend charter high school	0.21	0.010
	(0.11)	
Attend alternative high school	−0.19	−0.003
	(0.24)	
Drop out before high school	−0.36	−0.004
	(0.26)	
Transfer before high school	−0.45*	−0.047
	(0.10)	
Not promoted to high school	−0.19	−0.012
	(0.14)	

Note. N = 83,550. Schools = 151. Robust *SE* in parentheses. Effect sizes shown in standard deviation units for continuous outcomes and in average probability differences for categorical outcomes. MSA = Maryland School Assessment; SD = standard deviation units; APD = average probability difference.
*p < .05 for two-tailed tests with null of 0.

measures BCPSS entrance criteria schools consider during the admissions process. However, the importance of these findings cannot be overstated as the average graduation rate of both categories of selective entrance criteria high schools surpasses the average rate of any other category.

Finally, the findings that preadolescent debaters are more likely to be Black and qualify for free or reduced-price meals, after adjusting for all other covariates, suggest that UDL may be a culturally appropriate intervention for a population who may not be well served by existing structures. For example, the appendix tables, which provide information on the magnitude and statistical significance of the covariates on the outcomes of interest, indicate that Black students are predicted to score 4.7 points lower in reading and 6.5 points lower in math on the eighth-grade MSA compared to their White peers. Students who qualify for free or reduced-price meals are also predicted to score lower on these assessments, more likely to be chronically absent in Grades 4–8, and less likely to attend an entrance criteria high school. Thus, this study provides evidence of a program that not only attracts marginalized students but influences their academic achievement and engagement outcomes as well, a goal many educational interventions likely share.

Findings should be interpreted in light of study limitations. Primarily, if there are unobservable characteristics that influence both preadolescent debate participation and the outcomes of interest, estimates of the ATT will be biased. While all students in the sample were potentially able to participate in the Elementary and Middle School BUDL at their respective schools, unmeasured factors could prohibit a student's ability to participate (i.e., transportation to and from tournaments). There is likely some degree of omitted variable bias in the propensity score model because BERC does not have information on parent characteristics. Thus, it is possible that unobserved or omitted variables threaten the assumption that treatment and control groups are identical at baseline. However, the study's use of predebate outcome measures as covariates greatly curb this threat. For example, any unobserved characteristics associated with standardized test scores or attendance rates, such as parent characteristics, are also likely related to these variables measured in the third grade, before participation in BUDL is possible. The propensity score model presented in Table 2 successfully balanced the data across all covariates and the doubly robust estimation provides some assurance against model misspecification. To be sure, without a randomized control trial, it is impossible to fully account for selection into a program. Future research should utilize sensitivity analysis to examine the extent to which unmeasured confounding could influence these estimates.

Furthermore, this study's findings are only applicable to the clearly defined causal state of participating in an Elementary and Middle School BUDL tournament and they do not illuminate specific mechanisms or aspects of this causal state that are attributable to the estimated effects. An examination of the properties of debate and how they may be similar to other activities or educational interventions, such as its cooperative and competitive structure, is a needed area of future research.

Nevertheless, this study adds to the growing literature on debate participation in significant ways. First, unweighted comparisons between preadolescent debaters and non-debaters reveled demographic differences between the two groups, particularly in terms of sex, race, special education services received and free or reduced-price meals qualification, as well as differences in pre-debate achievement and engagement measures. Accounting for sample differences using inverse probability weighted techniques mitigates observed selection bias in this cross-sectional study. Second, because there are likely individual-level differences in the expectations of benefits from participation between those that participate in debate and those that do not, this study addresses concerns about differential treatment effect bias by focusing on the treatment group as the target population parameter. Finally, although there is a mounting body of research that suggests participation in debate is associated with increases in positive outcomes for high school students, this research constitutes the first quantitative study to examine these relationships among elementary and middle school students.

This study's findings are also unique considering the relatively limited budget of many UDLs. Like most districts with a UDL, the Baltimore City Public School District does not contribute any funds to BUDL, which reportedly spends $1,000 per student per year on average and relies on volunteers as well as donations to pay for its staff, provide training, and run tournaments. The extent to which this program is low-cost compared to other educational interventions is debatable, but one cannot help but wonder the potential range of benefits extracurricular debate could provide to inner-city areas if invested in fully by administrators and school leaders. As mentioned previously, a large number of low-income urban youth do not participate in any extracurricular activities (Schwarts, Capella, & Seidman, 2015), and minority students have been understudied in the extracurricular activity literature as a whole (Fredricks & Simpkins, 2012). Accordingly, studies such as this are critical to the ongoing local and national policy debates about the impact of extracurricular activities, especially for urban and preadolescent students, two groups where the opportunity to participate is limited. Ensuring that all students acquire the requisite skills to succeed in life is an urgent goal that must be addressed, and as findings from this article suggest, UDLs may provide a compelling solution to the often-cited shortcomings of urban schools. Must policymakers and practitioners focus their efforts in closing the achievement gap within the confines of the school day? Or can addressing the inequality present in extracurricular participation lead to more equal outcomes?

ORCID ID

Daniel Shackelford ![ORCID] https://orcid.org/0000-0002-3306-302X

REFERENCES

Abbott, R. D., O'Donnell, J., Hawkins, J. D., Hill, K. G., Kosterman, R., & Catalano, R. F. (1999). Changing teaching practices to promote achievement and bonding to school. *American Journal of Orthopsychiatry*, *68*(4), 542–552.

Anderson, S., & Mezuk, B. (2015). Positive youth development and participation in an urban debate league: Results from Chicago public schools, 1997-2007. *Journal of Negro Education*, *84*(3), 362.

Balfanz, R., & Byrnes, V. (2012). *The importance of being in school: A report on absenteeism in the nation's public schools*. Baltimore: Johns Hopkins University Center for Social Organization of Schools.

Balfanz, R., Herzog, L., & Mac Iver, D. J. (2007). Preventing student disengagement and keeping students on the graduation path in urban middle-grades schools: Early identification and effective interventions. *Educational Psychologist*, *42*(4), 223–235.

Breger, B. (2000). Overview of the urban debate program. *Rostrum*, *75*(14).

Broh, B. A. (2002). Linking extracurricular programming to academic achievement: Who benefits and why? *Sociology of Education*, *75*(1), 69.

Chang, H. N., Bauer, L., & Byrnes, V. (2018, September). *Data matters: Using chronic absence to accelerate action for student success*. n.p.: Attendance Works and Everyone Graduates Center.

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.

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

Denault, A. S., & Poulin, F. (2009). Intensity and breadth of participation in organized activities during the adolescent years: Multiple associations with youth outcomes. *Journal of Youth and Adolescence*, *38*, 1199–1213. doi:10.1007/s10964-009-9437-5

Durham, R. E., Stein, M., & Connolly, F. (2015). *College opportunities and success: Baltimore City graduates through the class of 2014*. Baltimore: Baltimore Education Research Consortium.

Eccles, J. S., & Barber, B. L. (1999). Student council, volunteering, basketball, or marching band what kind of extracurricular involvement matters? *Journal of Adolescent Research*, *14*, 10–43. doi:10.1177/0743558499141003

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

Feldman, A., & Matjasko, J. L. (2005). The role of school-based extracurricular activities in adolescent development: A comprehensive review and future directions. *Review of Educational Research*, *75*, 159–210.

Fine, G. A. (2001). *Gifted tongues: High school debate and adolescent culture*. Princeton, NJ: Princeton University Press.

Fredricks, J., & Simpkins, S. (2012). Promoting positive youth development through organized after-school activities: Taking a closer look at participation of ethnic minority youth. *Child Development Perspectives*, 6(3), 280–287. doi:10.1111/j.1750- 8606.2011.00206.x

Gibbs, B. G., Erickson, L. D., Dufur, M. J., & Miles, A. (2015). Extracurricular associations and college enrollment. *Social Science Research*, 50, 367–381.

Gottfredson, D. C., Cross, A., & Soulé, D. A. (2007). Distinguishing characteristics of effective and ineffective after-school programs to prevent delinquency and victimization. *Criminology & Public Policy*, 6(2), 289–318.

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, 312, 1900–1902.

Johnson, D. W., Johnson, R. T., & Stanne, M. B. (2000). *Cooperative learning methods: A meta-analysis*. Retrieved from https://pdfs .semanticscholar.org/93e9/97fd0e883cf7cceb3b1b612096c27aa 40f90.pdf

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.

Kena, G., Hussar, W., McFarland, J., de Brey, C., Musu-Gillette, L., Wang, X., & Dunlop Velez, E. (2016). *The condition of education 2016* (NCES 2016-144). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Kieffer, M. J., & Marinell, W. H. (2012). *Navigating the middle grade: Evidence from New York City*. New York: The Research Alliance for New York City Schools.

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

Mac Iver, M. A. (2010). *Gradual disengagement: A portrait of the 2008-2009 dropouts in the Baltimore City Schools*. A Research Report by the Baltimore Education Research Consortium, Baltimore, MD.

Mahoney, J. L., Cairns, B. D., & Farmer, T. W. (2003). Promoting interpersonal competence and educational success through extracurricular activity participation. *Journal of Educational Psychology*, 95, 409–418.

McNeal, R. B. (1995). Extracurricular activities and high-school dropouts. *Sociology of Education*, 68(1), 62–80.

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

Mezuk, B., Bondarenko, I., Smith, S., & Tucker, E. (2011). Impact of participating in a policy debate program on academic achievement: Evidence from the Chicago Urban Debate League. *Educational Research and Reviews*, 6, 622–635.

Mitchell, G. (1998). Pedagogical possibilities for argumentative agency in academic debate. *Argumentation Advocacy*, 35, 41–56.

Morgan, S. L., & Todd, J. J. (2008). A diagnostic routine for the detection of consequential heterogeneity of causal effects. *Sociological Methodology*, 38, 231–281.

Morgan, S. L., & Winship, C. (2015). *Counterfactuals and causal inference: Methods and principles for social research*. Cambridge, UK: Cambridge University Press.

National Association of Urban Debate Leagues. (2016). *Our results*. Retrieved from http://urbandebate.org/Our-Results

National Governors Association. (2010). *The common core state standards*. Retrieved from http://www.corestandards.org/the-standards.

National Institute on Out-of-School Time. (2003). *Fact sheet on school-age children's out of school time*. Wellesley, MA: Wellesley College, Center for Research on Women.

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. doi:10.1207/s1532480xads0902

Porter, A., McMaken, J., Hwang, J., & Yang, R. (2011). Common core standards: The new U.S. intended curriculum. *Educational Researcher*, 40, 103–116.

Quinn, J. (1999). Where need meets opportunity: Youth development programs for early teens. *The Future of Children*, 9(2), 96–116. Retrieved from http://www.jstor.org/stable/10.2307/1602709

Robins, J. M., & Rotnitzky, A. (2001). Comment on the Bickel and Kwon article, "Inference for semiparametric models: Some questions and an answer." *Statistica Sinica*, 11, 920–936.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41–55.

Rubin, D. B. (1973). Matching to remove bias in observational studies. *Biometrics*, 29, 159–183.

Schwartz, K., Cappella, E., & Seidman, E. (2015). Extracurricular participation and course performance in the middle grades: A study of low-income, urban youth. *American Journal of Community Psychology*, 56, 307–320.

Shackelford, D. T., Ratliff, S., & Mezuk, B. (2018). *Participating in a high school debate program and college matriculation and completion: Evidence from the Chicago Debate League*. Manuscript submitted for publication.

Shulruf, B. (2010). Do extracurricular activities in schools improve educational outcomes? A critical review and meta-analysis of the literature. *International Review of Education*, 56(5), 591.

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.

Snipes, J., & Horowitz, A. (2008). *Advancing adolescent literacy in urban schools*. Washington, DC: Council of the Great City Schools.

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, C. B. (2009, April). *Cities in Crisis 2009: Closing the graduation gap - educational and economic conditions in America's largest cities*. Bethesda, MD: Editorial Projects in Education Research Center.

U.S. Census Bureau. (2014). *A child's day: Living arrangements, nativity, and family transitions: 2011*. Washington, DC: U.S. Department of Commerce Economics and Statistics Administration. Retrieved from https://www.census.gov/newsroom/press-releases/2014/cb14-224.html

Winkler, C. (2011). To argue or to fight: Improving at-risk students' school conduct through urban debate. *Controversia*, 7(2), 76–90.

AUTHOR

DANIEL SHACKELFORD is a final-year PhD graduate student at the Johns Hopkins University School of Education, 2800 N. Charles Street, Baltimore, MD, 21218; *danielshackelford@jhu.edu*. His research focuses on the factors related to school achievement, particularly for low-income students and students of color in urban areas.

Appendix

Table A1
Ordinary Least Squares Estimates of Variables on Continuous Outcomes

Variables	MSA Reading Grade 8	MSA Math Grade 8	Attendance Rate (Grades 4–8)
Debate participation	6.119*	4.340*	2.201*
	(0.441)	(0.522)	(0.181)
Age	1.399*	2.079*	−0.110*
	(0.0255)	(0.0301)	(0.0105)
Male	−5.384*	−2.563*	−0.576*
	(0.146)	(0.173)	(0.0599)
American Indian	−2.717	−2.248	0.597
	(1.420)	(1.680)	(0.583)
Asian	−0.464	6.690*	1.455*
	(0.825)	(0.977)	(0.339)
Hispanic	−0.896	−2.792*	1.064*
	(0.611)	(0.723)	(0.251)
Black	−4.690*	−6.494*	1.803*
	(0.330)	(0.391)	(0.136)
English language learner	4.372*	10.04*	1.928*
	(0.557)	(0.659)	(0.229)
Special education services	−11.19*	−9.104*	−0.391*
	(0.187)	(0.221)	(0.0767)
Free or reduced-price meals	−5.461*	−8.633*	−0.441
	(0.359)	(0.425)	(0.247)
MSA reading Grade 3	0.227*	0.484*	0.0201*
	(0.00317)	(0.00376)	(0.00130)
MSA math Grade 3	0.412*	0.180*	0.0156*
	(0.00381)	(0.00450)	(0.00156)
Attendance rate (Grades K–3)	0.0671*	0.292*	0.545*
	(0.0125)	(0.0148)	(0.00513)
Constant	121.0*	79.87*	27.22*
	(1.872)	(2.215)	(0.769)
R-squared	.509	.474	.160

Note. $N = 83,550$. Schools = 151. Robust *SE* in parentheses. MSA = Maryland School Assessment.
*$p < .05$ for two-tailed tests with null of 0.

Table A2
Logit-Coefficients of Variables on the Binary Indicator Chronic Absenteeism

Variables	Logit-Coefficients
Debate participation	−0.877*
	(0.0716)
Age	0.0480*
	(0.00311)
Male	0.138*
	(0.0179)
American Indian	−0.157
	(0.164)
Asian	−0.444*
	(0.131)
Hispanic	−0.272*
	(0.0811)
Black	−0.465*
	(0.0405)
English language learner	−0.754*
	(0.0821)
Special education services	0.156*
	(0.0218)
Free or reduced-price meals	0.318*
	(0.0511)
MSA reading Grade 3	−0.00432*
	(0.000483)
MSA math Grade 3	−0.00545*
	(0.000400)
Attendance rate (Grades K–3)	−0.134*
	(0.00348)
Constant	14.51*
	(0.260)
Model chi-square	15,102.66
Degrees of freedom	13

Note. $N = 83,550$. Schools = 151. Robust *SE* in parentheses. MSA = Maryland School Assessment.
*$p < .05$ for two-tailed tests with null of 0.

Multinomial Logit-Coefficients of Variables on Ninth-Grade High School Destination
(Base = Traditional High School)

Variables	Selective General	Selective Career Tech	Charter	Alternative	Drop Out	Transfer	Not Promoted
Debate participation	0.772*	0.295*	0.242*	−0.197	−0.359	−0.452*	−0.198
	(0.0605)	(0.0674)	(0.184)	(0.235)	(0.264)	(0.104)	(0.133)
Age	0.112*	0.0220*	−0.101*	0.0918*	0.0863*	−0.0582*	−0.458*
	(0.00454)	(0.00414)	(0.00498)	(0.0112)	(0.0121)	(0.00380)	(0.00840)
Male	−0.631*	−0.132*	−0.156*	0.251*	0.0210	−0.0186	0.0433
	(0.0240)	(0.0233)	(0.0304)	(0.0623)	(0.0648)	(0.0221)	(0.0378)
American Indian	0.115	0.277	0.425	0.844	−1.595	−0.0943	0.219
	(0.244)	(0.316)	(0.317)	(0.623)	(1.012)	(0.186)	(0.286)
Asian	0.631*	0.291	0.0698	1.285	0.123	0.292*	0.742*
	(0.133)	(0.247)	(0.321)	(0.753)	(0.420)	(0.127)	(0.160)
Hispanic	0.0874	0.219	−0.462*	−0.0494	−0.260	−0.248*	−0.183
	(0.0922)	(0.138)	(0.190)	(0.631)	(0.270)	(0.0806)	(0.116)
Black	0.0993	0.926*	0.423*	0.953*	−0.686*	−0.651*	−0.708*
	(0.0526)	(0.0759)	(0.0882)	(0.231)	(0.134)	(0.0451)	(0.0710)
English language learner	0.460*	−0.506*	−0.490*	−2.915*	−0.186	0.292*	0.125
	(0.0838)	(0.125)	(0.166)	(1.040)	(0.272)	(0.0753)	(0.109)
Special education services	−0.904*	−0.339*	0.113*	−0.0529	−0.492*	−0.175*	−0.273*
	(0.0411)	(0.0308)	(0.0354)	(0.0709)	(0.0867)	(0.0275)	(0.0477)
Free or reduced-price meals	−0.702*	0.337*	−0.0326	0.784	−1.440*	−1.380*	−1.232*
	(0.0879)	(0.0764)	(0.109)	(0.414)	(0.129)	(0.0515)	(0.0804)
MSA reading Grade 3	0.0248*	0.0115*	0.00133*	−0.00191	0.00202	0.00604*	0.00465*
	(0.000549)	(0.000538)	(0.000644)	(0.00146)	(0.00160)	(0.000489)	(0.000736)
MSA math Grade 3	0.0187*	0.00391*	0.00276*	−0.00696*	0.00450*	0.00741*	0.00412*
	(0.000631)	(0.000633)	(0.000763)	(0.00181)	(0.00196)	(0.000579)	(0.000846)
Attendance rate (Grades K–3)	0.0542*	0.0274*	0.0148*	−0.0253*	−0.0579*	−0.0113*	0.0418*
	(0.00259)	(0.00217)	(0.00261)	(0.00398)	(0.00345)	(0.00174)	(0.00396)
Constant	−20.81*	−10.04*	−2.075*	−9.000*	−1.337	0.00430	3.898*
	(0.408)	(0.414)	(0.507)	(1.509)	(1.064)	(0.333)	(0.554)

Note. N = 83,550. Schools = 151. Model chi-square = 32,844.65. df = 91. Robust *SE* in parentheses. MSA = Maryland School Assessment.
*$p < .05$ for two-tailed tests with null of 0.