

Full Length Research Paper

Debate participation and academic achievement among high school students in the Houston Independent School District: 2012 - 2015

Tomohiro M. Ko¹ and Briana Mezuk^{1,2*}

¹Department of Epidemiology, School of Public Health, University of Michigan, 1415 Washington Heights, Ann Arbor, MI 48109, United States.

²Institute for Social Research, University of Michigan, 426 Thompson St, Ann Arbor, MI 48104, United States.

Received 4 February, 2021, Accepted 17 May, 2021

Competitive debate programs exist across the globe, and participation in debate has been linked to improved critical thinking skills and academic performance. However, few evaluations have been able to adequately address self-selection into the activity when examining its impact on achievement. This study evaluated the relationship between participating in a debate program and academic performance among high school students (N=35,788; 1,145 debaters and 34,643 non-debaters) using linked debate participation and academic record data from the Houston Independent School District. Academic performance was indicated by cumulative GPA and performance on the SAT college entrance exam. Selection into debate was addressed using propensity score methods informed by sociodemographic characteristics and 8th grade standardized test scores to account for pre-debate achievement. Debate participation was associated with 0.66 points (95% Confidence Interval (CI): 0.64, 0.68) higher GPA, 52.43 points (95%CI: 50.47, 54.38) higher SAT Math, and 57.05 points (95% CI: 55.14, 58.96) higher SAT Reading/Writing scores. Findings suggest that competitive debate is associated with better academic outcomes for students.

Key words: Achievement, program evaluation, testing, observational research, after school/co-curricular.

INTRODUCTION

There are persistent gaps in academic achievement and college-readiness in urban, public school districts, especially among lower income and minority students (Banerjee, 2016). Policy makers and educators have advanced extracurricular learning to address these achievement disparities (Marsh and Kleitman, 2002). However, there is limited quantitative evidence supporting

the effectiveness of extracurricular programs at improving academic outcomes for lower income and/or minority secondary school students, especially regarding college-readiness. Research is particularly needed in districts that predominantly serve Latino/Hispanic students, the fastest growing group in K-12 schools (US Department of Education, 2020). De facto segregation by race and

*Corresponding author. E-mail: bmezuk@umich.edu. Tel: 734.615.9204.

ethnicity remains in the US public school system. According to the US Department of Education, 95% of Hispanic and 96% of Black students attend a school that is at least 25% racial/ethnic minority; in comparison, only 52% of non-Hispanic white students attend a school that is at least 25% racial/ethnic minority (US Department of Education, 2020).

Competitive debate is a co-curricular activity centered on the communication of evidence-based argumentation. Pairs of students work together to debate both sides of policy-relevant topics (e.g., government support for renewable energy), and in the process practice academic skills including reading and interpreting complex non-fiction text, developing and responding to arguments orally and in writing, collaborative and cooperative learning, and time management (Mitchell, 1998). Debate leagues continue to grow worldwide, with international tournaments drawing debaters from up to 60 countries (English-Speaking Union, 2020). In addition, there is a large body of qualitative evidence supporting the positive impact of debate on critical thinking skills, school engagement, and personal development (Louden, 2010).

There are major challenges to isolating and quantifying the impact of extracurricular activities on academic performance. While there is an extensive literature, both quantitative and qualitative, describing the salience of a wide range of extracurricular activities (e.g., music, sports, theater) for adolescent development (Eccles et al., 2003; Eccles and Barber, 1999; Gibbs et al., 2015; Marsh and Kleitman, 2002), the causal evidence that *specific* activities enhance academic performance is limited. This stems from two methodologic issues that are challenging to address: First, the identification of an appropriate comparison group (Marsh and Kleitman, 2002); this is difficult because program evaluators often only have data on students who participate in the activity. Second, an adequate means to account for self-selection into the activity (Hunt, 2005); programs often only have data on students once they have begun participating, meaning that they cannot account for pre-activity academic performance when evaluating the impact of the program. Large academic administrative data systems, which can be linked to information regarding participation in extracurricular activities, provide an opportunity for addressing both of these limitations (Mezuk et al., 2011).

Using such large administrative data systems, a handful of studies have quantitatively evaluated the relationship between participating in a policy debate league and academic achievement in urban school districts. Mezuk et al. (2011) found that Chicago high school students who participated in debate were more likely to graduate from high school, performed better on the ACT college entrance exam, and gained more in GPA over the course of high school than comparable students who did not participate (Mezuk et al., 2011). A more recent report found that debate was associated with gains in standardized test scores and lower likelihood of

absenteeism among middle school students in Baltimore (Shackelford, 2019). Both the Mezuk et al. (2011) and Shackelford (2019) studies used propensity score methods to account for the non-random assignment (that is, self-selection) of students into debate programs; both identified that better-achieving students were more likely to self-select into debate, but that debate participation was still associated with academic outcomes even after accounting for this self-selection. In addition, other quantitative reports have examined the relationship between debate participation and indicators of psychosocial development (e.g., self-efficacy, civic engagement, etc.) and have reported positive correlations (Anderson and Mezuk, 2015; Kalesnikava et al., 2019). In sum, quantitative studies of debate participation in urban school districts show that while there is differential self-selection into debate, consistent with all extra-curricular activities (Hunt, 2005), debate participation is still associated with better academic performance after accounting for this self-selection.

The present study aims to extend this work by assessing the relationship between debate participation and indicators of academic achievement and college-readiness among a large sample of high school students from a district that serves a predominantly Hispanic/Latino student population. Data come from the Houston Urban Debate League (HUDL) and the Houston Independent School District (HISD), the largest school district in Texas, with records spanning 2012 to 2015. We use quasi-experimental propensity score methods to account for the non-random assortment of students into debate to attempt to isolate the influence of participation in this activity on academic achievement.

METHODS

Data sources

Two sources of de-identified data on three 9th grade cohorts (2012/13 through 2014/15) of students were merged to form the sample: 1) academic records from HISD and 2) debate participation records from HUDL. The analytic sample consisted of all HUDL participants ("debaters") during this period as indicated by debate tournament participation records. The comparison sample of non-debaters was created via a 30% random sample of 9th grade students who did not debate from each academic year (2012/13 to 2014/15), which equated to approximately 11,000 students from each 9th grade cohort. The resulting total sample for this analysis was 35,788 students, which consisted of 1,145 debaters (that is, students who participated in at least one debate tournament) and 34,643 non-debaters.

All demographic and academic performance variables were derived from HISD administrative records. Sociodemographic characteristics included sex, age in 9th grade, race (coded as Hispanic/Latino, Black, non-Hispanic white, Asian, Native American, and other for analysis), cohort year (2012/13, 2013/14 and 2014/15), and whether the student qualified for free/reduced cost lunch, which served as a proxy of economic disadvantage. Finally, to account for differential self-selection of students into debate as a function of academic performance, we indexed pre-debate (that is, 8th grade) achievement by performance on the Reading and Math

sections of the State of Texas Assessments of Academic Readiness (STAAR) test, a state-wide standardized exam (Texas Education Agency, 2021). While the exact percentiles on the STAAR sections vary year to year, for 8th grade, scores between 1700 – 1759 on the Reading and scores between 1700 – 1828 on the Math section are indicative that the student “Meets” academic readiness thresholds for those subjects; higher values indicate that the student “Masters” those subjects (Texas Education Agency, 2020).

Outcome assessment

We examined two academic outcomes: cumulative GPA (that is, last recorded GPA for each student, modeled as a continuous variable) and performance on the Math and Evidence-based Reading/Writing sections of the SAT college entrance exam. For individuals who took the SAT multiple times only the highest score was used. The format of the SAT changed during the study period (The College Board, 2015); from 2005 to 2016 the SAT was scored out of a total of 2400 points with three sections (Math, Critical Reading, and Writing) each worth 800 points. We converted these to the current (2016 – present) SAT format, which includes two sections (Math and Reading/Writing Sections) which are each worth 800 points (for a total possible score of 1600 points), according to College Board concordance guidelines (The College Board, 2016). The SAT has identified benchmarks that represent “college-readiness” (that is, a 75% likelihood of attaining at least a “C” in first semester college course related to each section); these are scores of ≥ 480 for the Reading/Writing section and ≥ 530 for the Math section (The College Board, 2016). SAT performance was examined as both a continuous outcome (that is, average expected score on each section) and as a binary outcome (that is, met college readiness benchmark for the section).

Treatment of missing data

Data in this study all come from administrative sources (e.g., debate tournament records and administrative school records) and as such, for some variables there is substantial missing data. As these data are unlikely to be missing completely at random, including only cases with complete data on all covariates ($n=16,704$) in our analysis would have resulted in a biased sample (Leyrat et al., 2019). To address this missing data problem, we used Multiple Imputation with Chained Equations (MICE) (van Buuren and Groothuis-Oudshoorn, 2011). We imputed 10 complete datasets from the original data, with a maximum of 10 iterations per imputation, using the R MICE package (Version 3.6.0). We verified the plausibility of the imputed values (e.g., ensuring there were no cases of implausible age in 9th grade) using diagnostic plots comparing marginal distributions of observed and imputed data.

Analysis

First, we compared the sociodemographic characteristics of debaters and non-debaters using Chi-squared tests for categorical variables and t-tests for continuous variables. This comparison clarifies to what degree debaters differed from students who did not debate, including differential self-selection into the activity, and provides a metric to assess the reach of the program (that is, which types of students are engaging in the debate league, and which are not).

Next, we used inverse probability of treatment weighting (IPTW) (Austin and Stuart, 2015) to account for selection bias in our estimates of the relationship between debate participation and the two outcome indicators of academic achievement (SAT

performance and GPA). IPTW addresses selection bias by weighting each observation in the dataset by the inverse of the probability (that is, propensity) they debated (e.g., students who are very likely to have debated, and did in fact debate, are down-weighted and students who are very unlikely to have debated, but did in fact debate, are up-weighted). This weighting creates a “pseudo-population” in which debaters and non-debaters are balanced based on their observed characteristics. In this manner, IPTW generates estimates of the debate-achievement relationship that are less biased than those that would be generated from standard multivariable regression (Austin and Stuart, 2015).

To generate the propensity score (that is, probability that a student debated), we used a two-step process: First, we fit a logistic regression model predicting debate participation (1=yes, 0=no) from observed socio-demographic characteristics (that is, sex, age in 9th grade, race, 9th grade cohort/year, and free/reduced lunch) and pre-debate achievement (that is, 8th grade STAAR reading and math scores) within each of the 10 imputed datasets. Next, from this logistic regression model, we estimated the predicted probability (that is, propensity, possible range: 0 (very unlikely to debate) to 1 (very likely to debate)) for each student in the sample. We generated the IPT weight for each student by taking the inverse of this probability (1/predicted probability of debate participation).

We then used this IPTW to fit regression models of debate predicting academic achievement (that is, GPA and SAT performance) using a two-step procedure: We fit a generalized linear model for each of the 10 imputed datasets estimating the effect of debate participation on each outcome (that is, GPA, SAT Math Score, and SAT Reading/Writing Score), using IPTW and adjusting for sex, age in 9th grade, race, ninth grade cohort, free/reduced lunch status, and 8th grade STAAR reading and math scores. Three alternative specifications of this model were considered: (1) unadjusted for all covariates while using IPTW, (2) unadjusted for all covariates using IPTW with the propensity score function including all interaction terms, and (3) adjusted for all covariates using IPTW with the propensity score function including all interaction terms. However, model fit was poor for the alternative models and the R^2 was consistently highest for the fully adjusted model using IPTW with no interaction terms in the propensity score function. Finally, parameter estimates (beta coefficients) and standard errors were then pooled across the 10 imputed datasets into a single set of values for each indicator of achievement.

All data analysis was conducted in R Studio (3.5.2) and all p-values refer to two-tailed tests. This study was reviewed and deemed exempt from human subjects regulation by the Institutional Review Board at the University of Michigan. It was approved by the Office of Research and Accountability at HISD.

RESULTS

As shown in Table 1, nearly two-thirds of the sample was Hispanic/Latino and three-quarters qualified for free/reduced lunch, a proxy indicator of socioeconomic disadvantage. This is consistent with the overall demographics of the HISD (Houston Independent School District, 2021), indicating that our sample was representative of the district as a whole. Debaters were slightly younger in 9th grade and were more likely to be female and Asian or non-Hispanic White compared to non-debaters; there was no difference in free/reduced lunch status. While 8th grade STAAR test scores were significantly higher for debaters, consistent with differential self-selection of higher-achieving students into the activity, even among debaters these higher scores

Table 1. Characteristics of Houston Independent School District High School students by debate participation.

| Parameter | Overall (n=35,788) | Debaters (n=1,145) | Non-debaters (n=34,643) | Test statistic | df | p |
|--|-----------------------|-----------------------|----------------------------|----------------|-------|--------|
| Sex | | | | | | |
| Female | 17255 (48.2%) | 639 (55.8%) | 16616 (48.0%) | 27.0 | 1 | <.0001 |
| Male | 18533 (51.8%) | 506 (44.2%) | 18027 (52.0%) | | | |
| Age in 9th Grade | | | | | | |
| Mean (SD) | 14.5 (0.836) | 14.1 (0.480) | 14.5 (0.841) | 26.0 | 1203 | <.0001 |
| Missing | 131 (0.4%) | 131 (11.4%) | 0 (0%) | | | |
| Race | | | | | | |
| Asian | 1275 (3.6%) | 109 (9.5%) | 1166 (3.4%) | 137.8 | 4 | <.0001 |
| Black | 9089 (25.4%) | 303 (26.5%) | 8786 (25.4%) | | | |
| Hispanic/Latino | 22982 (64.2%) | 637 (55.6%) | 22345 (64.5%) | | | |
| Other | 334 (0.9%) | 17 (1.5%) | 317 (0.9%) | | | |
| Non-Hispanic white | 2108 (5.9%) | 79 (6.9%) | 2029 (5.9%) | | | |
| Ninth grade cohort | | | | | | |
| 2012-2013 | 11407 (31.9%) | 395 (34.5%) | 11012 (31.8%) | 4.1 | 2 | 0.13 |
| 2013-2014 | 12099 (33.8%) | 380 (33.2%) | 11719 (33.8%) | | | |
| 2014-2015 | 12282 (34.3%) | 370 (32.3%) | 11912 (34.4%) | | | |
| Free/reduced lunch status | | | | | | |
| No | 9252 (25.9%) | 310 (27.1%) | 8942 (25.8%) | 0.9 | 1 | 0.35 |
| Yes | 26536 (74.1%) | 835 (72.9%) | 25701 (74.2%) | | | |
| STAAR reading score | | | | | | |
| Mean (SD) | 1740 (312) | 1770 (196) | 1730 (315) | 5.6 | 1004 | <.0001 |
| Missing | 8500 (23.8%) | 289 (25.2%) | 8211 (23.7%) | | | |
| STAAR Math score | | | | | | |
| Mean (SD) | 1730 (322) | 1750 (237) | 1730 (325) | 2.0 | 774.8 | 0.05 |
| Missing | 10709 (29.9%) | 446 (39.0%) | 10263 (29.6%) | | | |
| SAT Math section score | | | | | | |
| Mean (SD) | 471 (108) | 525 (112) | 469 (108) | 15.9 | 1135 | <.0001 |
| Missing | 14140 (39.5%) | 108 (9.4%) | 14032 (40.5%) | | | |
| SAT reading/writing section score | | | | | | |
| Mean (SD) | 470 (107) | 534 (109) | 467 (106) | 19.4 | 1137 | <.0001 |
| Missing | 14140 (39.5%) | 108 (9.4%) | 14032 (40.5%) | | | |
| GPA | | | | | | |
| Mean (SD) | 2.61 (0.958) | 3.38 (0.800) | 2.58 (0.951) | 32.8 | 1264 | <.0001 |
| Missing | 5668 (15.8%) | 9 (0.8%) | 5659 (16.3%) | | | |

P-value refers to comparison between debaters and non-debaters. Test statistic refers to a Chi2 for categorical variables and t-test for continuous variables.

were still only in the “meets” academic readiness category.

Using IPTW to account for self-selection into debate, the average cumulative GPA for debaters was 0.66 points (95% Confidence Interval (CI): 0.64, 0.68) higher

than comparison students. Similarly, debate participation was associated with 52.43 points (95% CI: 50.47, 54.38) higher score on the Math and 57.05 points (95% CI: 55.14, 58.96) higher score on the reading/writing section of the SAT. As shown in Figure 1, debate participants

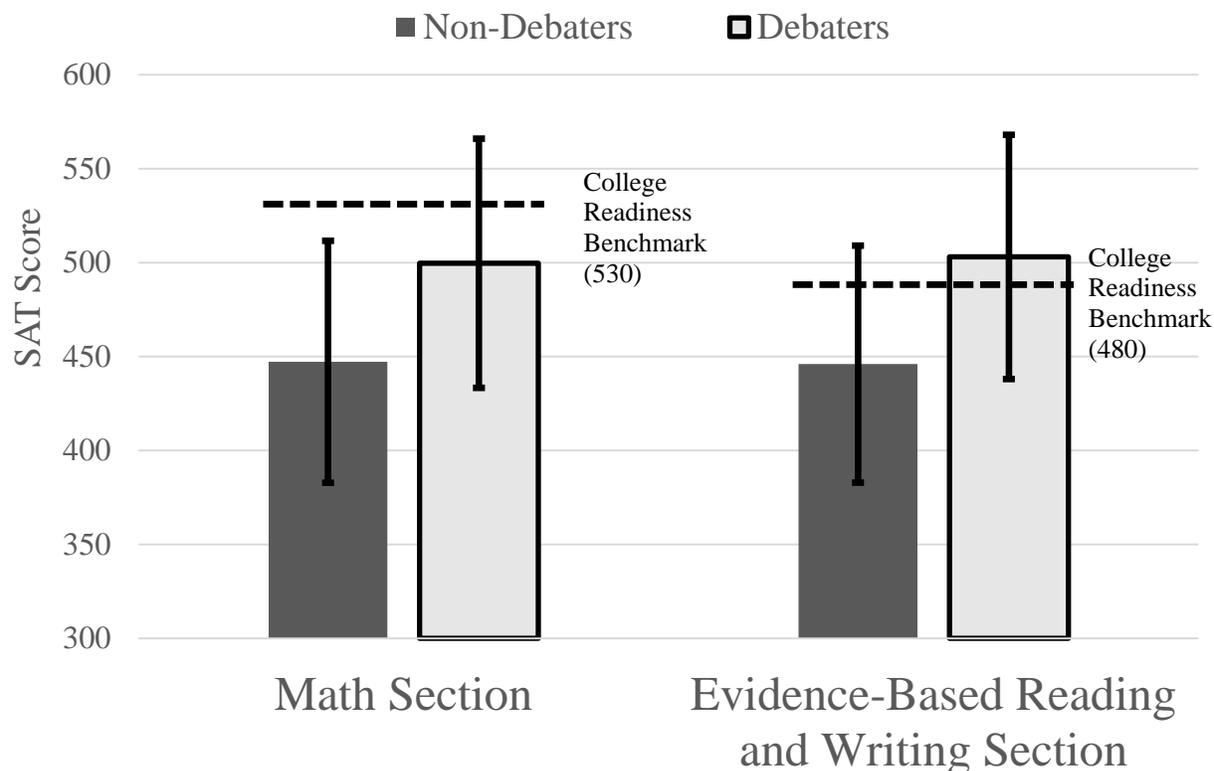


Figure 1. Average SAT performance by debate status among high school students in the Houston Independent School District, 2012/13 to 2014/15. IPTW adjusted average SAT performance by debate participation status. Estimates are pooled from imputed data. IPTW estimated using sociodemographic characteristics and 8th grade standardized test scores. Error bars indicate 95% confidence intervals. College-readiness benchmarks are those scores provided by the SAT to indicate a 75% likelihood of attaining at least a C in first semester courses related to the section (e.g., a quantitative-oriented course for the Math SAT).

Table 2. Sensitivity analyses examining the relationship between debate participation and GPA and SAT performance under three analytic scenarios

| Outcome | Complete case linear regression analysis (n=16,704) | 95% confidence interval | Multiple imputation linear regression analysis (n=35,788) | 95% confidence interval | Multiple imputation IPTW analysis (n=35,788) | 95% confidence interval |
|---------------------------|---|-------------------------|---|-------------------------|--|-------------------------|
| | Beta | | Beta | | Beta | |
| Cumulative GPA | 0.57 | [0.51, 0.63] | 0.60 | [0.55, 0.65] | 0.66 | [0.64, 0.68] |
| SAT Math Score | 40.65 | [33.82, 47.47] | 52.86 | [47.29, 58.45] | 52.43 | [50.47, 54.38] |
| SAT Reading/Writing score | 47.70 | [41.02, 54.39] | 61.02 | [55.57, 6.65] | 57.05 | [55.14, 58.96] |

Three analytic scenarios: (1) Complete case analysis (no imputation, standard linear regression modeling), (2) Multiple imputation (pooled across 10 imputed datasets, standard linear regression modeling) and (3) Inverse probability of treatment weighting (with multiple imputation).

were significantly more likely to meet the college-readiness benchmark on the Reading/Writing (Odds ratio: 1.18, 95% CI: 1.13, 1.23) section, but not the Math section, of the SAT.

The substantive impact of our analytic decision to use IPTW on our inferences is shown in Table 2. This table illustrates the estimates from 1) Complete case analysis

(that is, not using MICE) using standard generalized linear models (that is, not using IPTW), 2) Imputed data using standard linear models (that is, not using IPTW), and 3) Imputed data analyze using IPTW models. Across all three of these modeling approaches, debate participation was significantly associated with both GPA and SAT outcomes; the results of the IPTW show that the

relationship between debate and academic achievement was robust to differential self-selection based on observed sociodemographic characteristics and 8th grade (pre-debate) achievement as indicated by the STAAR standardized test performance.

DISCUSSION

Competitive academic debate programs exist in thousands of communities around the globe, including recent growth in urban school districts in the United States (International Debate Education Association, 2017). Prior research has described the benefits of debate participation for outcomes such as critical thinking skills (Green and Klug, 1990; Kennedy, 2007), as well as self-efficacy and various indicators of social/emotional development (Anderson and Mezuk, 2015; Fine, 2004; Kalesnikava et al., 2019), that are in turn correlated with school engagement (Bellon, 2000). The present study, which is one of the largest quantitative evaluations of debate participation and achievement among high school students conducted to date, extends this work by providing robust evidence of the benefits of debate on academic performance and college readiness. These findings are consistent with those of prior studies in Chicago (Mezuk, 2009; Mezuk et al., 2011), which found that debate participants were more likely to reach college-readiness benchmarks on the ACT college entrance exam; this study, which is the first to examine the relationship between debate participation and performance on the SAT college entrance exam, similarly found stronger effects on the Reading/Writing versus Math sections of the test. Findings are also consistent with research among middle school students in Baltimore (Shackelford, 2019), which found positive impacts of debate on school engagement and standardized test scores entering into high school. In sum, this study adds to the growing literature showing that debate participation is associated with improved academic outcomes for adolescents in large urban districts.

Findings should be interpreted considering study strengths and limitations. Consistent with prior work on debate, and extra-curricular activities in general, there was differential self-selection of students with stronger academic performance in middle school into this high school debate program (Hunt, 2005; Mezuk et al., 2011). While this study used propensity score weighting to account for this self-selection when estimating the relationship between debate participation and achievement, the validity of IPTW methods to mimic an experimental design requires strong, and generally untestable, assumptions about unmeasured confounders and measurement error. Therefore, while our approach reduces the bias that such threats to validity introduced to our inferences, we cannot exclude the possibility of residual confounding due to unmeasured factors (e.g., participation in other extra-curricular activities in high

school, parental/familial characteristics, non-cognitive skills such as grit (Heckman et al., 2006; Im et al., 2016; Shelly, 2011)). Strengths include the large sample with a diverse racial/ethnic study body, longitudinal design, and indicators of pre-debate achievement to minimize the bias introduced by self-selection into debate through IPTW methods.

The Hispanic/Latino population is the largest ethnic minority group in the United States, currently representing approximately 27% of K-12 public school students (US Department of Education, 2020). This is one of the first quantitative studies to examine the relationship between debate participation and academic outcomes in a predominantly Latino/Hispanic school district, and these findings are consistent with prior work examining co-curricular activities and school engagement among Latino/Hispanic students. For example, Diaz (2005) reported that Latino high school students who engaged in more extracurricular activities reported higher levels of school engagement, although this was a general phenomenon and not specific to any particular activity (Diaz, 2005). Similarly, LeCroy and Krysik (2008) reported that having a higher number of pro-academic peers were associated with both higher GPA and more school engagement among Latino middle school students (LeCroy and Krysik, 2008). As the number of Hispanic/Latino students grows, debate leagues have worked to ensure their programming is accessible to these students; for example, several leagues offer Spanish language debate competitions (e.g., leagues in Minnesota (Minnesota Urban Debate League, 2021) and New York (Zimmerman, 2019)).

In sum, the present study adds to the literature illustrating the role of time-intensive, academically-oriented extra-curricular activities like debate for supporting school achievement for students in urban districts (Mariana et al., 2006). It demonstrates the potential of large administrative data systems to support rigorous evaluations of the impact of such programs on student achievement at scale (Mezuk et al., 2011). When viewed in combination with the large body of qualitative and ethnographic work that has explored the various ways that competitive debate relates to adolescent development (Asad and Bell, 2014; Branham, 1995; Fine, 2004), these findings emphasize the salience of this activity for student engagement with learning both inside and outside the classroom (Louden, 2010).

CONFLICT OF INTERESTS

The authors have not declared any conflict of interests.

ACKNOWLEDGEMENTS

This work was supported by a grant from the CITI foundation, in collaboration with the National Association

for Urban Debate Leagues (NAUDL). The NAUDL had no role in the design, analysis, or interpretation of these findings, or in our decision to publish them.

REFERECES

- Anderson S, Mezuk B (2015). Positive Youth Development and Participation in an Urban Debate League: Results from Chicago Public Schools, 1997-2007. *The Journal of Negro Education* 44(3):362-378.
- Asad AL, Bell MC (2014). Winning to Learn, Learning to Win: Evaluative Frames and Practices in Urban Debate. *Qualitative Sociology* 37(1):1-26.
- Austin PC, Stuart EA (2015). Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Statistics in Medicine* 34(28):3661-3679.
- Banerjee PA (2016). A systematic review of factors linked to poor academic performance of disadvantaged students in science and maths in schools. *Cogent Education* 3(1):1178441.
- Bellon J (2000). A Research-Based Justification for Debate Across the Curriculum. *Argumentation and Advocacy* 36(3):161-175.
- Branham RJ (1995). "I Was Gone on Debating": Malcolm X's Prison Debates and Public Confrontations. *Argumentation and Advocacy* 31(3):117-137.
- Diaz JD (2005). School Attachment Among Latino Youth in Rural Minnesota. *Hispanic Journal of Behavioral Sciences* 27(3):300-318.
- Eccles JS, Barber BL (1999). Student Council, Volunteering, Basketball, or Marching Band: What Kind of Extracurricular Involvement Matters? *Journal of Adolescent Research* 14(1):10-43.
- Eccles JS, Barber BL, Stone M, Hunt J (2003). Extracurricular Activities and Adolescent Development. *Journal of Social Issues* 59(4):865-889.
- English-Speaking Union (2020). World Schools Debating Championships. ESU. <https://www.esu.org/competitions/wscd/>
- Fine GA (2004). Adolescence as Cultural Toolkit: High School Debate and the Repertoires of Childhood and Adulthood. *The Sociological Quarterly* 45(1):1-20.
- Gibbs BG, Erickson LD, Dufur MJ, Miles A (2015). Extracurricular associations and college enrollment. *Social Science Research* 50:367-381.
- Green CS, Klug HG (1990). Teaching Critical Thinking and Writing through Debates: An Experimental Evaluation. *Teaching Sociology* 18(4):462-471.
- Heckman JJ, Stixrud J, Urzua S (2006). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics* 24(3):411-482.
- Houston Independent School District (2021). General Information / Facts and Figures. <https://www.houstonisd.org/Page/41879>
- Hunt HD (2005). The Effect of Extracurricular Activities in the Educational Process: Influence on Academic Outcomes? *Sociological Spectrum* 25(4):417-445.
- Im MH, Hughes JN, Cao Q, Kwok O (2016). Effects of Extracurricular Participation During Middle School on Academic Motivation and Achievement at Grade 9. *American Educational Research Journal* 53(5):1343-1375.
- International Debate Education Association (2017). About IDEA. Retrieved from <https://idebate.org/about-idea-nw/>
- Kalesnikava VA, Ekey GP, Ko TM, Shackelford D, Mezuk B (2019). Grit, growth mindset and participation in competitive policy debate: Evidence from the Chicago Debate League. *Educational Research and Reviews* 14(10):358-371. <https://doi.org/10.5897/ERR2019-3707>
- Kennedy R (2007). In-Class Debates: Fertile Ground for Active Learning and the Cultivation of Critical Thinking and Oral Communication Skills. *International Journal of Teaching and Learning in Higher Education* 19(2):183-190.
- LeCroy CW, Krysik J (2008). Predictors of Academic Achievement and School Attachment among Hispanic Adolescents. *Children and Schools* 30(4):197-209. <https://doi.org/10.1093/cs/30.4.197>
- Leyrat C, Seaman SR, White IR, Douglas I, Smeeth L, Kim J, Resche-Rigon M, Carpenter JR, Williamson EJ (2019). Propensity score analysis with partially observed covariates: How should multiple imputation be used? *Statistical Methods in Medical Research* 28(1):3-19.
- Louden AD (2010). Navigating Opportunity: Policy Debate in the 21st Century. International Debate Education Association.
- Marsh HW, Kleitman S (2002). Extracurricular school activities: The good, the bad, and the nonlinear. *Harvard Educational Review* 72(4):464-514.
- Mezuk B (2009). Urban Debate and High School Educational Outcomes for African American Males: The Case of the Chicago Debate League. *The Journal of Negro Education* 78(3):290-304.
- 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(9):622-635.
- Minnesota Urban Debate League (2021). Spanish Debate—Community Topic. Minnesota Urban Debate League. <https://www.augsburg.edu/urbandebateleague/sdl/>
- Mitchell GR (1998). Pedagogical Possibilities for Argumentative Agency in Academic Debate. *Argumentation and Advocacy* 35(2):41-60.
- Moriana JA, Alós F, Alcalá R, Pino MJ, Herruzo J, Ruiz R (2006). Extracurricular activities and academic performance in secondary students. *Electronic Journal of Research in Educational Psychology* 4(8):12.
- Shackelford D (2019). The BUDL Effect: Examining Academic Achievement and Engagement Outcomes of Preadolescent Baltimore Urban Debate League Participants. *Educational Researcher* 48(3):145-157.
- Shelly B (2011). Bonding, Bridging, and Boundary Breaking: The Civic Lessons of High School Student Activities. *Journal of Political Science Education* 7(3):295-311.
- Texas Education Agency (2020). 2017–2018 STAAR Raw Score Conversion Tables. Texas Education Agency. <https://tea.texas.gov/texas-schools/accountability/academic-accountability/performance-reporting/2017-2018-staar-raw-score-conversion-tables>
- Texas Education Agency TE (2021, January 8). Testing. Texas Education Agency. <https://tea.texas.gov/student-assessment/testing>
- The College Board (2015). Compare SAT Specifications. SAT Suite of Assessments. <https://collegereadiness.collegeboard.org/sat/inside-the-test/compare-old-new-specifications>
- The College Board (2016). The College and Career Readiness Benchmarks for the SAT Suite of Assessments. SAT Suite of Assessments. <https://collegereadiness.collegeboard.org/about/scores/benchmarks>
- US Department of Education (2020). The Condition of Education—Preprimary, Elementary, and Secondary Education—Elementary and Secondary Enrollment—Racial/Ethnic Enrollment in Public Schools. https://nces-ed-gov.proxy.lib.umich.edu/programs/coe/indicator_cge.asp
- US Department of Education (2020). Bar Chart Races: Changing Demographics in K–12 Public School Enrollment | White House Hispanic Prosperity Initiative. <https://sites.ed.gov/hispanic-initiative/2020/07/bar-chart-races-changing-demographics-in-k-12-public-school-enrollment/>
- van Buuren S, Groothuis-Oudshoorn K (2011). Mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software* 45(1):1-67.
- Zimmerman A (2019). This bilingual debate team in NYC is fighting 'English-only norms' at national competitions. Chalkbeat New York. <https://ny.chalkbeat.org/2019/9/19/21121785/this-bilingual-debate-team-in-nyc-is-fighting-english-only-norms-at-national-competitions>