What's in a Game?

Examining Academic Achievement and Engagement Outcomes of Baltimore Urban Debate League Participants

by

Daniel Shackelford

A dissertation submitted to Johns Hopkins University

in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

March, 2019

ABSTRACT

The following dissertation research contributes to the understanding of whether participation in extracurricular debate influences the academic trajectories of urban youth and is organized as three separate but related journal articles. It reviews the research around Urban Debate Leagues, as well as the literature on academic games in general, and proposes a theoretical argument for their importance in student learning. Despite major advances in the research justification for Urban Debate Leagues, a population largely absent from analysis thus far is elementary and middle school students. This dissertation research fills this gap in the literature by examining observational data from a diverse school district to understand how preadolescent participation in an Urban Debate League is related to a variety of student outcomes.

Among a 10-year longitudinal sample of Baltimore City Public School System students, results show that the effect of preadolescent Baltimore Urban Debate League participation for debaters is 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. These results are based on models that mitigate observed selection bias by using inverse probability of treatment weighting to create statistically matched samples for comparison. Sensitivity analysis techniques are also utilized to determine the magnitude of hidden bias that would change inferences about the treatment effects of debate participation on high school outcomes both for preadolescent debaters (those that participate in grades 4-8) as well as for adolescent debaters (those that participate in grades 9-12). By taking these methodological approaches, the analysis in this dissertation advances the empirical evidence for extracurricular

debate and indicates that particular students may be perfectly positioned to respond positively to its interactive structure and content.

To the author's knowledge, this paper constitutes the first quantitative study on debate for elementary and middle school students, and policy implications for educational interventions that seek to attract low-income youth of color in urban areas and influence their trajectories at earlier stages of student development are discussed.

COMMITTEE MEMBERS

Advisor and Primary Reader: Stephen Morgan

Committee Member and Secondary Reader: Faith Connolly

Committee Member and Ph.D. Program Director: Marc Stein

Head External Committee Member: Lingxin Hao

External Committee Member: Robert Slavin

To My Detroit Debaters

ACKNOWLEDGMENTS

This dissertation research would not have been possible without the support from the Baltimore Education Research Consortium. Through their partnership with the Baltimore City Public School System, they patiently worked with me in gaining access to the administrative data necessary for this project as well as provided me with space within their offices to use throughout analysis. I would also like to acknowledge the National Association for Urban Debate Leagues and the Baltimore Urban Debate League for the data on debate participation. I am especially indebted to Executive Director Coleen Reyes, who graciously accommodated any programmatic requests without reservation, and Dr. Briana Mezuk, whose pioneering studies on debate inspired me and paved the way for this dissertation research.

Additionally, I would like to thank each of the members of my dissertation committee for providing me with extensive personal and professional guidance throughout my time at Johns Hopkins University. In particular, I would like to express my deep appreciation and gratitude to Dr. Marc Stein, who encouraged me in my first semester to pursue and carryout this research, and Dr. Faith Connolly, who has overseen each stage of this project and has been nothing but supportive throughout its development and completion. Last but not least, I would like to thank my advisor Dr. Stephen Morgan, who has been an outstanding mentor and friend to me. He has taught me more than I could ever give him credit for here, and it has truly been an honor to study under such a brilliant and compassionate scholar.

I would also like to thank my parents, Jess and Bernie Shackelford, and the rest of Team Shackelford for being the best family anyone could ask for. This includes my own former debate coaches, Carol and Mike Shackelford, whom I will be forever grateful for introducing me to the game that changed my life. Finally, I would like to thank my boyfriend Mark Dow for his love and support throughout my Ph.D. program. Thank you for everything.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Must policymakers and education 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? The following dissertation research contributes to the understanding of whether participation in extracurricular debate influences the academic trajectories of urban youth. It is organized as three separate but related journal articles.

The first paper reviews the research around Urban Debate Leagues, as well as the literature on academic games in general, and proposes a theoretical argument for their importance in student learning. While most reform efforts in education focus on curriculum in disciplinary classes, academic games such as debate can serve as a constructive intervention for students who are currently not well served by the existing school system. The paper documents how aspects of traditional education are highly inconsistent with what is known about student motivation and adolescent development, and argues that academic games can change the norms and values of students to better encourage education. Instead of simply focusing on the content taught in schools, this paper suggests that how it gets taught is equally important.

Specifically, this theoretical review unpacks the nature of extracurricular debate and contends that its structure is consistent with what prior experimental studies suggest make for authentic learning. The aims of the paper are threefold. First, in response to the growing rates of participation in Urban Debate Leagues, this chapter seeks to situate these programs into the larger discussion on education reform in this country. Second, because games as means by which children learn are often an overlooked resource for educators, the paper examines the theory around why academic games are not only more fun and engaging for students than traditional classroom practices, but also more in line with what psychologists know provides for a

consistent platform of student learning. Finally, the paper highlights a key characteristic of academic games that has been linked to achievement outcomes, the combination of intrateam cooperation paired with interteam competition, by examining the research studies on the Teams-Games-Tournaments learning technique. The paper concludes that Urban Debate Leagues, through the features they share with this technique and other academic games, may provide a compelling solution to address the often-cited shortcomings of urban schools for some students.

Indeed, the research evaluating student outcomes among debate participants shows promising results. But despite major advances in the research justification for Urban Debate Leagues, 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. In order to truly assess the impact participating in extracurricular debate has on student outcomes, Urban Debate Leagues should be studied throughout the stages of student development, which includes the crucial period of the elementary and middle school years. The second paper of this dissertation fills this gap in the literature by examining observational data from a diverse school district to understand how preadolescent participation in an Urban Debate League is related to a variety of student outcomes.

Among a 10-year longitudinal sample of Baltimore City Public School System students, the paper finds that the effect of preadolescent Baltimore Urban Debate League participation for debaters is 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. These results are based on models that mitigate observed selection bias 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. Analysis also includes 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 by providing additional protection against model misspecification and any imbalance that remains after applying weights derived from the propensity scores. Finally, 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 average effect for the treatment group as the target parameter of interest. By taking these methodological approaches, the analysis in this paper advances the empirical evidence for extracurricular debate and indicates that particular elementary and middle school students may be perfectly positioned to respond positively to its interactive structure and content. To the author's knowledge, this paper constitutes the first quantitative study on debate for this age group, and 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.

Of course, without a randomized control trial, it is impossible to fully account for selection into debate. An important approach in evaluating evidence for causation in observational research is sensitivity analysis, which can provide investigators a method to assess how robust estimates of a causal effect are to potential unobservable treatment selection patterns. Thus, for the final paper, sensitivity analysis techniques are utilized to determine the magnitude of hidden bias that would change inferences about the treatment effects of debate participation. It

expands upon the prior study by examining previously unexplored outcomes of high school attendance and SAT scores both for preadolescent debaters (those that participate in grades 4-8) as well as for adolescent debaters (those that participate in grades 9-12) to see whether treatment effects of participation persist and remain insensitive to unobserved confounding. It proceeds by offering estimates of treatment effects on the treated assuming that treatment selection ignorability holds for a set of observed covariates that include measures of achievement and engagement assessed prior to participation. Then, these estimates are analyzed to see how sensitive interpretations are to potential unmeasured selection effects.

The particular methodology of the study is noteworthy in many respects. First, this analysis provides a variety of estimates, summarized in a contour plot, that includes which pairs of parameters result in non-significant or null treatment effects. Second, the approach assesses how potential sensitivity parameter compare to observed measured covariates. Finally, the technique incorporates treatment weights to target estimands such as the average treatment effect on the treated. Results show that preadolescent and adolescent debate participation is associated with increases in all high school outcomes and that positing an unmeasured confounder as predictive as the strongest observed covariate would not substantively change the interpretations of estimates of debate participation on SAT verbal reasoning scores and high school attendance rate. However, the treatment effects for SAT mathematics scores are not as resistant to potential unmeasured confounding.

Taken as a whole, this dissertation research has many implications for policymakers and educational practitioners. First, to the extent that they can be compared, estimates of preadolescent debate participation are larger than the effects found for adolescent participation in high school regardless of the outcome of interest. This finding is line with research that shows

how educational programs and interventions that target a younger population show greater gains over time and it potentially suggests a shift in the way schools provide access to participation for students at younger ages, as most extracurricular activities are offered primarily during a student's high school years. Second, a large number of low-income urban youth do not participate in any extracurricular activities, and minority students have been understudied in the extracurricular activity literature as a whole. Accordingly, research such as this is critical to the ongoing local and national policy debates about the impact of academic extracurricular activities, especially for urban and preadolescent students, two groups where the opportunity to participate is limited. Additionally, these studies find that debaters are more likely to be black and qualify for free or reduced-price meals, after adjusting for all other observed covariates, which suggests that Urban Debate Leagues may be a culturally appropriate intervention for a population that typically has lower levels of educational attainment. Thus, Urban Debate Leagues have the potential to not only attract marginalized students, but influence their academic achievement and engagement outcomes as well, an objective many educational interventions likely share.

Ensuring that all students acquire the requisite skills to succeed in life is an urgent goal that must be addressed, and improving urban education requires more than traditional reforms to close achievement gaps. As findings from this dissertation suggest, Urban Debate Leagues may be a promising means to improve educational trajectories for some of the country's most needy students.

I. WHAT'S IN A GAME? AN ARGUMENT FOR URBAN DEBATE LEAGUES

Introduction

Pronounced disparities across urbanicity, income, and race remain in achievement outcomes throughout the United States. 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). Education in urban areas is often further characterized by student discipline problems, poor health in children, and limited access to supplemental learning activities (Thompson, Ransdell, & Rousseau, 2005). As education gaps both within and between schools continue to grow, classrooms across the country are increasingly filled with students of varying skill levels. A problem created by such disparities is challenging the advanced student while keeping the disadvantaged student from falling further behind.

Curriculum reforms are increasingly touted as the answer to these troubling statistics. Large amounts of money have been invested in solutions that have emphasized offering all students the same content, such as the Common Core curriculum, in order to ensure consistent and universal standards (Sleeter, 2012; Brooks & Dietz, 2013). Furthermore, a rise in Direct Instruction has been seen as a way to instill this content in a streamlined way (Stockard, 2010; Beatty, 2013). Despite these efforts, achievement gaps persist. A look at the most recent NAEP scores, better known as "The Nation's Report Card," suggests that the reforms over the past 15 years under the No Child Left Behind era had little to no impact on achievement as income based gaps in fourth and eight grade data remain relatively unchanged (US Department of Education, 2017). Thus, problems in academic performance may not stem from what we are teaching

students, but how. For the content in a curriculum is of little importance if students are not motivated to learn, and it seems as though many students today are coming to school placing a relatively low value on academic achievement while instead focusing on peer acceptance and socialization.

This phenomenon, however, is nothing new. In his 1961 book *The Adolescent Society*, James Coleman documents how aspects of traditional education are highly inconsistent with what is known about student motivation and adolescent development. For instance, he explains that the interpersonal structure of rewards, with its heavy emphasis on testing and grades, forces students to view fellow classmates as competitors rather than teammates, and produces a response on the part of an adolescent social system that effectively impedes the process of education. While what he characterizes as "interpersonal competition in scholastic matters" between students generates social pressure not to engage and excel, "interscholastic competition" between schools has quite the opposite effect. In his critique, Coleman proposes a novel idea: schools should replace the individualized competition for grades with interscholastic academic games consisting of systematically organized tournaments. With this shift in emphasis, the perplexing factors that face many classrooms – increasing diversity of student skill levels and values placed on peer socialization – become assets.

Coleman knew it would take some "considerable inventiveness" to find the best forms of group competition to change the cultural norms of schools, and he suggested debate teams as a possible solution. Nearly five decades later, recent research on the National Association of Urban Debate Leagues, a program that aims to provide access to competitive debate to low-income urban school districts, may provide evidence of an interscholastic game that changes student attitudes toward learning and influences student achievement. The following reviews both the

research around Urban Debate Leagues, as well as the literature on academic games in general, and proposes a theoretical argument for their importance in student learning.

Specifically, the aims of this paper are threefold. First, in response to the growing rates of participation in Urban Debate Leagues, this paper seeks to situate these programs into the larger discussion on education reform in this country. Second, because games as means by which children learn are often an overlooked resource for educators, this paper examines the theory around why academic games are not only more fun and engaging for students than traditional classroom practices, but also more in line with what psychologists know provides for a consistent platform of student learning. Finally, this paper highlights a key characteristic of academic games that has been linked to achievement outcomes, the combination of intrateam cooperation paired with interteam competition, by examining the research studies on the Teams-Games-Tournaments learning technique. While the gold standard of randomized control trials are increasingly difficult to come by in educational research, these experimental findings offer informative answers to questions around the generalizability of treatment effects of Urban Debate Leagues and also help one determine what's in a game and why is it needed in education today.

The Rise of Urban Debate Leagues

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. Most tournaments utilize a "power-matched" structure where teams with identical win-loss records pair off against each other. 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, the activity 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 one of the only comprehensive ethnographic studies on debate, Fine (2001) concludes that it "is a valuable training ground for adolescents" and that "our educational system would be more successful in its goal of producing competent citizens if all, or many, students had the opportunity to participate in this activity." (p.18). Unfortunately, given the state of many large urban school systems in the United States, the cost of providing debate programs is often prohibitive. In light of this fact, the first Urban Debate League program began in 1985 as a partnership between the Atlanta Public School system and Emory University in order to foster educational improvement and expand the benefits of debate to underserved populations of impoverished minorities (Winkler, 2011).

In 1997, the concept spread to New York and the 'Urban Debate Program' was founded from funding from George Soros's Open Society Institute. He believes that debate "provides urban youth with the skills they need to actively participate as citizens in an open society, so that their voices are heard and their opinions are considered in public discourse, both in their communities and beyond" (Soros Foundation, 2003). National in scope, the program sought to extend seed grants to university debate programs that in turn would provide outreach to high

schools in their immediate locale. Following the establishment of the initial leagues in Atlanta and New York, another 12 leagues were launched between 1997 and 2003 in Baltimore, Chicago, Detroit, Kansas City, Los Angeles, Newark, Providence, the San Francisco Bay Area, St. Louis, Seattle, Tuscaloosa, and Washington, D.C. (Breger, 2000).

Since then, Urban Debate Leagues have been established on the premise that they improve outcomes for underserved students in depressed urban areas. National tournament events and year-end awards banquets have been incorporated to engage students to be part of their local league, and some leagues have even expanded to include elementary and middle school divisions. Today, the National Association of Urban Debate Leagues 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).

Urban Debate Leagues have been positioned as an innovative approach among the multitude of efforts aimed at improving education, and the research evaluating student outcomes among the organization's participants provides promising results. A study of participants in Chicago shows that high school students who debate have higher 12th grade 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 than non-debaters (Shackelford, Ratliff, & Mezuk, 2018). Positive effects have also been found for elementary and middle school participation during preadolescence. Among a 10-year sample of Baltimore City Public School

System students, the effect of preadolescent Baltimore Urban Debate League participation for debaters is associated with increases in standardized test scores, a decreased likelihood of chronic absenteeism, and an increased likelihood of attending an entrance criteria high school (Shackelford, 2019).

These quantitative findings are in line with years of qualitative research that illustrates the positive impact this activity has for students. After interviewing urban students in Baltimore, DeLuca (2016) explains that debate can serve as an "identity project" for students that inspire the grit necessary to accomplish a variety of goals. Winkler's (2011) study on the Milwaukee and Atlanta Urban Debate Leagues utilizes student voice to further provide a glimpse into the theory of change behind participation. When asked to explain why debate supports school engagement, one participant noted: "Since joining debate, I am more interested in going to college… Debate makes me believe I could succeed in life." (p. 86).

Other interpretations suggest that extracurricular debate participation can have a positive effect on academic performance indirectly as a result of non-cognitive and social benefits. For instance, Zaff and colleagues (2003) argue that extracurricular activities provide students with a safe space during the high-risk after-school hours, and that the adult supervision inherent in these activities provides support needed for positive youth development. Kaufman and Gabler (2004) propose a cultural capital theory whereby students informally share desirable attributes related to school success (e.g., knowledge about college admissions processes, social norms regarding education). Another hypothesized pathway stems from students simply being exposed to academically orientated and engaged peers for prolonged periods of time (Eccles et al., 2003; Gibbs et al., 2015). In general, these studies posit that benefits of activity participation are conjectured to stem from elements inherent to all extracurricular programs.

But unlike most extracurricular activities, Urban Debate Leagues offer a set of skills that closely relate to many scholastic goals such as public speaking, reading, and critical thinking. For example, the first writing standard for grades 9 and 10 states "students should be able to write arguments and support claims in an analysis of substantive topics or texts using valid reasoning and relevant 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).

In debate, students write and develop their cases by reviewing the relevant evidence for a topic, synthesize that information into persuasive arguments, organize those arguments into logical sequences, and weigh their arguments against those made by the opposing team. These skills are in line with critical thinking proficiencies assessed in students, often characterized by education scholars' as the ability to assess claims and make objective judgments on the basis of well-supported reasons (Wade & Tarvis, 1987). As opposed to simply reiterating facts or syntax, debate utilizes the entirety of Bloom's taxonomy for cognitive learning: comprehension, application, analysis, synthesis, and evaluation. These domains encapsulate both the writing and the critical thinking process (Olson, 1984), and to the extent that debate may improve a student's critical thinking ability, it may in turn improve their writing skills in discernible ways. Accordingly, it is plausible that Urban Debate Leagues influence students' academic achievement more so than extracurricular activities in general. While debate is in fact a game, it is distinctive insofar as it reinforces the same academic writing and language skills that are the focus of school curricula. For this reason, a review of the literature on academic games and student learning is warranted.

Academic Games and Student Learning

Games as means by which children learn have received far less attention than they deserve. While it's true that games are often used by teachers of early grades as a way for children to socialize, they are typically seen as secondary to the primary goal of education, and are later abandoned in favor of more "serious" approaches to learning. But what's in a game and how are its components related to learning at all stages of student development?

In 1968, Coleman defined the essential properties of a game as follows:

(1) Its basic elements are players or actors, each striving to achieve a goal.
(2) It is limited to a small, fixed set of players.
(3) Its rules limit the range and define the nature of legitimate actions of the players
(4) Through the rules, it establishes the basic order, sequence, and structure within which the actions take place
(5) It is delimited in time as well as extensivity, with an end defined by the rules
(6) Its rules constitute a temporary suspension of some of the ordinary activities of life and rules of behavior by substituting for them these special time-and-space delimited ones (p. 62-63).

These traits outline at least two reasons why games should be of interest to education practitioners. The first is their requirement for action rather than merely passive observation.

As summarized above, a game partitions off a set of players, a set of allowable actions, and establishes a framework within which the action takes place. In this way, games function as a simulation of sorts in which children practice with components of life itself and must interpret appropriate rules and norms to act accordingly. Child psychologist Jean Piaget (1951) was one of the first to document this relationship after perceptive observations of children playing a game of marbles. He argued that the learning of the rules in a game is analogous to the learning of the nature of a social order. Piaget believed that all learning emerges from action and that individuals construct and reconstruct their knowledge of the world as a result of interactions with their environment. He called this a "joy of being of the cause" (p.149). As all people have undoubtedly felt, there is an inherent joy in any successful action. Unfortunately, traditional

schools seldom provide adolescents the authority or responsibility for active participation. In fact, students crave interaction and are often bored when required to play a passive role. This shift, from passive to active, summarizes the primary reason that academic games should be utilized in education.

Coleman echoes this sentiment and equates games to "plays within the larger play of life" with a structure of action that is learned and constitutes the framework into which information is fitted, just like in reality: "The player, as a consequence, has a natural screening device for information as well as a natural basis for choosing what information to seek out" (Coleman, 1968, pg. 69). In other words, participants seek only the information they need to play the game. For this reason, the learning that takes place in a game is intrinsically motivated. Students do not learn information in a game because a teacher says it is important or because they want to receive a good grade. Rather, they learn its content, rules, and structure because it is what they will need for action within the game.

In comparing these distinct motivations for learning, one begins to see the relatively counterintuitive features of the traditional school model: one that is increasingly characterized, at least in many urban schools, by Direct Instruction and rarely provides students opportunities for active participation. For example, the lectures that typify the traditional schooling experience contain few of the life-like properties that are inherent in academic games. Instead, they are merely a manifestation of the simplistic belief that information is transmitted by communication and repetition aids learning. Freire (1993) describes this type of education as suffering from a narration sickness that creates a hierarchy between teacher and student. For him, authentic education is not "carried out by A for B… but by A with B" (p.82), and he critiqued any model

of teaching that viewed students as nothing more than containers to be filled by teachers depositing content. Freire wanted students to not simply reproduce words, but create their own.

This definition of authentic education supports learning through an activity that is student-centered like debate, which reverses the narration pattern of traditional education. Advocates of the Urban Debate League movement contend that the focal point of education is recast as students become the primary speakers while teachers and other adults become the listeners. Debaters often energetically assume responsibility for their own learning by conducting original research and crafting their own materials for debate (Wade, 1999). Additionally, studies consistently show that critical thinking ability, measured by a variety of assessments, significantly improves through competitive forensics participation (Allen, Berkowitz, Hunt, & Louden, 1999; Keefe, Hart, & Norton, 1982). This increase is theorized to be the result of debate's dialectal nature (Warner & Bruschke, 2001). For example, the habit of questioning the claims of others as well as thinking through the possible objections of one's own claims allows debaters to develop the mental faculties needed to become active consumers of information instead of passive receptors, and recent research finds that participation in the activity is positively associated with gains in growth mindset among urban high school students (Kalesnikava, Ekey, Ko, Shackelford, & Mezuk, 2019). Moreover, Mitchell contends that, "an essential part of the debate process involves citizens empowering themselves to invent, clarify, and amplify their viewpoints in public forums" (Mitchell, 1998, p. 50). In this manner, debate inherently serves to empower the student.

Mehta (2017) reached similar conclusions when he examined participation in debate teams and argues that, in many high schools, there is deeper learning happening in peripheral activities than in core disciplinary classes. He too describes these activities as opportunities to

"play the whole game at the junior level" where apprenticeship-type learning provides students with a clear purpose and a chance to connect their identities to real-world domains of professional practice. Debate requires that students role-play as policymakers and weigh the pros of cons of issues ranging from alternative energy incentives to human rights in China. Its simulated structure of action is designed to mirror, as much as possible, the motives and interests of a real person in such a situation. In this way, debate serves as a good vehicle for learning academic content.

Debate may be an effective educational tool for younger students as well, as research suggests that elementary and middle school students may be perfectly positioned to respond positively to its interactive structure and content. Psychologists identify the preadolescent period as critical to the development of industry competency, identity, and autonomy (Erikson, 1968). During this stage, students learn to read and write language, get along with peers in school, develop empathy and acceptance of others, and begin to understand the concept of self-efficacy (O'Connell, Boat, & Warner, 2011). Vygotsky's zone of proximal development (1981) explains that children can imitate a variety of actions that go well beyond the limits of their own capabilities and it suggests that learning is most effective when students interact and cooperate with peers in a shared environment. This line of reasoning extends to recent work on "activity settings theory" which posits that an individual's cognitive, emotional, and behavioral development are influenced by shared activities with others (O'Donnell & Tharp, 2012). Thus, through practice and collaboration with coaches and fellow teammates, debate is an activity that capitalizes on a preadolescent student's growing sense of autonomy as well as their need for connection with peers. It is undoubtedly because games are constructed with these necessities of life that they play such an important part in the learning and socialization of children.

Of course, the goal of any game is not learning itself. Aside from the intrinsic "joy of being the cause," many are motivated by extrinsic rewards. Indeed, children ultimately learn the skills and content required by any given game because it is this information that allows them to win. Thus, the second reason academic games deserve the attention of education practitioners and policymakers is because the inherent competitive rewards they hold for mastery are the same kinds of motivations that shape many student behaviors.

In debate, wins and losses are recorded after every round and the best teams at each tournament receive trophies. This competition provides a built-in incentive for participation in the activity. For example, when citing their decision to remain engaged in debate, approximately 25% of students sampled in the Chicago Debate League said, "I enjoy the competition" (Kalesnikava et al., 2019). Scholars have long documented how this integrated competition promotes achievement as it's "competitive nature motivates students with little pre-existing interest in education to research, read, and test ideas because they want to win" (Collier, 2004, p. 26). In other words, the competitive structure of tournaments offers incentives for using literacy skills. Warner and Bersche (2001) note how many underachieving students, once "gone on debating" find themselves wanting to learn and develop their skills because debate provides them an opportunity to showcase their newfound knowledge and demonstrate mastery.

Coleman describes this process as a kind of "learning to be motivated" (1968, p. 64). In academic games, students are not competing against an abstract measure of success; rather, they are motivated to act in a given direction to achieve a clearly defined goal. In the traditional classroom however, teachers seem to operate under the assumption that a child is already motivated to learn academic content, and education is often conceptualized as an activity that occurs after students have been motivated. Indeed, when students fail to learn their lessons, it is

often said that they are not motivated to learn and subsequently cannot be taught. Thus, the kind of learning that occurs in a game is not only logically prior to the kind of learning that occurs in the standard information-transmission model of traditional classrooms, but it is complimentary to what we know is consistent with adolescent development.

This perspective implies a number of points about the use of academic games in schools. First, it suggests that appropriate games for learning are those in which winning, or attainment of the goal, is facilitated by the content knowledge or skills that schools aim to teach. Games with goals unrelated to content will not in themselves make the child motivated to learn mathematics or history. Second, the competitive aspect that is inherent to the activity may attract students who may not otherwise be disposed to educationally related undertakings to engage in school. Thus, while the effect of academic games on student outcomes for children who are already highly motivated in school should be subtle, a more direct and powerful impact may be seen for students often described as "unmotivated."

The quantitative research on Urban Debate Leagues offers support for this view. For instance, debate participation has been found to be associated with higher ACT Reading scores among Black males, but not among a sample of students who debated overall (Mezuk, 2009). This finding provides evidence of a constructive intervention for a student population whose literacy scores often otherwise remain stagnant, and the author concluded that this subgroup of students might particularly benefit from the activity. In follow-up analysis, the effect that debate participation had on the likelihood of graduating high school was marginally stronger for black and Hispanic students relative to their white peers (Mezuk et al., 2011). These findings, paired with the theory that youth who are at greater risk for poor academic outcomes benefit most from participation in academic games, make the argument for debate all the more persuasive.

Although academic games may vary widely in their particulars, they share key features that offer a platform for depth. They not only provide the goal for which the content is relevant, but the very nature of games guarantees that the player will be motivated to act toward that goal. Viewed in this light, it becomes much more evident why investments in curriculum seem to have little effect on what a child learns. For if the essential task of education is to teach students how to be motivated, then we must ensure that this carried out prior to, and concurrent with, any reform in content.

Experimental Studies on the Teams-Games-Tournaments Learning Technique

While it seems that there are particular aspects of Urban Debate Leagues that facilitate learning, it is difficult to assess the extent of potential causal claims. The foremost threat to internal validity for research on any extracurricular activity stems from the voluntary nature of participation of these programs. 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, & Seole, 2007). Consequently, it is tough to untangle causal relationships from selection effects between voluntary extracurricular activity participation and student outcomes. Although prior research on debate has adjusted for observed pre-existing characteristics through a variety of propensity score matching techniques (Mezuk et al., 2011; Shackelford, 2019), it is impossible to fully account for selection into a program without a randomized control trial in which students are assigned to participate in debate or not. Fortunately, a sizeable body of experimental research has been generated to evaluate student motivation and achievement outcomes among academic game structures that share many similarities to debate. In particular, the methodological approach taken by investigators of the Teams-Games-Tournaments (TGT) cooperative learning technique

may provide informative answers to questions around the internal and external validity of treatment effects for programs like Urban Debate Leagues.

TGT is a cooperative intervention that creates and structures active competition among individuals from different groups. It was the first of several Johns Hopkins learning methods developed at the Center for Social Organization of Schools and can be characterized as an actualization of Coleman's belief that team competition should be used in schools to give education the kind of appeal that sports and other peer-supported activities enjoy. As its name suggests, the technique's structure consists of three basic elements: (1) teams of students organized by equivalent academic levels, (2) games of skill relating to content material, and (3) tournaments where students represent their teams and compete against students from other teams. Each element involves a unique transformation of traditional classroom learning practices that may prove beneficial to student achievement. The team component is expected to create an interdependence among students that has the potential to increase the importance of, and support for, doing well academically. The gaming task provides an authentic and highly engaging performance setting that encourages active participation, while the tournament organization fortifies the competitive incentive structure with rewards more concrete than abstract grades. Furthermore, in the TGT configuration, teams compete against others of similar past performance that allows both low-achieving and high-achieving students the chance to succeed. The research on TGT has focused on answering whether the technique, when compared to more traditional instructional approaches, results in higher levels of student motivation and achievement, and if so, which components are responsible.

The initial set of randomized control trials on the TGT technique implemented its structure in a variety of subject areas and grade levels 3-12. For instance, investigators first

examined the technique with the academic math game *EQUATIONS* (Edwards, DeVries, & Snyder, 1972). In a sample of 7th grade students, two classes (one of "low" ability and "average" ability) were assigned to the TGT treatment while two classes were assigned to a business-as-usual control condition. In the experimental condition, students played the academic game *EQUATIONS* twice a week with feedback contingent upon how well teams performed in the game competition. In the control condition, students received feedback based on their individual performance in drills and quizzes. Results indicate that that experimental treatment significantly increased students' mathematics achievement on the computational SAT subtest over that of a traditionally taught class. Additionally, investigators observed that posttest scores were more dependent on ability level for the control classes than for the experimental classes, and concluded that the TGT treatment tended to reduce differential learning rates among classes of different ability levels.

While the design of the study did not allow investigators to determine which aspects of the TGT design resulted in significantly greater learning, they did observe several key phenomena in the experimental classes. First the game's competition succeeded in "turning on" students who had not been putting forth any effort and may have provided students without prior interest in mathematics a reason to learn the content. Second, students had an incentive to help fellow team members to improve their team's performance, and an increased level of peer tutoring was observed. Thus, the incorporation of a TGT technique into the course curriculum may have caused students to translate their increased interaction due to the game's structure into peer tutoring on the subject material at hand.

To test this theory, experimenters explicitly examined whether the TGT technique improved classroom social processes in follow-up studies on 3rd grade classrooms (DeVries &

Edwards, 1973; DeVries, Mescon, & Shackman, 1975). Like the prior study, students were randomly assigned to a treatment group, which participated in organized tournaments around instructional games, or a control group, which consisted of daily teacher lectures and worksheets. A student self-report survey was administered that included questions such as "which students have helped you with your language arts work?" and from this this questionnaire, measures of cohesion and social isolation were developed (the later, for example, was indicated by individuals who were not listed as a friend or helped by any more than one of their classmates). Results found significant positive TGT effects on both measures of classroom social processes with the treatment group reporting higher levels of cohesion among students and lower numbers of social isolates in the classroom. These findings demonstrate that the TGT technique can work effectively with young children despite its complex use of both cooperative and competitive social structures.

In order to determine the relative contribution of each of these components, investigators developed their next set of studies as two-tiered experiments where students were randomly assigned to either an individual or group practice session as well as either an individual or group competitive reward system (Slavin, DeVries, & Hulten, 1975; Hulten & DeVries, 1976). Students in the individual practice sessions were not permitted to help each other with classroom assignments whereas students in the group practice condition were encouraged by the teacher to work together on classwork. Similarly, in the individual competition reward structure, students competed in academic games by themselves, while students in the group competition reward structure played these games as representatives of a team. Because students in the group practice/group competition condition were able to work with their fellow teammates, the group

competition reward system provided an extra incentive for intrateam tutoring. However, no such incentive existed for the group practice/individual competition condition.

In addition to a standardized test of academic achievement, dependent variables included four student attitude scales: (1) perceived probability of game success, (2) incentive value of game success, (3) peer pressure for game success, and (4) game satisfaction. These studies found that the relationship a student had with competition success and achievement was stronger when competition was structured at the group rather than the individual level. For instance, Hulten and DeVries (1976) indicate that students in the group competition reward structure improved significantly more on the SAT than students in the individual competition group. Analysis of the attitude scales provides some explanation of this team effect as group competition students attached more importance to game success and reported higher levels of peer group interest as well as peer pressure to do well at the game than did students who participated in the competition individually. These findings suggest that team rewards lead students to develop norms that stress success of the academic task and that this individual performance is actually reinforced by peers. However, group practice outcomes did not differ significantly in their academic performance or attitudes towards success regardless of competition reward structure. Consequently, the results of these studies show that a group competition reward system is a necessary component in TGT learning strategies and seems to be one way the technique redirects student values to support instead of oppose achievement in the classroom.

The finding that group rewards are necessary in order for cooperative learning structures to have a significant impact on achievement has been corroborated in the extant research on cooperative learning techniques. In a meta-analysis of 122 studies, researchers found that cooperation with intergroup competition is more effective at promoting student achievement than

both interpersonal competition and individualist efforts, and that the one variable that seems to be most effective in the cooperative design is the reward variable (Johnson, Marayama, Johnson, Nelson, & Skon, 1981). Slavin's (1983) research supports this analysis and concludes that only cooperative learning methods that provide group rewards consistently increase student achievement more than controlled methods. Thus, while the specific Teams-Games-Tournaments moniker may not be necessary, it stresses important factors that have been shown to increase academic achievement. Mainly, it is the combination of both intrateam cooperation paired with interteam competition that makes the TGT technique, and methods like it, such an innovative and exciting approach to student learning.

TGT should not be considered a major curriculum revision; rather, it takes traditional curriculum materials and places them in different reward (teams instead of individuals) and task structures (games and tournaments instead of lectures and worksheets). It was developed to address the problems that teachers face in motivating students who place a higher value on peer socialization than academics no matter the content. Because most high achieving students are able to learn with the individual reward structure in the conventional classroom environment, it is not surprising that TGT techniques have been shown to have more profound effects on lower-performing students. The data suggests that these students who participate in TGT structures perceive a more drastic change in the importance they attach to success as well as in the relationship they have with content material. For example, in a recent study, TGT was found to be more effective than the control condition in promoting positive math attitudes (Ke & Grabowksi, 2007). Such a change in norms is an important precursor to increased educational outcomes. To be sure, while engagement and content are both important to students, too often do policymakers and practitioners focus on the later without first addressing the former.

Since Coleman suggested that schools mobilize peer support for effective academic performance through the use of team competition around academic tasks, countless TGT studies have demonstrated that efforts that incorporate cooperation with intergroup competition result in higher achievement. These studies meet strong methodological criteria that include random assignment of students to treatment and frequent rotation of participating teachers across experimental conditions. The generalizability of results have been consistently replicated in a wide variety of settings, subject areas, and grades, and include recent positive findings from experimental studies around the globe (Wyk, 2011; Salam, Hossain, & Rahman, 2015; Annurwanda, 2018). Because the effects of TGT persist across a variety of games, identifying the internal structure of the technique may prove beneficial for educational interventions that seek to engage students. Instead of simply focusing on the content taught in schools, this research suggests that how it gets taught may be equally important for student outcomes. Debate is one such activity that utilizes a cooperative incentive structure in which students are rewarded based on their performance as a team, and to such a degree, Urban Debate Leagues can be classified as a TGT cooperative learning technique. Considering debate's unique academic attributes paired with the aforementioned benefits from the structure of the TGT learning technique, the argument for its theorized influence on student learning is hard to deny.

Conclusion

In the United States of America, the growing income achievement gap remains one of the most significant issues in education. Without a doubt, substantial resources have been devoted to closing the gap, and yet it remains pervasive and persistent (Reardon, 2011). Increasingly, these efforts are driven by curriculum reforms and high-stakes testing (Au, 2011; Korhnhaber &

Orfield, 2011). But perhaps these more traditional approaches to reform are based on a faulty notion of how children can and should learn. After nearly five decades of attempts by educators and others to reduce the achievement gap with only limited success, it may be time for a more nuanced view. As Coleman showed long ago, the problem most schools face is that learning and achievement, regardless of content, are not what matter most to adolescents. Particularly in the preteen and teen years, students often value interacting with each other more than anything else, and research suggests that as children get older, their interests and attitudes toward school in general tend to decline (Anderman & Maehr, 1994). Recent proposed solutions in school reform often overlook these facts and consequently omit the most crucial step in learning: the necessary condition of student motivation.

This paper argues that academic games fulfill precisely this condition and that by shifting the competitive structure of learning environments through their inherent components, educators can change the norms and values of students to better encourage academics. Specifically, this theoretical review unpacked the nature of extracurricular debate and contends that its structure is consistent with what prior experimental studies suggest make for authentic student learning. Urban Debate Leagues, through its features it shares with the Teams-Games-Tournament (TGT) cooperative learning technique, show remarkable promise and may provide a compelling solution to address the often-cited shortcomings of urban schools for some students. While most reform efforts in education focus on disciplinary classes, academic games, such as debate, can serve as a constructive intervention for students who are currently not well served by the existing school system.

Despite the unique benefits that academic games offer for urban adolescents, these youth tend to report lower levels of participation in a variety competency-enhancing activities

(Schwarts, Capella, & Seidman, 2015). In many urban schools, teachers are expected to rely on predetermined, scripted curriculum materials to shape their instructional practices (Milner, 2013), and games are often the first to go when budgets fall short. For example, a wide range of research documents the differences between race, place, and class when examining participation rates in extracurricular activities. Not only are white students more likely to be involved in extracurricular activities than black and Hispanic students (Darling, Caldwell, & Smith. 2005), but urban students spend two-thirds less time in structured activities outside of school compared to their suburban counterparts (Larson, Richards, Sims, & Dworkin, 2001). This inequity in participation is closely related to a lack of access for many students. While all families value extracurricular activities (Chin & Phillips, 2004), income and time constraints serve as barriers that lower participation rates for many communities (Pederson, 2005; Quinn, 1999).

However, the advent of new Urban Debate Leagues demonstrates that when debate opportunities exist in under-served schools, students tend to flock to them. Improving urban education requires more than traditional reforms to close achievement gaps, and one cannot help but wonder the potential range of benefits academic games could provide to inner-city areas if invested in fully by administrators and school leaders. Unfortunately, as is the case for many resources in education, academic games like debate are not equally distributed or accessible across schools. If they are to become widespread, proponents must find a way to convince education practitioners that academic games are worthy of their time and expenditures. As it stands, Coleman's words may be most persuasive:

> "If we refuse to accept as inevitable the educational unconcern of the adolescent culture, then this poses a serious challenge. For to change the norms, the very foci of attention, of a cultural system is a difficult task… Yet if the challenge can be met, if the attention of the adolescent culture can be directed toward, rather than away from, those educational goals which adults hold for children, then this provides a far more fundamental and satisfactory solution to the problem of focusing teenagers' attention on learning." (1959, p.348).

Now all we have to do is listen.

References

Allen, M., Berkowitz, S., Hunt, S., & Louden, A. (1999). A meta-analysis of the impact of forensics and communication education on critical thinking. *Communication Education, 48(*1), 18–30.

Anderman. E. M., & Maehr, M. L. (1994). Motivation and schooling in the middle grades. *Review of Educational Research, 64*, 287-309.

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.

Annurwana, P., (2018). The effect of Teams Games Tournament on mathematics self-efficacy in junior high schools. SHS Web of Conferences 42 (79). https://doi.org/10.1051/shsconf/20184200079 GC-TALE 2017

Au, W. (2011). Teaching under the new Taylorism: High-stakestesting and the standardization of the 21st century curriculum. *Journal of Curriculum Studies, 43,* 25-45.

Beatty, B. (2011). The dilemma of scripted instruction: Comparing teacher autonomy, fidelity, and resistance in the Froebelian Kindergarten, Montessori, direct instruction, and success for all. *Teachers College Record, 113*(3), 395-430.

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

Brooks, J. G., & Dietz, M. E. (2013). The dangers & opportunities of the Common Core. *Educational Leadership, 70*(4), 64-67.

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.

Coleman, J. S. (1968). "Academic games and learning," National Association of Secondary School Principals Bulletin, 52:62-72.

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

Darling, N., Caldwell, L. L., & Smith, R. (2005). Participation in school-based extracurricular activities and adolescent adjustment. *Journal of Leisure Research, 37*, 51–76. doi:10.1007/ s10964-005-7266-8.

DeLuca, S., Clampet-Lundquist, S., & Edin, K. (2016). *Coming of Age in the Other America.* New York: Russell Sage.

DeVries, D. L., & Edwards, K. J. (1973). Learning games and student teams: Their effects on classroom process. *American Educational Research Journal, 10*, 307-318.

DeVries, D. L, Mescon, T., & Shackman, S. L. (1975) "Teams-Games- Tournament m the elementary classroom a replication " *Report No 190*, Center for Social Organization of Schools, Johns Hopkins University.

Eccles, J. S., Barber, B. L., Stone, M., & Hunt, J. (2003). Extracurricular activities and adolescent development. *Journal of Social Issues, 59*, 865–889.

Edwards, K. J., DeVries, D. L., & Snyder, J. P. (1972). Games and teams: A winning combination. *Simulation and Games, 3*, 247-269.

Erikson, E. H. (1968). *Identity: Youth and Crisis (No. 7)*. New York: WW Norton & Company.

Fine, G. A. (2001). *Gifted Tongues: High School Debate and Adolescent Culture*. Princeton, NJ: Princeton University Press.

Freire, P. (1993). *Pedagogy of the Oppressed*. New York: Continuum Publishing Corporation.

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.

Hulten, B. H. & DeVries, D. L. (1976). "Team Competition and group practice: Effects on student achievement and attitudes " *Report No 212*, Center for Social Organization of Schools, Johns Hopkins University.

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.

Kalesnikava, V. A., Ekey, G. P., Ko, T., Shackelford D. T., & Mezuk, B. (2019). *Grit, growth mindset and participation in competitive policy debate: Evidence from the Chicago Debate League.* Manuscript submitted for publication.

Kaufman, J., & Gabler, J. (2004). Cultural capital and the extracurricular activities of girls and boys in the college attainment process. *Poetics, 32*, 145–168.

Ke, F., & Grabowski, B. (2007). Game playing for math learning: Cooperative or not? *British Journal of Educational Technology, 38*(2), 249–259.

Keefe, C., Harte, T., & Norton, L. (1982). *Introduction to debate*. New York: Macmillan.

Kena, G., Hussar, W., McFarland, J., de Brey, C., Musu-Gillete, 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.

Kornhaber, M. L. & Orfield, G. (2001). High-stakes testing policies: examining their assumptions and consequences. In G. Orfield and M. L. Kornhaber (eds), *Raising standards or raising barriers? Inequality and high-stakes testing in public education* (New York: Century Foundation Press), 1–18.

Larson, R. W., Richards, M. H., Sims, B., & Dworkin, J. (2001). How urban African-American young adolescents spend their time: Time budgets for locations, activities, and companionship. *American Journal of Community Psychology, 29*, 565–597.

Mehta, J. & Fine, S. (2017). "Why the periphery is often more vital than the core." *Ed. Harvard Education Magazine*, Winter 2017.

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*, 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.

Milner, H. R. (2013). Scripted and narrowed curriculum reform in urban schools. *Urban Education, 48*, 163-170.

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

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.

O'Connell, M. E., Boat, T., & Warner, K. E. (2011). *Preventing Mental, Emotional, and Behavioral Disorders Among Young People: Progress and Possibilities*. Washington, D.C.: The National Academic Press.

O'Donnell, C. R., & Tharp, R. G. (2012). Integrating cultural community psychology: Activity settings and the shared meanings of intersubjectivity. *American Journal of Community Psychology, 49*(1), 22–30.

Olson, C.B. (1984). Fostering critical thinking skills through writing. *Educational Leadership (42)* 28-39.

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.

Piaget, J. (1951). *Play, dreams, and imitation in childhood*. New York: Norton.

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

Reardon, S. F. (2011). The widening academic achievement gap between rich and poor: New evidence and possible explanations. In G. J. Duncan & R. J. Murnane (Eds.), *Whither opportunity? Rising inequality, schools, and children's life chances* (pp. 91–115). New York: Russell Sage Foundation.

Salam, A., Hossain, A., & Rahman, S. (2015). Effects of using Teams Games Tournaments (TGT) cooperative technique for learning mathematics in secondary schools of Bangladesh. *Malaysian Online Journal of Educational Technology 3*(3), p35-45.

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. (2019). The BUDL effect: Examining academic and engagement outcomes of preadolescent Baltimore Urban Debate League participants. "Forthcoming." *Educational Researcher.*

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.

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

Slavin, R. E., DeVries, D. L, & Hulten, B. H. (1975). "Individual vs. team competition: The interpersonal consequences of academic performance " *Report No 188*, Center for Social Organization of Schools, Johns Hopkins University.

Sleeter, C. (2012). Confronting the marginalization of culturally responsive pedagogy. *Urban Education, 47*, 562–584. doi:10.1177/0042085911431472

Snipes, J., Horowitz, A, (2008). *Advancing Adolescent Literacy in Urban Schools*. Washington, DC: Council of the Great City Schools.

Soros Foundation Youth Initiative. (2003). Retrieved http://www/ soros.org/initiatives/youth/focus_areas/urban_debate.

Stockard, J. (2010). Improving elementary level mathematics achievement in a large urban district: The effects of direct instruction in the Baltimore city public school system. *Journal of Direct Instruction, 10*, 1-16.

Thompson, S., Ransdell, M., & Rousseau, C. (2005). Effective teachers in urban school settings: Linking teacher disposition and student performance on standardized tests. *Journal of Authentic Learning, 2*, 22-34.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, (2017). *National Assessment of Educational Progress (NAEP), various years, 1990–2017 Mathematics and Reading Assessments.*

Vygotsky, L. (1981). *Learning through interaction: The study of language development.* Cambridge: Cambridge University Press.

Wade, M. M. (1999). Urban debate leagues and educational reform. *The Forensics Educator, 14*(1), 37-41.

Wade, C., & Tavris, C. (1987). *Psychology* (1st ed.). New York: Harper & Row.

Warner, E., & Bruschke, J. (2001). 'Gone on debating': Competitive academic debate as a tool of empowerment. *Contemporary Argumentation and Debate, 22*, 1-21.

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

Wyk, M. M. V. (2011). The effects of Teams-Games-Tournaments on achievement, retention, and attitudes of economics education students. J*ournal of Social Science, 26*(3), 183-193.

Zaff, J. F., Moore, K. A., Papillo, A. R., & Williams, S. (2003). Implications of extracurricular activity participation during adolescence on positive outcomes. *Journal of Adolescent Research, 18*, 599–630.

II. THE BUDL EFFECT: EXAMINING ACADEMIC ACHIEVEMENT AND ENGAGEMENT OUTCOMES OF BALTIMORE URBAN DEBATE LEAGUE PARTICIPANTS

Introduction

Student participation in extracurricular activities has been linked to many positive outcomes (Eccles & Barber, 1999; Denault & Poulin, 2009a; 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 (Pederson, 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.

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 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, 2019) than non-debaters.

But despite major advances in the research justification for Urban Debate Leagues (UDL), 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, Hawkins, Hill, Kosterman, & Catalano, 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 8th grade reading and math standardized test scores, attendance rates, and 9th 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 Urban Debate League 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 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, & Stane, 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, Marayama, 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.

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' non-cognitive 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 an Urban Debate League 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 9th 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 4-8 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 Urban Debate League from a diverse school district may influence student outcomes.

Positive findings will be noteworthy as they may outline to policymakers appropriate and effective means to influence certain 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 eight 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 an Urban Debate League may improve academic achievement and engagement outcomes during this period for particular students, thereby preventing them 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 2004-05 to 2013-14 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 Baltimore Urban Debate League during preadolescence (grades 4-8).

The outcome variables of interest in this study include standardized 8th grade reading and math test scores from the Maryland School Assessment, average attendance rate in grades 4-8, and 9th grade high school destination. The Maryland School Assessment (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 2003-04 to 2013-14 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 3rd grade are used to account for a student's pre-debate achievement while scores from the 8th grade are used to measure a student's academic achievement at the end of preadolescence.

BERC enrollment data keeps 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 US Department of Education indicate that nearly 8 million students in the United States were chronically absent in the 2015-16 school year. Chronic absenteeism has been shown to increase achievement gaps at the elementary, middle, and high school levels (Balfanz & Brynes, 2012), and in Baltimore, analysis from BERC in 2009 revealed that nine in ten 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 cut off point) is also examined as an outcome.

The final outcome variable of interest in this study is 9th 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 over-age and severely under-credited 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 9th 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 9th 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 50-80%, 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 9th 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 pre-debate outcome measures, various demographic and background variables will also be used as covariates during analysis. These include, 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, free or reduced-price meals qualification, and whether a student transferred BCPSS schools in the first three years of elementary school.

Methods

Because this study only includes students who attended a school that participated in the Elementary and Middle School Baltimore Urban Debate League, 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, & Seole, 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 pre-existing 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 often 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 9th 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 pre-debate measures of academic achievement (3rd grade reading and math test scores) and engagement measures (average K-3 attendance rates) to attempt to mitigate 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 non-debaters 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:

$$\text{For } d_i = 1\text{: } w_i, \text{ATT} = 1,$$

$$\text{For } d_i = 0\text{: } w_i, \text{ATT} = p_i / 1 - p_i$$

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 (Rubin 1973, Morgan & Todd, 2008).

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 preadolescence among those students who typically participate. These regressions were restricted to the sample of students within the range of propensity scores that contain students in both the treatment and control groups. This resulted in 611 non-debaters (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 accounted for clustering by school when generating standard errors.

Results

Table 1 displays a summary of descriptive statistics of covariates and outcomes of interest for debaters and non-debaters. 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 non-debaters. For example, the average standardized difference in means of covariates between the treatment and control group is 0.1515 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 non-debaters 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, more likely to have transferred BCPSS schools in the first three years of elementary school, and more likely to have higher pre-debate attendance and 3rd grade achievement as measured by MSA test scores than non-debaters.

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 0.1515 to 0.0005 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 Maryland State Assessment reading and math test scores. Before accounting for sample differences, preadolescent debaters scored approximately 14 points higher on average than non-debaters on both the 8th grade reading and math Maryland State Assessment (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: $b = 6.38$, $p < 0.001$ and math: $b = 4.52$, $p < 0.001$). In standard deviation units, the effect of preadolescent debate participation for debaters is an approximate 21% and 13% increase for 8th 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 non-debaters. This difference of 3% is about one third of a standard deviation unit. The ATT estimate for preadolescent debate participation is 1.87%, or approximately one-fifth of a standard deviation unit ($p < 0.001$). Converting the

attendance rate outcome into a binary indicator of chronic absenteeism with 90% attendance used as the cut-off point, a statistically significant relationship remains (b = -0.86, $p < 0.001$). Average probability differences demonstrate interpretable effect sizes. For example, the average probability of being chronically absent is 10% lower for debaters than for non-debaters.

9th 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 2004-05 to the 2013-14 school years, approximately 80% attended a BCPSS high school in the 9th 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. The remaining 20% of students in the sample either dropped out before the 9th grade (1%), transferred out of the Baltimore City Public School System (15%), or were held back from attending the 9th 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, $p < 0.001$) and selective career tech entrance criteria schools (b = 0.29, $p < 0.001$). The average probability of attending a selective general entrance criteria high school is approximately 12% higher for debaters than non-debaters while the average probability of attending a selective career tech entrance criteria school is approximately 2% higher for debaters than non-debaters. A statistically significant negative relationship was also found for the odds of transferring out of BCPSS (b = -0.45, $p < 0.001$).

Appendix Tables 1, 2, and 3 present findings from ordinary least squares, logistic, and

multinomial logistic regression models, respectively, which can be used to understand the associations that demographic and pre-debate achievement covariates have with outcomes of interest.

Discussion

The key findings from this study are that preadolescent debate participation in an Urban Debate League had statistically significant relationships on many academic achievement and engagement outcomes among debaters. Preadolescent debate participation was associated with a 6.38 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 are not 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 9th 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 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 Urban Debate Leagues 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 approximately 7 points lower in reading and 9 points lower in math on the 8th 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, as well as are 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 Baltimore Urban Debate League 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 pre-debate 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 3rd 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 Baltimore Urban Debate League 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. As such, results are only applicable to the portion of the student population with similar observed characteristics as the treatment group. Approaches that interpret analysis as the average causal effect of debate participation are plagued with an unknown amount of selection bias. However, targeting the specific demographic that participates in the activity is suitable for administers of programs as estimates apply to the population of students they serve. 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 Urban Debate Leagues. 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 academic 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 paper suggest, Urban Debate Leagues may provide a compelling solution to the often-cited shortcomings of urban schools for some students. 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?

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, *The 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). T*he importance of being in school: A report on absenteeism in the nation's public schools.* Johns Hopkins University Center for Social Organization of Schools.

Balfanz, R., Herzog, L., & MacIver, 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). *Data Matters: Using Chronic Absence to Accelerate Action for Student Success*, Attendance Works and Everyone Graduates Center, September 2018 .

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, MD: 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*. (online) Retrieved from https://pdfs.semanticscholar.org/93e9/97fd0e883cf7cceb3b1b612096c27aa40f90.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-Gillete, 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, NY: 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, 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. *The 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*:231-81.

Morgan, S. L. & Winship, C. (2015). *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge: 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*:920-36.

Rosenbaum, P. R. & Rubin, D. B. (1983). "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika 7*0:41-55.

Rubin, D. B. (1973). "Matching to Remove Bias in Observational Studies.: *Biometrics 29*: 159-83.

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. (2019). *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. (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. *The Review of Economics and Statistics, 92*(2), 284-301.

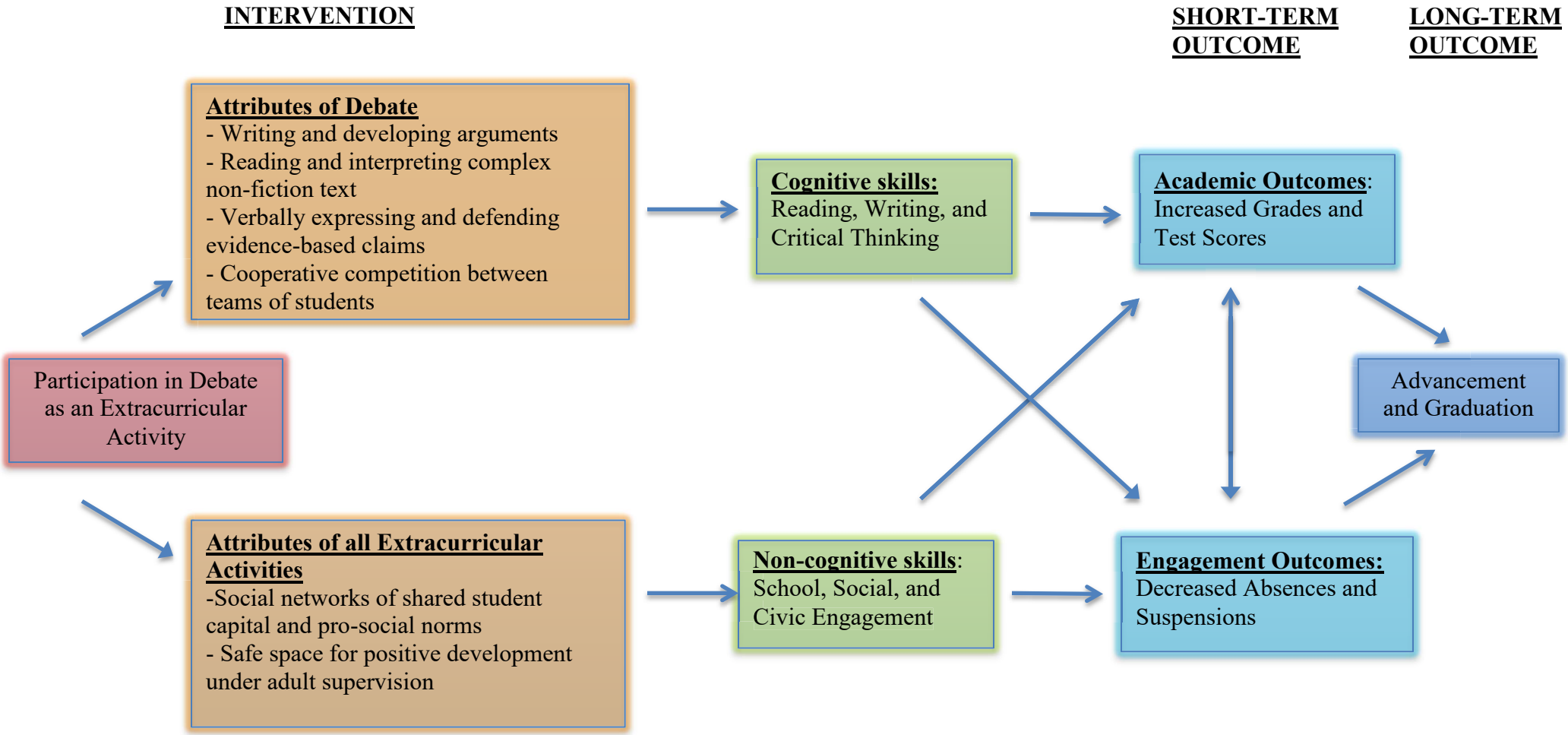
Swanson, C. B. (2009). *Cities in Crisis 2009: Closing the Graduation Gap - Educational and Economic Conditions in America's Largest Cities. Editorial Projects in Education Research Center*. April 2009.

U.S. Census Bureau. (2014). *A Child's Day: Living Arrangements, Nativity, and Family Transitions: 2011* 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.

Figures

Figure 1. *Conceptual Model for the Theory of Change behind Participation in Debate as an Extracurricular Activity*



Notes: 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.

Tables

Table 1. *Summary of descriptive statistics for debaters and non-debaters*

	Preadolescent Debate Sample (n = 2,263)		*Non-Debate Sample (n = 81,906)*	
VARIABLES	*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	
Transferred Schools (Grades1-3)	0.354		0.345	
Pre-debate 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	
9th 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	

Notes: $N = 84{,}169$. 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 = 0.1515 and the average standardized difference in standard deviations of continuous covariates = 0.0804.

Table 2. *Logit-Coefficients for a propensity score model of preadolescent debate participation*

VARIABLES	Logit-Coefficients
Age	0.0240
	(0.0154)
Male	-0.331*
	(0.0506)
American Indian	0.417
	(0.464)
Asian	-0.0872
	(0.301)
Hispanic	0.00140
	(0.312)
Black	0.543*
	(0.149)
English Language Learner	-0.149
	(0.188)
Special Education Services	-0.284*
	(0.0720)
Free or Reduced-Price Meals	0.404*
	(0.169)
Transferred Schools (Grades 1-3)	0.129*
	(0.0552)
Attendance Rate (Grades K-3)	0.0306*
	(0.00455)
MSA Reading Grade 3	0.00675*
	(0.00139)
MSA Math Grade 3	0.00372*
	(0.00111)
Constant	-11.83
Model Chi-square	427.57
Degrees of Freedom	13

Notes: $N = 84{,}169$. Robust se in parentheses, and * $p < 0.05$ for two-tailed tests with null of 0.

Table 3. *Balance achieved by weighting*

	Preadolescent Debate Sample (n = 2,263)		*Non-Debate Sample (n = 81,906)*	
COVARIATES	*Mean*	*SD*	*Mean*	*SD*
Age	20.50	3.28	20.49	3.40
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	
Transferred Schools (Grades 1-3)	0.354		0.353	
Pre-debate 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

Notes: N = 84,169. Means and standard deviations are weighted by the estimated ATT weight in order to demonstrate achieved balance. The average standardized difference in means of covariates = 0.0005 and the average standardized difference in standard deviations of continuous covariates = 0.0490.

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	6.38*	0.205
	(0.52)	
MSA Math Grade 8	4.52*	0.127
	(0.69)	
Attendance Rate (Grades 4-8)	1.87*	0.198
	(0.13)	
CATEGORICAL OUTCOMES	*Logit-Coefficients*	*Effect Size in APD*
Chronic Absenteeism (Grades 4-8)	-0.86*	-0.099
	(0.08)	
9th Grade High School Destination		
(Base Outcome = Attend Traditional High School)		
Attend Selective General High School	0.75*	0.123
	(0.07)	
Attend Selective Career Tech High School	0.29*	0.015
	(0.07)	
Attend Charter High School	0.21	0.010
	(0.11)	
Attend Alternative High School	-0.17	-0.003
	(0.23)	
Dropout Before High School	-0.37	-0.004
	(0.27)	
Transfer Before High School	-0.45*	-0.047
	(0.11)	
Not Promoted to High School	-0.17	-0.012
	(0.16)	

Notes: $N = 83{,}558$. Robust se in parentheses, and * $p < 0.05$ for two-tailed tests with null of 0.
Effect sizes shown in standard deviation units for continuous outcomes and in average probability differences for categorical outcomes.

Appendix Table 1. *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	6.201*	4.340*	1.852*
	(0.519)	(0.695)	(0.129)
Age	1.286*	1.958*	-0.134*
	(0.0436)	(0.0720)	(0.0179)
Male	-5.341*	-2.491*	-0.541*
	(0.178)	(0.216)	(0.0690)
American Indian	-2.717	-2.526	0.499
	(1.420)	(1.527)	(0.578)
Asian	-0.688	6.500*	1.415*
	(0.826)	(1.156)	(0.360)
Hispanic	-2.009*	-3.369*	0.979*
	(0.719)	(1.012)	(0.302)
Black	-7.007*	-8.811*	1.643*
	(0.588)	(0.929)	(0.217)
English Language Learner	4.861*	10.573*	1.959*
	(0.904)	(0.989)	(0.204)
Special Education Services	-11.428*	-9.466*	-0.420*
	(0.332)	(0.415)	(0.105)
Free or Reduced-Price Meals	-9.064*	-12.776*	-0.239
	(0.952)	(1.194)	(0.200)
Transferred Schools (Grades 1-3)	-1.113*	-0.779*	-0.606*
	(0.198)	(0.277)	(0.0901)
MSA Reading Grade 3	0.412*	0.181*	0.0175*
	(0.00917)	(0.00985)	(0.00191)
MSA Math Grade 3	0.225*	0.479*	0.0200*
	(0.00538)	(0.00937)	(0.00160)
Attendance Rate (Grades K-3)	0.0818*	0.313*	0.545*
	(0.0183)	(0.0238)	(0.0127)
Constant	128.696	87.997	27.897
R-squared	0.542	0.510	0.174

Notes: $N = 83{,}558$. Robust se in parentheses, and * $p < 0.05$ for two-tailed tests with null of 0.

Appendix Table 2. *Logit-coefficients of variables on the binary indicator chronic absenteeism*

VARIABLES	
Debate Participation	-0.854*
	(0.0784)
Age	0.0538*
	(0.00527)
Male	0.128*
	(0.0200)
American Indian	-0.134
	(0.194)
Asian	-0.499*
	(0.136)
Hispanic	-0.274*
	(0.0971)
Black	-0.491*
	(0.0483)
English Language Learner	-0.759*
	(0.0711)
Special Education Services	0.146*
	(0.0265)
Free or Reduced-Price Meals	0.313*
	(0.0782)
Transferred Schools (Grades 1-3)	0.232
	(0.234)
MSA Reading Grade 3	-0.00499*
	(0.000511)
MSA Math Grade 3	-0.00551*
	(0.000483)
Attendance Rate (Grades K-3)	-0.131*
	(0.00348)
Constant	14.039
Model Chi-square	3726.00
Degrees of Freedom	14

Notes: $N = 83{,}558$. Robust se in parentheses, and * $p < 0.05$ for two-tailed tests with null of 0.

Appendix Table 3. *Multinomial logit-coefficients of variables on 9th grade high school destination (base = traditional high school)*

VARIABLES	Selective General	Selective CT	Charter	Alternative	Drop Out	Transfer	Not Promoted
Debate Participation	0.769*	0.295*	0.221	-0.189	-0.361	-0.452*	-0.185
	(0.0687)	(0.0692)	(0.124)	(0.226)	(0.273)	(0.107)	(0.158)
Age	0.108*	0.0205*	-0.102*	0.0993*	0.0855*	-0.0595*	-0.464*
	(0.0101)	(0.00578)	(0.00736)	(0.0109)	(0.0130)	(0.00657)	(0.00858)
Male	-0.611*	-0.128*	-0.154*	0.240*	0.0264	-0.0146	0.0542
	(0.0205)	(0.0278)	(0.0333)	(0.0679)	(0.0655)	(0.0228)	(0.0367)
American Indian	0.0428	0.376	0.508	0.923	-1.514	-0.127	0.130
	(0.256)	(0.370)	(0.417)	(0.692)	(1.056)	(0.182)	(0.290)
Asian	0.693*	0.385	0.158	1.290	0.180	0.283	0.715*
	(0.163)	(0.265)	(0.372)	(0.736)	(0.517)	(0.153)	(0.202)
Hispanic	-0.0764	0.298	-0.395	-0.0508	-0.154	-0.315*	-0.352
	(0.246)	(0.138)	(0.259)	(0.584)	(0.341)	(0.124)	(0.148)
Black	0.0110	0.926*	0.607*	0.953*	-0.491*	-0.719*	-0.876*
	(0.0153)	(0.195)	(0.243)	(0.263)	(0.139)	(0.0774)	(0.106)
English Language Learner	0.419*	-0.529*	-0.519*	-2.915*	-0.190	0.284*	0.0691
	(0.122)	(0.128)	(0.204)	(0.906)	(0.341)	(0.127)	(0.131)
Special Education Services	-0.883*	-0.332*	0.119*	-0.0790	-0.504*	-0.164*	-0.240*
	(0.0452)	(0.0417)	(0.0398)	(0.0661)	(0.0956)	(0.0291)	(0.0556)
Free or Reduced-Price Meals	-0.760*	0.304*	-0.0624	0.968	-1.361*	-1.528*	-1.488*
	(0.117)	(0.0970)	(0.116)	(0.419)	(0.168)	(0.0804)	(0.0917)
Transferred Schools Grades (1-3)	-0.334*	-0.137*	-0.140*	0.261*	-0.340	-0.0698	-0.668*
	(0.0342)	(0.0288)	(0.0356)	(0.628)	(0.745)	(0.0415)	(0.0489)
MSA Reading Grade 3	0.0194*	0.00408*	0.00285*	-0.00720*	0.00434	0.00762*	0.00447*
	(0.00106)	(0.000748)	(0.000931)	(0.00178)	(0.00243)	(0.000743)	(0.00109)
MSA Math Grade 3	0.0249*	0.0114*	0.00142*	-0.00316*	0.00168	0.00620*	0.00455*
	(0.000920)	(0.000677)	(0.000679)	(0.00143)	(0.00178)	(0.000657)	(0.00101)
Attendance Rate (Grades K-3)	0.0526*	0.0259*	0.0132*	-0.0238*	-0.0599*	-0.0109*	0.0361*
	(0.00422)	(0.00265)	(0.00281)	(0.00441)	(0.00391)	(0.00178)	(0.00387)
Constant	-24.69	-11.24	-3.06	-2.06	-0.37	-2.06	1.88

Notes: N = 83,558. Model Chi-square = 39,957.40. df = 98. Robust se in parentheses, and * $p < 0.05$ for two-tailed tests with null of 0.

III. THE BALTIMORE URBAN DEBATE LEAGUE AND HIGH SCHOOL OUTCOMES: A SENSITIVITY ANALYSIS

Introduction

Is there an effect of extracurricular activity participation on student outcomes? This question has been asked countless times in the education literature and while an extensive body of research documents strong associations between student participation in extracurricular activities and a variety of outcomes (Eccles & Barber, 1999; Denault & Poulin, 2009a; Feldman & Matjasko, 2005), association does not equal causation. For example, it is possible that an unobserved covariate, like motivation, influences both a student's decision to participate in extracurricular activities and his or her academic achievement. Thus, these positive associations may not be informative about whether increasing participation will causally affect student outcomes later in life and the pressing question remains: can the association between extracurricular activities and educational attainment be given a causal interpretation or is participation in these programs simply another indication of an already motivated student?

As is illustrated above, the foremost threat to internal validity for research on extracurricular activities concerns the voluntary nature of participation of these programs. Students motivated to join extracurricular activities are also those who tend to be more positively oriented to school than their peers (Gottfredson, Cross, & Seole, 2007; McNeal, 1995). Consequently, it is difficult to untangle genuine causal relationships between voluntary extracurricular activity participation and student outcomes from selection effects. While removing selection bias is relatively easy for studies that can randomize who receives treatment, investigators rarely have the same luxury in the social sciences, where prohibiting access to educational interventions aimed at improving student outcomes is frequently not possible.

Attempts to account for selection effects are made by "controlling" for as many variables as possible. However, important covariates that may confound the relationship between the exposure and the outcome often remain unmeasured. Such unobserved confounding can bias estimates of the effect of the intervention and consequently undermine the conclusions made in many observational studies.

Fortunately, sensitivity analysis techniques can help assess the extent to which an unmeasured variable might influence treatment effects. The objective of a sensitivity analysis is to determine the magnitude of hidden bias that would change inferences about an effect (Rosenbaum, 1987). If the treatment effect does not differ substantially under a range of plausible values for unobserved confounding, then conclusions of a study may be immune to violations of treatment selection ignorability.

The current paper examines the causal relationship between one particular extracurricular activity, debate, and a variety of high school outcomes among a diverse urban school district. Specifically, analysis aims to contribute to the understanding of whether participation in the Baltimore Urban Debate League influences SAT scores and high school attendance both for preadolescent debaters (those that participate in grades 4-8) as well as for adolescent debaters (those that participate in grades 9-12). Because the Baltimore Urban Debate League prohibits try-outs and provides every interested student an opportunity to participate, a randomized control trial is an ill-suited study design to evaluate the program. Informative empirical analysis is still possible, however, as checks for sensitivity provide policymakers as well as educational practitioners a method to assess how robust evidence for the activity is to violations in assumptions often made in observational research.

The following study proceeds by providing estimates of treatment effects on the treated for preadolescent and adolescent debate participation assuming that treatment selection ignorability holds for a set of pre-treatment observed covariates that include measures of achievement and engagement assessed prior to participation. Then, these estimates are analyzed to see how sensitive interpretations are to maintained assumptions. Not only does this approach help build confidence in the results of research on extracurricular debate by differentiating between effects that are relatively immune to potential unmeasured confounding versus those that may change significantly, but it also serves to advance sensitivity analysis as a useful complement to any observational study for which it is unclear if treatment selection ignorability has been satisfied.

Background

Debate as an extracurricular activity has been positioned as an innovative approach aimed at improving student educational outcomes. Because it offers a set of skills that closely relates to many scholastic goals such as public speaking, reading, and critical thinking, Urban Debate Leagues throughout the country have been established on the premise that they improve schooling for underserved students in depressed urban areas. Today, the National Association of Urban Debate Leagues 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 indicates that high school students who debate have greater gains in cumulative grade point average over the course

of high school, are more likely to graduate high school, and are more likely to be college ready in reading and English relative to comparable peers who do not participate in debate (Mezuk, 2009; Mezuk, Bondarenko, Smith, & Tucker, 2011). Follow-up analysis finds 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, 2019) than non-debaters.

Positive effects have also been found for elementary and middle school participation during preadolescence. 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 8th grade standardized test scores, a decreased likelihood of chronic absenteeism, and an increased likelihood of attending a selective entrance criteria high school (Shackelford, 2019). These results were based on models that mitigate observed selection bias by accounting for sample differences using inverse-probability-treatment weighting and address concerns about differential treatment effect bias by focusing on the average effect for the treatment group as the target parameter of interest.

However, these prior studies assume that all confounding covariates, variables that predict both treatment and outcome, are observed. To illustrate, figure 1 depicts a causal graph in which the effect of treatment on an outcome of interest is confounded only by observed variables. Formally, this graph provides an instance in which treatment selection is ignorable such that appropriate adjustment on X would accurately identify the causal effect of Z on Y. Instead, suppose that a relevant covariate that predicts both the treatment and the outcome is unobservable and therefore omitted from analysis. Figure 2 shows an instance where treatment selection is not ignorable because the causal effect of Z on Y is confounded both by observed

variables in X and an unobserved variable U. Thus, effect estimates that only adjust for X are subject to confounding bias (Morgan & Winship, 2015).

There is likely some degree of omitted variable bias in the studies outlined above because school administrative datasets are limited in their access to certain measures, such as parent characteristics or student motivation. These unobserved measures threaten the assumption that treatment and control groups are identical at baseline and, consequently, that treatment selection is ignorable. If these same characteristics influence both debate participation and the outcomes of interest, treatment effect estimates will be biased (Lin, Psaty, & Konmal, 1998).

In the face of such potential unmeasured confounding, an important approach in evaluating evidence for causation is sensitivity analysis. Based largely on the work of Rosenbaum and his colleagues (1983), these methods seek to answer how robust or sensitive estimates of a causal effect are to the potential effects of unobservable treatment selection patterns. Sensitivity analysis considers how strong unmeasured confounding would have to be to explain away observed associations and it can be useful in assessing a plausible range of values for the causal effect of the treatment on the outcome corresponding to assumptions concerning the relationship between the unmeasured confounder and observed variables in the model (Rosenbaum, 1986).

In general, the typical steps involved in sensitivity analysis consist of first offering a provisional point estimate of the treatment effect by conditioning on observed covariates assuming that treatment selection ignorability holds. Then, often through simulation, display how the provisional point treatment effect estimate would be expected to change assuming pairs of values for the effects of an unobserved covariate on treatment as well as on the outcome. This usually includes reporting the values that would prevent one form rejecting the null hypothesis of

no effect. Finally, using external information on what is known about the unobserved confounder, assess whether or not the pairs of values seem reasonably likely given other observed covariates. If the treatment effect estimate under plausible levels of confounding differs substantially from the original estimate, then the effect can be deemed sensitive to violations of treatment selection ignorability. If not, then one can confidently stick to the substantive interpretations of the treatment effect and conclude that the point estimate is sufficiently unlikely to be equal to zero.

Prior studies have shown how influential sensitivity analysis can be in assessing how robust an association is to potential unmeasured confounding. For example, in a study on the effect of neighborhood context on the odds of completing high school, Harding (2003) concludes that an unobserved confounder would have to be more powerful than family income in terms of associations with the treatment and outcome in order to produce non-significant treatment effects. The utility of sensitivity analysis has led many authors in the counterfactual tradition to advocate for its use as a supplement to all observational studies intended to produce evidence for causality (VanderWeele, 2015).

Present Study

As mentioned previously, prior work on the Elementary and Middle School Baltimore Urban Debate League shows evidence of an average treatment effect on the treated for a variety of end of middle school outcomes that include attendance rate and standardized 8^{th} grade reading and math test scores (Shackelford, 2019). The present study expands upon the previous manuscript by examining high school outcomes from two different samples of BUDL participants: students who participate during elementary and middle school (EMS) and students

who participate in high school (HS). This first sample is examined to see whether treatment effects of participation during elementary or middle school persist and remain insensitive to unobserved confounding on attendance and standardized math and reading test scores in high school. This sample consists of 71,089 Baltimore City Public School System Students who both attended a school that participated in the EMS Baltimore Urban Debate League from the 2004-05 to 2013-14 school years and reached the age of 18 by fall of 2018. Of this sample 1,929 students participated in BUDL exclusively during preadolescence (Grades 4-8). The second sample consists of 95,969 BCPSS students who attended a school that participated in the HS Baltimore Urban Debate League during the same 10-year period, of which 2,710 students participated in BUDL exclusively during adolescence (Grades 9-12).

Informed by the extant research on debate participation, my primary hypothesis is that participation in BUDL, regardless of time of "treatment," will be associated with academic achievement and engagement outcomes in high school. Furthermore, if participation in BUDL is protective against declines in school performance, I expect the strongest benefits to be from students that participate during younger ages as opposed to participation later in life as findings suggest that the time spent in elementary and middle school are particularly salient periods to alter student trajectories (Kieffer & Marinell, 2012). 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 those entering the 9th 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. Thus, this literature on how

educational programs and interventions that target younger populations show greater gains over time suggests that early participation in BUDL during preadolescence (elementary and middle school) will be more positively associated with these outcomes than participation later in life during adolescence (high school).

Data come from de-identified yearly administrative student-level records from BCPSS in partnership with the Baltimore Education Research Consortium (BERC) that houses the 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). In essence, this study can be conceptualized as a program evaluation for every student who participated in the Baltimore Urban Debate League from 2003-2014.[a]

The outcome variables of interest in this study include scores from the verbal reasoning and mathematics sections of the SAT as well as average attendance rate in high school. SAT scores will be used to measure a student's academic achievement toward the end of their high school career as well as their college readiness. The test consists of two 800-point sections: the verbal reasoning section, which measures evidence-based critical reading and writing, and the mathematics section, which measures content that includes algebra, geometry, statistics, and probability. BERC houses SAT test score data for every BCPSS student who takes the test.

BERC enrollment data also keeps track of each student's days absent and days present for every school year attended in BCPSS. Because schools vary in the total number of days in

[1]A majority of BUDL students in this time frame exclusively participated in either the EMS or HS faction of the program. A total of 191 students participated in BUDL during Grades 4-8 as well as Grades 9-12. Due to the limited sample size of this group, as well as the goal of the current study to distinguish between participation at various stages of student development, this sample is not analyzed.

session per year, an attendance rate percentage equaling 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 high school attendance rate (grades 9-12) was created to serve as an outcome variable and two rates were created to account for a student's attendance immediately before participation. For the purposes of this study, an average attendance rate for grades K-3 is used to adjust for student participation prior to possible participation in the EMS BUDL for the preadolescent sample and an average attendance rate in grades 4-8 is used to adjust for student attendance prior to possible HS BUDL participation for the adolescent sample.

In the education field, the availability of pre-treatment outcome measures on individual students is critical for obtaining precise impact estimates (Bloom, Richburg-Hayes & Black, 2007; Schochet, 2010). For the purposes of this study, scores on the Maryland School Assessment (MSA) are used to account for pre-debate verbal reasoning and math achievement. The MSA was a test of reading and math achievement given to all Maryland students in grades 3-8 between the 2003-04 to 2013-14 school years. Similar to the sections of the SAT, the reading MSA tests a student's general reading processes and text comprehension, while the math MSA tests algebra/patterns, geometry/measurement, and statistics/probability. Because these tests were given at times prior to possible debate participation, MSA reading and math scores from the 3rd and 8th grade are used to account for a student's pre-debate achievement for the preadolescent and adolescent sample respectively.

Aside from the aforementioned pre-debate outcome measures, various demographic and background variables will also be used as covariates during analysis. These include age in 2018,

race-ethnicity (coded as Asian, Black, Hispanic, or White), and binary indicators of sex, English Language Learner status, special education services received, free or reduced-price meals qualification, and whether a student transferred BCPSS schools in the years prior to participation. However, this study is motivated to address a lack of confidence in the assumption commonly invoked with observational data that one can adjust for all confounders. It seems implausible that an administrative dataset could measure all covariates that might affect both a students' choice to participate in the Baltimore Urban Debate League and subsequent academic and engagement outcomes. Accordingly, this predicament is a prime candidate for sensitivity analysis as it is important to understand how sensitive results from this study might be to a potential unmeasured confounder; for instance, how predictive would such a confounder have to be to remove support for the effect of debate participation on end of high school outcomes?

Methods

The specific sensitivity analysis method used in this study was developed by Carnegie, Harada, and Hill (2016). Their technique starts by assuming that ignorability of a binary treatment *Z* is satisfied with the addition of a confounder *U* (see figure 2). They advocate a parametric approach that allows for independent specification of associations between the unmeasured confounder and the treatment and the unmeasured confounder and the outcome. For a binary treatment variable, the complete-data likelihood is

$$Y|X,U,Z \sim N\left(X\beta^{y} + \zeta^{y}U + \tau Z, \sigma^{2}_{y*xuz}\right),$$

$$Z|X,U \sim Bernoulli\left(\Phi(X\beta^{z} + \zeta^{z}U)\right),$$

$$U \sim Bernoulli(\pi^{u})$$

where continuous outcome *Y* is fully determined by a linear function of observed covariates *X*, unobserved confounder *U*, and the binary treatment variable *Z*. Conveniently, the sensitivity parameters ζ^{y} (the association between unobserved *U* and outcome *Y*) and ζ^{z} (the association between unobserved *U* and treatment *Z*) are easily interpretable as regression coefficients from a linear regression and a probit regression respectively (Dorie et al., 2016). Their algorithm determines ranges of the sensitivity parameters that could inform the treatment effect estimate by generating candidate realizations of *U* given its conditional distribution. The relationships that this confounder has with the treatment and response are manipulated through a simulation to identify how large these sensitivity parameters need to be in order for a treatment effect to substantively change. Again, one cannot know with absolute certainty the strength of the unmeasured confounder's associations, but this method specifies many different values and determines how the estimate is affected by each value (Carnegie et al., 2016).

To gain an overall picture of the sensitivity of the treatment effect, estimates computed by this analysis algorithm can be summarized in a contour plot to reveal the combinations of parameters that drive the treatment effect to become non-significant or have no effect. Consequently, this method not only adheres to recent research that encourages observational studies to report an "E-value," or the minimum strength of association that an unmeasured confounder would need to have with both the treatment and the outcome to fully explain away an effect (VanderWeele & Ding, 2017), but it also provides an objective way to directly assess the sensitivity of a given study to violations of the ignorability assumption.

However, this specific simulation approach is distinctive in a few ways. First, the algorithm draws values for the sensitivity parameters from the distribution of the unobserved confounder conditional on observed data. Because an unmeasured confounder cannot account for

any more variability in treatment or outcome than is observed, valid sensitivity parameters are determined by examining the residual variance after conditioning on observed variables. In this way, the unmeasured confounder *U* is assumed to be independent of the measured confounders *X* and only represent the portion of the unobserved covariate not explained by observed covariates.

Second, this method calibrates the magnitude of the sensitivity parameters relative to the corresponding coefficients for observed confounders in the data (from the regressions of *Y* on *Z* and *X* and from the regression of *Z* on *X)* to serve as a benchmark. To create a common scale for these coefficients (among themselves and relative to the coefficients that serve as the sensitivity parameters) all continuous covariates are standardized to have a mean of zero and standard deviation of one. As an additional aide, the algorithm identifies the treatment effect estimate obtained with sensitivity parameters equivalent to the pair of observed coefficients that are farthest from the origin. This allows an investigator to determine whether a treatment effect remains assuming that the associations form an unmeasured confounder are as strong as the most predictive covariate. In a wide range of settings, the most predictive covariate is often a baseline measure of the outcome of interest (Bloom et al., 2007). Thus, if hypothetical sensitivity parameters do not change the sign or significance of the treatment effect even when given the strength associated with pre-treatment outcome measures, than evidence for the activity may be particularly immune to unmeasured confounding.

Finally, their approach allows an investigator to incorporate treatment weights to target estimands such as the average treatment effect on the treated (ATT). These weights are estimated using logistic regression to predict the propensity that one receives treatment conditional on observed variables, an approach identical to the technique utilized in traditional-inverse-probability treatment weighting. Targeting the ATT accounts for differential treatment-effect

bias, which assumes that the associations between experiencing treatment and any observed outcomes may differ across subgroups. This self-selection on the individual-level causal effect renders the average treatment effect for those that typically do not receive treatment, as well as the average treatment effect overall, unidentifiable (Morgan & Winship, 2015). The following incorporates these weights when presenting regression estimates for provisional treatment effects. To this end, results are only applicable to the portion of the student population with similar observed characteristics as the treatment group and should not be interpreted as the average causal effect of debate participation for all students. Finally, analysis utilizes sensitivity analysis technique described above to examine how sensitive these results are to unobserved confounding.

Results

Table 1 displays a summary of descriptive statistics of covariates and outcomes for debaters and non-debaters in both samples. On average, students who end up being debaters score higher on standardized assessments and attend school at higher rates prior to participation in BUDL. These differences show evidence of baseline bias and consequently highlight the need to adjust for these observed covariates in subsequent analysis.

Table 2 presents propensity score models predicting likelihood of debate participation among observed covariates for both samples and shows statistically significant associations between some covariates and selection into debate. In both samples debaters are more likely to be female, more likely to be black (as opposed to white), and are more likely to qualify for free or reduced-priced meals than non-debaters, on average.

Table 3 presents the means and standard deviations of covariates when applying the ATT

weight. It demonstrates that the weight constructed from the estimated propensity scores successfully balance the data as no statistically significant differences in observed covariates remain in the weighted samples between debaters and non-debaters.

Table 4 summarizes the results from ATT-weighted regression models predicting the relationship between debate participation and high school outcomes in both samples. These estimates are presented both as raw coefficients and standard deviation units. For both EMS BUDL participation as well as HS BUDL participation, debate is associated with statistically significant increases in all outcomes.

Figures 3-8 report how sensitive these estimates are to potential unmeasured confounding. ATT-weighted provisional estimates in standard deviation units are displayed above the horizontal axis. Because the axes of the contour plot represent regions of no confounding, with one or both the sensitivity parameters set to zero, these treatment effect estimates match the provisional estimates in Table 4. All observed predictors are labeled. Covariates with a positive coefficient on the outcome are identified with a plus sign and covariates with a negative coefficient are identified with an inverted triangle and have been transformed through multiplication by -1.

Each plot combines four different types of contours that provide a basis for an informed discussion on treatment effects. The black contours each represent the combinations of sensitivity parameters that lead to the same estimated treatment effect. For all six figures, more than half of the plot corresponds to significant, positive treatment effect estimates when targeting the ATT. The level of unobserved confounding required to drive the estimates to non-significance are represented by the blue lines while the red curve represents the contour along which the treatment effect estimate is reduced to zero. Finally, the gray contour plot gives the

treatment effect estimate when the posited unmeasured confounder is given the strength of the observed covariate furthest from the origin.

For example, the gray contour lines in figures 3 and 4 are below the lower blue line labeled “N.S” which indicates that for the effect of debate participation on SAT Verbal scores to be driven to non-significance, the sensitivity parameters of an unmeasured confounder would have to be larger than the coefficients for the most predictive observed covariate. For the adolescent sample, the greatest predictor is the pre-debate participation MSA reading score. Thus, even if a potential unobserved covariate were as predictive of reading and writing achievement on the SAT as prior reading and writing achievement on a standardized test in Grade 8, the inclusion of the variable would not change the sign or statistical significance of the estimated treatment effect on the treated.

However, in regards to SAT Math scores, the gray contour lines in figures 5 and 6 lie slightly within the blue lines that bracket the region of non-significance at the 5% level. For the adolescent sample, observed covariates also are within this region. Therefore, it seems plausible that an unobserved confounder might be equally strong. However, it is important to remember that this confounder would have be as strong as these predictors and associated with both the treatment and the outcome through pathways independent of the observed covariates to drive the estimated treatment effect on the treated to non-significance.

Finally, figures 7 and 8 depict sensitivity analysis results for debate participation on high school attendance rate. The display of the coefficients for the observed confounders in these plots allows one to see that in order for either no treatment effects or a non-significant treatment effects to represent the truth, the unmeasured confounder would have to have considerably greater predictive power than the observed covariates. The gray contour lines indicate that the

inclusion of an omitted variable which is as predictive as the most strongest observed covariate would have a relatively modest impact on both provisional treatment effects on the treated, moving the estimates to 0.16 and 0.13 for preadolescent and adolescent debate participation respectively.

Table 5 summarizes the primary results given by the gray contour lines in each sensitivity analysis plot. Positing an unmeasured confounder as predictive as the strongest observed covariate among a set that includes pre-outcome measures would not substantively change the interpretations of provisional ATT estimates of debate participation on SAT verbal reasoning scores and high school attendance rate as estimates remained positive and statistically significant. However, the treatment effects for SAT math scores are not as immune to potential unmeasured confounding. While debate participation may lead to higher math achievement, the evidence for causality is not nearly as strong as it is for verbal reasoning achievement and high school attendance.

Discussion

The observational study is a setting of great importance in educational research. While it has been widely used in empirical work to address whether extracurricular participation can positively influence student outcomes, a central concern with using non-experimental data to inform policy decisions is that one can never be certain that efforts to adjust for confounders are adequate when trying to isolate a treatment effect. As a consequence of the nonrandom assignment of treatment, treated and controlled subjects may not be comparable prior to treatment, and differences between outcomes may reflect either effects of the treatment or this lack of comparability, that is, that some third factor related to both the treatment and outcome

might explain their association with no true causal effect (Rosenbaum, 1987). Rather than abandon the goal of causal inference all together, sensitivity analysis can serve as a middle ground to help illuminate the extent to which findings are susceptible to violations of treatment selection ignorability when implementing a randomized experiment is infeasible.

This study found that most treatment effects of debate participation on end of high school outcomes among debaters persist and remain robust to ignorability violations. With the treatment group targeted as population parameter of interest, adolescent debate participation in high school was associated with a 16.69-point increase on the SAT reading and writing section, a 10.73-point increase on the SAT math section, and a 3.42 percent increase in high school attendance rate . These effects are noteworthy due to the study's use of pre-debate outcome measures used as covariates. Research finds that assessments that occur prior to treatment are typically the single best covariate for explaining the outcome's variation at a later date (Snochet 2008; Bloom et al. 2005). This was certainly the case among all observed covariates for the adolescent high school sample. These measures reflect many different unobservable factors that influence future outcomes. For example, parent characteristics or intrinsic student motivation, two potential unobserved confounders, are also likely associated with student achievement and engagement measured prior to participation. Thus, in order for the treatment effects on treated found for high school debate participation to change their substantive interpretation, any potential unobserved covariates would have to be as strong than the predictive value of these pre-debate outcome measures.

Significant positive treatment effects for the treated on high school outcomes were also found for preadolescent debate participation: an increase of 17.11 points in SAT verbal scores, 14.69 points in SAT math scores, and 3.78 percentage points in attendance rate. These estimates

were larger than the ATT effects found for adolescent participation in high school. However, because the pre-outcome measures are further removed from the outcomes of interest (measured at grade 3 instead of grade 8) they are less predictive, and treatment effects may be more sensitive to unmeasured confounding. Nevertheless the findings of positive associations between early participation are noteworthy as they outline to policymakers potentially appropriate and effective means to influence certain students' overall academic trajectories at earlier stages of development. The National Institute on Out-of-School Time (2003) estimates that approximately eight million children between the ages of 5 and 14 are unsupervised after school, but despite the lack of access to extracurricular activities for younger students, preadolescent students appear to be more open to participation in extracurricular activities than high school students (Denault & Poulin 2009b). While extracurricular activities are offered primarily during a student's high school years, results from this study potentially suggest a shift in the way schools provide access to participation for students at younger ages, particularly for activities that are aligned with academic goals.

Regardless of the time at which students participate in debate, larger treatment effect sizes were found for SAT reading and writing scores than SAT math scores among debaters. This finding may not be surprising as debate is an activity that focuses on informational text comprehension. However, the positive relationship with math scores suggest that debaters may gain skills that are not explicitly practiced in the activity indirectly through increases in school engagement outcomes. The positive relationship with debate participation and student attendance rate supports this interpretation. While this study is the first to examine high school attendance rate and SAT scores as outcome variables, its findings are in line with prior research which shows that participating in debate is robustly associated with other academic achievement

measures in high school (Mezuk, 2009; Mezuk et al., 2011). Additionally, the finding that debaters are more likely to be black and qualify for free or reduced-price meals, after adjusting for all other observed covariates, suggest that Urban Debate Leagues may be a culturally appropriate intervention for some students who may not be well served by existing structures.

This study is not without its limitations. Like most sensitivity analysis techniques, the method utilized in this study relies on parametric assumptions that the outcome is structured linearly, and that treatment effects do not vary with levels of the unobserved covariate after conditioning on observed measures. This potential for model misspecification may introduce its own form bias, and future research should potentially structure the outcome in more complex patterns. However, the developers of the technique have demonstrated its efficacy through a variety of simulations and have outlined the benefits of the approach.

First, while some criticize sensitivity analysis as being too subjective to the values of the parameters that an investigator specifies, this analysis provides a variety of estimates, summarized in a contour plot, that include which pairs of parameters result in non-significant or null treatment effects. This provides any interested reader the opportunity to judge whether he or she thinks a plausible independent confounder of that magnitude is likely present and, accordingly, how sensitive the conclusions of a study are to unmeasured confounding. This transparency does not afford investigators the option to simply choose a few potential parameters that conveniently provide a corrected estimate consistent with prior results. Second, the approach allows one to assess how potential sensitivity parameters compare to measured covariates. Adjusting for baseline characteristics, especially measures of the outcomes of interest prior to treatment, is important when examining the impacts of educational interventions on student achievement (Schochet, 2010). In this study, strong pre-treatment measures of standardized

reading and math achievement, as well as prior attendance rate, provide reference for how strong an independently associated confounder would have to be to explain away a provisional effect estimate. Finally, because there are likely individual-level differences in the expectations of benefits from treatment between those that participate in debate and those that do not, the method used in this study makes it possible to address concerns about differential treatment effect bias by focusing on the treatment group as the target population parameter. Targeting the specific demographic that typically participates in the activity is suitable for administers of programs as estimates apply only to the population students that they serve. Approaches that interpret estimates as average treatment effects for all students are subject to unknown amounts of additional selection bias

Educational researchers have long been interested in whether participation in programs contributes to higher academic outcomes for students. However, the answer to this question is often hampered by selection bias. Given the widespread concern with the strong and untestable assumption of ignorability, it is surprising that sensitivity analysis methods have not become more common among applied researchers when pursuing causal questions with non-experimental data. Sensitivity analyses, such as the one carried out in this study, can supplement many observational studies by determining how robust findings are to unmeasured confounding. In this case, the empirical analysis above strengthens the evidence for participation in debate and indicates that Urban Debate Leagues may be a promising means to improve educational trajectories for some of the country's most needy students.

References

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.

Bloom, H. S., Richburg-Hayes, L., & Black, A. R. (2007). Using covariates to improve precision for studies that randomize schools to evaluate educational interventions. *Educational Evaluation and Policy Analysis, 29*(1), 30–59.

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

Carnegie, N.B., Harada, M., Hill, J. (2016). Assessing sensitivity to unmeasured confounding using a simulated potential confounder. *Journal of Research on Educational Effectiveness; 9(*3):395–420.

Denault, A. S., & Poulin, F. (2009a). 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.

Denault, A., & Poulin, F. (2009b). Predictors of adolescent participation in organized activities: A five-year longitudinal study. *Journal of Research on Adolescence, 19*(2), 287–311. doi:10. 1111/j.1532-7795.2009.00597.

Dorie, V., Carnegie, N. B., Harada, M. and Hill, J. (2016). A flexible, interpretable framework for assessing sensitivity to unmeasured confounding. *Statistics in Medicine 35*, 53-70.

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.

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.

Harding, D. (2003). Counterfactual models of neighborhood effects: the effect of neighborhood poverty on high school dropout and teenage pregnancy. *American Journal of Sociology 109*: 676–719.

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

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

Lin, D., Psaty, B., & Kronmal, R. (1998). Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics, 54*(3), 948-963. doi:10.2307/2533848.

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, MD.

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. *The 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. & Winship, C. (2015). *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge: Cambridge University Press.

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

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.

Rosenbaum, P. R. (1986). Dropping out of high school in the United States: An observational study. *Journal of Educational Statistics 11*(3):207-24. doi: 10.3102/10769986011003207.

Rosenbaum, P.R. (1987). Sensitivity analysis for certain permutation inferences in matched observational studies. *Biometrika 74*(1):13-26. doi: 10.1093/biomet/74.1.13.

Rosenbaum, P. R. & Rubin, D. B. (1983). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. J*ournal of the Royal Statistical Society*, Series B 11, 212-218.

Schochet, P. Z. (2010). The late pretest problem in randomized control trials of education interventions. *Journal of Educational and Behavioral Statistics, 35*(4), 379–406.

Shackelford, D. T. (2019). The BUDL effect: Examining academic and engagement outcomes of preadolescent Baltimore Urban Debate League participants. "Forthcoming." *Educational Researcher.*

Shackelford, D.T., Ratliff, S., & Mezuk, B. (2019). *Participating in a high school debate program and college matriculation and completion: Evidence from the Chicago Debate League.* Manuscript submitted for publication.

VanderWeele, T. J. (2015). *Explanation in Causal Inference: Methods for Mediation and Interaction*. New York: Oxford University Press.

VanderWeele, T. J. & Ding, P. (2017). Sensitivity analysis in observational research: Introducing the E-value. *Annals of Internal Medicine 167*(4):268-74. doi: 10.7326/M16-2607.

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

Figures

Figure 1. *Causal graph in which the effect of treatment Z on outcome Y is confounded only by observed variables X (treatment selection is ignorable).*

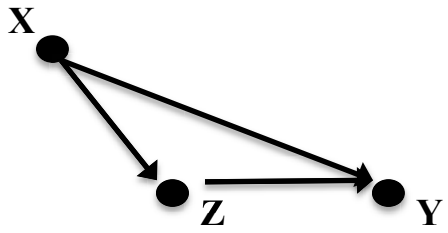


Figure 2. *Causal graph in which the effect of treatment Z on outcome Y is confounded by observed variables X and an unobserved variable U (treatment selection is non-ignorable).*

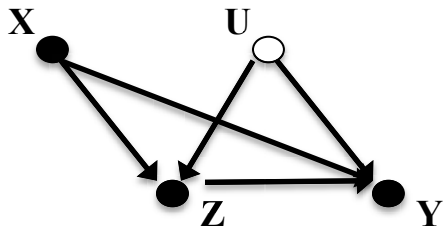


Figure 3. *Sensitivity analysis results that target the ATT of preadolescent (EMS) debate participation on SAT Verbal scores.*

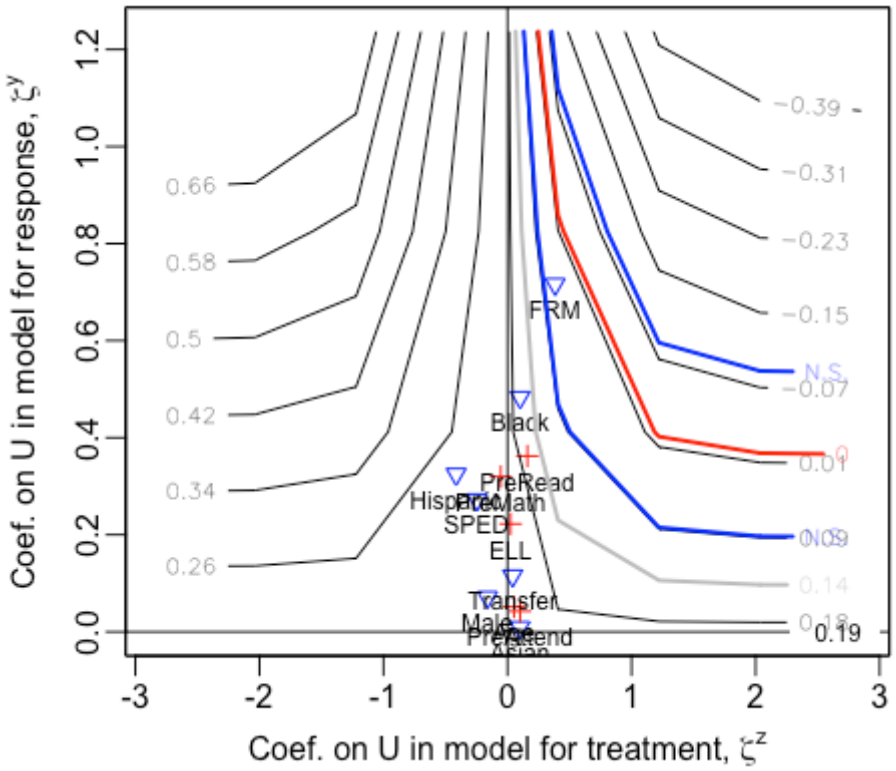


Figure 4. *Sensitivity analysis results that target the ATT of adolescent (HS) debate participation on SAT Verbal scores.*

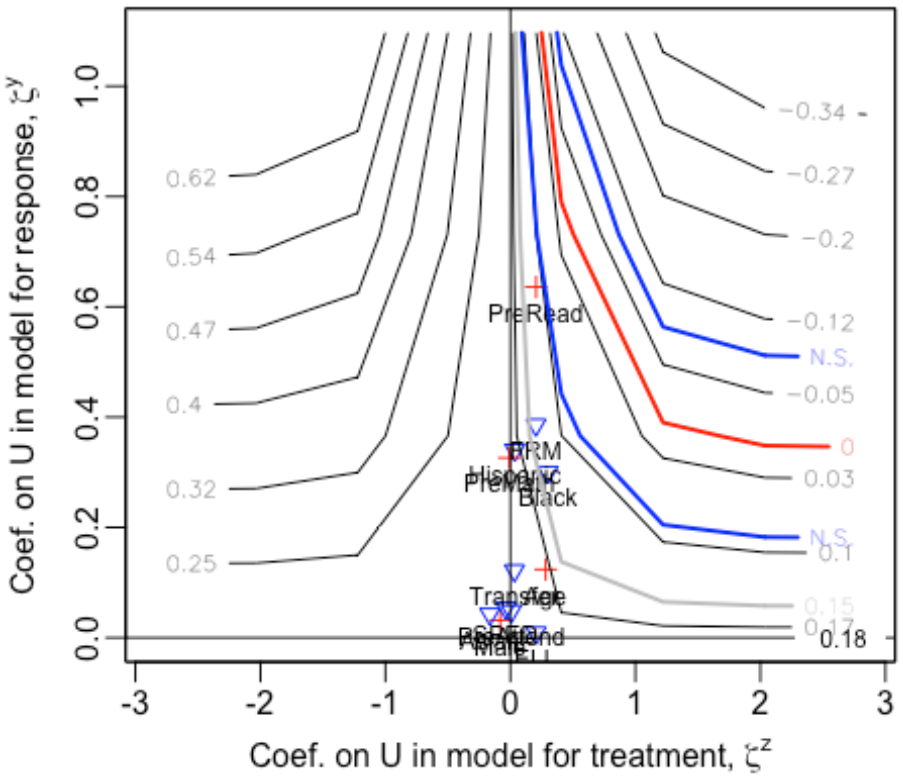


Figure 5. *Sensitivity analysis results that target the ATT of preadolescent (EMS) debate participation on SAT Math scores.*

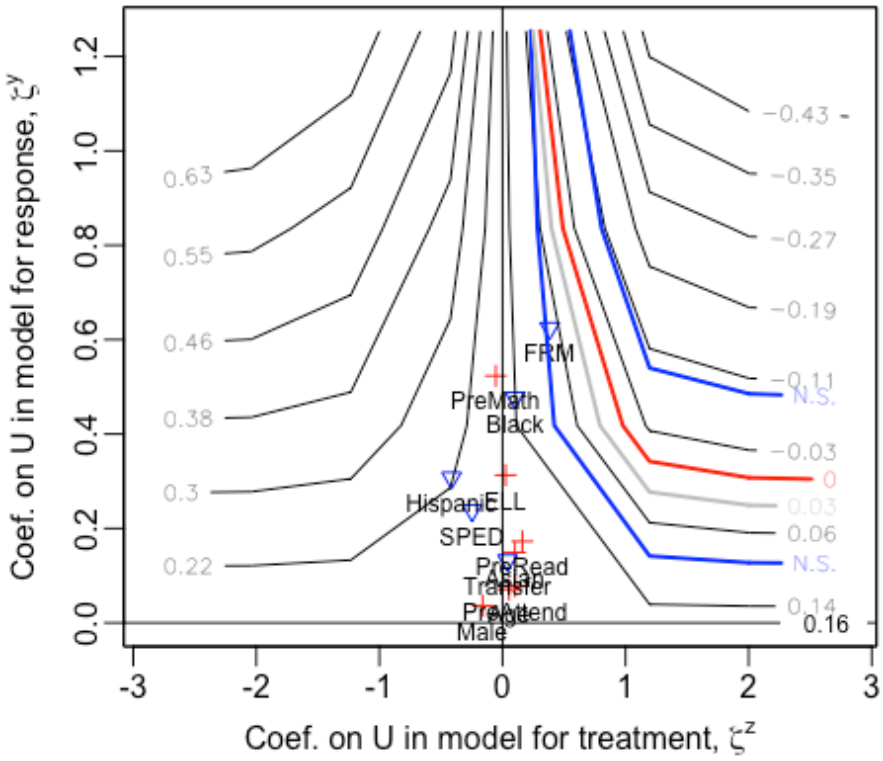


Figure 6. *Sensitivity analysis results that target the ATT of adolescent (HS) debate participation on SAT Math scores.*

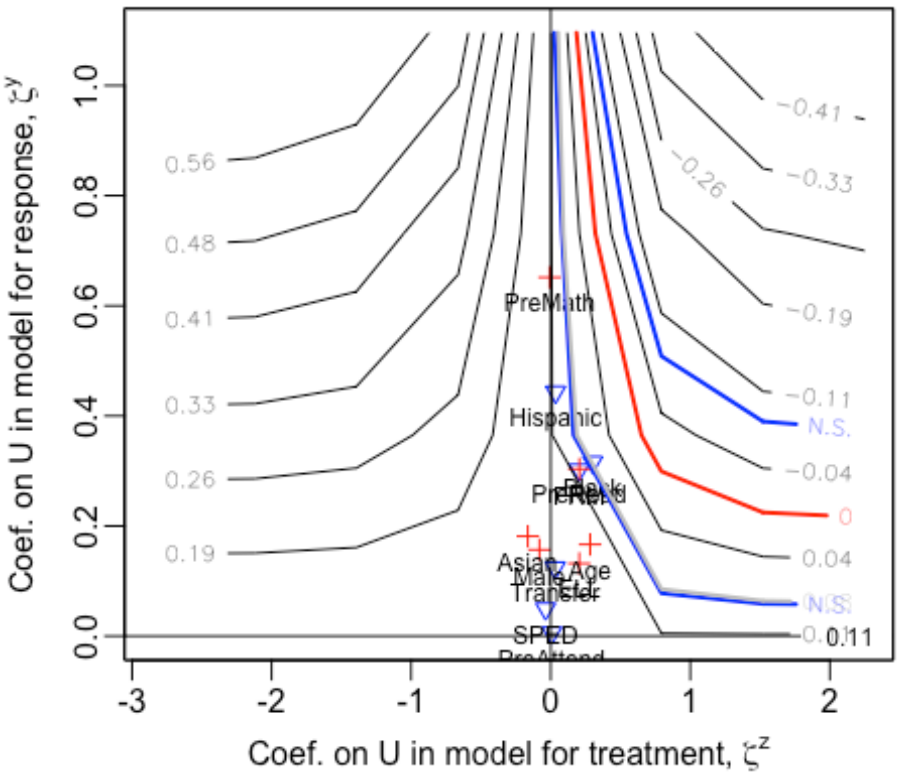


Figure 7. *Sensitivity analysis results that target the ATT of preadolescent (EMS) debate participation on high school attendance rate.*

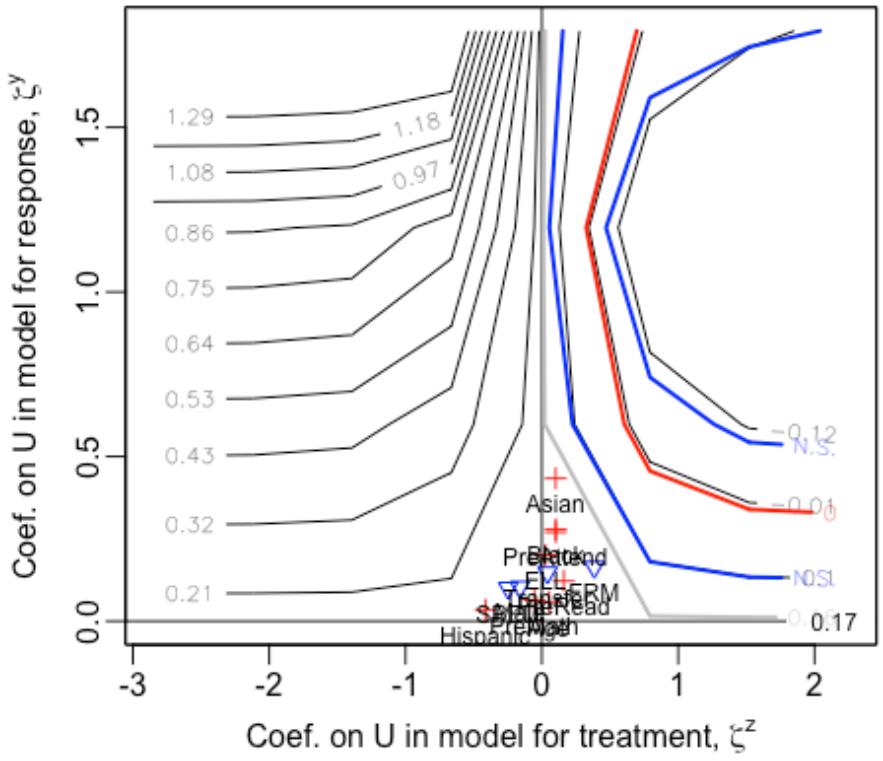


Figure 8. *Sensitivity analysis results that target the ATT of adolescent (HS) debate participation on high school attendance rate.*

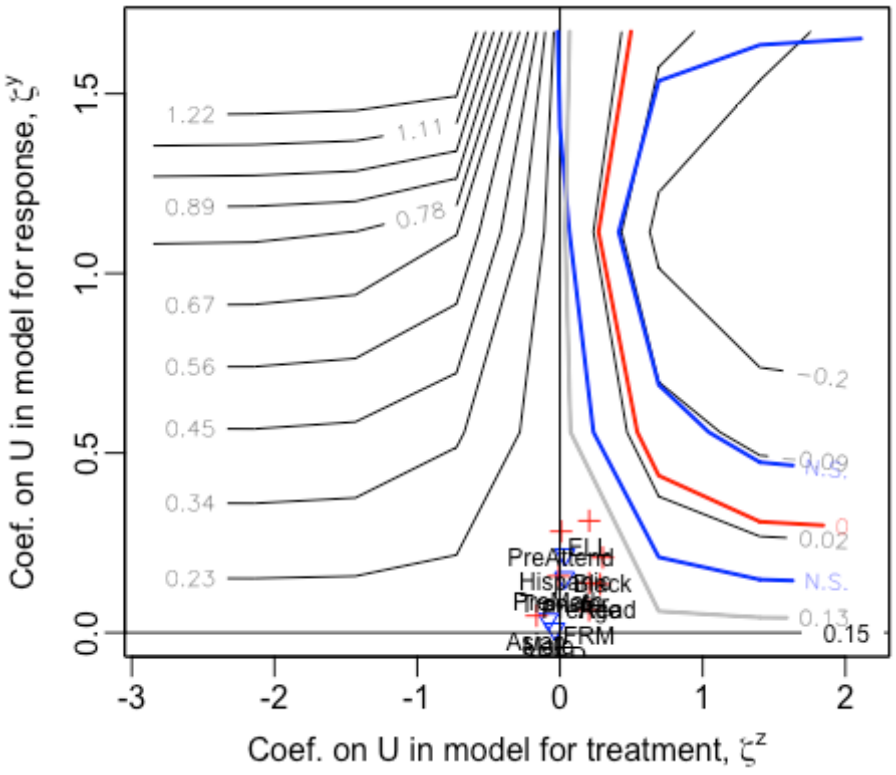


Table 1. *Summary of descriptive statistics for debaters and non-debaters in both samples.*

	Preadolescent Sample				*Adolescent Sample*			
	EMS Treatment (n = 1,929)		*EMS Control (n = 69,160)*		*HS Treatment (n = 2,710)*		*HS Control (n = 93,259)*	
VARIABLES	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
Covariates								
Age	21.35	2.74	21.70	2.89	25.21	3.69	23.67	3.88
Male	0.400		0.511		0.418		0.500	
Race and Hispanic origin								
Asian	0.006		0.009		0.007		0.011	
Hispanic	0.012		0.026		0.008		0.036	
Black	0.925		0.870		0.913		0.876	
White	0.057		0.095		0.072		0.077	
English Language Learner	0.016		0.029		0.010		0.040	
Special Education Services	0.140		0.235		0.146		0.215	
Free or Reduced-Price Meals	0.961		0.947		0.903		0.890	
Previously Transferred Schools[a]	0.359		0.344		0.218		0.220	
Pre-debate measures[b]								
MSA Pre-Read Score	405.04	30.55	394.54	30.68	397.97	30.34	389.61	31.69
MSA Pre-Math Score	392.01	38.18	379.85	39.75	392.80	39.22	386.34	39.09
Pre-Attendance Rate	94.67	5.34	93.36	6.67	92.18	8.97	91.46	8.87
Outcomes								
SAT Verbal Score	399.39	96.13	362.89	91.76	394.05	103.32	359.25	94.37
SAT Math Score	385.52	97.75	351.64	93.93	377.39	103.78	348.94	96.20
Attendance Rate (Grades 9-12)	83.55	19.45	77.00	22.86	86.37	15.45	78.30	23.56

Notes: a. Binary measure indicating whether a student transferred BCPSS schools during grades 1-3 for the preadolescent sample and grades 6-8 for the adolescent sample. b. MSA Pre-Read and Pre-Math scores measured at grade 3 for the preadolescent sample and at grade 8 for the adolescent sample. Pre-Attendance Rate consists of grades K-3 for the preadolescent sample and grades 4-8 for the adolescent sample.

Table 2. *Logit-Coefficients for propensity score models of debate participation.*

	Preadolescent Sample (n =71,089)	*Adolescent Sample (n = 95,969)*
VARIABLES	Logit-Coefficients	Logit-Coefficients
Age	0.0121	0.293*
	(0.0181)	(0.0195)
Male	-0.149*	-0.115*
	(0.0559)	(0.0531)
Asian	-0.191	-0.202
	(0.306)	(0.220)
Hispanic	-0.435	0.0578
	(0.282)	(0.319)
Black	0.243*	0.352*
	(0.114)	(0.111)
English Language Learner	0.109	0.143
	(0.238)	(0.417)
Special Education Services	-0.278*	-0.0465
	(0.0822)	(0.0712)
Free or Reduced-Price Meals	0.436*	0.301*
	(0.145)	(0.143)
Previously Transferred Schools	0.139*	0.0898
	(0.0571)	(0.0546)
MSA Pre-Reading Score	0.00716*	0.0135*
	(0.00142)	(0.00130)
MSA Pre-Math Score	0.00314*	0.00202
	(0.00127)	(0.00125)
Pre-Attendance Rate	0.0335*	0.00118
	(0.00486)	(0.00377)
Constant	-11.55	-13.67
Model Chi-square	402.55	438.18
Degrees of Freedom	12	12

Notes: Robust se in parentheses, and * $p < 0.05$ for two-tailed tests with null of 0.

Table 3. *Balance achieved by weighting.*

	Preadolescent Sample				*Adolescent Sample*			
	EMS Treatment (n = 1,929)		*EMS Control (n = 69,160)*		*HS Treatment (n = 2,710)*		*HS Control (n = 93,259)*	
COVARIATES	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
Age	21.35	2.74	21.35	2.87	25.21	3.69	25.21	3.62
Male	0.400		0.400		0.418		0.418	
Race and Hispanic origin								
Asian	0.006		0.006		0.007		0.007	
Hispanic	0.012		0.012		0.008		0.008	
Black	0.925		0.925		0.913		0.913	
White	0.057		0.057		0.072		0.072	
English Language Learner	0.016		0.016		0.010		0.010	
Special Education Services	0.140		0.140		0.146		0.146	
Free or Reduced-Price Meals	0.961		0.961		0.903		0.903	
Previously Transferred Schools[a]	0.359		0.359		0.218		0.218	
Pre-debate measures[b]								
MSA Pre-Read Score	405.04	30.55	405.09	31.63	397.97	30.34	398.00	29.14
MSA Pre-Math Score	392.01	38.18	392.05	39.96	392.80	39.22	392.84	36.93
Pre-Attendance Rate	94.67	5.34	94.67	4.87	92.18	8.97	92.18	7.96

Notes: a. Binary measure indicating whether a student transferred BCPSS schools during grades 1-3 for the preadolescent sample and grades 6-8 for the adolescent sample. b. MSA Pre-Read and Pre-Math scores measured at grade 3 for the preadolescent sample and at grade 8 for the adolescent sample. Pre-Attendance Rate consists of grades K-3 for the preadolescent sample and grades 4-8 for the adolescent sample. Means and standard deviations are weighted by the estimated ATT weight in order to demonstrate achieved balance.

Table 4. *Estimates for the average treatment effect for the treated of debate participation on high school outcomes.*

	Preadolescent Sample (n =71,089)		*Adolescent Sample (n = 95,969)*	
OUTCOMES	*Variable-scaled Coefficients*	*Effect Size in SD*	*Variable-scaled Coefficients*	*Effect Size in SD*
SAT Verbal Score	17.11*	0.19	16.69*	0.18
	(1.64)		(1.89)	
SAT Math Score	14.69*	0.16	10.73*	0.11
	(1.92)		(1.86)	
Attendance Rate (Grades 9-12)	3.78*	0.17	3.42*	0.15
	(0.47)		(0.67)	

Notes: Robust se in parentheses, and * $p < 0.05$ for two-tailed tests with null of 0.

Table 5. *Summary of sensitivity analysis for ATT estimates of debate participation on high school outcomes.*

	Preadolescent Sample (n =71,089)		*Adolescent Sample (n = 95,969)*	
OUTCOMES	*ATT EFFECT SIZE*		*ATT EFFECT SIZE*	
	Before Inclusion of Potential U	*After Inclusion of Potential U*	*Before Inclusion of Potential U*	*After Inclusion of Potential U*
SAT Verbal Score	0.19*	0.14*	0.18*	0.15*
SAT Math Score	0.16*	0.03	0.11*	0.08
Attendance Rate (Grades 9-12)	0.17*	0.16*	0.15*	0.13*

Notes: * $p < 0.05$ for two-tailed tests with null of 0.

BIOGRAPHY

Daniel Shackelford was born in 1990.

Daniel did his undergraduate work at the University of Utah where he majored in Political Science and Social Science Teaching, and minored in History. After graduating, he moved to Detroit to teach high school social studies through the Teach for America program. Daniel founded a debate team at the high school he taught at and served as the Head Coach for two years. He earned his Masters Degree in Educational Studies: Urban Pedagogies from the University of Michigan. In 2015, Daniel began his Ph.D. at Johns Hopkins University.