Gender Disparities in Competitive High School Debate

New Evidence from Comprehensive Tournament Results Data

Daniel Tartakovsky[*]
Econ 980b: Education in the Economy
May 9, 2016

[*] A.B. Candidate in Economics, Harvard College, Class of 2017. Thank you to Chris Palmer for providing Tabroom.com data and to Professor Claudia Goldin, Professor Lawrence Katz, and Priya Shanmugam for helpful comments on earlier drafts. Any errors are mine.

Abstract

This paper examines a comprehensive dataset of results from competitive high school debate competitions. On average, I find that women are about 4 percentage points less likely than men to win preliminary debate rounds. A win-loss gap of 2.5 to 3 percentage points persists despite controlling for individual and tournament characteristics, suggesting that the results are not solely driven by differences in debate experience or the quality of debate programs. An analysis of the 2016 graduating cohort suggests that part of the gender gap a result of differential attrition: Women who debate at least once as sophomores are 2.5 percentage points less likely than men to debate as juniors. Moreover, the gap appears to be much larger in rounds 1 and 2 of a tournament than in other preliminary rounds, though it is unclear whether this is a result of judge biases, aggregate preparation differences between male and female competitors, or other reasons. Finally, there is no evidence that being assigned a female judge improves female debaters' performance in specific rounds.

I. Introduction

Anecdotal evidence suggests the existence of a serious gender gap in competitive success in the high school debate community.[1] Only one woman finished in the top sixteen in the Lincoln Douglas division at the 2016 Tournament of Champions, the premier national debate competition held at the end of April each year in Lexington, Kentucky.[2] Discussions of gender disparities abound (see, e.g., Timmons & Boyer 2013). Yet few have comprehensively, quantitatively examined the determinants of success in Lincoln Douglas debate. While Shin (2016) demonstrates a positive association between gender and competitive success in Lincoln Douglas tournaments, further research is needed to determine whether this association is partly driven by confounding variables and to examine potential solutions. There are many possible theoretical explanations for gender-related differences in performance: Judges may be biased against women, male debaters may train more or have better access to coaches, women who experience overt sexism or more latent hostility may choose to leave the activity, etc. Without an empirical analysis, it is impossible to either quantify the gender gap or explore its potential underlying mechanisms.

---

[1] To clarify, in this paper, "gender gap" refers to gender identity since the data I use asks for self- or coach-reported gender. The data does not distinguish between biological sex and gender identity. It is probably safe to interpret summary statistics as applicable to either a "biological sex gap" or a "gender gap" given the small size of the population whose biological sex and gender identity differ, but there is no way to know this with certainty. Moreover, I exclude observations where gender is labeled "Other" due to concerns about reporting accuracy and sample size. For instance, only 5 debate rounds in my sample were judged by a judge whose gender was labeled "Other." While some of these observations may capture students or judges who identify as genderqueer or otherwise, others may simply be errors in data entry. Gates & Steinberger (2010) have documented substantial errors in gender reporting on U.S. Census data that pose large challenges to studying same-sex couples, and similar problems apply to this paper's dataset.

[2] See "Final Places in Lincoln Douglas," Tournament of Champions, 2016, Tabroom.com.

This paper comprehensively quantifies the gender gap in high school Lincoln Douglas debate for the first time. I use an individual-level dataset of results from highly competitive national debate competitions in the Lincoln Douglas format. Although this paper deals with the somewhat narrow topic of gender gaps in one segment of a competitive high school extracurricular activity, it lies at the intersection of several parts of the labor economics literature. First, it relates to previous research on inequality in outcomes in both academic competitions (e.g., Smith 2013) and to research on sex-based hiring (e.g., Goldin and Rousse 2000). Second, the paper ties into research on the economics of education. In particular, previous studies have linked instruction by a same-gender teacher to improved student performance and teachers' perceptions of students (Dee 2007). Others have found mixed or negative results (Antecol et al 2012). This paper studies the effect of a female judge on female students in a high-stakes academic environment.

Competitive debate is also a particularly attractive setting in which to examine gender stereotypes. Many of the factors cited as barriers to female success in the workplace, including criticism of attire, voice, and tone, may exist in debate because of its focus on unavoidably subjective evaluations of communicative and argumentative skills. This paper's focus on national-level debate tournaments combines education and competition and thus bridges the gap between the two sets of academic literature.

Using data from 89 major debate competitions, I find that women are about 4 percentage points less likely than men to win debate rounds and 2.5 percentage points more likely to stop debating on the National Circuit after their sophomore year of high school. An overall win-loss gap of 2.5 to 3 percentage points persists despite controlling for time-invariant individual and tournament characteristics as well as tournament year but seems concentrated in the first two

rounds of competitions. It is difficult to causally interpret the finding that women do worse in the first two rounds of a debate tournament than in later rounds. However, being assigned a female judge does not seem to improve female debaters' performance, suggesting that the causes of the female performance gap are deeper than merely a lack of female judges. Results are robust to various specifications and control variables.

The rest of the paper is structured as follows. Section II gives a brief overview of Lincoln Douglas Debate, Section III describes the data, Section IV presents and interprets summary statistics, and Section V outlines my empirical approaches and presents results. Section VI concludes.

II. Overview of Lincoln-Douglas Debate

Lincoln-Douglas Debate is one of the three most popular debate styles in the United States, the others being Public Forum Debate and Policy Debate. Lincoln-Douglas (LD) is unique in that it is uses one-on-one format while other debate events involve teams of two. Traditionally, LD has emphasized debate topics that deal with ethical controversies such as whether animals deserve rights. These topics are called resolutions. In recent years, LD at the national level has focused more on policy controversies.[3] Resolutions typically last for a period of two months.[4] Students debate the same topic at tournaments held within a given two-month period. Some tournaments have multiple divisions (Novice, Junior Varsity, and Varsity). In general, debaters in their first year of debate may choose to compete in the Novice division, and debaters in their first two years of debate may choose to compete in the Junior Varsity division.

The National Circuit

I limit my analysis to the National Circuit, a group of tournaments held in various states that offer "bids" to the National Tournament of Champions (TOC). The Tournament of Champions is the culmination of the National Circuit season and is held at the end of each April in Lexington, Kentucky. In order to qualify to the TOC, debaters must typically earn two such bids by performing well at National Circuit tournaments.[5] The most competitive tournaments

---

[3] For instance, this year's January/February topic is "In the United States, private ownership of handguns ought to be banned." See "Past Lincoln-Douglas Debate Topics."

[4] September/October, November/December, January/February, and March/April. The end-of-year national championship, the Tournament of Champions, uses the January/February topic.

[5] There are some exceptions. First, a few students each year who have only earned one bid apply for an "at-large" qualification to the TOC. A committee of debate coaches evaluates and accepts applicants. In 2016, 13 at-large applicants were accepted. See "2016 Accepted At-Larges." Second, students can also qualify by earning one "ghost bid" and one regular bid. A ghost bid is earned when two students from the same school are paired against each other in the round before bids are awarded (i.e., in quarterfinals if the tournament offers bids to all semifinalists). The

generally offer bids to those who reach octafinals (the top sixteen) , while others offer bids for debaters who reach quarterfinals, semifinals, or finals. While tournaments on the national circuit are by no means identical to each other, many of the same competitors and judges appear at different tournaments to qualify to the TOC and to gain additional competitive exposure. It is thus relatively easy to track individuals on the circuit over time.

Individual Debates

Individual debates have the same general format. Table 1 lists the name of each speech as well as the speech's length and general purpose. In aggregate, debaters each have 13 minutes of speaking time per debate. At the end of the debate, the judge gives one debater the win and the other the loss. The judge also assigns "speaker points," generally on a scale of 25-30.[6] The judge then explains his or her reasoning to debaters and lets debaters ask him or her questions about the debate. Speaker points are used to determine speaker awards and distinguish among debaters with the same win-loss record. Speaker points allow judges to reward or penalize debaters for their speaking styles, creative argumentative strategies, or whatever other criteria the judge finds relevant separate from the win/loss.

Preliminary Rounds

While individual tournaments differ slightly in their design, most have six preliminary debates followed by elimination rounds. The first two preliminary debates are paired randomly ("preset"). In other words, there are no restrictions on who can debate whom, except that in even-numbered rounds, debaters will defend the side of the topic opposite to that which they

---

debater who advances to semifinals earns a regular bid, while the quarterfinalist earns a "ghost bid" that converts to a regular bid when that student receives another regular bid.

[6] Judges can give speaker points as low as 0 but this is almost never done. In my dataset the average is around 28.

defended in the previous debate. In other words, someone who argues on the affirmative (in favor of the topic) in round 1 will argue for the negative (against the topic) in round 2 and vice-versa. At tournaments with six preliminary rounds, each debater thus debates three times on each side. After the two preset rounds, the rest of the preliminary rounds are "power-paired." Power-pairing means (1) debaters only debate others with their same win-loss record and (2) debaters with the same record are paired high-low according to their previous total speaker points.[7] In other words, debaters who win their first two debates have a 2-0 record and, in round 3, can only debate others with a 2-0 record. The 2-0 debater with the highest speaker points will debate the 2-0 debater with the lowest speaker points, and so on. At the end of six preliminary rounds, debaters with a 4-2 record or better generally advance to elimination rounds and are ranked in an elimination round bracket according to their record and their speaker points.[8] Figure 1 provides a visual depiction of the typical tournament format where the first elimination round is a Round of 16.[9] Figure 2 shows an example of a tournament bracket from the 2013 TOC.

Elimination Rounds

Elimination rounds are sudden-death: Debaters who lose are immediately eliminated from the tournament. Elimination debates are almost always judged by a panel of three judges.

---

[7] Having the same win-loss record as someone is referred to as being in the same "bracket." If there is an odd number of debaters in a bracket, one person will be "pulled up" or "pulled down" to debate someone in a different bracket. If there is an odd number of debaters in the 0-win bracket, one debater in the bracket will usually receive an auto-win (a "bye").

[8] A debater's highest and lowest speaker points are usually excluded for the purposes of determining seeding and speaker awards. This is meant to correct for extreme outliers.

[9] Frequently there will be a "run-off" round to determine the Round of 16. In other words, if there are 18 debaters with 4-2 records or better, the bottom 4 will debate each other to determine which two will take the 15th and 16th slots. The same could occur with the Round of 32 depending on the tournament's size.

The winner of a coin-flip chooses which side of the resolution to defend.[10] If a tournament offers bids at the quarterfinal level, debaters who reach quarterfinals receive one bid to the Tournament of Champions. While limited evidence suggests that gender disparities are most stark in elimination rounds, given time and data constraints I mostly analyze preliminary rounds in this paper.[11]

[10] The exception is that if two debaters oppose each other during preliminary rounds, they will be locked into debating each other on the opposite side during elimination rounds.

[11] I specifically examine rounds 1-6 in order to make results comparable across tournaments (i.e. I exclude results from round 7 of tournaments that have 7 preliminary rounds).

III. Data

Tabroom.com is a comprehensive platform for managing debate tournaments.[12] Tournament directors use Tabroom to set up tournament webpages, hire judges, and communicate with tournament participants and coaches. Since judges use Tabroom to submit their decisions during a tournament, the website stores debaters' names, schools, wins and losses, and speaker points, as well as judges' names. While Tabroom accommodates many debate formats, I focus on Lincoln Douglas debate. First, Lincoln Douglas is a one-on-one format, which simplifies analyzing gender disparities by avoiding team events with mixed-gender competitors and more complicated forms of ranking. Second, appropriately matching schools that are the same but labeled slightly differently, which occurs frequently in the data, is easier with a smaller number of schools and reduces the likelihood of errors in matching.[13]

In addition to renaming schools to ensure debaters within the same program are grouped correctly, I alter the dataset in two ways. First, I select a sample of tournaments. Second, I adopt several approaches to label missing genders for debaters and judges.

In order to include a comprehensive sample of competitions, I cross-reference the Tabroom.com dataset with a list of National Circuit tournaments that offer bids for the 2015-16 season listed on the Tournament of Champions website.[14] Although many smaller tournaments use other tournament tabulation software, all but one of the tournaments that offer bids to

---

[12] Tabroom.com is now the most popular system for National Circuit LD tournaments, but some tournaments use a different system called Joy of Tournaments, so not all competitions appear in the data.

[13] Since I am familiar with many of the schools on the national circuit, I can manually verify the accuracy of the data. Appendix A contains detailed information on how I adjusted school names in the data.

[14] While the Tabroom data does not explicitly label tournaments as on or off the National Circuit, only 51 tournaments offer bids to the Tournament of Champions, so it is easy to manually compare the dataset with the list of tournaments.

quarterfinalists or round of 16 participants appear in the Tabroom.com data. Appendix C contains a comprehensive list of tournaments I use.

Since Tabroom allows but does not require coaches to indicate the gender of debaters or judges, about 21% of observations for debaters and 50% for judges are initially missing gender labels. I adopt three strategies to assign genders to missing observations. First, I use 1990 Census data containing about 5,500 common baby names. The Census data corresponds to people who were 25 years old in 2015, which is a reasonable approximation for judges (who are often college or graduate students) as well as debaters.[15] In cases where the same name appears in both the male and female Census lists, I assign the more common gender associated with the name. Second, I merge the Tabroom data with a list of common names of South Asian origin I found on Github, a website where programmers and researchers can share code and datasets. Third, I manually assign gender in what I believe are clear-cut cases.[16] Table 2 contains results from these steps. After the three procedures, 99% of debaters and 96% of judges have assigned genders. The vast majority of the improvement is due to the official Census data.

[15] Although debaters are 14-18 years old rather than 25, this would only be a problem if a substantial fraction of names were unisex *and* the popularity of those unisex names being assigned to men as opposed to women switched between 1990 and 2000. This seems to be a minor issue since only about 1% of Americans have a unisex name (see "Most Common Unisex Names").

[16] The list of names I assign manually is available upon request. These cases include what are likely to be misspellings or abbreviations of debater and judge names by coaches (e.g., "Micahel" becomes "Michael," "Catheri" becomes "Catherine," and "Danella" becomes "Daniella").

IV. Summary Statistics

General Summary Statistics

Table 3 reports summary statistics for men and women separately. Column 3 reports results for a t-test for difference in means between the two groups. The table contains three notable results. First, men comprise about 60% of the competitors and an even higher fraction of total observations, since men compete in 42 preliminary rounds while women compete in about 35. Second, men win a higher fraction of debates: There is a 3.7 percentage point male-female win gap in preliminary rounds. Finally, the performance gap in elimination rounds is even larger. Men are 12 percentage points more likely to win an elimination round than women.

Attrition Over Time?

While it is impossible to understand the determinants of this gender gap without an empirical model that includes potential confounding variables, the above evidence suggests that there are in fact substantial performance differences by gender. One explanation for differential performance may be that women are less likely to continue debating for all four years of high school. If male debaters are more likely than female debaters to persist in the activity, all else being equal they will accumulate more experience and perform better. Two data points suggest that this is the case. First, men graduate on average 2 months earlier than women.[17] Moreover, 46% of current high school freshmen are female compared to only 33% of those who graduated high school last year. However, this statistic could be misleading: If more women have begun competing in Lincoln-Douglas debate over the last few years, we would expect there to be relatively more young female debaters even absent differential attrition. To resolve this issue, I

[17] The average male debater's graduation year is 2016.39 compared to 2016.55 for the average female debater ($p<0.001$). This translates to approximately a 2-month gap.

restrict the sample to the cohort of debaters graduating from high school in 2016. I then calculate a "participation gap," defined as the difference between fraction of male and female debaters who, conditional on having debated as sophomores, also debate as juniors and as seniors. Table 4 shows that women who debated in at least one tournament as sophomores are about 2.5 percentage points less likely than men to debate as juniors. However, the participation gap does not seem to grow from junior to senior year. There is thus some evidence that women are more likely to quit National Circuit Lincoln-Douglas debate than men. Still, the lack of an increase participation gap between junior and senior year cautions against too strong of an interpretation of these results. Moreover, it is impossible to see whether women switch to a different kind of debate or stop debating entirely, so the information is imperfect.

A Larger Performance Gap in Preset Rounds?

Table 5 reports a final set of summary statistics that divides rounds into those that are randomly paired ("preset" rounds 1 and 2) and those that are power-paired (rounds 3 through 6). The male-female win-loss gap is 5.7 percentage points in rounds 1 and 2, twice as large as the gap in rounds 3 through 6. While it is impossible to know whether this represents a meaningful difference without accounting for sources of omitted variable bias, there are at least two possible explanations for this pattern. The first, which I call the "preparedness thesis," holds that large differences in preparation for tournaments would, other things equal, lead to larger win-loss gaps by gender in rounds 1 and 2 than in later debates. Since power-paired rounds match debaters according to skill as determined by their previous performance at the tournament, if women are less prepared to compete at a given tournament they would perform worse in preset rounds and

better in power-paired rounds.[18] The second possible explanation, which I call the "judge bias thesis," holds that judge discrimination against women would result in larger performance gaps in rounds 1 and 2 relative to later rounds. Suppose a female debater loses a preset round due to a judge's unconscious bias against her. All else equal, in a later power-paired round the female debater's skill level should exceed that of a male opponent whose record would reflect his skill alone as opposed to a some weighted average of skill and harm from past judge discrimination.[19] In other words, power-pairing would undo some of the effect of judge biases by pairing female debaters against less skilled opponents in subsequent rounds.[20] While it is impossible to definitively determine which of these two effects dominates or whether both are significant, controlling for graduation year, school, and other variables that serve as proxies for pre-tournament preparation could help shed light on whether the "preparedness" or "judge bias" thesis is more plausible.

In sum, basic summary statistics suggest that there is a substantial gender-based performance gap at Lincoln-Douglas tournaments. Of course, absent controlling for confounding variables, these relationships are subject to omitted variable bias. To see whether the gender gap is robust to the inclusion of controls, I proceed by implementing two empirical strategies.

---

[18] Note that I use "preparedness" very broadly so as to encompass factors that may affect female performance but nevertheless represent barriers to female success outside of the control of female debaters themselves. For instance, male debaters could have more experience due to their age, receive more attention from their school's coaches, feel more welcome or comfortable at debate competitions, have access to better research due to larger social networks, etc. These factors could lead men to perform better than women even if judges are unbiased.

[19] In other words, past judge discrimination forces women into a lower bracket, which makes it easier for women to win future rounds because their opponents will tend to be worse.

[20] This "undoing" of bias may not occur in rounds where women debate other women, assuming women face sexism in similar ways. However, since tournaments are on average 60% male, the effect of judge biases against women should result in a lower gender gap in rounds 3 through 6 than in rounds 1 and 2 as long as women debate at least one man in those rounds.

V. Empirical Strategy and Results

To begin, I use a linear probability model to determine the association between gender and in-round performance when controlling for a set of covariates.[21] The probability that an individual *i* at tournament *j* in year *t* wins preliminary round *r* is a function of the individual's sex (F), whether the round is a preset round (S), whether the judge is female (U), and other individual (X) and tournament (Z) factors:

$$(1) \qquad \boldsymbol{P_{ijtr} = f\left(X_{it}, F_i, U_{jtr}, Z_{jtr}, S_r\right)}$$

I use this model to measure the interaction coefficient $\boldsymbol{\rho}$ between preset rounds $\boldsymbol{S_r}$ and gender $\boldsymbol{F_i}$. Individual controls $\boldsymbol{X_{it}}$ include school and graduation year. Tournament controls $\boldsymbol{Z_{jtr}}$ include tournament and year. The full model is as follows:

$$(2) \qquad \boldsymbol{P_{ijtr} = \beta F_i + +\sigma S_r + \rho(F_i \times S_r) + X_{it}\theta_1 + Z_{jtr}\theta_2 + \varepsilon_{ijtr}}$$

Table 6 presents results for this specification.[22] In the baseline case (6.1), women are about 3.8 percentage points less likely to win debates than men. Controlling for graduation year shrinks the gap to about 3.5 percentage points (6.2), and once school fixed effects are included the coefficient on female is no longer significant (6.3). On the other hand, once one controls for

[21] One shortcoming of a linear probability model is that it is not restricted to probabilities below 0 or above 1 and can thus perform poorly at low and high outcome values. While a Probit model constrains the outcome variable to the [0,1] range by applying a standard normal CDF, it is more difficult to interpret coefficients in a Probit regression. Moreover, from a theoretical standpoint it seems very unlikely that the coefficients on control variables in my regressions could predict an outcome variable outside of the [0,1] range, since the differences in win-loss percentages between any large groups (i.e., men and women or freshmen and sophomores) while large would never be something like 90 percentage points. Linear regression is thus a reasonable choice to model this data.

[22] $\boldsymbol{\theta_1}$ and $\boldsymbol{\theta_2}$ are vectors of coefficients on individual and tournament controls, respectively.

preset round and the interaction between female and preset (6.4), there remains a 3-point win-loss gap between men and women in preset rounds. This gap remains significant at the 1% level when controlling for school, tournament, year, and graduation year. The gap's robustness to these controls does not necessarily show that individual characteristics such as previous experience are irrelevant but does make the "judge bias" thesis more plausible under the assumption that the existing controls are good proxies for previous debate training.

Although the linear probability model includes a set of individual controls, it does not account for all unobserved, time-invariant individual heterogeneity. For instance, an individual's race and income may be correlated with competitive success.[23] Following Goldin & Rouse (2000), I adopt an individual fixed-effects framework, which can be thought of as following individual debaters across time and exploiting variation in variables such as judge gender to determine their effects on the individual's debate performance. Since we are interested in the effect of judge gender on female performance as well as whether women perform worse in preset rounds even when controlling for time-invariant individual characteristics, I use the following full model:

$$(3) \qquad \boldsymbol{P_{ijtr} = \alpha + \delta U_{jtr} + \gamma\left(F_i \times U_{jtr}\right) + \sigma S_r +}$$

$$\boldsymbol{\rho(F_i \times S_r) + X_{it}\theta_1 + Z_{jtr}\theta_2 + \varepsilon_{ijtr}}$$

The coefficient on the interaction between $\boldsymbol{F_i}$ and $\boldsymbol{U_{jtr}}$, $\boldsymbol{\gamma}$, measures the change in probability that a woman will win in front of a female judge relative to a male judge.[24] The coefficient on the

[23] Note that one's household income will likely vary over long periods of time, but an individual fixed-effects model can be useful if there is little variation in most debaters' household income within a three or four-year period.

[24] I do not use the term $\boldsymbol{F_i}$ by itself on the right hand side because it is perfectly determined by the left hand side variable. This model does not allow one to examine the effect of "gender" itself

interaction between $\boldsymbol{F_i}$ and $\boldsymbol{S_r}$, $\boldsymbol{\rho}$, measures the change in probability that a woman wins a preset round relative to a man.

Table 7 contains results for the individual fixed-effects model. Having a female judge is not significantly associated with female performance under any of the specifications. Nor do female judges reduce the gender gap in preset rounds, as shown by the insignificant coeffient on the triple interaction term between debater gender, judge gender, and round type. One possible interpretation of this would be that within-tournament anti-female judge bias is less important than exogenous factors such as previous debate experience. However, if judge biases are a product of how women are generally perceived (e.g., through societal stereotypes), one would not necessarily expect female judges to be unbiased arbiters. In other words, the lack of a statistically significant relationship between being assigned a female judge for a debate and winning the debate is impossible to interpret causally. Still, relative to a man, a woman is 2.58 points less likely to win a preset round. The results from the individual fixed-effects model thus lend further credence to the idea that the gender gap persists in preset rounds despite controlling for time-invariant unobservable variables that a standard linear regression would not capture.

VI. Conclusion

Summary statistics and two empirical models yield similar conclusions: men are substantially more likely than women to win debate rounds at National Circuit Lincoln Douglas debate competitions. The basic linear model puts the gap at 4 percentage points while the full linear regression with individual, tournament, and year fixed effects, as well the individual-fixed effects model, suggest a gap of 2.5 to 3 percentage points. That the gap narrows with the addition

---

because for gender does not vary for the same individual over time, so in that sense the linear model may be more informative to capture any remaining effect of gender not captured by the set of control variables.

of controls is not necessarily a positive sign for those concerned about female success, for including a control variable for debaters' graduation year seems to account for some of this narrowing. That accounting for one's graduation year reduces the negative association between being a women and winning a debate round suggests that women tend to be younger (and therefore less experienced) than male debaters, perhaps because they are less likely to continue debating past their sophomore year. A simple analysis of debaters graduating in 2016 who debated at least once as sophomores confirms that women are 2.5 percentage points less likely than men to debate the following season.

To be clear, these findings are suggestive rather than causal and leave many questions unanswered. For instance, the research does not shed light on whether judge biases are an important factor at debate competitions. The lack of evidence that having more female judges improves female performance suggests that tournament administrators and coaches concerned with female success should not expect that to eliminate the gender gap by hiring female judges. Of course, female representation at tournaments may have positive effects on female debaters outside the narrow context of individual wins and losses associated with the judge's gender. Future research could explore this possibility by tracking female success across different kinds of competitions, perhaps examining whether women benefit from there being a higher fraction of female judges, from peer effects, or from other policies tournaments adopt to reduce the gender gap.[25] The richness of Tabroom.com data would also allow researchers to compare gender gaps across different debate formats, including speech events and team debate formats, to understand to what extent the gender gap is Lincoln Douglas-specific. Gaining a more nuanced

[25] An example of a peer effect would be female debaters becoming friends with each other and helping each other succeed or feeling more comfortable in spaces where they are more represented.

understanding of why men enjoy more competitive success than women would help stakeholders in the community minimize the likelihood that women are denied opportunities and treated unfairly in an activity that is meant to empower everyone instead of leaving some people behind.

References

"2016 Accepted At Larges." 2016. Tournament of Champions. University of Kentucky.

Antecol, Heather, Ozkan Eren, and Serkan Ozbeklin. March 2012. "The Effect of Teacher Gender on Student Achievement in Primary School: Evidence from a Randomized Experiment." IZA Discussion Paper Issue 6453.

Dee, Thomas S. 2007. "Teachers and the Gender Gaps in Student Achievement." *Journal of Human Resources* 48(3): 529-554.

"Final Places in Lincoln Douglas," Tournament of Champions, 2016, Tabroom.com.

Flowers, Andrew. "The Most Common Unisex Names in America: Is Yours One of Them?" *Five Thirty Eight*, 10 Jun. 2015.

Gates, Gary J. and Michael D. Steinberger. 2010. "Same-Sex Unmarried Partner Couples in the American Community Survey: The Role of Misreporting, Miscoding and Misallocation." Williams Institute Working Paper.

Goldin, Claudia and Cecilia Rouse. Sep. 2000. "Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians." *American Economic Review* 90(4): 715-740.

"Past Lincoln-Douglas Debate Topics." 2016. National Speech and Debate Association.

Shin, Ariel. May 2016. "A Statistical Analysis of the Gender Gap." *VBriefly*. Victory Briefs Institute.

Smith, Jonathan. 2013. "Peers, Pressure, and Performance at the National Spelling Bee," *Journal of Human Resources* 48(2): 265-285.

Timmons, Cindi and Bekah Boyer. Jan. 2014. "Women in Debate: Working Toward a More Complete Picture." *VBriefly*. Victory Briefs Institute.

United States Census Bureau. "Frequently Occurring Names in the U.S. – 1990." Census.gov. Oct. 1995.

Figure 1. Lincoln Douglas Standard Tournament Format

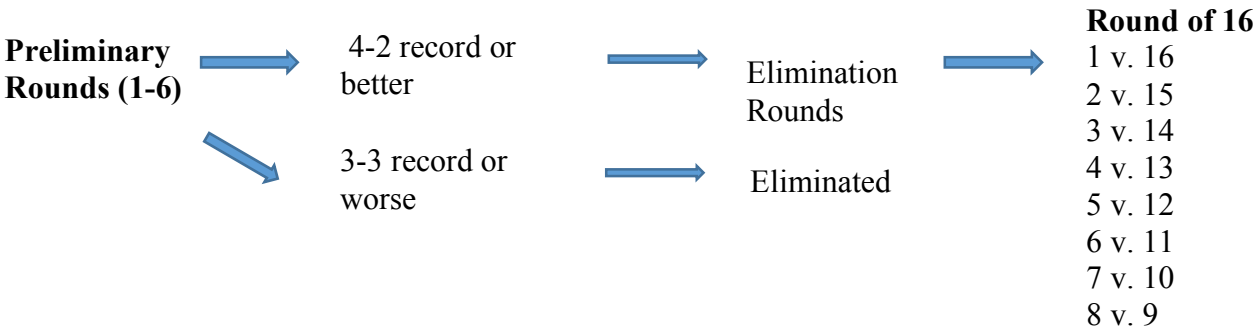

Figure 2. Sample Bracket (2013 Tournament of Champions)

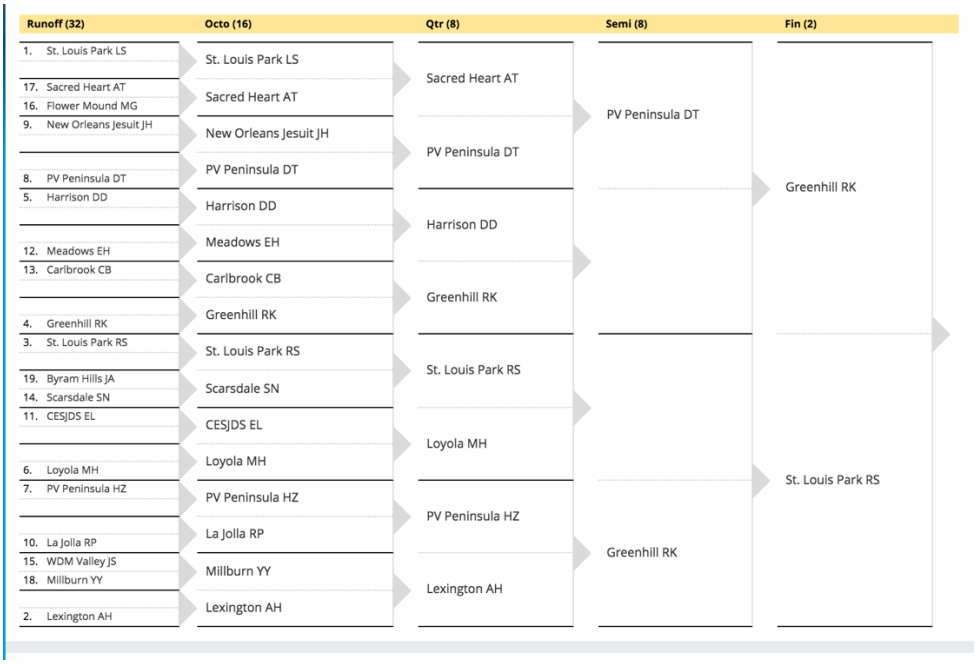

Table 1. Lincoln-Douglas Debate Format

Speech	Length (Minutes)	Speech Description
Aff Constructive (AC)	6	Affirmative establishes thesis and supporting arguments
Cross-Examination (CX)	3	Negative asks affirmative questions about the AC
Neg Constructive (NC)	7	Negative establishes thesis and answers the AC
Cross-Examination (CX)	3	Affirmative asks the negative questions about the NC
First Aff Rebuttal (1AR)	4	Affirmative defends arguments and answers the NC
Neg Rebuttal (NR)	6	Negative answers 1AR responses and concludes the debate [a]
Second Aff Rebuttal (2AR)	3	Affirmative answers NR and concludes the debate
Preparation Time	4 [b]	Debaters can use prep time in between speeches to prepare strategies [c]

Note: Table 1 lists the name, length, and a short description of each speech in the high school Lincoln-Douglas debate format.

[a] Debaters typically use final speeches to compare arguments and give the judge reasons to vote for their side.

[b] Some tournaments also use 5 minutes of prep time and a few use 3 minutes.

[c] Affirmatives typically take some time before the 1AR and some before the 2AR to plan their speeches. The same is true for negatives before the NC and NR.

Table 2. Effectiveness of Gender Assignment Procedures

	% Observations with Gender Labeled			
	Debaters	**Increase**	**Judges**	**Increase**
Initial Dataset	78.6%		49.1%	
+ 1990 Census Data	92.9%	+ 14.3 pp	88.7%	+ 39.6 pp
+ Indian Names List	95.2%	+ 2.3 pp	91.3%	+ 2.6 pp
+ Manual List	98.8%	+ 3.6 pp	96.1%	+ 4.8 pp

Note: Each number in the "increase" column denotes additional percentage point gain in observations labeled with gender in the Tabroom.com results dataset using the row's respective procedure (1990 Census data, a list of Indian names, and the manual addition of gender labels).

Table 3. Summary Statistics: Performance of National Circuit Lincoln Douglas Debaters by Gender

	(1) Men	(2) Women	(3) Difference in Means
Number of Competitors [a]	2,734	1,932	N/A
Fraction of Competitors	0.5859	0.4141	N/A
Average Number of Preliminary Rounds Won [b]	2.95 (0.0070)	2.72 (0.0085)	0.23*** (0.011)
Fraction of Preliminary Rounds Won [b]	0.515 (0.0025)	0.478 (0.0032)	0.0370*** (0.0040)
Fraction of Elimination Rounds Won	0.537 (0.0027)	0.418 (0.0417)	0.118** (0.0499)
Average Speaker Points [c]	28.07/30 (0.0071)	27.98/30 (0.0089)	0.0895*** (0.0115)
Average Number of Preliminary Rounds Debated	42.34 (0.185)	35.52 (0.219)	6.82*** (0.292)

* p<0.05, ** p<0.01, *** p<0.001

Note: Table 3 reports summary statistics for high school Lincoln Douglas debate tournament results on Tabroom.com. [a] The total number of identifiable unique competitors (those whose genders are either labeled or can be inferred using Census data) is 4,666. [b] Restricted to National Circuit tournaments with 6 preliminary rounds. While one would expect the average number of prelim rounds won to be exactly 3, some debaters drop out or leave the tournament early, likely resulting in a left-skewed win distribution. [c] Speaker points are generally awarded on a scale of 0-30; in practice, the scale is about 25-30, with 27.5-28 being an average varsity debater. Standard errors in parentheses.

Source: Tabroom.com National Circuit Lincoln Douglas Debate competition results for a sample of 89 tournaments from spanning the 2011-12 to 2015-16 seasons. See Appendix C for list of tournaments.

Table 4. Debate Participation of 2016 Graduate Cohort Over Time

(Conditional on Having Debated as Sophomores)

	(1) Men	(2) Women	(3) Difference in Means
Fraction Debating as Juniors	0.872	0.847	0.0246***
	(0.00341)	(0.00473)	(0.00572)
Fraction Debating as Seniors	0.585	0.560	0.0243***
	(0.00502)	(0.00653)	(0.00823)
Number in Cohort	692	476	

* p<0.05, ** p<0.01, *** p<0.001

Note: Table 4 reports the fraction of high school Lincoln Douglas debaters graduating in 2016 who debated as juniors and as seniors, *conditional* on having debated as sophomores. Restricted to first six preliminary rounds at National Circuit tournaments listed in Appendix C. Column 3 contains results from t-tests for difference in means with standard errors in parentheses. I define "debating as sophomores" as students graduating in 2016 who debate in at least one preliminary round from August 1, 2013 through July 31, 2014. "Juniors" is the equivalent but for the August 1, 2014 to July 31, 2015 period. "Seniors" is the equivalent but from August 1, 2015 to present. The total number of observations is 15,397.

Source: Tabroom.com National Circuit Lincoln Douglas Debate competition results for a sample of 89 tournaments from spanning the 2011-12 to 2015-16 seasons. See Appendix C for list of tournaments.

Table 5. Performance of National Circuit Lincoln Douglas Debaters by Gender: Preset versus Power-Paired Rounds

	(1)	(2)	(3)
	Men	Women	Difference in Means
Fraction of Rounds Won	0.515	0.478	0.0369***
	(0.0025)	(0.0032)	(0.0040)
Fraction of Rounds Won (Power Paired) [a]	0.511	0.484	0.0274***
	(0.0030)	(0.0039)	(0.0050)
Fraction of Rounds Won (Randomly Paired) [a]	0.522	0.465	0.0566***
	(0.0042)	(0.0054)	(0.0068)
Observations [b]	41,112	24,582	Total: 65,964

* p<0.05, ** p<0.01, *** p<0.001

Note: Table 5 reports sample means of various performance metrics for high school Lincoln Douglas debaters by gender. Column 3 contains results from t-tests for difference in means. [a] Rounds 1 and 2 of tournaments are almost always "Randomly Paired" or "Preset," meaning debaters are paired against each other randomly. Rounds 3 onward are "Power Paired," meaning debaters are paired against each other on the basis of their previous record and speaker points. For instance, a debater with a 4-0 record generally would only debate someone who also has a 4-0 record. [b] Each observation denotes an individual's win or loss in a specific debate (not the individual).

Source: Tabroom.com National Circuit Lincoln Douglas Debate competition results for a sample of 89 tournaments from spanning the 2011-12 to 2015-16 seasons. See Appendix C for list of tournaments.

Table 6. Determinants of Competitive Success in High School Lincoln Douglas Debate (OLS Model): Preset Rounds

Binary dependent variable: Individual wins debate round				
Indep. Variables	(1) Baseline Model	(2) + Graduation Year	(3) + Divisions & School Effects (Baseline: Novice)	(4) + Preset Round & Interaction
Female	-0.0379**	-0.0352**	-0.0151	-0.0037
	(0.0044)	(0.0045)	(0.0127)	(0.0128)
JV x Female			0.0072	0.0064
			(0.0171)	(0.0171)
Varsity x Female			-0.0198	-0.0208
			(0.0134)	(0.0132)
Preset				0.0130**
				(0.0035)
Female x Preset				-0.0308**
				(0.0086)
Obs. [a]	65,964	65,964	65,958	65,958
Adj-R2	0.00	0.00	0.05	0.05
Individual Controls	No	Graduation Year Only	Yes	Yes
Tourn. Effects	Yes	Yes	Yes	Yes
Year Effects	Yes	Yes	Yes	Yes

* p<0.05, ** p<0.01, *** p<0.001

Note: Table 6 reports the coefficients from an ordinary least squares (OLS) regression. The dependent variable is whether the debater won the round. Individual controls include school fixed effects and graduation year. Standard errors in parentheses are clustered at the tournament-round level. [a] Six observations are missing school information, so including school as an individual control slightly reduces the sample size.

Source: Tabroom.com National Circuit Lincoln Douglas Debate competition results for a sample of 89 tournaments from spanning the 2011-12 to 2015-16 seasons. See Appendix C for list of tournaments.

Table 7. Determinants of Competitive Success in High School Lincoln Douglas Debate (Individual Fixed-Effects Model): Judge Gender & Preset Rounds

Binary dependent variable: Individual wins debate round				
	(1)	(2)	(3)	(4)
Indep. Variables	Baseline Model	+ Preset and Interaction	+ Judge Preset Interaction	+ Division Interaction (Baseline: Novice)
Female x Judge Female	-0.0073 (0.0093)	-0.0074 (0.0093)	-0.0026 (0.0108)	-0.0020 0.0108)
Preset		0.0169** (0.0059)	0.0169** (0.0059)	0.0158** (0.0059)
Female x Preset		-0.0311** (0.0093)	-0.0266* (0.0107)	-0.0258* (0.0107)
Female x Preset x Female Judge			-0.0138 (0.0146)	-0.0159 (0.0146)
JV x Female				0.0087 (0.0207)
Varsity x Female				-0.0066 (0.0196)
Obs.	65,964	65,964	65,964	65,964
Adj-R2	0.08	0.08	0.08	0.09
Fixed Effects [a]	Yes	Yes	Yes	Yes

* $p<0.05$; ** $p<0.01$

Note: Table 7 reports the coefficients from a linear regression tracking individuals across tournaments. The dependent variable is whether the debater won the round. Standard errors in parentheses are clustered at the individual level. [a] All columns include individual, tournament, and year fixed effects.

Source: Tabroom.com National Circuit Lincoln Douglas Debate competition results for a sample of 89 tournaments from spanning the 2011-12 to 2015-16 seasons. See Appendix C for list of tournaments.

Appendix A: School Variable Construction

This section outlines the procedure I use to construct the school variable. In general, using the original school variable without modification would result in the same debater being assigned to different schools. Whether it is because tournaments sometimes use shorter or longer names for schools or because coaches themselves enter their school name inconsistently across tournaments, it is often the case that schools are labeled similarly but not identically. For example, "Palos Verdes Peninsula" and "PV Peninsula" both refer to the same high school.

From a statistical standpoint, it is crucial to construct the school variable properly since a debater's likelihood of success is almost certainly related to the quality of his or her high school's debate program. To ensure that the school variable is properly constructed, I adopt both a general coding procedure (applied to all schools) and a manual approach for special cases.

General procedure applied to all schools:

- I format school names in proper case. In other words, "smith" and "SMITH" become "Smith."
- I remove special characters (periods, commas, parentheses, dashes) as well as extra, leading, and trailing blank spaces. For example, if the same debater is denoted as being from "Harvard-Westlake School" at one tournament and "Harvard Westlake School" at another, I replace the dash with a space so that the school labels are identical across tournaments.
- I remove "High School," "School," "HS," and "Country Day." This ensures "PV Peninsula" and "PV Peninsula High School" are labeled identically.
- Since some debaters compete as "independent entries" if, for instance, their school does not approve their debate travel, I remove "Independent" from the end of school names. "Torrey Pines Independent" thus becomes "Torrey Pines."

Manual changes:

Original variable contains	Resulting school name
Boxborough	Acton Boxborough
Delray	Am Her Delray
Plantation	Am Her Plantation
AofHL	AofHL
Bronx	Bronx Science
"Cambridge" or "Cambridge Rindge"	CRLS
Chaminade College Preparatory	Chaminade College Prep
Churchill	Churchill

Original variable contains	Resulting school name
Dsm Roosevelt	Des Moines Roosevelt
East High	East
"Flintridge Sha" or "Fsha" or "Flintridge Sacred Heart Academy"	Flintridge Sacred Heart
Foothill High	Foothill
"Harvard-Westlake" or "Hwms"	Harvard Westlake
Graham Brown	James Graham Brown
Loyola of Los Angeles	Loyola
"Jm" or "Jmhsdebate"	John Marshall
Mountain View	Mountain View
Nfa	Newburgh Free Academy
Skyview	Sky View
Southside Ctr for Intl	Southside
"Torrey" or "Torrey Falcons"	Torrey Pines
Wdm	West Des Moines Valley
Exact name is "The Woodlands"	Woodlands
Exact name is "University" [26]	University FL

Appendix B: Note On Properly Identifying Individuals

Tournaments generally distinguish debaters within schools by using first and last initials. So, Jane Doe from John Smith High School would be "John Smith JD" on a tournament pairing. However, any problem with the original school label also affects debaters' individual labels. For example, "John Smith High JD" and "John Smith JD" would almost certainly be the same person but treated differently in the dataset. As a result, I do not use this variable to distinguish among individuals.

Fortunately, Tabroom.com assigns a unique 5- or 6-digit identifier to each individual that does not depend on the specific tournament pairing label. So "John Smith JD" and "John Smith High JD" might both have identifier "12345." So I simply use this numeric identifier to track individuals. I operate under the assumption that this identifier is accurate over the sample.

[26] There is also "University Prep" and "University Ohio," which I leave unchanged as they are different schools. However, a two-way table reveals that "University" and "University FL" contain the same debaters, so I rename "University" to "University FL."

Appendix C: List of Tournaments

Tournament	Season	State	Number of Competitors (Varsity)
Alta	2014-15	UT	101
Alta	2015-16	UT	116
Apple Valley	2014-15	MN	116
Apple Valley	2015-16	MN	133
Barkley Forum	2014-15	GA	75
Barkley Forum	2015-16	GA	109
Beltway	2013-14	MD	35
Beltway	2014-15	MD	70
Beltway	2015-16	MD	50
Berkeley	2013-14	CA	255
Berkeley	2014-15	CA	243
Berkeley	2015-16	CA	262
Blake	2011-12	MN	153
Blake	2014-15	MN	145
Blake	2015-16	MN	97
Bronx	2011-12	NY	152
Bronx	2014-15	NY	168
Bronx	2015-16	NY	165
College Prep	2014-15	CA	110
College Prep	2015-16	CA	169
Colleyville Heritage	2015-16	TX	46
Columbia	2012-13	NY	84
Columbia	2013-14	NY	89
Columbia	2014-15	NY	76
The Glenbrooks	2011-12	IL	149
The Glenbrooks	2014-15	IL	152
Golden Desert	2013-14	NV	89
Golden Desert	2014-15	NV	120
Golden Desert	2015-16	NV	107
Grapevine	2015-16	TX	97
Greenhill	2014-15	TX	92
Greenhill	2015-16	TX	79
Harvard-Westlake	2014-15	CA	108

Tournament	Season	State	Number of Competitors (Varsity)
Harvard	2013-14	MA	278
Harvard	2015-16	MA	247
Holy Cross	2015-16	LA	36
Lexington	2012-13	MA	116
Lexington	2013-14	MA	122
Lexington	2014-15	MA	132
Lexington	2015-16	MA	104
Loyola	2013-14	CA	48
Loyola	2014-15	CA	70
Loyola	2015-16	CA	60
Meadows	2014-15	NV	70
Meadows	2015-16	NV	64
NDCA Nationals	2013-14	UT	43
NDCA Nationals	2014-15	NV	48
Newark	2012-13	NJ	54
Newark	2013-14	NJ	56
Newark	2014-15	NJ	72
Newark	2015-16	NJ	54
Princeton	2011-12	NJ	115
Princeton	2012-13	NJ	100
Princeton	2013-14	NJ	113
Princeton	2014-15	NJ	140
Ridge	2011-12	NJ	70
Ridge	2013-14	NJ	84
Ridge	2014-15	NJ	84
Ridge	2015-16	NJ	70
Scarsdale	2013-14	NY	80
Scarsdale	2014-15	NY	62
Scarsdale	2015-16	NY	72
St Mark’s	2014-15	TX	98
St Mark’s	2015-16	TX	96
Stanford	2013-14	CA	201
Stanford	2014-15	CA	159
Stanford	2015-16	CA	108
TOC	2012-13	KY	75
TOC	2013-14	KY	92

Tournament	Season	State	Number of Competitors (Varsity)
TOC	2014-15	KY	88
USC	2013-14	CA	46
USC	2014-15	CA	28
USC	2015-16	CA	62
VBT	2013-14	CA	115
Valley	2014-15	IA	124
Valley	2015-16	IA	128
Voices	2014-15	CA	97
Voices	2015-16	CA	95
Wake Forest	2015-16	NC	57
Yale	2011-12	CT	137
Yale	2013-14	CT	131
Yale	2014-15	CT	153
Yale	2015-16	CT	138
Total Number of Tournaments: 89			

Note: As there is no direct variable for size of the debate pool, I use the number of people debating in round 1 of the tournament as the measure of the number of competitors.