

Analyzing Gender & Performance In Competitive Environments With Machine Learning: A High School Debate Case Study

Samarth Chitgopekar

February 2022

Abstract In this paper, I examine the relationship between gender and performance in high school Public Forum, a 2 on 2 style of debate where judges subjectively decide the winner, for 3 seasons from 2019 to 2022. I use a variety of metrics to gauge success, including a composite score, speaker points, and win rates. I classify the gender of competitors using Natural Language Processing on a training set of ethnically-representative names based on data provided by the National Speech and Debate Association. I introduce Gender Dominance, a calculus-based approach to quantifying the gender disparity (in terms of overrepresentation and underrepresentation) in any given range of a performance metric. I find that, year-over-year, all examined performance metrics have shown a decrease in the gender disparity to the point where, for the 2021-22 season, they do not indicate any statistically significant difference in performance between males or females. Finally, I identify several factors that this decrease could be attributed to, including debater-led advocacy (both in and out of round) and tournament-led advocacy regarding gender discrimination. My research does not guarantee that any portion of rounds are free from discrimination. Instead, it compares trends between gender and performance while offering possible explanations for causality, which future research should work to prove with the ultimate goal of creating a general plan of attack to address gender disparities within any competitive field.

1 Introduction

1.1 The Significance Of Gender Disparities

Since the turn of the 21st century, great social efforts have been made to promote gender equality. Many of these have been successful to an extent: female workforce participation and pay has increased, women now make up more of the college population, and conversations surrounding gender have become more commonplace. However, there are still noticeable disparities. For example, there is a compensation deficit between males and females in which men are paid 11.3% more for the same work with the same qualifications.

Examining the change of these disparities over time, then, proves to be valuable when seeking to understand the state of gender equality today. We should choose to examine structured competitive environments because they generally yield quantifiable statistics that we can use to analyze trends with respect to gender (eg. win rate).

1.2 Choosing High School Public Forum

There are many choices to consider when choosing a competitive environment to analyze. In this section, I'll explain why I chose Public Forum debate at the high school level.

Sports are intrinsically competitive by nature, but oftentimes carry extensive separation between gender. For example, the National Basketball Association has a separate league for females, the Women's National Basketball Association. Many try to compare factors such as pay across the league, but because there are fundamental differences in competition, they often fall short (Robinson (2021)). This also manifests itself across other professional sports. Recently, the U.S. Women's National Team sued U.S. Soccer for discrimination, citing a similar pay gap (Burhan (2021)).

Public Forum debate solves these shortcomings since there are no separate divisions¹ by gender. This means that, on average, any given team will face the same difficulty of opponent as another team competing at similar tournaments—regardless of gender. Additionally, because there are only 2 speakers on a team, we can easily track relationships with gender since there are only 3 possible combinations² of sex.

There are other forms of debate that offer these benefits, such as Lincoln-Douglas or Policy Debate³. However, Public Forum is the most popular style today, which is critical when gathering a large enough dataset to perform queries on. Additionally, this dataset already exists as a reproducible, community-maintained, open-sourced resource (Chitgopekar (2021)). Debate isn't prominent as a professional sport and Public Forum isn't well-established at the collegiate level⁴. So, high school data was used.

1.3 Structure of Public Forum Debate

¹None of the tournaments I used (well-established national competitions) had gender-specific divisions. See all of the tournaments in the dataset in 6.1.

²The combinations are: Male-Male, Male-Female, Female-Female.

³The National Speech and Debate Association provides an exhaustive list of all formats at <https://www.speechanddebate.org/competition-events/>.

⁴Collegiate Public Forum is newly established and does not have widespread participation. For more information, reference the official league's site at <https://www.collegepublicforum.org/>.

1.3.1 Rounds

A Public Forum round begins with 2 teams, each made up of a 1st and 2nd speaker, and a judge. In certain rounds (generally at state championships or elimination rounds), there will be a panel of 3 (though it can be any odd number greater than 1) judges. The judge(s) will always choose a team to win and draws cannot be given. At most tournaments⁵, each individual debater is assigned speaker points by the judge. These are used as tiebreakers for clearing teams into elimination rounds or for speaker awards. See 6.4 for the structure of a Public Forum round.

1.3.2 National Circuit Debate

The high school Public Forum national circuit is broadly defined as a collection of all the tournaments recognized as ‘bid tournaments’ to the University of Kentucky’s Tournament of Champions⁶, which is held annually in April. These are the tournaments that are traditionally studied in the field (eg. Nie & Yi (2020)), and are what I will analyze in this paper.

A standard bid tournament has 5-7 preliminary rounds, in which competitors can face teams not from their home school. The first two rounds are randomly matched. After that, the most common way the remaining rounds are structured is by record, in which teams with the same win-loss record are matched up⁷.

After the preliminary rounds, the tournament will clear teams that have more wins than losses to elimination rounds. This second set of rounds is why most National Circuit tournaments will be spread out across multiple days. Elimination rounds go until finals⁸ where 2 teams face off. The i th round before finals has a number of teams given by 2^{1+i} and number of rounds given by $\frac{2^{1+i}}{2}$. Usually there are a number of competitors matching the criteria for clearing that cannot be expressed by 2^x , meaning that there’s no immediate way to get a ‘clean’ starting elimination round. In this case, a partial elimination round is used that consists of the appropriate number of debates such that, after half of competing teams are eliminated, the remaining debaters can participate in a full elimination round. The lower seeded debaters (by win-loss record) are chosen to participate in these rounds.

The Tournament of Champions has 2 divisions, a gold and silver. The gold level requires 2 gold bids to attend, and the silver level requires 2 silver bids to attend⁹. Bid tournaments allocate a varying amount of bids (dependent on their respective difficulties) at certain elimination rounds. If the n th elimination round offers a gold bid, the $n-1$ th round offers a silver bid. All competitors debating in a bid round receive the bid (winning at the level is not required)¹⁰.

If a gold bid is earned, the team is not given the additional silver bid. Overall, I scrape an average of 70 tournaments per year across the 3 years studied. The dataset only uses tournaments hosted on Tabroom¹¹, as this site is by far the most widely used and provides the most data. Other services include SpeechWire¹².

⁵Generally elimination rounds don’t have speaker points scored because there’s no need for them as a tiebreaker (teams have already been cleared from preliminary rounds) and speaker awards are usually distributed before the elimination rounds start.

⁶More information is available at <https://uktoc.org/>.

⁷Some tournaments have entry pools with odd numbers of competitors, in which case there will be a team receiving a bye (not debating) each round. They are awarded a win.

⁸Occasionally, teams agree to not debate finals and ‘Co-Champion’ the tournament. This generally occurs because all bids have already been awarded.

⁹The Tournament of Champions also awards a set number of ‘At-Large’ bids to select teams who did not meet the participation requirements.

¹⁰Debaters who faced a team from their home school immediately before an elimination round that awarded a bid are given that bid, which is called a ‘ghost-bid’. This includes teams in this situation at silver bid elimination rounds (they will receive a gold bid).

¹¹Tabroom is available at <https://www.tabroom.com/index/index.mhtml>.

¹²Speechwire is available at <https://www.speechwire.com/index.php>.

1.3.3 Local Circuit Debate

I do not examine any local circuit results in this paper, but they are still crucial to understanding the dynamics of Public Forum debate, as it is where the vast majority of debaters compete. Even national circuit debaters participate in local circuit tournaments between bid tournaments.

The local circuit generally consists of one-day tournaments with 4-5 preliminary rounds. The local circuit is heavily fragmented and tends to be much more inconsistent with the posting of results. Many of them are ran via paper sheets with round postings. However, there are still some hosted on Tabroom and SpeechWire.

These tournaments give awards based off win-loss records with speaker points as a tiebreaker between teams with the same records. Separate awards may be given to top speakers.

2 Literature Review

2.1 Overview

Research in the relationship between competitiveness and gender has indicated that, in patriarchal societies, women become less competitive around puberty—immediately preceding the age range of high school debate (Andersen & Ertac (2013)). This should indicate the presence, to some extent, of a performance gap. When examining high school debate specifically, literature has focused on gender’s relationship with win rates, participation, and attrition across the Lincoln-Douglas and Public Forum formats (eg. Tartakovsky (2017)). More research has been done on legacy debate styles, such as Policy Debate and Lincoln-Douglas, than Public Forum (which is relatively new).

After controlling for a variety of factors, previous work has shown that male-male teams are 2.5-3% (Lincoln-Douglas) and 37.6% (Public Forum) more likely to win a preliminary round than female-female teams (Tartakovsky (2016) and Abbott (2018) respectively), women are 38% less likely to participate in Public Forum (Abbott (2018)), and female debaters are 30.34% more likely to quit (Nie & Yi (2020)).

The effect of program strength (eg. powerhouse schools) and the gender makeup of debaters at the program on the aforementioned relationships have also been examined (eg. Nie & Yi (2020)).

2.2 Contributions

In this paper, I contribute to the existing literature by incorporating 5 distinct factors into my analysis.

First, I examine speaker points, which are subjectively assigned by judges to all 4 debaters in a round. In doing so, I look at both raw speaker points and adjusted speaker points, which remove outliers (see 3.2 and 3.3 respectively).

Second, I also study win rates in elimination rounds, not just preliminary rounds. This perspective helps us see if, by having a panel of judges decide the winner, any gender gaps can be shrunk.

Third, I use tournaments from the 2019-20, 2020-21, and 2021-22 seasons. The most recent research in the field’s latest season used is 2019-20 (Nie & Yi (2020)). This allows us to examine fresh data, which we can compare to previous research¹³.

Fourth, our dataset exists as a reproducible, community-maintained, open-sourced resource that collects the most statistical indicators of any in the field. I leverage this unique insight by implementing the OTR Score (Vaidyanathan & Chitgopekar (2021)), a composite ranking factor that is the first of its kind in this field, as it allows us to award tournament difficulty, elimination round performance, preliminary round performance, and opponent difficulty in one metric (see 3.1 for more information).

Fifth, I am the first to opt for gender classification using machine learning in lieu of the census-based probability model. This enables classification of names not collected in the census (see 3.3 for more information).

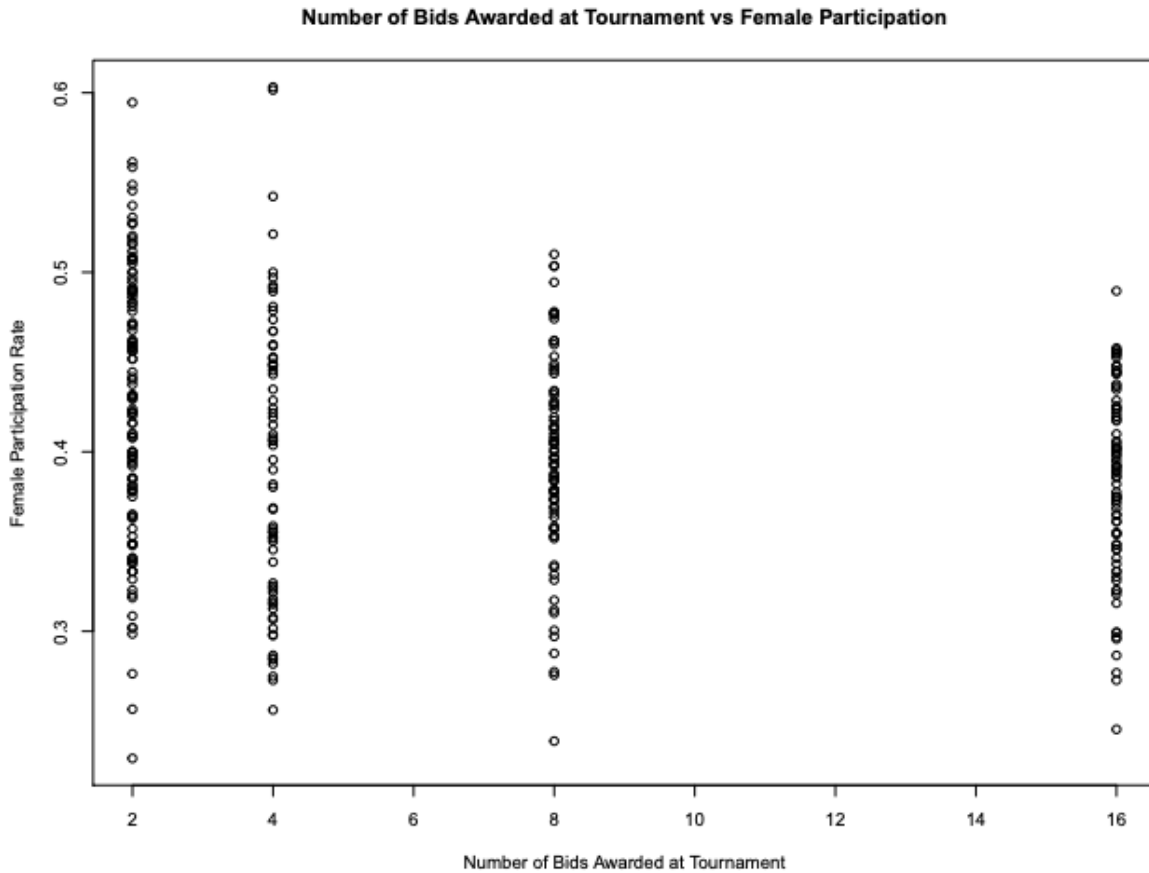
¹³See 5.1.1 for limitations on comparisons of our dataset with previous studies.

3 Methodology

Each relationship we examine has an independent axis representing various sequential ranges for the factor we're testing gender's relationship with. We then have the dependent axis represent the percentage of debaters of a particular gender that are present in that range. 'Higher' ranges (those that are farther from the origin) indicate higher performance for all trends we examine.

We then create a best-fit curve using a 7th degree polyfit (with least squares). Normally, we'd expect gender to be held as the independent variable with the selected performance indicators being on the dependent axis. However, this setup enables us to understand a factor, which I will name 'Gender Dominance', at various levels of performance. The former would only give us 2 lines with indiscernable points along them, similar to the one below:

Nie & Yi (2020) Appendix B: Figures (1)



We can understand gender dominance as:

“The mean percentage difference between the representation of one gender over another throughout a particular range of a given metric.”

The best-fit curves produced by this model allow us to understand gender dominance quantitatively with:

$$\text{Gender Dominance (\%)} = \frac{1}{b-a} \int_a^b \text{GenderBestFit}(r) - \text{OppositeGenderBestFit}(r) dr$$

Here, our interval is $a < r < b$ where $(a, b) \in R$ and both best fit curves, $\text{GenderBestFit}(r)$ (the best fit of the chosen gender) and $\text{OppositeGenderBestFit}(r)$ (the best fit of the other gender) exist over the interval. Even though our independent axis is plotted at specific ranges, our polyfits are defined at any value for our independent metric. We refer to this metric as r .

Later, I will qualitatively describe changes in the Gender Dominance factor based on the plots of the curves of best fit. A positive Gender Dominance indicates that the chosen gender is overrepresented in the range compared to the other gender. A negative Gender Dominance indicates that the chosen gender is underrepresented in the range compared to the other gender. The magnitude of the Gender Dominance factor indicates the extent to which the representation gap exists in that range—when it is 0, neither gender is dominant over the interval.

In the interest of consistency, going forward I will describe Gender Dominance with respect to Males with:

$$\text{Male Dominance (\%)} = \frac{1}{b-a} \int_a^b \text{MaleBestFit}(r) - \text{FemaleBestFit}(r) dr$$

3.1 The OTR Score

The OTR Score (Vaidyanathan & Chitgopekar (2021)) is a composite ranking factor that accounts for tournament difficulty, elimination round performance, preliminary round performance, and opponent difficulty. It is the first of its kind to be used in this field and enables a more holistic understanding of performance.

Participating at any national circuit tournament will give a team a OTR_{comp} , where $OTR_{comp} \geq 0$. Each compensation is given by:

$$OTR_{comp} = \frac{\# \text{ of Preliminary Round Wins}}{\# \text{ of Preliminary Rounds}} * \text{Mean Opponent Wins} * \text{Tournament Boost} * \text{Elimination Boost}$$

In calculating the compensation for a specific tournament, we can input a tournament difficulty boost factor (which is constant for all attending the tournament) bounded by:

$$0 < \text{Tournament Boost} \leq 2^{14}$$

Similarly, we can repeat with an elimination round boost factor (which varies based on independent performance) given by:

Elimination Boost = $i + 1$ (where i is the number of elimination rounds debated in).

Finally, we calculate $OTR_{score} = \sum OTR_{comp}$.

3.2 Raw Speaker Points

Speaker points are assigned by judges and are designed to reward speaking ability. Teams who speak well but lose the round due to strategic factors have an avenue for success. For each team, we averaged their preliminary round speaker points as provided to compute Raw Speaker Points. Tournaments who didn't score speakers were excluded from the calculation.

¹⁴The initial value for our tournament boost factor is determined by the number of bids the tournament offers (see 6.3 for a list of these values). At the end of the season, scores are recalculated using tournament boost factors proportional to the number of bids a tournament's entry pool ended up earning. For more detailed information on the recalculation process, reference Vaidyanathan & Chitgopekar (2021).

3.3 Adjusted Speaker Points

Raw speaker points (see 3.2 for more information) can be greatly impacted by outliers. We study Adjusted Speaker Points, which attempt to remove outliers from the calculation, to see how the impact of outlier judges differs across gender. If we noticed that a particular gender got a significantly greater boost from outlier removal (as compared to Raw Speaker Points), we can conclude that they encountered judges who acted in an irrational manner (as compared to the majority of judges) at a higher rate (further discussed in 5.2).

Outliers are removed via Interquartile Range with an outlier constant of 2. Adjusted Speaker Points for a specific tournament are derived by first filtering for an updated list of speaker points with the aforementioned Interquartile Range calculation. Then, the adjusted speaking average for that tournament is calculated. The final Adjusted Speaker Points value is the average of all recorded Adjusted Speaker Points.

3.4 Dataset Information

I use the dataset provided by Tournaments.Tech¹⁵, which is the most popular Public Forum debate ranking site and receives regular updates with community feedback. The dataset and source code is made available to the public under the MIT License¹⁶.

Tabroom.com is able to provide useful statistics, but oftentimes tournament directors miss publishing certain results/metrics. The API used to derive the dataset checks for all of these factors and works around them to ensure accurate data is collected (Chitgopekar (2021)). For example, if the average number of opponent wins isn't provided, it is recalculated to ensure the parameter's existence.

For these reasons, the dataset is a unique and integral part of the research conducted in this study.

3.5 Classification of Gender

Gender¹⁷ is not provided on Tabroom.com and there is no official way of finding the gender of any given debater. However, previous research (eg. Nie & Yi (2020)) has adopted a technique to estimate the gender of debaters. With this approach, the probability of an individual (given their first name) being a given gender is:

$$P(\text{gender}, \text{first name}) = \frac{\text{<first name>'s appearances in <gender>'s name dataset}}{\text{<first name>'s appearances across both gender's datasets}}$$

The name dataset is generally comprised of U.S. Census name data and a combination of other name datasets to better represent the ethnicities of competitors—such as the incorporation of an Asian-only name dataset.

My approach deviates from this technique and instead opts for a Machine Learning based solution. Specifically, I employ the use of Natural Language Processing (NLP) via Python 3's Natural Language Toolkit (NLTK) Library (detailed by Bird (2004)) and an ethnically representative¹⁸ training dataset to create a classifier function. Upon testing with a known dataset, the classifier was over 75% accurate. This method allows us to classify names not found in our dataset, as NLP focuses on pattern recognition.

¹⁵Tournaments.Tech is available at <https://tournaments.tech>.

¹⁶Accurate as of 2/22/2022.

¹⁷It is critical to note that this paper does not distinguish between biological sex and gender, and that they are understood to be interchangeable in the context of this research.

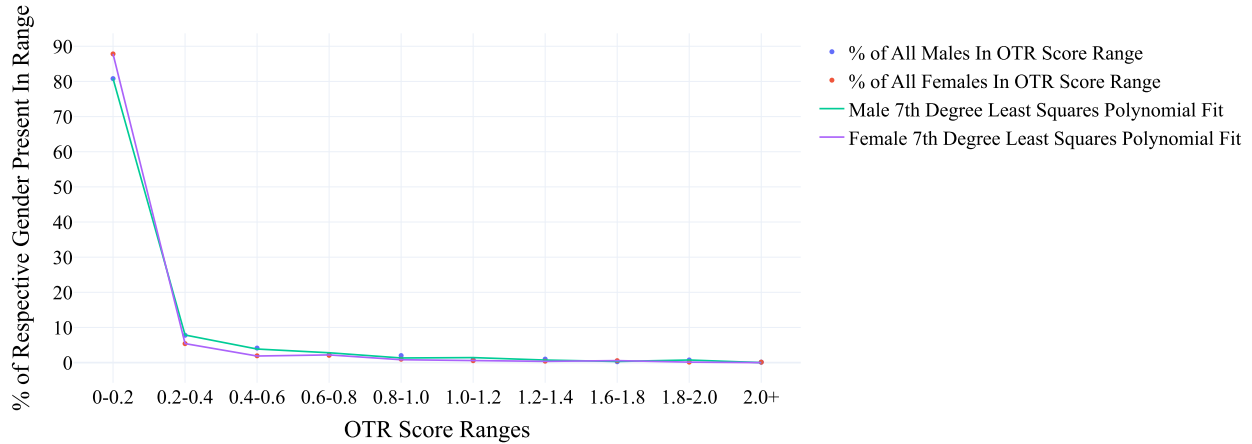
¹⁸The National Speech and Debate Association provides access to a membership database at <https://www.speechanddebate.org/membership-database/> which contains the demographic information used to craft the training set.

4 Results

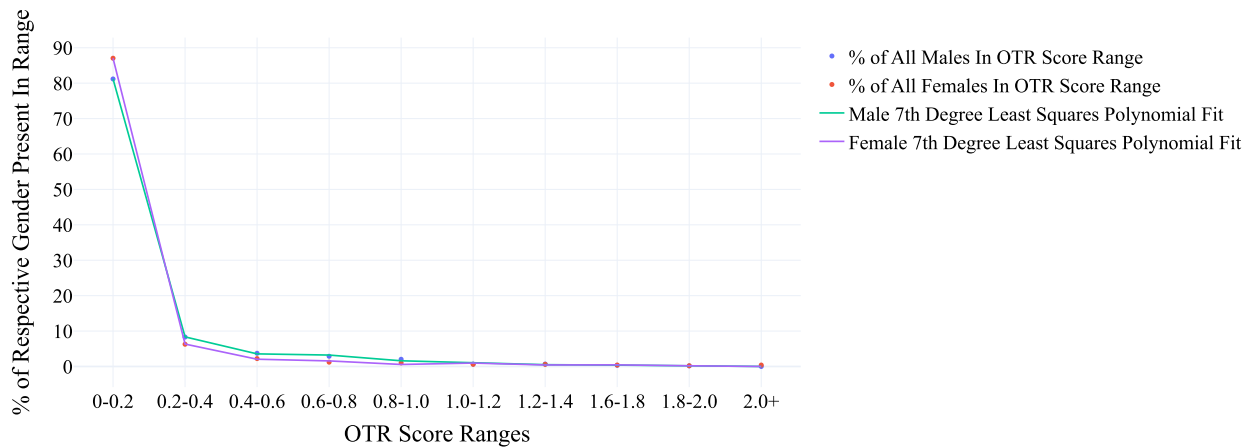
In this section, I present the Gender Dominance graphs for various performance metrics for each season studied. I qualify how Male Dominance has changed over various seasons for each performance metric, ultimately concluding whether or not the trends indicate more or less of a gender disparity in that metric.

4.1 OTR Score

4.1.1 2020-21 OTR Score v. Gender



4.1.2 2021-22 OTR Score v. Gender



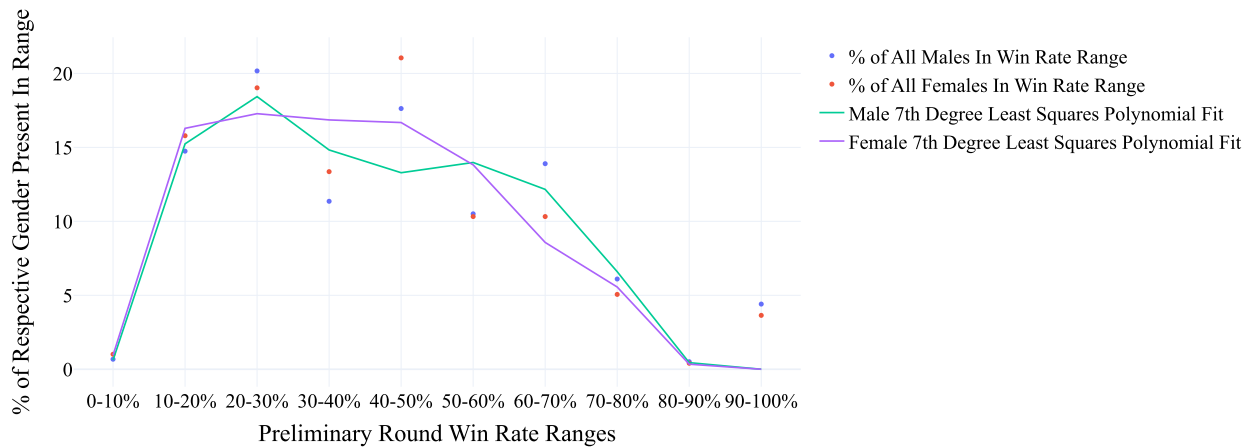
There is no statistically significant change in the intervals for the Male Dominance Factor is positive, though we can see that between the 2020-21 and 2021-22 seasons, more males were present in the $0 < r < 0.4$ range even though females still dominated there. Additionally, females increased their representation in the $0.4 < r < 0.6$ range. Since average OTR Scores are always well below 0.2 and the amount of debaters dies of exponentially with an increase in OTR Score, these two changes have led to more equal performance by gender under the OTR Score metric.

4.2 Preliminary Round Win Rate

4.2.1 2020-21 Preliminary Round Win Rate v. Gender



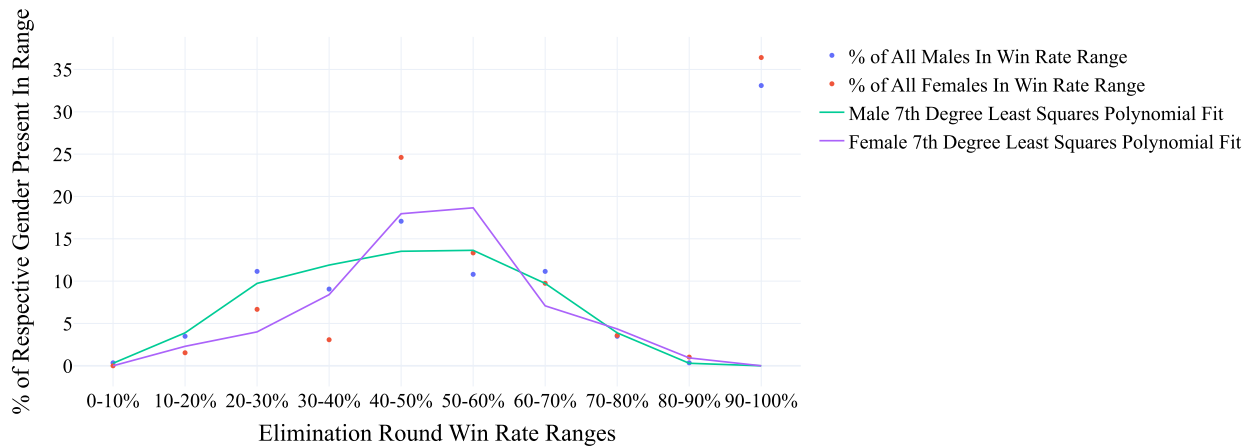
4.2.2 2021-22 Preliminary Round Win Rate v. Gender



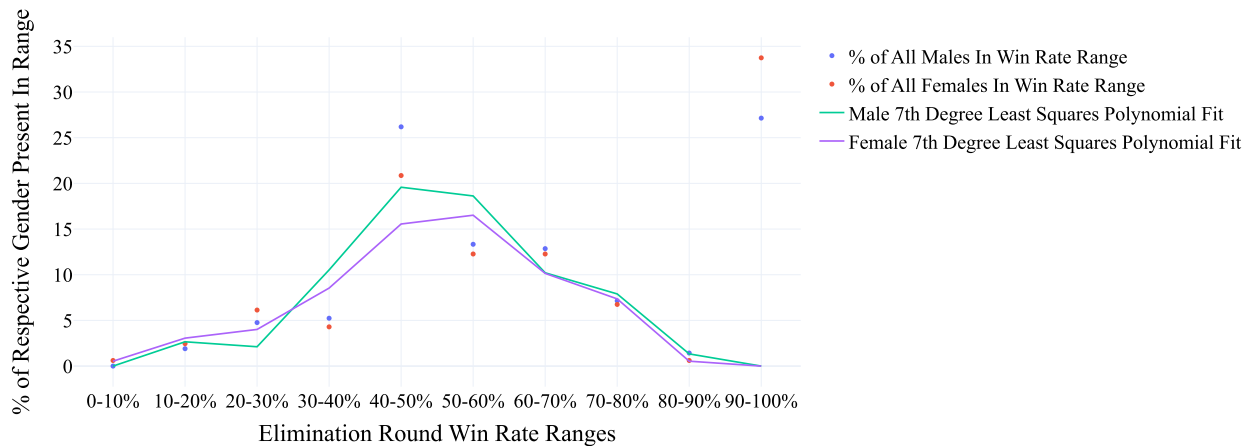
There are multiple statistically significant changes in the intervals of interest for the Male Dominance Factor. From the 2020-21 season to the 2021-22 season, males lowered their representation in the $35\% < r < 60\%$ range. The dataset has the average near 44%, which is why this is a key deviation. Moreover, though males still dominate the $60\% < r < 90\%$ range, it is key to note that the magnitude of the dominance has decreased significantly at the very top end range of $80\% < r < 90\%$ and that the lower bound of the range used to begin at 45%—an improvement of over 15% in a season. These two changes indicate more equal performance by gender under the Preliminary Round Win Rate metric.

4.3 Elimination Round Win Rate

4.3.1 2020-21 Elimination Round Win Rate v. Gender



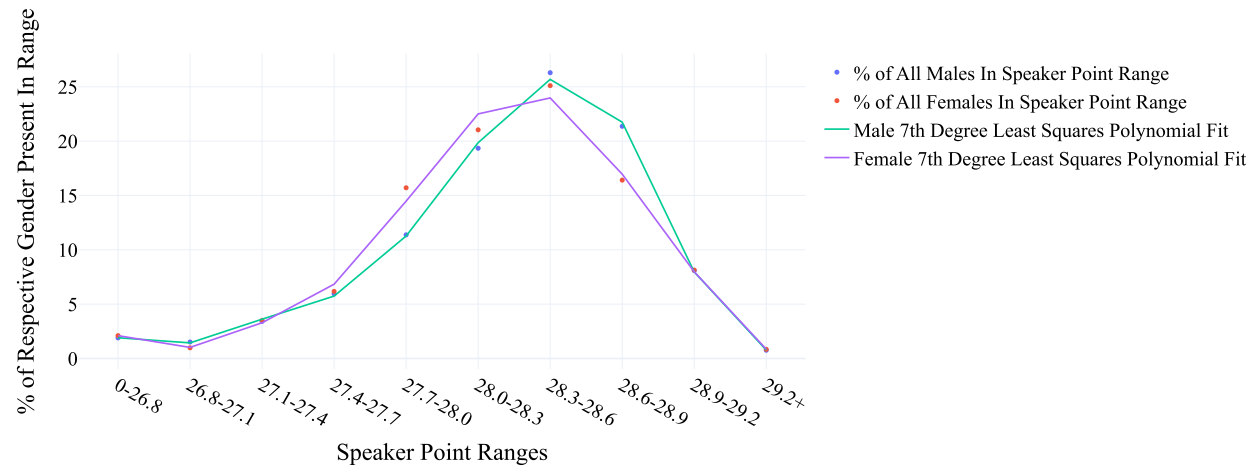
4.3.2 2021-22 Elimination Round Win Rate v. Gender



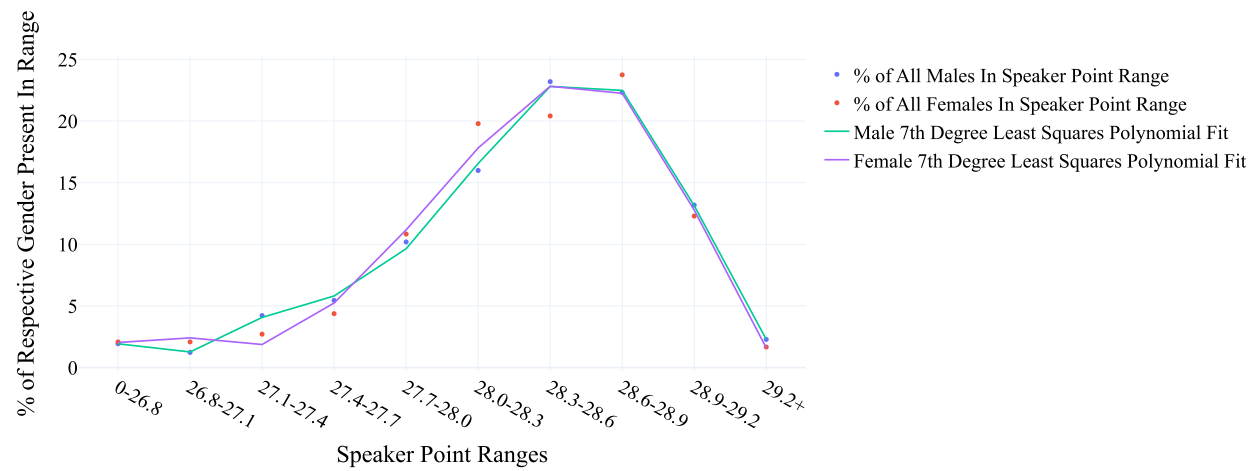
There are multiple statistically significant changes in the intervals of interest for the Male Dominance Factor. From the 2020-21 season to the 2021-22 season, males lowered their representation in the $0\% < r < 45\%$ range while increasing their dominance in the $45\% < r < 70\%$ range. Previously, males were overrepresented in less-performant ranges and underrepresented in more-performant ranges. Now, males and females are performing at a similar rate across the entire range. These two changes indicate more equal performance by gender under the Elimination Round Win Rate metric.

4.4 Raw Speaker Points

4.4.1 2020-21 Raw Speaker Points v. Gender



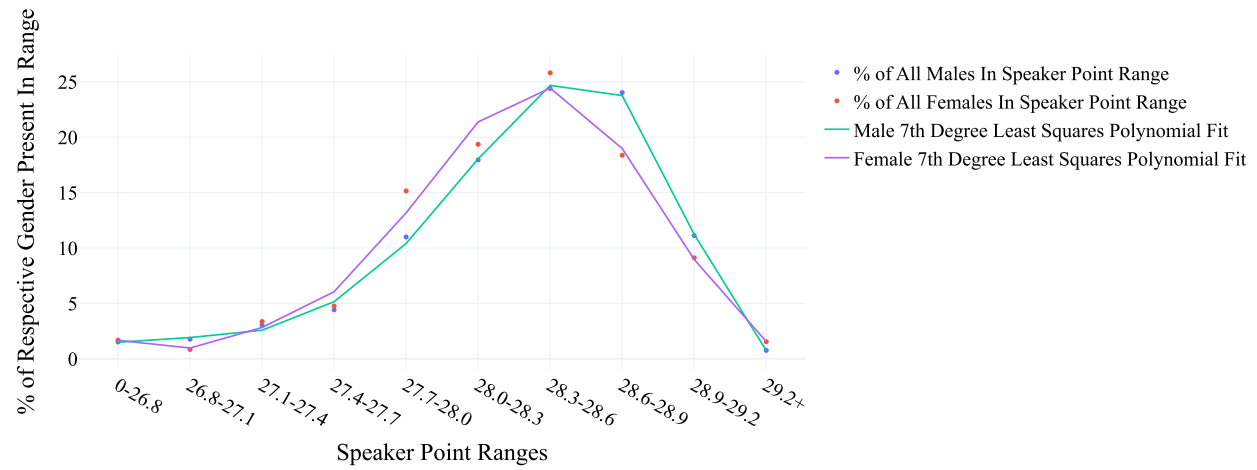
4.4.2 2021-22 Raw Speaker Points v. Gender



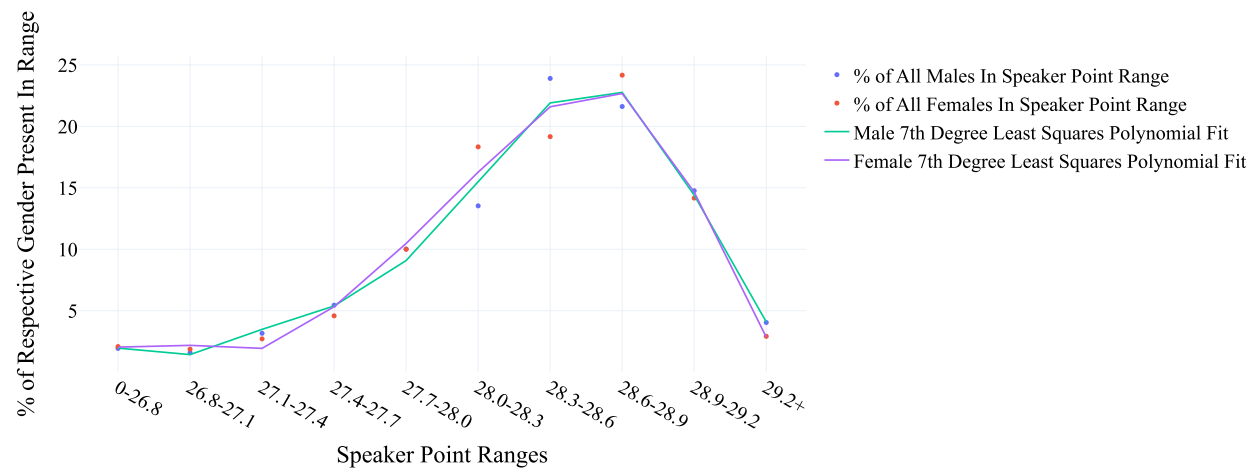
There are multiple statistically significant changes in the intervals of interest for the Male Dominance Factor. From the 2020-21 season to the 2021-22 season, males increased their representation in the $27.4 < r < 28.6$ range while decreasing their dominance in the $28.6 < r < 29.2+$ range. Previously, males were underrepresented in less-performant ranges and overrepresented in more-performant ranges. Now, males and females are performing at a similar rate across the entire range. These two changes indicate more equal performance by gender under the Raw Speaker Points metric.

4.5 Adjusted Speaker Points

4.5.1 2020-21 Adjusted Speaker Points v. Gender



4.5.2 2021-22 Adjusted Speaker Points v. Gender



There are multiple statistically significant changes in the intervals of interest for the Male Dominance Factor. From the 2020-21 season to the 2021-22 season, males increased their representation in the $27.1 < r < 28.6$ range while decreasing their dominance in the $28.6 < r < 29.2+$ range. Previously, males were underrepresented in less-performant ranges and overrepresented in more-performant ranges. Now, males and females are performing at a similar rate across the entire range. These two changes indicate more equal performance by gender under the Adjusted Speaker Points metric.

5 Discussion

5.1 Limitations

5.1.1 Comparisons With Previous Research

Comparing my results with relevant literature is something I do throughout this paper. However, it is key to understand that my gender classification model, which exploits Natural Language Processing, differs from previous approaches that utilized Census data to classify names (detailed in 3.5).

Moreover, I am the first to employ the use of the Gender Dominance factor in my analyses to understand trends over time. I compare calculated averages for various metrics (available in 6.2), which should provide a more standardized comparison.

Finally, I use a different, peer-reviewed dataset for my research. I find this to be a more effective approach (detailed in 3.4), and it seems that no other literature in the field (eg. Abbott (2018), Tartakovsky (2016) & Tartakovsky (2017), Nie & Yi (2020)) shares a dataset. This is the first paper of its kind published alongside the necessary code and data to reproduce results or calculate them for various additional years.

This means that comparisons with previous analyses will not be completely accurate on a raw numerical level, but trends and their respective evolution should persist and can be expected to be reasonably comparable. Regardless, this paper's contributions give it many unique reasons to be preferred, including: speaker point analysis, a uniquely high-quality dataset, controlling for tournament and opponent difficulty, and analyzing elimination round performance.

5.1.2 Speaker Points (Raw and Adjusted)

Though the vast majority of the tournaments I analyzed included speaker points, it is important to note that a small, statistically insignificant minority do not.

Additionally, conventions for awarding speaker points vary by tournaments. I do not find that this makes a statistically significant impact because we are still able to analyze the relative changes in the speaker points between gender.

5.2 Conclusion: The State of Gender In Public Forum

From a basic level of analysis, we can see that averages for all measured metrics have become closer over the past 3 years. Using statistical averages for the current 2021-22 season, we see that there is no advantage for either gender under the OTR Score metric, a 0.01 point advantage for females under the Raw Speaker Points metric, a 0.02 point advantage for females under the Adjusted Speaker Points metric, a 1% advantage for females under the Preliminary Round Win Rate metric, and no advantage for either gender in the Elimination Round Win Rate metric (detailed in 6.2). Though there is a net advantage for females, due to the magnitude of the advantage I conclude that the average across a multitude of performance metrics indicate no statistically significant gender disparity.

Moreover, upon examining the Gender Dominance factor across the measured performance metrics, we increasingly see more gender equality qualitatively, as over time, the lines of best fit for both genders have trended closer to another. This means that both genders are performing relatively equally across all ranges of performance (eg. win rates, OTR Score, speaker points) and that there are no outliers 'blowing up' the

average. Even at the upper end (the most performant groups), we see that this trend holds true. In fact, Abbott discovers similar results, finding that females' expected number of rounds 1 increased by 17.34% between 2015 and 2016 (Abbott (2018)).

5.3 Attribution

There are several factors that could be responsible for the diminishing gender disparity in national circuit Public Forum debate over time. The first is student-led advocacy regarding disparities within debate. For example, the Beyond Resolved Foundation is a student-ran organization dedicated to raising awareness about gender and racial issues across all forms of debate (BRF (2019)). This can also take place in-round. In the final round of the 2021 Gold Tournament of Champions, arguably one of the most prestigious rounds in all of debate, a team chose to cede the round after their opponents ran a critical argument, commonly referred to as a 'K', about troubles faced by transgender debaters on the circuit (one of the debaters was transgender). Instead of debating, the round was turned into an open conversation about issues and solutions for the problem—the significance of which can't be understated due to the importance of the round (Kentucky (2021)). These arguments are becoming increasingly more common on the circuit, including critiques on gender disparities. Raising awareness is key to addressing these issues. Second, tournament-led advocacy is also a factor that could be responsible for the diminishing gender gap. An example of a tournament with a structure set up for attempting to ensure discrimination-free debate can be seen below:

Harvard (2022) Tournament Equity Policy

Equity Policies. The tournament strives to foster an environment promoting inclusivity and respect. Participants should uphold those values. The tournament subscribes to the NSDA's Conduct Policies. The tournament will abide by Harvard College's policies and procedures for sexual and gender-based harassment and discrimination. They are available here: <https://oge.harvard.edu/policies>. Per its regulations, Harvard College will designate a representative of its Title IX office to be on call during the tournament to receive any complaints directly or indirectly by means of the tournament director. The tournament will also abide by policies governing codes of conduct within the Harvard community, including those prohibiting racial discrimination and harassment. Resources for reporting incidents of discrimination and harassment, in addition to reporting to the Tournament Director, are available here: <https://diversity.college.harvard.edu/harassment-and-discrimination-resources>

Tournament Ombuds Committee. The tournament will designate an ombuds committee of three individuals to receive complaints and concerns regarding diversity, equity, and inclusion issues. You may contact the ombuds committee by emailing ombuds@harvarddebate.org. Please note that this is not the appropriate place to report sexual or racial discrimination and harassment as detailed in the prior section; those reports should be made to the Tournament Director directly who will report to the appropriate Harvard offices. Additionally, concerns about tournament procedure and tabulation should be reported to your event's tab room.

Harvard University's Debate Tournament provided an equity committee that is responsible for addressing in-tournament discrimination. Additionally, they clearly stated that no discrimination on any basis would be permitted. This type of action can certainly deter judges with explicit biases. There is a strong correlation with the increase of both student-led and tournament-led advocacy/awareness regarding the gender disparity and the actual decrease of the gender disparity itself. However, more targeted research regarding the efficacy of both attributions discussed here is needed before showing any causality.

Another point of interest I identify is that females' advantage doubles when going from Raw Speaker Points to Adjusted Speaker Points (detailed in 3.3), which is due to encountering irrational judges at a higher rate. This data serves as a starting point for further research into whether or not these outlier judges are more frequent for females due to discrimination and the extent to which the asymmetrical Adjusted Speaker Point boost is attributable to gender-based discrimination. It is not guaranteed that this phenomenon is even statistically significant. Though the Adjusted Speaker Point advantage for females is twice that of males, the difference is still 0.01 points, so future literature will need to create a framework to determine this, perhaps by studying speaker points from older data.

5.4 Applicability In Other Fields

This research indicates the gender disparities are not inherently permanent and can indeed diminish to the point of statistical insignificance in competitive fields. Though I've given several possible factors that could be responsible for this trend, future research should focus on their efficacies in order to show causality. Then, the impact of incorporating these actions (after appropriate causality has been shown) can be analyzed in other competitive environments, hopefully creating a general, battle-tested plan of attack to be used in any field to minimize gender disparities.

6 Appendix

6.1 Tournaments Scraped

Tournament Name	Season
The Paradigm Dowling Catholic	2019-20
New York City Invitational Debate and Speech Tournament	2019-20
Peach State Classic	2019-20
Barkley Forum for High Schools	2019-20
Lexington Winter Invitational	2019-20
The Tradition	2019-20
Pennsbury Falcon Invitational	2019-20
The Milo Cup at Millard North	2019-20
James Logan Martin Luther King Jr Invitational	2019-20
Tim Averill Invitational online	2019-20
Holy Cross Navy and Old Gold Debate and Speech Exhibition	2019-20
Holiday Classic	2019-20
SF Roosevelt Sweetstakes	2019-20
Colleyville Heritage Winter Invitational	2019-20
Mid America Cup	2019-20
2021 Sunvite	2019-20
Bethel Park Black Hawk Invitational	2019-20
Arizona State HDSHC Invitational	2019-20
48th Annual Laird Lewis Invitational at Myers Park HS	2019-20
Jack Howe Memorial Tournament	
Raymond B Furlong Tournament at Saint James	2019-20
Peninsula Invitational	2019-20
University of Michigan HS Debate Tournament	2019-20
Nano Nagle Classic Formerly Voices	2019-20
Tournament of Champions (Silver)	2019-20
The Princeton Classic	2019-20
Cavalier Invitational at Durham Academy	2019-20
Virtual Scarsdale Invitational Scarvite	2019-20
Tournament of Champions (Gold)	2019-20
ETHS Superb Owl	2019-20
35th Annual Stanford Invitational	2019-20
47th Annual Harvard National Forensics Tournament	2019-20
John Edie Holiday Debates Hosted by The Blake School	2019-20
Ridge Debates	2019-20
The Cougar Classic at the University of Houston	2019-20
Katy Taylor TFA TOC NIETOC Fall Classic	2019-20
National Speech and Debate Season Opener Hosted by UK	2019-20
Columbia University Invitational	2019-20
Golden Desert Debate Tournament at UNLV 2021	2019-20
Bingham PF and Policy Invitational	2019-20
Millard West Wildcat Online Debate Tournament	2019-20
The Ed Long Invitational at The Hockaday School	2019-20

Tournament Name	Season
Cal Invitational UC Berkeley	2019-20
Yale University Invitational	2019-20
Apple Valley MinneApple Debate Tournament	2019-20
Badgerland Chung vitational	2019-20
Villiger 41	2019-20
46th University of Pennsylvania Tournament	2019-20
Nova Titan Invitational	2019-20
Grapevine Classic	2019-20
Virtually the Saints Classic	2019-20
The 46th Winston Churchill Classic TOC and NIETOC Qualifier	2019-20
Isidore Newman School Invitational	2019-20
Seattle Academy Invitational	2019-20
Alta Silver and Black Invitational	2019-20
Lakeland Westchester Classic 2021	2019-20
Glenbrooks Speech and Debate Tournament	2019-20
Florida Blue Key Speech and Debate Tournament	2019-20
John Lewis SVUDL Invitational formerly SCU	2019-20
Dempsey Cronin	
Plano West TFATOC Qualifier	2019-20
The Longhorn Classic Online	2019-20
Cal State Fullerton Invitational	2019-20
The Paradigm Dowling Catholic	2021-22
New York City Invitational Debate and Speech Tournament	2021-22
Three Rivers TOC NIETOC at Upper St Clair	2021-22
Peach State Classic	2021-22
John Lewis SVUDL Invitational formerly SCU	2021-22
DempseyCronin	
Barkley Forum for High Schools	2021-22
Saints Classic	2021-22
Lexington Winter Invitational	2021-22
The Tradition	2021-22
Pennsbury Falcon Invitational	2021-22
Capitol Beltway Fall Classic	2021-22
Mount Vernon Invitational and NIETOC TOC Qualifier	2021-22
James Logan Martin Luther King Jr Invitational	2021-22
Fall Tournament	2021-22
Lewis and Clark Invitational	2021-22
Tim Averill Invitational online	2021-22
Holy Cross Navy and Old Gold Debate and Speech Exhibition	2021-22
SF Roosevelt Sweetstakes	2021-22
47th University of Pennsylvania Tournament	2021-22
Colleyville Heritage Winter Invitational	2021-22
Mid America Cup	2021-22
Stephen Stewart Middle and High School Invitational	2021-22
Bellaire Forensic Tournament	2021-22
Arizona State HDSHC Invitational	2021-22
Jack Howe Memorial Tournament	2021-22

Tournament Name	Season
Raymond B Furlong Tournament at Saint James	2021-22
Peninsula Invitational	2021-22
University of Michigan HS Debate Tournament	2021-22
Yale University Invitational 2021	2021-22
Palm Classic	2021-22
Cavalier Invitational at Durham Academy	2021-22
Virtual Scarsdale Invitational Scarvite	2021-22
The Longhorn Classic	2021-22
Plano West Classic	2021-22
Nano Nagle Classic and Nano Nagle RR	2021-22
John Edie Holiday Debates Hosted by The Blake School	2021-22
Ridge Debates	2021-22
Katy Taylor TFA TOC NIETOC Fall Classic	2021-22
Golden Desert Debate Tournament At UNLV	2021-22
National Speech and Debate Season Opener hosted by UK	2021-22
Columbia University Invitational	2021-22
The Ed Long Invitational at The Hockaday School	2021-22
The 47th Churchill Classic TOC and NIETOC Qualifier	2021-22
Apple Valley MinneApple Debate Tournament	2021-22
Badgerland Chung vitational	2021-22
Nova Titan Invitational	2021-22
Sunvite	2021-22
University of Houston Cougar Classic	2021-22
Grapevine Classic	2021-22
Jean Ward Invitational hosted by Lewis Clark College	2021-22
49th Annual Laird Lewis Invitational at Myers Park HS	2021-22
The CSUF Invitational	2021-22
Isidore Newman School Invitational	2021-22
Chicago Classic	2021-22
Puget Sound High School Tournament	2021-22
Silver and Black	2021-22
Princeton Classic	2021-22
Glenbrooks Speech and Debate Tournament	2021-22
Florida Blue Key Speech and Debate Tournament	2021-22
LSW Silver Talon	2021-22
Villiger 42 Saint Josephs University	2021-22
La Costa Canyon Winter Classic	2021-22
48th Annual Harvard National Forensics Tournament	2021-22

6.2 Mean Performance Indicators by Gender

When examining data for win percentage ranges, keep in mind that mixed teams (Male-Female) are not studied in this paper. Additionally, many tournaments stop debating as early as the Quarterfinal round,

leaving ‘Co-Champions’. Both of these factors explain why the win rate do not blend to 50% in any given year’s preliminary and elimination rounds.

6.2.1 2019-20 Season

Gender	OTR Score	Raw Speaker Points	Adj. Speaker Points	Prelim. Round Win %	Elim. Round Win %
Male	0.17	28.30	28.36	51%	52%
Female	0.09	28.20	28.26	39%	46%

6.2.2 2020-21 Season

Gender	OTR Score	Raw Speaker Points	Adj. Speaker Points	Prelim. Round Win %	Elim. Round Win %
Male	0.15	28.29	28.34	44%	51%
Female	0.11	28.25	28.30	41%	55%

6.2.3 2021-22 Season

Gender	OTR Score	Raw Speaker Points	Adj. Speaker Points	Prelim. Round Win %	Elim. Round Win %
Male	0.12	28.32	28.36	43%	56%
Female	0.12	28.33	28.38	44%	56%

6.3 Initial OTR Score Tournament Boost Factors

Tournament Gold Bid Round	Tournament Boost ₀
Finals	1
Semifinals	1.25
Quarterfinals	1.55
Octafinals	2

6.4 Structure Of A Public Forum Round

1. Teams flip a coin, the winner chooses their side (PRO/CON) or speaking order (1st/2nd). The loser gets their choice in whatever wasn't chosen by the winner.
2. The 1st speaker from the team speaking 1st in round gives a 4 minute constructive speech.
3. The 1st speaker from the team speaking 2nd in round gives a 4 minute constructive speech.
4. There is a cross-examination period¹⁹ between both first speakers for 3 minutes.
5. The 2nd speaker from the team speaking 1st in round gives a 4 minute rebuttal speech.
6. The 2nd speaker from the team speaking 2nd in round gives a 4 minute rebuttal speech.
7. There is a cross-examination period between both second speakers for 3 minutes.
8. The 1st speaker from the team speaking 1st in round gives a 3 minute summary speech.
9. The 1st speaker from the team speaking 2nd in round gives a 3 minute summary speech.
10. All 4 debaters engage in a grand cross-examination for 3 minutes.
11. The 2nd speaker from the team speaking 1st in round gives a 2 minute final focus speech.
12. The 2nd speaker from the team speaking 2nd in round gives a 2 minute final focus speech.
13. The judge writes a reason for decision that includes the winning team and speaker points (if instructed by the tournament) for all debaters.

7. References

- Abbott, A. (2018). *Gender inequality in speech and debate: To what extent are female competitors statistically disadvantaged in high school public forum debate?* Retrieved from https://drive.google.com/file/d/1-UpUVxGRU8cAn8SKvwXaPf7Y_pXAZZtW/view
- Andersen, S., & Ertac, S. (2013). Gender, competitiveness, and socialization at a young age: Evidence from a matrilineal and a patriarchal society. *The Review of Economics and Statistics*, 95, 1438–1443.
- Bird, S. (2004). NLTK: The natural language toolkit. *Linguistic Data Consortium at the University of Pennsylvania*.
- BRF. (2019). *Beyond resolved: home*. Retrieved from <https://www.beyond-resolved.org/>
- Burhan, A. (2021). Alex morgan says u.s. Soccer's offer of equal pay at reduced level is 'not good enough'. *Forbes*. Retrieved from <https://www.forbes.com/sites/asifburhan/2021/09/15/alex-morgan-believes-offer-of-equal-pay-at-reduced-level-is-not-good-enough/?sh=6d61e428542f>
- Chitgopekar, S. (2021). *Http-samc/tabroom-api: Dataset for PF tournaments on tabroom.com*. Retrieved from <https://github.com/http-samc/tabroom-API>
- Harvard. (2022). *2021 UKTOC public forum gold finals*. Retrieved from https://www.tabroom.com/index/tourn/index.mhtml?tourn_id=21532
- Kentucky, U. of. (2021). *2021 UKTOC public forum gold finals*. Retrieved from https://www.youtube.com/watch?v=7iG_0V9dGRo
- Nie, A., & Yi, A. (2020). *An empirical study of gender differences in competitive high school debate*. Retrieved from <https://ssrn.com/abstract=3715996>

¹⁹In cross-examinations, it is custom for the first speaking team to get the first question, after which speakers alternate until time runs out.

- Robinson, D. (2021). Pay disparity between NBA, WNBA is a numbers game. *Deseret News*. Retrieved from <https://www.deseret.com/sports/2021/4/7/22367915/pay-disparity-between-nba-wnba-is-a-numbers-game>
- Tartakovsky, D. (2016). *Gender disparities in competitive high school debate new evidence from comprehensive tournament results data*. Retrieved from https://www.vbriefly.com/wp-content/uploads/2016/05/Tartakovsky_Tabroom_Analysis.pdf
- Tartakovsky, D. (2017). *Gender differences in reactions to setbacks: Evidence from high school debate competitions*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3061621
- Vaidyanathan, A., & Chitgopekar, S. (2021). *The OTR score*. Retrieved from <https://tournaments.tech/methodology>