Disparate Judge Feedback for Male and Female Debaters: Evidence from High School Public Forum

Eliza Spinna

Stuyvesant High School

January, 2019
Abstract

Recent research has shown that males are disproportionately represented in the speech and debate community in both participation and achievement. My study analyzes whether women receive less, or less meaningful, feedback on their debate rounds from judges. I use content analysis to analyze public forum debate ballots from the 2016 New York State Championship tournament. I categorize the feedback males and females receive into “argumentation,” “presentation,” and “empty” comments. My research concludes that men receive more comments for all categories on their debate rounds. This may indicate that judges pay more attention to male debaters and are more willing to give them feedback. Receiving less feedback than their male peers may discourage women from further participation in debate competitions and impede their progress towards achieving their educational and career goals.
1. Introduction

Public Forum Debate (PFD) is a competitive high school forensics event across the United States. It was created by the National Speech and Debate Association in 2002. PFD is structured around holding a debate between two teams of two competitors each. Each team argues in support of or against a given topic.

PFD is a unique form of debate in that it is designed to be judged by a layperson, or a person with no formal debate experience. PFD thus removes technical aspects of debate inherent to other formats such as speed-reading, which requires extensive coaching to accomplish. By increasing the accessibility of debate and membership, involvement in PFD has skyrocketed to become the most popular event sponsored by the National Speech and Debate Association (NSDA).

Despite these developments, a significant gender gap exists within the debate community. Most members of the speech and debate community accept that males disproportionately represent the debate community in both participation and achievement (Abbott, 2018).

The creation of PFD has come at a cost: lay judges do not have training in debate and thus may not have a strong grasp on the styles of argumentation used by debaters. At the end of the round, the judge is required to fill out a ballot in which the judge provides feedback on each individual debater and the reason each team won or lost the round. The NSDA guidelines encourage lay judges to write comments on the substance of the round, including the argumentation, evidence, and rhetoric used by the debaters. However, some judges make comments about superficial aspects of the competition, such as the physical appearance or voice quality of the debaters. This practice is generally discouraged by the NSDA and tournament directors as it does not constitute productive feedback (Abbott, 2018).

Research on disparities in how debaters are evaluated has significant implications for the debate community. Joshua Feinzig and Natalie Atyeo’s research on the gender disparity in another debate event, Lincoln Douglas Debate, explains, “When individual rounds are affected by judging bias, the activity
ceases to incentivize genuine participation and the educational benefits of the activity diminish. Students, coaches, and schools are much less likely to participate in debate if subjectivity is prevalent.”

In the past few years, several studies of the gender composition of debate have been conducted by debaters and coaches. On March 7, 2018, Allan Abbott published a statistical analysis of the gender disparity in Public Forum Debate called “Gender Inequality in Speech and Debate” (Abbott, 2018). His research indicates that only 38% of competitors are female, a significant imbalance, but one that is improving at a rate of about 11% annually. He also found that, on average, male teams win at a rate that is 37.6% greater than female teams, and 15.83% higher than male-female teams. Thus, female debaters are statistically underrepresented and disadvantaged in PFD. Abbott’s research added to a growing body of literature advocating for nationwide efforts to increase inclusively within the activity (Abbott, 2018; Feinzig and Atyeo, 2011).

Abbott’s research does not distinguish between the two ways females are excluded from the activity. The first is that female high schoolers do not join debate in the first place. J. Cinder Griffin and Holly Jane Raider, two of the first researchers to study high school debate through the lens of gender, write in their study “Women in High School Debate” that:

Structural barriers endemic to the forensics community dissuade female ninth graders from entering the activity. Recruitment procedures and initial exposure may unintentionally create a first impression of the activity as dominated by men. By and large, it is a male debater or a male debate coach that will discuss the activity with new students for the first time. Additionally, most debate coaches are men. This reinforces a socially proven norm to prospective debaters, that debate is an activity controlled by men (Griffin and Raider, 1989).

The second is that even when females do join, they drop out at higher rates than male debaters. Griffin and Raider find that “while entry barriers are formidable, female attrition rates affect the number of females in the activity most significantly.” They ultimately conclude that “novice, female debaters
have few role models and, consequently, are more likely to drop out than their male counterparts; resulting in an unending cycle of female attrition” (Griffin and Raider, 1989).

Of these two factors in the rate of female participation, my study will focus on the latter; only females who have already joined the activity participate in rounds and receive ballots. Incentivizing females to join debate is fruitless if females believe they will not be judged on the same criteria as their male peers.

While no prior research has been conducted on the comments females receive on their ballots, research has been dedicated to collecting data on attitudes toward females in speaking roles by Dr. Linda Carli in 2001. The research indicates that observers have held women in speaking roles to a lower standard than men and expect women to perform poorly. Ultimately, Carli concludes that females must not only work to overcome their opponents, but also to surpass societal assumptions about gender (Carli, 2015). Debate events diverge from the individual interactions Carli examines in that they provide judges with the ability to create side-by-side comparisons of debaters, thus further enhancing and revealing these expectations. The ballot provides an easy way to study these judgments.

2. Methodology

In my study, I use ballots from the 2016 New York State Championship Tournament competition. They are posted on Tabroom.com, a website debaters use to receive and review online ballots. Each ballot is handwritten by a judge and scanned to one .pdf file. Such a ballot is required to be submitted for every debate round in each division (varsity, junior varsity, and novice) in this tournament.

I chose this specific tournament to study because it is conducted in the format for which PFD was designed: the tournament is known for inviting mostly lay judges who are parents of debaters. I analyzed the novice (first-year division) as it has a higher percentage of female debaters in the pool than the junior varsity or varsity divisions. The novice division in the 2016 New York State Championship tournament
competition was 58% male and 42% female. In addition, the novice division has the largest and most complete set of ballots.

The tournament hosts 150 varsity debaters (75 teams) in PFD. Each team debates four rounds. A total of 150 ballots are thus published each year for varsity PFD. Each ballot describes an individual debate and thus evaluates four debaters. Therefore, in the data I used from this tournament, 600 individual debaters are evaluated in total (four per ballot per round).

For this tournament, I found that a debater’s individual evaluation on a ballot ranges from zero to nine comments. I categorized each of these comments into two larger code categories: “argumentation” and “non-argumentation” comments, and used content analysis to categorize the individual comments into one of these categories. I followed E. R. Babbie’s guidelines for creating coding categories from “The Practice of Social Research.”

Argumentation comments focus on the logic and rhetoric in the round and the quality of the argumentation and evidence introduced. Judges are directed to make these comments by tournament directors because they provide meaningful feedback on how debaters can improve their argumentation going forward. Listed below are common examples of argumentation comments found on 2017 New York State Championship Tournament competition ballots.

**Examples of Argumentation Comments**
This argument was effective because your evidence was more recent and more credible.
I found this argument to be unconvincing because of the lack of analysis.
Your arguments should be more developed and are missing a critical piece of evidence.
I agree with your argument that reducing poverty is more important than your opponent’s argument about national debt.

My second category of comment is “non-argumentation” comments, which I subdivided into three subcategories: “presentation” comments, “argumentation” comments, and “empty” comments.
Presentation comments have to do with a debater’s style of argumentation and use of aggression, but not the argumentation itself. These comments may give debaters constructive feedback on how to visually improve their style, but can be discouraging because these comments neglect to acknowledge the argumentative preparation a debater has done. Also, broad comments calling a debater “too aggressive” do not give the debater meaningful and specific feedback on how to improve this quality.

Because the idea of presentation can be nebulous, I ensured consistency by defining a set of words and phrases that fall into this category. Listed below are common examples of presentation comments found on 2017 New York State Championship Tournament competition ballots.

**Examples of Presentation Comments**
Your style of argumentation was too aggressive.
Calmer presentation would be appreciated
Slow down/speed up your speaking.
Your confidence is good.
Body language good
Holding a computer a certain way is distracting.

My second category of non-argumentation comment is the empty comment. Empty comments provide generic feedback that does not direct the debater on specific aspects of their performance to be improved. These comments may boost a debater’s confidence if they provide praise, but may also confuse debaters who do not understand why the comment was assigned. Like presentation comments, empty comments do not provide a clear avenue as to how the debater can improve their performance. Listed below are common examples of empty comments found on 2017 New York State Championship Tournament competition ballots.

**Examples of Empty Comments**
Good points.
Strong rebuttal.
Weak speech.
Speaker was average.
Lastly, the most superficial "appearance" comments relate to a debater's vocal tone or pitch, or to their physical appearance (such as clothing, make-up, and hair). These are uncommon because tournament directors discourage them. Appearance comments do not give debaters any means to improve their speech, and may discourage debaters who believe that they are being evaluated by factors beyond their control or irrelevant to the substance of the debate. Listed below are examples of appearance comments found on 2016 and 2017 New York State Tournament competition ballots.

**Examples of Appearance Comments**
Reduce your vocal fry.
Fix your collar.
Your voice is too high/low/loud.
Dress more formally/casually.

The chart below delineates the categories into which I sorted the comments.

```
Type of Comment
  | Argumentation Comments | Non-Argumentation Comments |
  |                       |                           |
  | Appearance Comments   | Presentation Comments     |
  | Empty Comments         |
```

In the study, one comment is counted as a single unique observation. Comments about the whole round were counted for all speakers. Comments about the partnership were counted for both speakers in the partnership.

The study was conducted in two phases. For guidelines as to how to conduct my content analysis, I followed E. R. Ballie's Babbie's instructions in "The Practice of Social Research."
The first phase was the codebook construction, during which I examined 100 ballots from the 2017 New York State Championship Tournament competition. This phase enabled me to assemble my codebook guidelines for which comments fit into each category, so that when I conducted the second phase of data collection (with the 2016 New York State Championship Tournament competition ballots), there was less categorization coding uncertainty. I chose the 2017 New York State Championship Tournament competition for the standardization set because 100 ballots were available online, whereas the 2016 New York State Championship Tournament competition had slightly larger set of 125 ballots, so I used those for the second phase of the content analysis.

Each time I made a decision about whether a comment fits into one of these categories, I added it to my codebook of acceptable comments to use as a reference during phase two.

Since this study used content analysis, I asked four peers to use the list of phrases I compiled to code each comment on 20 of the 2017 New York State Tournament ballots. They were provided with my codebook. Two of the peers were debaters (one male, one female) and two were non-debaters (one male, one female). I coded each comment on the same 20 ballots independently. My peers' decisions on which category each comment fit into matched mine, thus implying that my codebook was clear and standardized. My peers also agreed that the categories were clearly defined.

After I defined the categories and listed examples of each comment, I moved on to phase two, in which I coded the comments from the ballots of the 2016 sample. For each ballot, I analyzed each phrase and coded it into the empty, presentation, appearance, or argumentation categories. I also calculated the total comments offered for each debater.

I was able to identify a debater's gender based on the school they attended (more than two thirds of the competition pool attends same-gender private schools) or based on my own knowledge (I personally know, and know the gender of, most debaters on the circuit). The gender composition at the 2016 New York State Championship Tournament competition was 58% male and 42% female. Since I
did not have a large sample of gender non-binary debaters, I only categorized and evaluated the comments of male and female debaters. There was a single gender non-binary debater in the pool, whose results I did not factor into either the male or female groups.

3. Results

My hypothesis was that female debaters would receive fewer total and argumentation comments than male debaters. I also hypothesized that females would receive more presentation and empty comments than males. However, the results indicate that male debaters receive more comments in every category.

As shown in Table 1: Total Comments, female debaters received zero comments 11.69% of the time, whereas males received zero comments 7.10% of the time. At the other end of the spectrum, male debaters received five or more comments 16.19% of the time, whereas female debaters received five or more comments only 4.44% of the time. Thus, male debaters were far more likely to receive a ballot with several comments on their performance, whereas female debaters were not.

I merged the five through nine comment groups because the difference between receiving five comments and nine comments is minimal. Either way, the debater is given ample feedback on how to improve. On the other hand, the difference between receiving zero and five comments is significant.

Table 1: Total Comments

<table>
<thead>
<tr>
<th>Gender</th>
<th>0 comments</th>
<th>1 comment</th>
<th>2 comments</th>
<th>3 comments</th>
<th>4 comments</th>
<th>5+ comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>11.69%</td>
<td>20.56%</td>
<td>24.60%</td>
<td>25.40%</td>
<td>13.31%</td>
<td>4.44%</td>
</tr>
<tr>
<td>Male</td>
<td>7.10%</td>
<td>17.90%</td>
<td>25.28%</td>
<td>20.45%</td>
<td>13.07%</td>
<td>16.17%</td>
</tr>
</tbody>
</table>

Chi-Square = 23.32, df = 5, p < .000
Argumentation comments followed a similar trend. As shown in Table 2: Argumentation Comments, male debaters received five or more comments on their argumentation 6.53% of the time, whereas female debaters received five or more argumentation comments only 1.61% of the time. Similarly, male debaters received four comments on their argumentation 4.83% of the time, whereas females received four comments 3.23% of the time. However, male and female debaters received 0 comments at a roughly equal rate of 29.84 and 29.55 percent of the time, respectively.

Table 2: Argumentation Comments

<table>
<thead>
<tr>
<th>Gender</th>
<th>0 comments</th>
<th>1 comment</th>
<th>2 comments</th>
<th>3 comments</th>
<th>4 comments</th>
<th>5+ comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>29.84%</td>
<td>25.00%</td>
<td>15.55%</td>
<td>21.77%</td>
<td>3.23%</td>
<td>1.61%</td>
</tr>
<tr>
<td>Male</td>
<td>29.55%</td>
<td>25.00%</td>
<td>19.03%</td>
<td>15.06%</td>
<td>4.83%</td>
<td>6.53%</td>
</tr>
</tbody>
</table>

Chi-Square = 12.43, df = 5, p=.029

Male and female debaters received presentation comments at roughly similar rates, and the differences were not statistically significant. However, it is worth noting that male debaters received three or more presentation comments 3.41% of the time, whereas female debaters received three or more presentation comments 1.21% of the time. See Table 3: Presentation Comments.

Table 3: Presentation Comments

<table>
<thead>
<tr>
<th>Gender</th>
<th>0 comments</th>
<th>1 comment</th>
<th>2 comments</th>
<th>3+ comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>68.95%</td>
<td>24.19%</td>
<td>5.65%</td>
<td>1.21%</td>
</tr>
<tr>
<td>Male</td>
<td>68.47%</td>
<td>22.73%</td>
<td>5.40%</td>
<td>3.41%</td>
</tr>
</tbody>
</table>
Chi-Square = 2.97, df = 3, p = .396

Empty comments were given more frequently to male debaters at a statistically significant rate. As shown in Table 4: Empty Comments, male debaters received three or more empty comments 4.26% of the time, whereas female debaters received three or more empty comments 2.82% of the time. Similarly, male debaters received two empty comments 11.65% of the time, whereas female debaters received empty comments 6.85% of the time. On the other end of the spectrum, female debaters received zero empty comments 78.23% of the time, whereas male debaters received empty comments 61.65% of the time.

Table 4: Empty Comments

<table>
<thead>
<tr>
<th>Gender</th>
<th>0 comments</th>
<th>1 comment</th>
<th>2 comments</th>
<th>3+ comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>78.23%</td>
<td>12.10%</td>
<td>6.85%</td>
<td>2.82%</td>
</tr>
<tr>
<td>Male</td>
<td>61.65%</td>
<td>22.44%</td>
<td>11.65%</td>
<td>4.26%</td>
</tr>
</tbody>
</table>

Chi-Square = 18.69, df = 3, p < .000

The final category was appearance comments. These were given very infrequently, and only to male debaters. In one of the cases, the judge noted that the appearance comment did not factor into the debater’s win/loss decision or individual speaker score. As shown in Table 5: Appearance Comments, male debaters received appearance comments 1.42% of the time, whereas female debaters received no appearance comments throughout the entire tournament. The results were approaching significance according to the chi-square test.

Table 5: Appearance Comments
<table>
<thead>
<tr>
<th>Gender</th>
<th>0 comments</th>
<th>1 comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>100.00%</td>
<td>0%</td>
</tr>
<tr>
<td>Male</td>
<td>98.58%</td>
<td>1.42%</td>
</tr>
</tbody>
</table>

Chi-Square = 3.55, df = 1, p = .059

Thus, in all categories, male debaters tended to receive more comments on the upper end of the scale and female debaters received fewer.

4. Conclusion and Future Studies

The results of this study show that male debaters receive more comments across all categories. This surprised me because I had predicted male debaters would receive more total and argumentation comments, but female debaters would receive more presentation, empty, and appearance comments.

Male debaters received more total comments, which may be attributed to judges paying more attention to male debaters. Specifically, judges may give more thought to the argumentation of male debaters, and thus write more comments in the argumentation category for male debaters. This trend likely results in female debaters feeling as though their argumentative additions to the round are not valued by judges. Male debaters, on the other hand, can build on the feedback they receive and use it to improve their arguments, thus creating a self-perpetuating cycle wherein male debaters are given ample opportunities to grow and reap the intellectual benefits of the activity.

The results in the empty category surprised me because I had predicted that females would receive more empty feedback on their speeches as judges focused on the argumentation of their male peers. However, judges were more willing to heap empty feedback onto the male debaters. Comments such as “your speech was interesting” appeared frequently. These comments do not give male debaters
any avenues to improve their debating, but they signal to male debaters that their presence is acknowledged. Empty comments may give male debaters a sense of general confidence in their efficacy within the round. It may be worth conducting a future study of whether these empty comments tended to be positive or negative, and if male and female debaters received more positive or negative empty comments. Receiving the empty comments “speaker was interesting” or “speaker was below average” may have different impacts on a debater’s self-esteem and self-efficacy, which could be clarified by further research.

The presentation comments were not statistically significant and male and female debaters received them in roughly even quantities, except that male debaters did receive 3+ comments slightly more frequently than females. This trend of equalizing the number of presentation comments given to male and female debaters is likely due to tournament directors’ active discouragement of the usage of presentation comments on the ballot. This may imply that the debate community is drifting away from presentation-based judging and towards argumentation-based judging.

Lastly, the appearance comments were given very infrequently, likely because tournament directors discourage judges from giving them. In addition, the debate community has become increasingly aware of the dangers of commenting on a debater’s appearance. Judges are very hesitant to comment on a woman’s appearance, likely because of the backlash they would receive of the appearance comment was brought to the attention of the tournament director. One notable appearance comment was “fix your collar,” which the judge noted did not have an impact on the male debater’s speech, but still felt was important to comment on. One implication for future study is how judges give appearance comments based on a debater’s race; the comment about the collar was given to an Asian-American debater. While this may simply be a coincidence, it is worth looking into whether a debater’s race has an impact on the content of their comments.
It also may be worth looking into the impact of socio-economic class on a debater's performance and judge feedback. The 2016 New York State Championship tournament hosts over two thirds of its debaters from same-gender private schools, which may have had an impact on the results of this research. Wealthy debaters have more time and money to buy clothing in the "western business attire" style often recommended, so this may have been a factor in the low number of appearance comments. It also may be informative to study the impact of attending an all-girls private school on female retention rates after novice year. It is possible that being surrounded by female peers, partners, and leaders increases the likelihood of a novice debater remaining on the team.

Another potential follow-up study would be to examine the differences between how male and female judges evaluate debaters. It is possible that a large proportion of male judges contributed to the high percentage of comments received by male debaters. It is also possible that female judges may pay more attention to female debaters' argumentation. If this is true, then tournament directors may have a special responsibility to ensure that debaters are being judged by equal numbers of male and female judges, and to educate male and female judges on potential biases implied by assigning different quantities of comments to debaters of different genders.

Overall, I think that these findings can make an impact on the debate community and cause judges to be more mindful of how they evaluate debaters. I am excited to add to the body of research being conducted on how gender impacts the activity of debate. In addition, I believe this research project will provide insights with broader applicability to the understanding of how females are perceived and evaluated in their critically formative teenage years, and how such perceptions and evaluations may impede their progress towards achieving educational and career goals in later years.
Works Cited


