

---

September 2020

## Rummaging Through Cumes: What Existing Results Data for LD at the NFA National Tournament Can Tell Us About Tournament Design

Joseph Dudek  
*University of the Pacific*

Follow this and additional works at: <https://cornerstone.lib.mnsu.edu/ndcieproceedings>



Part of the [Speech and Rhetorical Studies Commons](#)

---

### Recommended Citation

Dudek, J. (2010). Rummaging through cumes: What existing results data for LD at the NFA National Tournament can tell us about tournament design. In D. Cronn-Mills, & L. Schnoor (Eds.), Conference Proceedings: National Developmental Conference on Individual Events 2010 (pp.41-45). Minnesota State University, Mankato.

This is brought to you for free and open access by Cornerstone: A Collection of Scholarly and Creative Works for Minnesota State University, Mankato. It has been accepted for inclusion in Proceedings of the National Developmental Conference on Individual Events by an authorized editor of Cornerstone: A Collection of Scholarly and Creative Works for Minnesota State University, Mankato.

## Rummaging Through Cumes: What Existing Results Data for LD at the NFA National Tournament Can Tell Us About Tournament Design

Joseph Dudek  
University of the Pacific

### Abstract

The top seeded LD competitor entering out-round competition has lost their first out-round in each of the last two NFA National Tournaments. This anecdote highlights the immensely dynamic nature of the event and, perhaps, begins to question the sanctity of out-round seeding. In order to better understand the nature of preliminary rounds, their importance in deciding who will advance to out-rounds, and how many of them are actually necessary, it is imperative that we dissect the results from recent tournaments and work toward creating a more fair and competitive tournament. In this paper, I use results data from the 2009 and 2010 NFA National Tournaments to understand how rounds 5 and 6 impact out-round seeding for both individuals and their teams so that we can consider carefully the effect of moving to a 4-round tournament.

### Introduction

The spring, 2009 NFA business meeting at Drury University introduced to the community the continuing problem of finding a manageable and pedagogically useful way to implement LD debate at the national tournament. NFA President Larry Schnoor noted that this section of the business meeting was designed to encourage discussion and ideas and numerous members provided suggestions for how to best solve the issue. While the various approaches presented represented very thoughtful consideration of some issues associated with education and budget, note that very little of the discussion was motivated by collected and analyzed data. It is our hope that this paper provides some profiles of existing data on LD that will help create an even more informed debate on the topic.

As the sudden emergence of this debate would predict, there has been relatively little research conducted on how LD should be implemented at the National Tournament. The vast majority of research surrounding LD deals with more blatantly controversial issues like judging philosophy (see Bile 1996; Burkholt & Diers 2004), debate theory (see Abrams & Novak 1997), and event accessibility (see Shelton & Patterson 1997; Minch 2002; Millsap & Millsap 2006). Very little of this kind of research is useful in assembling useful suggestions for administrative changes to the event akin to those suggested during the aforementioned business meeting.

This paper, therefore, attempts to close this gap in the literature by formulating realistic administrative solutions based upon their predicted impact on debate itself. This is done by evaluating the importance of late prelim rounds on prelim seeding and evaluating how that change in seeding is likely to affect out-round performance. Maintain that if a shorter tournament is a viable alternative, it would help make room for several of the suggestions posed by attendees

of the business meeting.

### Method

In this paper, treat the preliminary rounds (prelims) as an evaluative tool designed to determine the caliber of debaters for use in selecting the best 32 debaters to enter out-rounds. As such, this tool is subject to questions of reliability and validity, even if those concepts take on slightly different forms in context.

Because of the remarkable accessibility brought about by the digital publication of the 2009 NFA LD results, it became possible to construct a reasonably simple computer program to parse that data and begin to analyze it deeply. As a result, the following statistical analysis is done exclusively on the 2009 data. The 2010 data, though published in digital form, was not compatible with text parsing making its analysis vastly more arduous. To accommodate this fact, the statistical analysis of the 2009 data will be followed by an anecdotal analysis of the 2010 data to address similarities and differences between the sets.

### The Construct

As the National Tournament is designed at some foundational level to find and award the best competitor in any given event, it seems reasonable to suppose that the purpose of L.D. prelims is to sort debaters based on skill. In this conception of the event, the prelims become a measurement tool designed to evaluate debater skill. The most skilled debaters are then selected to engage in a single-elimination tournament to establish a champion.

It is important to note that, while much of our discussion will surround debater skill, the construct is not necessary for the statistical analysis to be useful. The analysis of data below discusses real numbers and stable predictions, regardless of what the motivator of those predictions is. The construct simply acts as a justification for the nonspuriousness of the relationships established and as a foundation for our hypotheses. If debater skill is a thing that exists, than it ought to impact how quickly debaters arrange themselves by skill in prelims and how accurately prelims predict out-round success.

### Reliability

We use the term 'reliability' to refer to the power of prelims to hold rankings relatively constant after a certain number of rounds have been finished. If prelims are designed to accurately rank debaters in terms of skill, then the debaters' rank should become relatively steady as the number of rounds increases. This notion of reliability deviates from most commonly accepted approaches to the topic (Schutt 2009:135-8). That said, it is the only available mechanism to evaluate reliability absent a second sample or another existing metric for evaluating a debater's skill and is at least

conceptually related to a foundational notion of test-retest reliability.

In order to give several useful profiles of the data, I compute the mean distance traveled (MDT) along the rankings by debaters between any two rounds. I predict that the MDT will decrease as the number of rounds increases and that average distance traveled will be particularly low for top-tier debaters during the last three rounds.

**Validity**

Validity usually describes the extent to which a measurement tool is actually measuring what it set out to understand. I am looking to see if this tool is actually picking out top-tier debaters so that they can compete against each other in out-rounds. This is a remarkably difficult task, as there does not seem to be a quantifiable metric for debater skill.

I, therefore, ground our meaning of validity in the formulation of a useful criterion. In this case, I am looking to see whether prelim ranking is a reasonable predictor of out-round success. As Carmines and Zeller (1979) note, criterion-related validity “has the closest relationship to what is meant by the everyday usage of the term” (p. 17). Our community tends to share the notion that those who are successful at national out-round competition tend to be among the most skilled debaters at the tournament. Moreover, data about prelims as a predictor of out-round success can be useful to policymakers within the event even absent our construct.

In order to allow for quantitative analysis of ranking data, I assign ranks to debaters based on their placement during out-rounds as the maximum rank they could have been given which round they lost in. A semifinalist, for example, acquires a rank of 3, as only 2 debaters advanced farther than them.

Of particular interest to us is a category of debaters who would not have broken if the tournament ended after 4 rounds, but broke as a result of the final two rounds. If this group of debaters did particularly well in out-rounds, then it was of critical importance that they be in the top 32 for prelims to have effectively found the top debaters. In essence, I evaluate the validity of a 4-round version of the tournament as being inversely related to how far this group of debaters advanced as a result of rounds 5 and 6. In doing

so, I hypothesize that both the 4- and 6-round tournaments will be reasonable predictors of out-round success.

**Sample**

2009 was the first year after which NFA released all of the national’s results in a digital form. As this paper is meant only as a pilot study on relatively accessible data, the results from this national tournament is the entire sample (R83). Because I have a particular interest in those debaters who break to out-rounds as a result of prelim success, I break this sample into several subgroups. The first subgroup consists of the debaters who broke to out-rounds at the tournament which consists of the top 32 ranks after 6 prelims (R32). I then further bifurcate this group into R16 and R8, the top 16 and 8 debaters respectively. Our construct would indicate that R8 represents a uniquely skilled set of debaters.

**2009 Results**

Reliability tests demonstrated that MDT decreased as rounds progressed for every sample. A linear regression on R83 revealed that round number and MDT were inversely correlated with  $r^2=0.97$  and  $p < 0.005$  with the average debater moving only 9.24 places between rounds 5 and 6. Further analysis revealed that the average member of R16 moved only 12.5 spots between rounds 4 and 6, meaning they must have been in the top 32 after round 4. Additionally, the average member of R8 moved only 5.75 spots between rounds 4 and 6, indicating they were already at an elite ranking after round 4.

When correlating R83 MDT values with round number, the correlation yielded an unbelievable  $r = 0.98$ . This result suggests that there is a strong source of biased error in these MDT values. This error is best understood as the inability to change one’s rank during later rounds because of the diversity of records. A win when someone is 0-1 is much more likely to cause a drastic shift in their ranking than a win when someone is 3-2. Correcting for this error would require a complex application of combinatorics which is not prudent for our analysis. This biased error would not, however, be near enough to explain the immense rigidity demonstrated in R8.

When I compared prelim ranking with out-round ranking, I found several positive correlations. Round 4 rankings correlated positively with out-round rankings for R32 with  $r = 0.36$  and  $p < 0.02$ . Round 6 rankings correlated positively with out-round rankings for R32 with  $r = 0.41$  and  $p < 0.01$ .

|            | Round 1-2 | Round 2-3 | Round 3-4 | Round 4-5 | Round 5-6 |
|------------|-----------|-----------|-----------|-----------|-----------|
| <b>R83</b> | 15.96     | 13.44     | 12.52     | 11.46     | 9.24      |
| <b>R32</b> | 16.13     | 13.03     | 12.94     | 10.13     | 7.25      |
| <b>R16</b> | 11.94     | 11.38     | 8         | 9.94      | 6.19      |
| <b>R8</b>  | 13        | 7.13      | 5.25      | 5.25      | 3         |

Figure 1 – MDT by round for all samples.

## Discussion

This study found that prelims are a reasonable predictor of out-round success. I note that both 4- and 6-round versions of the national tournament predict out-round success within reasonable parameters. There are, however, some concerns that need to be addressed before concluding that a 4-round tournament would have been sufficient.

First, I should address a common concern that it takes several prelims to ensure that the best debaters have risen to their appropriate rank. Here, the data is very clear. R8 contained all four semifinalists and two of the four non-advancing quarterfinalists. Moreover, every member of R8 would have broken had the tournament been ended after *any round beyond the first*. There are two members of R8 who would not have been in the top 16 after four rounds and they lost in octo-finals and quarterfinals respectively. All of this suggests that it took extremely few rounds to isolate the most skilled debaters atop the rankings.

These results call into question a fundamental assessment of value at the national tournament. While this paper advocates that highlighting the best debaters should be the focus of the national tournament, it is reasonable to suggest that isolating and rewarding the top 32 debaters in a thorough and complete way is also a valuable task. Our data suggests that a choice between a 4- and 6-round tournament is fundamentally a choice between these two kinds of recognition with the 4-round tournament aimed solely at efficiently isolating the very best debaters to ensure that they are in out-rounds.

Next, it seems reasonable to contend that  $r$ -values of 0.41 and 0.36 fall below a significant threshold. Given the degrees of freedom in this calculation, that would be a difficult claim to justify. Moreover, these values for  $r$  are arbitrarily lowered by an inability to create a smooth ranking system for out-round results. Because *all* of the double-octo-finalists are ranked the same, there are large clumps in the data that arbitrarily skew the slope of the best-fit line against the correlation we're hoping to establish. Figure 2 (below) helps to illustrate this point by showing how the best-fit line dodges the most convincing pieces of data in the lower-left section of the scatter plot.

Finally, one might be tempted to argue that those individuals who advanced to the top 32 as a result of rounds 5 and 6 (who would not have broken in a 4-round tournament) had an important impact on out-rounds. The data does not support such a contention. Of the nine debaters for whom this was the case, seven of them lost their double octo-final round and the remaining two lost their octo-final round. This data suggests that out-rounds from quarterfinals on would not have been significantly affected by ending prelims early.

Figure 2 – Correlation data for criterion-validity analysis.

## Applying Data From 2010

The analysis of the 2010 data can only really be done at a 6-round level, as creating seedings for 4-round tournaments would require the data to be vastly more manipulable. That being said, the 2010 data does shine a very interesting light on the sanctity of the bracket in a 6-round tournament, something the community has not yet had a good chance to discuss.

Unlike the 2009 results, no member of the 2010 top 8 group advanced past quarterfinals. In fact, five of the eight lost in octo-finals or earlier. This includes the first- and second-seeded debaters who both lost their double octo-final rounds. One might notice that this is not unprecedented, as the first-seeded debater in 2009 also lost her first out-round debate. Several coaches on the circuit have correlated this early loss to the 6-0 first-seed being forced to debate the top-speaking 3-3 in double-octo-finals. After all, the top-speaking 3-3 seems much more dangerous in double-octo-finals than the bottom-speaking 4-2. This proved insufficient in 2010 when both the top 3-3 and the bottom 4-2 won their double-octo-final round. This year was particularly bad for a linear regression because the 32-seed won the entire tournament.

This tremendous variability among the top 16 (see Figure 4 below) suggests that the seeding system for out-rounds is not accurately serving as a predictor of success among top debaters. As there is little that can be done to change the seedings acquired by competitors, it seems reasonable to consider other policy implications of this obvious imperfection. First, the NFA LD tournament might consider breaking brackets for out-rounds, as there is not a good reason for forcing someone to retain their seed if that seed is an arbitrary variable. Second, this could serve as reasonable (albeit disheartening) evidence that the imperfections in a 4-round tournament are not unique to the smaller tournament, further justifying a shortened prelim schedule.

The most predictable and consistent part of the 2010 data was the out-round result for any debater seeded between 18 and 29. All of these debaters lost the double-octo-final rounds making the 17-, 31-, and 32-seed the only bottom-half debaters to emerge from the first elimination round. An inspection of each of these debaters' performance in rounds 5 and 6 shows a large number of either very high (at or above 28) or very low (at or below 22) speaker points awarded during those rounds. This would seem to hint that the 2009 data's demonstration of the ability of rank variability to predict out-round success is supported.

Below are two graphs that are particularly telling. Figure 3 shows how well seeding predicted performance in 2010 and demonstrates a trend line that looks remarkably similar to that in Figure 2. This must be because of the consistent losses by seeds 18-30, because Figure 4 shows an inverted relationship if we exclude this low-seed population.

Figure 3 – Correlation data for 2010 rankings

Figure 4 – Correlation data for 2010 rankings for the top 16

debaters

**Conclusion**

In many ways, it is not the goal of this paper to provide rigid conclusions. Instead, the paper concludes with a series of questions that the LD community at large ought consider in order to properly address worries identified in the introduction: Is there a reason why we have created a tournament with 6 prelim rounds instead of 4 or 8? Is that reason grounded in any LD-specific analysis?

- Does the lack of seeding sanctity exhibited during the 2010 nationals call into question NFA LD policy on breaking brackets?
- Does the strong correlation between consistency and out-round success justify a new ranking system that is based on something besides win/loss?
- And finally: What else would we like to learn from available data?

**References**

Abrams, R.A. and Novak, C.L. (1997). Lincoln-Douglas debate: Inherency revisited. *National Forensic Journal*, 15(3).

Bile, J.T. (1996). Propositional focus in dialectical perspective: A call for greater consistency in the theory of NFA Lincoln-Douglas debate. *National Forensic*

*Journal*.

Burkholt, M.J. and Diers, A.R. (2004). Judging philosophy: Reflected in the communication of judge's decisions in NFA Lincoln Douglas debate. *National Forensic Journal*, 22(2).

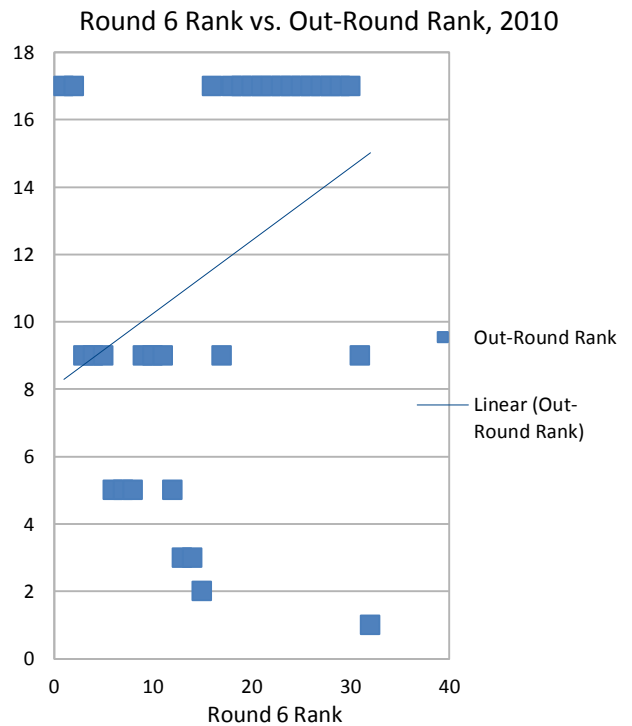
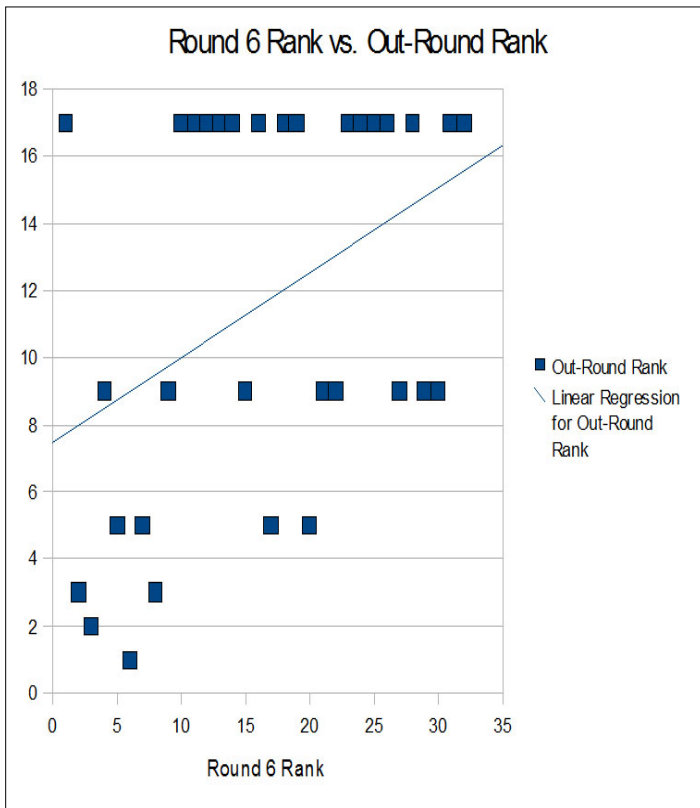
Carmines, E.G. and Zeller, R.A. (1979). *Reliability and validity assessment*. Thousand Oaks, CA: Sage.

Millsap, S. and Millsap S. (2006) Gender bias in NFA-LD: An examination of participation rates and a content analysis of NFA-LD ballots. *National Forensic Journal*, 24(1).

Minch, K.M. (2002). Grappling for original intent: Failed mission or misunderstood goals in NFA Lincoln-Douglas debate? *National Forensic Journal*, 20(1), 49-53.

Schutt, R.K. (2009). *Investigating the social world: The process and practice of research*. Thousand Oaks, CA: Pine Forge Press.

Shelton, M.W. and Patterson, J.W. (1997). Collegiate Lincoln-Douglas debate and high school champions: Implications of a survey of participants in the tournament of champions. *National Forensic Journal*, 15(2).



Round 6 vs. Out-Rounds for Top-16, 2010

