Interpreting the Oral Interpretation Judge: Content Analysis of Oral Interpretation Ballots

Daniel D. Mills*

One of the primary purposes behind forensics competition is the pedagogical benefits students receive from participation in the activity. One of the best methods of advancing a pedagogical base is through the ballots students receive at tournaments. Carey and Rodier (1987) state, "ballots play an important role in the educational process of literature performance and in the forensic activity" (p. 16), and Jensen (1988) notes, "... the benefits competitors receive during the actual tournament are achieved through the critic in each section of any given event" (p. 1). Comments students receive should provide insight into how a competitor might improve in a particular event. However, complaints about comments or a lack of comments are not unusual.

Pratt (1987) noted a ballot serves the two purposes of judging and coaching, and Allen (1983) noted:

One way or another critics must be held accountable for their estimates of performance. They must be able to defend the rankings they assign, and this means being able to justify specific hierarchical standards which are applied to performance. Otherwise we must accept that a great deal of judgment is, alas, based on impressions only, and that such impressions are based on conjecture and opinion as much as fact (p. 52-53).

This study examines written comments students receive on oral interpretation ballots. The first purpose of this study is purely pragmatic—to develop a better understanding of the criteria used to judge competitive oral interpretation events. A second purpose is to examine possible implications resulting from these current practices. While constructive guidelines concerning the judging of individual events have been suggested (Hanson, 1988; Lewis, 1984; Mills, 1983; Ross, 1984; Verlinden, 1987), it is also important to substantiate what is actually occurring in the judging of oral interpretation events.

Content analysis of individual event ballots has primarily been undertaken in original speaking events (Jensen, 1988; Pratt, 1987) and rhetorical criticism/communication analysis (Harris, 1987; Dean & Benoit, 1984). Carey and Rodier (1987) went the other direction and


DANIEL D. MILLS is a Doctoral Candidate in the Department of Speech Communication at the University of Nebraska, Lincoln, NE 68588. The author is indebted to Dr. Ann Burnett Pettus, Director of Forensics at the University of Nebraska-Lincoln, for her guidance and review of the manuscript.

Selected tables have been included in this study. A complete listing of all data is available by writing the author.
undertook a content analysis of oral interpretation ballots. They divided the content into five principal areas: quantity of comments, format of ballot, types of comments, valence, and advice. She used five pre-set categories: 1) selection; 2) presentational skills; 3) personal comments to contestants; 4) judge disclosure of personal preferences or judging style; and 5) comments written in the form of questions (p. 6). While this study has proven helpful as a starting point in determining what is written on oral interpretation ballots, a more thorough description of the comments is needed in order to understand what issues/areas are being addressed by judges.

Rather than use pre-set categories, this study allowed the categories to generate themselves from the comments. This process gives a clearer picture of what is being written on ballots. A comprehensive knowledge of ballot comments may prove beneficial in better preparing students for competition, and point out potential "problem areas" within the activity.

**Method**

Two hundred and fifty individual-event ballots were randomly selected for use in this study. Fifty ballots were randomly selected from each of the five areas of oral interpretation common on the intercollegiate forensic circuit in the 1989-1990 season: oral interpretation of poetry, oral interpretation of prose, dramatic interpretation, dramatic duo and program oral interpretation. Fifty ballots per event was predetermined in order that comparisons could be made across the five events. The ballots were written for students from a large midwestern university at invitational tournaments during the 1989-1990 season. While all the ballots are for students from the same university, this should not be a major limitation. It is assumed judges do not dramatically alter comments for students from any particular school. The predominant number of ballots are from the American Forensic Association district in which that university resides; however, a national representation is achieved with ballots written by judges at East and West coast invitational tournaments.

In order for comparisons to be made concerning the number of comments, the method used by Carey and Rodier (1987) was replicated. Each ballot was broken down into its "smallest unit possible . . . a generic comment like 'very enjoyable,' or 'good job,' was counted in order to identify the total number of comments being made," and "if a comment was a restatement, it was still counted twice" (Carey & Rodier, 1987, p. 5). This method is compatible with the syntactical method of unit analysis offered by Krippendorf (1980).
The development of the categories in this study followed Berelson's (1952) definition of "what is said," specifically, a subject-matter orientation. This study did not make use of any preset categorization system. Rather, a preliminary classification placed the comments into as many categories as necessary in accordance with the Berelson's (1952) perspective that categories are only limited by imagination. Fifty-eight categories emerged from this process. These categories were then reviewed with the purpose of discovering broader headings under which specific categories may be included (i.e., the major heading of "Variety" was developed to include general comments on variety along with more specific comments on vocal variety and nonverbal variety.) This process resulted in a final taxonomy of 19 categories.

The 19 categories listed in alphabetical order (and the specific categories from which they were formed) are:
1) **Blocking**—specific comments on physical movement (only relevant within dramatic duo);
2) **Character**—comments on acting, characterization, nonverbal characterization, vocal characterization, distinction between characters, nonverbal distinction between characters, vocal distinction between characters, interaction between characters, focal points of characters, miming action of characters, and thought process of characters;
3) **Decision**—specific justification by the judge for the rank and/or rate;
4) **Delivery**—vocal, conversational, pronunciation, enunciation, speed, tone, volume, vocal quality, nonverbs, movement, eye contact, facial expression, gestures, pacing, pause, and timing;
5) **Emotion**—emotional display and perceptions on feeling the emotion being conveyed;
6) **Familiarity with Material**—perceptions on how well the competitor knows the selection(s);
7) **Interpretation of the literature**—this was primarily suggestions on how specific line(s) should be interpreted;
8) **Introduction**—comments on content of introduction;
9) **Involvement**—student's engrossment with delivery or literature; the internalization of the material;
10) **Material**—author(s) name(s), author intent, cutting of material, how the story builds, offensive language in material, literary merit, material selected, teaser, theme, title(s);
11) **Pacifier**—comments intended to "pacify" a negative reaction by the student to either a comment or to the rank (i.e., "tough round," and "this was a really close round");
12) Pat on Back—comments which are "positive strokes" for the competitor (i.e., "You're a super competitor!");
13) Personal—comments of a personal nature to the competitor (i.e., "Good to see you out on the circuit again");
14) Presence—specific comments on the competitor's presence during the round;
15) Scriptbook—comments related to the speaker's use, misuse, or non-use of the script;
16) Technique—comments which specifically mentioned a interpreter's use of technique;
17) Time—time of the presentation and comments related to the time/length of the presentation;
18) Variety—comments specifically addressing nonverbal and vocal variety;
19) Visualization—comments on the competitor "seeing" what is described or happening in the story.

Each comment was also identified by its corresponding oral interpretation event, directionality of the comment, and the rank received on the ballot. Due to the standard practice of ranking "5" the last two speakers in a round of 6, all ballots with a rank of "6" were converted to a "5." The use of directionality is a common practice in content analysis and maybe used to determine criticism-praise (Emert & Barker, 1989). Directionality was employed to determine positive, constructive and neutral comments. Positive comments were deemed those remarks which supported the way a student handled any of the 19 categories (i.e., "I really like the way you present this character"); constructive comments offered advice for change or questioned a student on one of the 19 categories (i.e., "I think the man should be a little older" or "why is the kid so whiny?"); neutral comments carried no valence for reinforcement or change (these were commonly comments on time, title and theme).

Results

The 250 ballots yielded 2,596 comments, an average of 10.38 comments per ballot. The range of comments was from 0 to 28. The comments were distributed across the 50 ballots per event, with poetry interpretation having a mean of 11.38 comments per ballot, dramatic duo 10.72, prose interpretation 10.70, dramatic interpretation 9.62, and program oral interpretation with a mean of 8.90 comments per ballot (see Table 1).

In addressing all five oral interpretation events as a whole, Table 2 identifies what students and coaches can expect to see on ballots. The most commonly written comment was focused on material (25%). The
most frequent comment in "Material" dealt with how the material was cut (n = 214), and the material selected (n = 170) for competition. The second and third most frequently noted categories are close in total comments. "Character" was second with a total of 458 comments (17.64%) and "Delivery" was noted 426 times (16.41%).

The category receiving the least attention on all 250 ballots was "Decision." Only six comments were a specific reason for a rank and/or rate. Two examples of decision-based comments include, Top 3 showed more diversity and greater technical proficiency in performance" and "I went for a piece that had character interaction and variation in emotion."
Within each of the five oral interpretation events a rank order was also determined. Examination of the most frequently written comments for each event illustrates that all five events are highly similar. Comments on character, material, and delivery were the three most frequently mentioned comments in all five events. These three categories also consistently comprised the majority of comments for all the events. Character, material, and delivery comments made up 63.82 percent (n=307) of the average dramatic interpretation ballot, 61.68 percent (n=330) in prose, 60.26 percent (n=323) in dramatic duo, 59.78 percent in program oral interpretation, and 51.25 percent (n=307) in poetry.

While more variation was evident in the least commonly mentioned categories in the five events, certain categories were still dominant. Blocking was the lowest-ranked category in four of the five events. However, this low number of comments is easily accounted for—blocking was only evident within dramatic duo. The other category consistently on the low end of the spectrum is "Decision." Four of those decisions were written on poetry ballots, one in dramatic duo, and one in program oral interpretation.

A few of the categories in the mid-frequency range deserve attention. Students received a "pat on the back" from judges 159 times, but were offered "pacifiers" only 14 times. It is interesting to note that 45 comments were focused on the scriptbook and 18 comments specifically addressed technical performance. Aspects which may relate a link between the student and the literature (emotion, involvement, visualization) received only 6 percent (n = 166) comments.

The data collected also allowed for comparisons between the number of comments and the ranks received by the students. The results, as listed on Table 3, clearly reveal the higher a student ranks in a round, the more comments will be written on the ballot. A rank of "1" had a mean of 11.33 comments per ballot compared to a rank of "5" with a mean of 9.11, a difference of 2.22 comments per ballot. The biggest decrease in number of comments is seen between a rank of "3" and "4." While not substantiated, this difference may be linked to a "top 3, bottom 3" perception.

Table 3: Distribution of Comments According to Rank

<table>
<thead>
<tr>
<th>Rank</th>
<th>N Comment</th>
<th>N Ballots</th>
<th>Mean per Ballot</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>544</td>
<td>48</td>
<td>11.33</td>
</tr>
<tr>
<td>2</td>
<td>604</td>
<td>54</td>
<td>11.19</td>
</tr>
<tr>
<td>3</td>
<td>506</td>
<td>46</td>
<td>11.00</td>
</tr>
<tr>
<td>4</td>
<td>423</td>
<td>45</td>
<td>9.40</td>
</tr>
<tr>
<td>5</td>
<td>519</td>
<td>57</td>
<td>9.11</td>
</tr>
</tbody>
</table>
Examining the distribution of percentage of comments across ranks in terms of directionality reveals an interesting yet predictable outcome (see Table 4). The higher the rank, the more likely the student is to receive positive comments; the lower the rank, the more likely the student is to receive constructive comments. In fact, the correlation is extremely similar—a rank of "1" received an average of 60.29 percent of positive comments, while a rank of "5" received an average of 59.59 percent constructive comments. A rank of "1" received an average 27.21 percent of constructive comments, while a rank of "5" received 27.75 percent positive comments. In addition, a first place is the only rank where the positive comments dominated over the constructive comments—a domination of two-to-one. Second place rankings showed more of a balance, with constructive comments (47.85%) edging out positive comments (42.22%). A pattern emerged in neutral comments, with ranks of "1" and "5" being comparable with a high percentage, ranks of "2" and "4" comparable with the lower percentages, and a rank of "3" floating in between. However, no plausible reason could be determined for this pattern.

Table 4: All Five Interpretation Events: Distribution of Positive, Constructive and Neutral Comments According to Rank

<table>
<thead>
<tr>
<th>Rank</th>
<th>N Comments</th>
<th>N Pos</th>
<th>N Con</th>
<th>N Neu</th>
<th>% of Pos</th>
<th>% of Con</th>
<th>% of Neu</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>544</td>
<td>328</td>
<td>148</td>
<td>68</td>
<td>60.29%</td>
<td>27.21%</td>
<td>12.50%</td>
</tr>
<tr>
<td>2</td>
<td>604</td>
<td>255</td>
<td>289</td>
<td>60</td>
<td>42.22%</td>
<td>47.85%</td>
<td>9.93%</td>
</tr>
<tr>
<td>3</td>
<td>506</td>
<td>178</td>
<td>276</td>
<td>52</td>
<td>35.18%</td>
<td>54.55%</td>
<td>10.28%</td>
</tr>
<tr>
<td>4</td>
<td>423</td>
<td>131</td>
<td>250</td>
<td>42</td>
<td>30.97%</td>
<td>59.10%</td>
<td>9.93%</td>
</tr>
<tr>
<td>5</td>
<td>519</td>
<td>144</td>
<td>309</td>
<td>66</td>
<td>27.75%</td>
<td>59.54%</td>
<td>12.72%</td>
</tr>
</tbody>
</table>

Note: Pos = Positive Comments
Con = Constructive Comments
Neu = Neutral Comments

Discussion

This content analysis of oral interpretation events provides some interesting insight into the activity of competitive individual events. The 10.38 mean of comments written on ballots in this research is consistent with results from other studies both in oral interpretation and public address. Carey and Rodier's (1987) analysis of oral interpretation ballots averaged 10.72 comments per ballot, and Jensen's (1988) study of original public address events had a mean of 10.42.

In examining the distribution of comments per event, the high and low were poetry and program oral interpretation. The lower number of comments for program oral interpretation may be due to its experienc-
tial status as an individual event. Judges may not have written as many comments simply because they were not sure what to write. The high number of comments for poetry is more difficult to determine. Simple logic would assume that dramatic duo would have the highest number of comments, as a judge is writing for two speakers. Poetry may simply be a "favorite" event for judges on the intercollegiate circuit and thus receives greater attention through written comments.

What is more interesting than the mean number of comments per ballot is the content of those comments. The emphasis on material may seem to be a major concern on the intercollegiate forensic circuit, but a closer examination of the sub-categories from which "Material" was formed reveal that 150 of the comments were neutral in nature (primarily dealing with title, author's name, and theme) and 120 of 170 comments directed at the material selected were positive in directionality (i.e., "interesting story"). If these factors are taken into consideration, "Material" would fall to third under "Character" and "Delivery." The sub-categories of "Material" also reveal interesting facts on what are considered "touchy subjects" in competition. Author's intent and offensive language were an infrequent occurrence on ballots, only being mentioned once. The subject of literary merit was only noted four times.

The second and third most frequently mentioned categories, "Character" and "Delivery," raise other questions. The predominant dimensions composing these categories are technique-oriented. How the character acts, distinction between characters, focal points, enunciation, speed, volume, and movement ... are all primarily related to technical training. The one sub-category distinctly separated from this issue is the thought process of characters, but this was a rare comment, appearing only 11 times. The propensity toward technique is not a new problem in oral interpretation. Colley (1983) stated, "judging [is] reduced to a matter of technique, degree of slickness" (p. 44) and the Action Caucus in Oral Interpretation in Forensic Competition (1983) reported: "We see the same sort of undesirable reading or performance behaviors repeatedly in oral interpretation competition. We see slickness, showiness, and emphasis on technique" (p. 54). The argument could be made that the emphasis on technique is because students must first master these features before moving on toward an internalization of the Literature. I disagree with this view. Technique should be used as support for understanding and relating the material—not as the primary means of conveying a selection. I align myself with Colley (1983), who feels, "the content of the message is the important thing, not the techniques used to deliver the content. Technical display is not art" (p. 45).
On the other end of the spectrum are those comments which received little attention, the most problematic being the little attention given to the explication of a decision. The rank order of the 19 categories illustrates that one of the main concerns for a student, the reason they receive the rank/rate they did, is the least often written comment. Without a clear explanation for a rank/rate, a student must first interpret the literature in competition, and then "interpret" the judges' comments on the ballots. The question which comes to mind is whether judges have a clear and distinct reason for the ranks/rates they give, or if it is often just a "gut reaction." The missing judge's decision is mentioned by Carey and Rodier (1987), and Jensen (1988), both noting the need for this problematic area to be addressed. Jensen (1988) says, "it is also essential that a critic's comments clearly explain ratings and rankings given to a student in order to help that participant to grow, both educationally and competitively" (p. 8), and Carey and Rodier (1987) state, "there is often no clear logical or apparent reason for the rank or rate" (p. 16). Mills (1990) argues that one potential method for clearing up the lack of explicit judging decisions is to include it as a specific instruction with its own specific area on each ballot. Whether this would prove useful is waiting for a future research program to determine.

The distribution of comments according to rank reveals that the better a speaker is, the more comments s/he is likely to receive. Judges may wish to be more conscious of this proclivity to write more for the better competitor in the round. It seems logical that the ballot and its remarks would best serve the student receiving the "5" and "6" rather than the student receiving the first or second in the round. If the pedagogical purpose of forensics is to continue, further exploration in this area is warranted to determine if potentially negative consequences are resulting from this propensity. Additional research is possible by focusing on one specific event and breaking the categories down into their basic components. Ballot content analysis may also look for regional differences in type and quantity of comments, and differences between the coach/judge and the hired judge.

This study of oral interpretation individual event ballots has attempted to determine and to highlight some of the norms and their potential problems in judging comments. While judges are writing an adequate number of comments, they also tend to be emphasizing technique over understanding and internalization. Finally, there needs to be a concentrated effort to supply students with concrete reasons for the ranks and rates they receive on ballots. The pedagogical benefits of forensics are extensive, but they can only be maintained through introspective analysis of the event. This article is one step in that direction.
References


Mills, D. D. (1990). What should be done to improve oral interpretation events in individual events: Need to include a "reason for decision" on oral interpretation ballots. Presented at the Second Developmental Conference on Individual Events, Denver, CO.


