

“BEHIND THE NUMBERS”

How to decipher Judge’s Scores and Improve Your Ensemble

Phillip Haines, Presenter

“How will this clinic help me and my ensemble?”

- Provide some insight into how judges determine the scores given.
- Provide tools to interpret information provided on a recap.
- Knowledge of how groups are adjudicated can help in designing shows.
- Knowledge of the criterion reference system can better focus rehearsals and planning on “what’s tested”.
- Knowledge of information provided on a recap can help you to better evaluate your program’s relative strengths and weaknesses and aid in planning for future contests and seasons.
- Ultimately, this knowledge can help create a more successful program.
- “You have to know what’s on the test and how it’s scored in order to improve your grade.”

“How do judges determine the numbers they give?”

- Is it “Rank and Rate” or “Rate and Rank”? The process of determining the score:
 - “Impression” – first thoughts after completion of show (establishes the box)
 - “Analysis” – refines the number into the lower, middle or upper third of the box
 - “Comparison” – ranks the group in relation to competitors in the same “neighborhood”
- Some realities:
 - All tools are designed, first and foremost, to determine competitive placement with as much consistency and accountability as possible.
 - Judges tend to avoid the lowest number in a box or highest number in a box early in a contest.
 - Spreads may compress in larger contests.
 - Ultimately, all of the above is subject to the experience level of the adjudicator in the following areas:
 - Knowledge of the paradigms of each class and of the boxes within each class
 - Ability to numbers manage over the course of a contest
 - Ability to retain accurate memory of performances over time

“BEHIND THE NUMBERS”

- Numbers Management
 - The ability of a judge to accurately translate his or her impression of a group’s performance into a concrete number that is relative to the criterion reference material being used.
 - The ability to place competitors on a scale relative to each other that accurately reflects their relationship (this group is not only better but exactly this much better).
 - This requires that an adjudicator:
 - have a strong understanding of the paradigms of the activity.
 - use that understanding to translate a performance into a range of numbers based on what can often be considered somewhat nebulous criteria.
 - juggle numbers through the course of a contest to maintain appropriate relationships between competitors in such a way that the group’s individual score carries meaning, the group’s relative placement in the recap has meaning, and the overall score is indicative of where that group is compared to all other groups in the activity (whether that adjudicator or the group in question will ever see these groups).
 - THIS IS NOT EASY!!
- Achievement –The “What” and the “How”
 - “Achievement is the product of two elements that occur simultaneously: Content, which is what is being performed, and Performance, which is how well it is being performed (What + How = Achievement). Since neither content nor performance exists in isolation, the evaluation of achievement must contain a simultaneous examination of both elements.” Bands of America Official Procedures and Adjudication Handbook
 - What –the program (the “vocabulary” of music and visual components)
 - HOW - given the program, how well are the students performing?
 - THE CONTENT (what) is often the limiting factor in VIRTUALLY ALL captions and subcaptions!
 - While a judge can credit the “intent” without completely successful performance, credit cannot be given to what isn’t there in the first place.
 - Often, credit for the performance (how) is limited by the depth and breadth (or lack thereof) of the content (what).
 - This is not license to “overwrite”. The content must be suited to the training level of the ensemble.
 - **Be aware that ADLA does NOT credit vocabulary that is viewed as “beyond the expectations of the class”. Therefore, it is important to understand what is “appropriate” to the class in which you wish to compete. It is also important to remember that the ensemble’s classification is, ultimately, determined by the vocabulary presented.**

“BEHIND THE NUMBERS”

- The back of the score sheet provides tools for numbers management by:
 - Listing a set of standards on which to base comments AND scores.
 - Creating a “common language” for adjudicators, directors, staff and students.
 - Setting a range of numbers for continuity from show to show.
 - Establishing the basis for comparing two (or more) groups and determining the “winner”.
- If you are not familiar with the criterion reference on the back of the sheets, you don’t really know how you are being evaluated.
- “Contest Dynamics” – factors that may affect the scores given at a particular contest. Examples include:
 - Large contests with a wide variety of groups
 - Scores for lower achieving groups may be lower to allow room for the number of groups competing
 - Spreads between groups may be condensed due to the limitation of numbers available
 - Prelims/Semi Finals format (since the first priority for judges is to “get the right groups in finals”, non-finalists could receive lower numbers than they might normally receive and finalists could receive higher numbers)
 - Adjudicators with standards established in different areas or without clearly established paradigms (new to ADLA)
 - Judges who are used to higher achievement will score lower achieving groups lower and vice versa. In “judge speak”, this is called having different tolerances.
 - Judges without strongly established paradigms do not have a good reference for what scores mean and can start too low or too high (this does NOT necessarily mean that placement or spreads are affected – just score).

“BEHIND THE NUMBERS”

- Creating the wrong spread between competitors (usually too small a spread) compresses the scores of all other competitors in that range. Examples include:
 - Quality of first performing group (score not related closely enough to performance in either direction will have an effect on all other scores).
 - Spread to a significantly better performing group (if the spread is not large enough, scores will generally be lower and spreads will be compressed for the contest). This is one way to get “boxed in” (not enough number left to appropriately place competitors).
 - Because ADLA allows adjudicators to adjust scores throughout the contest, these are less likely to happen.
- Other examples of contest dynamics include:
 - Environmental factors
 - Misread by a judge
 - “Group dynamics” of adjudicators in a given contest

“What information is conveyed on a recap?”

- Understand Caption Weighting:
 - Music Analysis (40%)
 - Design (15%)
 - Performance (25%)
 - Visual Analysis (20%)
 - Design (10%)
 - Performance (10%)
 - Effect (40%)
 - Music (20%)
 - Overall (20%)

“BEHIND THE NUMBERS”

- Box Placement
 - This is the first decision an adjudicator makes!
 - Each box lists a specific level of achievement for the class based on descriptors
 - Middle of the box – meeting the criteria for the box
 - Lower third of the box – meeting some of the standards but not all (some standards may still be in the lower box)
 - Upper third of the box – meeting standards at a high level/approaching the standards of the next box (some criteria may be in the next box)
 - The difference between the top of one box and the bottom of the next box is NOT PRECISE. It is possible to fluctuate between two boxes for this reason.
- The ordinal and the spread
 - Ordinal = placement within class (or across class)
 - Spread = number of points separating ordinals and placements.
 - Spreads and their meaning (important note: The following is specific to ADLA. Other circuits/activities often use different ranges).
 - 1 – 2%: “Very comparable” - The units are essentially equal except that minor issues and “feelings” rate one over the other (*1 tenth or less is, for all intents and purposes a “broken tie”*).
 - 3 – 4%: “Minor differences” - There are subtle, but objective differences and nuances that make it clear that one unit is better than the other unit.
 - 5 – 7%: “Definitive differences” - There are at least 1 or 2 significant differences in the descriptives that are easily identifiable when comparing two units. *This spread begins to mark the point at which one group is quite clearly superior to another **for a particular performance.***
 - 8% or more: “Significant differences” - There are generally several significant differences in the descriptives when comparing two units. *One unit is approaching “another level” (or has already reached it).*

“BEHIND THE NUMBERS”

- THIS IS WHERE AN UNDERSTANDING OF THE SCORING PROCESS CAN PROVIDE THE MOST MEANING TO THE SCORES GIVEN!
 - Generally, look for design (content) scores to be higher than performance scores by a small percentage (2 to 3 points on a 100 point scale).
 - Deviations from this spread indicate areas for growth or concern
 - Larger number in performance than design (especially early season) – the performance level of the ensemble is exceeding the show’s content. This is definitely something you would want to discuss with the adjudicator in critique.
 - Tie or very small spread between design and spread early in the season – this indicates there may be limited room for growth over the season due to limitations in design.
 - Larger spread between design and performance – there may be flaws in performance or the performers may not be displaying the training required for the vocabulary they are asked to perform. This can be a good thing (meaning there is a lot of room for growth over the season) or a bad thing (if the show is written beyond the ability of the ensemble to achieve it).
 - If your ordinals in the design (“what”) captions are lower or you have a larger spread between you and your competitors, it’s possible that your ensemble will lose ground to your competitor(s) over time.
 - If your ordinals in the execution (“how”) captions are lower or you have a larger spread between you and your competitors, then your execution may be limiting your success and/or your show may not be written at the appropriate level for your group. It can also mean that, once your show is better performed, you will end up moving up in placement.
 - Ideally you are looking for commentary and numbers that indicate the design is appropriate and working. This produces the most potential for growth through the season.
- “Outliers” – scores and/or placements that vary significantly from the scores/placements of other adjudicators in a contest or between contests.
 - If the ordinals and/or scores are “all over the place”, this usually means the groups are not differentiated enough for the judges to make consistent decisions. Often it means there are flaws in achievement (design, execution or both) in most groups and the judge has to make decisions as to which group to “reward”. If this continues through the season, the groups are basically tied and it comes down the performances on a given night and the background and opinions of the adjudicators for a given contest.
 - Is the judge “boxed in” (scores between groups too close together preventing spreads because of a flaw in numbers management)? Look for consistency in ordinals.

“BEHIND THE NUMBERS”

- Is the judge more lenient in using the criteria or using scores unrelated to the criteria (“Happy judge”)? Again, look for consistency in ordinals.
- Are the score AND ordinal significantly lower than other judges and/or other contests? Look for commentary to back up the number. This sometimes happens when something sticks out in a negative way to a particular judge or it can be a “bad read”.
- Are the score AND ordinal significantly HIGHER than other judges and/or other contests? It’s human nature to believe this judge and assume all others are “wrong”. However, if this happens rarely, it’s likely a “bad read”.
- “Bad read” – an adjudicator who “misreads” elements in the show. SOMETIMES, THE “BAD READ” MEANS A JUDGE WHO PLACES THE GROUP HIGHER THAN NORMAL.
- In all cases, looking at the “big picture” (season long results) tends to put outliers in perspective.

“BEHIND THE NUMBERS”

- The final score
 - Represents the collective “opinion” of the three adjudicators
 - Determines your group’s “neighborhood” (the following is not intended to be a “precise” definition of spreads but more of a rough guide):
 - Within 2 points (up OR down) = your close competitors (could change order at any show)
 - Within 4 points (up OR down) = your competitors (the group that gets better quicker or achieves more of their show or has, ultimately, the better designed show, will “win” over time)
 - Between 4 and 8 points (up) = Your group may be able to compete with these groups over time but, all things being equal, they are likely to beat you each time you see them. You can learn a lot by studying what these groups do and how they are ahead. If you don’t know, make sure you find out (it’s always okay to ask!!).
 - Between 4 and 8 points (down) = at this point, they are not your competition. If they “catch” you, then your group has flaws (in design, training, and/or performance) or they had flaws early that they fixed.
 - 9 points or more = “out of your league” at this point. Trying to catch or pass these groups is not at all likely in the current season unless there is an extenuating reason for the difference (one group is incomplete but another is performing a completed product, for example). However, analyzing why they are where they are IS important. What are they doing that is allowing them to succeed and how can you and your ensemble learn from them?
- Comparing recaps over a season
 - Please note that, in comparing scores from week to week, you are, in essence, comparing “apples to oranges”.
 - Different judges for each contest will not always produce predictable results.
 - The criterion reference system contains a range of scores that apply to performances. What number is given, ultimately, depends as much on the dynamics of a particular contest as on the performance of a particular group. The box (and, ideally, the third of the box) shouldn’t change with similar performances but the actual number within that range may fluctuate significantly.

“BEHIND THE NUMBERS”

- Growth in score OVER THE SEASON.
 - Was it consistent with other groups in your class?
 - Did your group “fall behind” or exceed the growth in scores of others in your class or area?
 - Note that all comparisons are RELATIVE to other groups in your class or score area!
 - Did other groups “catch” you or did you move up”
 - If you stagnated and others moved up, look for reasons.
 - If you moved ahead, look for reasons.
 - In either case it’s often design that made the difference.
 - That being said, it is important to recognize the importance of training and designing a show that is achievable by the performers.
- Change in criterion reference box
 - If you started in box 3, did you move to box 4 over the season?
 - If you started in box 4, did you move to box 5 over the season? If not, what held you back?
 - How far into the box did you move? Were you near the top of the box by the end of the season? If so, what will it take to “break into” the next box in the following year?

“BEHIND THE NUMBERS”

“All this information is nice but my ensemble’s scores and ordinals are all over the place.

There is very little agreement between the judges in any given contest and little agreement in placement week to week. Why?

- First, remember that the three judges are judging very different criteria. Agreement should not be expected unless all areas of the design and performance are equal.
- Are the final scores relatively close (within four or five points among all groups)? This means the groups are very similar in achievement. There is not enough difference between the groups to clearly and consistently determine placement.
- This often happens in lower divisions and, to be honest, with weaker products/performances. Differentiating “shades of gray” is difficult and comparing flaws is difficult. One group may have a superior design but inferior performance and the next may have the opposite.
- This can also happen in a large contest when spreads get closer together to accommodate the number of competitors. With less room on the score sheet to differentiate groups, spreads narrow and placements vary more widely.

“Why did my group get a lower score than the previous week even though they performed MUCH better?”

- Did you place in the same box? In the same third of a box (or just below it)?
- Did you place in the same NEIGHBORHOOD?
- Did MOST groups IN YOUR NEIGHBORHOOD experience a similar result?
- Are there LIMITATIONS in your DESIGN or PERFORMANCE that are preventing growth?
- What captions helped/hurt you from one week to the next?
- Was your lower score the result of one judge?
- Was your higher score from the previous week the result of one judge?
- What other CONTEST DYNAMICS may have taken place?
 - Large versus smaller shows
 - Judging panel establishing a lower “baseline” score within a criterion reference box.
 - Different competitors and different levels of competitors can significantly alter scoring.

“BEHIND THE NUMBERS”

“Why is my ensemble placing lower at the end of the season than it did at the beginning of season”?

- Do the judges generally agree with each other through the season? If so, it's NOT the judges “fault” (don't “shoot the messenger”)
- Did other groups struggle early because their show wasn't complete or poorly performed and then “take off” when these areas were addressed?
- Are there LIMITATIONS in your DESIGN or PERFORMANCE that are preventing growth?
- Have you “maxed out” your show's potential? Look for scores in design versus performance.
- Have the judges indicated, in their commentary, that there are flaws in your show? Did you address them?
- Is your show too difficult to achieve at a consistently high level?
- Do the performers have sufficient training to achieve the vocabulary written for them?
- Are rehearsals structured to maximize performance at the end of the season or does the ensemble “peak” before then?