Throwing Away High and Low

Question of fairness

BY DOUG LOVELL
IAC Director-Northeast Region

Sometimes in aerobatics we hear suggested, as an alternative to straight averaging or "Fair Play," simply throwing away the high and low scoring judge for each pilot. The reasoning is:

1. This is done in other sports, including international Olympic sports. (see sidebar)

2. A judge who scores unfairly high or low will be caught by this procedure.

The method is intuitively appealing, especially given the spy vs. spy, east vs. west argument. In that argument, the west judge gives the west competitor maximum marks and the east competitor minimum marks. The east judge does vice-versa, giving maximum marks to the east competitor and minimum marks to the west competitor. Throw-away the high and the low scores for both competitors and, voila! All is right and fair.

That a method works in a single scenario does not prove the method is sound. Let's go beyond the east vs. west argument to consider what happens when you don't have an east vs. west scenario. When we think about the method more carefully, we find a number of problems. We'll start with the little things and get to the bigger, more complex issues.

First, if there is more than one unfairly high or low judge, the system falls flat. It fails.

Next, if there are only three judges, it leaves only one judge providing a score. If that judge is always the middle judge, then that judge decides the contest.

In a panel of fair judges, no judge wants to be the high or low judge. They want to be the middle judge. So they give middle scores. Now you've changed the scoring behavior of the judges in an undesirable way.

Finally, let's look at an example. This chart on the right lists the pilots in rank order according to the individual judge rankings. The chart on the adjacent page shows scores and results with the throw-away high and low scheme. The gray boxes are the discarded scores. The result rank values are on the right.

Looking at the high-low discard chart we see that pilot one, the winning pilot, got their score from judges one, four and five. Pilots two, six, eight, 10 and 12 got their scores from judges one, three and four. Pilots three and nine got their scores from judges three, four, and five.

In all, on a panel of five judges, there are 10 different sets of three judges from whom each pilot can get their score. On a panel of seven, there are 21 different sets of five judges. Every pilot could get their score from a dif-

Pilot	J1	J2	J3	J4	J5	rank
P1	1	1	1	1	3	1
P2	2	3	2	2	1	2
P3	3	5	3	3	7	3
P4	7	4	4	6	2	4
P5	12	2	5	5	11	5
P6	6	10	6	7	4	6
P7	4	6	10	10	8	7
P8	10	9	12	9	5	8
P9	5	11	7	10	9	9
P10	8	12	8	4	6	10
P11	9	7	9	12	12	11
P12	11	8	11	8	10	12
P13	13	13	13	13	13	13

Pilot	J1	J2	J3	J4	J5	score	rank
P1	2803.0	2701.5	2888.0	2864.5	2732.5	2800.0	1
P2	2744.0	2604.0	2781.0	2686.5	2808.5	2737.17	2
P3	2735.5	2578.0	2722.5	2634.5	2649.5	2668.83	3
P4	2612.5	2585.0	2676.0	2563.5	2745.5	2624.50	4
P5	2442.5	2632.5	2651.5	2601.5	2548.5	2594.17	7
P6	2657.0	2429.0	2636.0	2527.0	2710.5	2606.67	5
P7	2714.5	2517.5	2566.0	2488.0	2599.0	2560.83	9
P8	2554.5	2446.0	2506.5	2504.0	2706.0	2521.67	12
P9	2674.5	2405.5	2629.5	2488.0	2583.0	2566.83	8
P10	2595.5	2395.5	2605.0	2610.0	2673.5	2603.50	6
P11	2574.0	2499.5	2599.0	2344.5	2497.0	2523.50	11
P12	2539.5	2446.5	2553.0	2513.0	2562.0	2535.17	10
P13	1192.0	1137.0	911.5	954.0	1059.0	1050.70	13

ferent selection of judges. That makes the throwaway approach seem somewhat arbitrary.

Further, consider that we place three to five judges on the line. If each of them has a 75 percent likelihood of properly ranking two pilots, when all five of them rank the two pilots we increase the likelihood that they have it right to 90 percent. When you throw two away, the likelihood goes down to 84 percent. Throwing away judges is a hatchet approach to the problem of fairness.

The last problem with throwing away the high and the low is that the system breaks the Condorcet Criterion. A system that satisfies the Condorcet Criterion guarantees that the candidate (pilot) favored by the majority will always win.

Consider pilot five. The rankings chart shows that pilot five is clearly preferred by three judges (J2, J3, J4) over pilot six, and badly penalized by the remaining two (J1, J5). The high-low discard results show that throwing away the high and the low does not help pilot five at all. It hurts. Pilot five drops two places. Consider pilot 10. Pilot 10 is very happy. A different judge (but only one judge, J2), penalized pilot 10. One judge really liked pilot 10 (J4). The high and the low were thrown out, but pilot 10 still moves up to sixth place, ahead of pilot five.

Three judges, a majority, ranked pilot five at fifth place or above. Only two judges ranked pilot 10 at or above sixth place. Throwing away the high and low judges puts pilot ten at sixth place before pilot five at seventh place—not at all equitable. Now you might think that fifth, sixth, and seventh place aren't too important; but, remember that those could just as easily be first, second, and third. Throwing away the high and the low makes a complete mess of the results. IAC

Olympic Scoring

It isn't exactly true that Olympic sports throw away high and low scores. They do a number of things. In figure skating the ISU Judging System uses two separate scores—one from a technical specialist looking at slow motion video, and another from a panel of 12. For the 12 panel, they randomly select nine scores, then throw out the high and the low.

The United States Figure Skating Association does not do this. They use nine judges and use all nine of the scores.

Olympic diving uses a zero to 10 system, difficulty factor and straight averaging. Gymnastics uses a supervisor and an eight judge panel divided into D and E judges. The two judge D-panel decides the difficulty and overall content value for the performance, collaborating to resolve differences in their independent evaluations. The six judge E panel observes the performance and deducts for faults in execution and artistry. The scoring system throws-away the high and the low, using the middle scores. When the middle scores disagree by an amount greater than a given threshold, by table lookup, the supervisor adjudicates the scores in consultation with a jury chair and video replay. The supervisor and jury chair have a great deal of latitude to determine the score when there is lack of agreement between the middle judges. They are only looking at one set of zero to ten grades. They get to employ expert judgment and consultation with judges to decide what grade to give.