



## **Statistical Consequences of the ISU's New Proposed Cumulative Scoring System - Katherine Godfrey, PhD**

The ISU has proposed a new cumulative point-based system for scoring skating competitions, based largely on assigning point values to individual elements. The results will continue to be based on a random selection of judges from a panel. Although the final details are not available, the ISU published description of the system indicates that the new system will:

- place much of the control over what skaters will choose to do in the hands of the experts who decide how much each element is worth
- change the direction of the sport to favor attempts at hard elements not yet mastered over easier mastered elements, particularly if the hard elements are highly valued jumps
- produce unintended consequences as judges learn the new system, with the possibility that a judge may end up placing skaters not in the order they intended
- not provide any serious safeguards against biased judging from a single judge or from conspiracies to produce bloc judging, while making it possible that the majority opinion of the panel is not reflected in the results

The ISU announced the adoption of the new system on December 27, 2002 in [ISU Communication No. 1197](#). The new system is detailed in Rule 121 paragraph 3, extracted from the ISU Constitution and General Regulations 2002, which has been transcribed at <http://www.frogsonice.com/skateweb/articles/oc-project.shtml>.

The main features of the new system are:

1. Each element of a skater's program will be assigned a base value, similar in principle to the Degree of Difficulty for diving. The details of these values have not been made available. The elements will be entered into the computer system by a "group of experts"(not the judging panel).
2. As each element is performed, each judge will assign it a score on a 7-grade scale, based on how well the element is performed.
3. After the performance, each judge will also assign the performance 5 overall scores on a scale of 0 to 10 with increments of 0.5. These scores address some of the overall criteria

now specified for judging the current technical and especially the presentation marks, such as speed, sureness, choreography and interpretation.

4. A random subset of the judges (9 of 14 for major ISU competitions like Worlds, 7 of 9 or 10 for individual Grand Prix events like Skate America) will be drawn, and only their scores will be used to place the skaters.
5. The scores for each element and for each of the overall scores will be averaged across all the selected judges. For each individual element, the average will then be weighted according to the base values to get new weighted averages. Then all the averages will be added together to get the score for that skater for that phase of the competition. The final score for each skater will be the sum of the scores across all phases of the competition. The order of these totals will determine the order of finish.

Judges currently judge the overall performance by two sets of criteria, technical merit and presentation, and assign two overall marks to each performance to determine the skater's placement. Changing to a system where a major component of the scoring system is how a skater performs a single element in isolation is a major shift in the philosophy as well as mechanics of judging figure skating.

Those who assign the base values to the elements will have a tremendous impact on the direction of the sport. The relative values of various elements will influence skaters and their coaches to choreograph programs in light of this Code of Points: high-rated elements will be preferred over low-rated ones. If a high-rated element done badly will score higher than a lower-rated element done well because of the weighting, we have more splatfests in our future as skaters try for harder elements they have not truly mastered, hoping for at least partial credit.

The ISU has not published details of what constitutes an individual element. Based on elements detailed in current ISU scoring sheets, a typical free skate for a top men's singles competitor will have 5 to 7 solo jumps (quads or triples), 2 jump combinations, 4 spins of various types, and 2 footwork sequences, for a total of about 15 elements. Note that at least half the elements will be jumps or jump combinations. Even without assuming that the harder jump elements will likely get base values greater than any spin or footwork sequences, it's clear that jumping success will be an even larger determinant of the outcome than it is now.

Because the judges will be filling out running scorecards, grading each element as it comes, a corrupt or biased judge could push a skater up or down substantially simply by being lenient or strict on each element. We can't calculate the potential for this without knowing the details of the base values, but a judge who gives one skater 1 point less (after weighting) on each element and another skater 1 point more on each element than the rest of the panel will make the gap between them 3 to 5 points greater than if he'd agreed with the rest of the panel. For a panel of 9, 15 elements with an extra point each would be an extra 1.67 points averaged over the panel. Doing the same to lower the other skater's score would create a 3.3 point change in the difference in the skaters' scores. Similar half-point manipulation of the overall score components could add another 0.5 to the difference. In a close contest, this might be enough to change the final placement.

Even assuming that all the judges are judging honestly and diligently, constructing a score from adding up a lot of subscores will create more variation in the scores than the assignment of one

or two numeric scores would. Unless individual judges have the memory of Las Vegas card counters, they may even find that their final scores for each skater have inadvertently placed the skaters in the wrong order (e.g., they gave A a higher score than B even though they truly believed B's was the better all-round performance, because of the weighting the base-value assigned to some elements). Also, scores across the judging panel will tend to be more variable than ordinals are now, at least until judges become proficient in the system and adjust to the effect of the base-value weighting for elements. Note that this latter variability also makes it harder to detect actual cheating, no matter what statistical analyses are applied after the fact to the results.

The ISU has claimed that random selection of judges will make it impossible to create blocs, because the deal-makers would not know whom to approach. However, picking a random number of judges from a panel won't eliminate the ability to set up blocs. They may have to be bigger than before to ensure results, but they'll have the advantage of being undetectable once they're formed.

First of all, potential conspirators will probably have a good idea of who will vote for their favored competitor in any event, so they need only approach "swing" judges. Secondly, they won't want to call attention to themselves by trying to move a skater up or down many placements; instead, they'll want to concentrate on making sure a likely contender wins. Imagine that a panel of 14 is deemed by the conspirators to have 6 sure votes for their skater, and that 9 judges will be selected from the panel. If the conspirators can get 2 other judges to vote for their candidate, they have a 76% chance of getting at least 5 of their 8 judges on the final panel. So by getting a bloc only as large as they would need if all judges' scores counted, the conspirators have a 3 out of 4 chance of success. If they want to be even safer, a third cajoled judge will give them a 94% chance of success.

Selecting random judges from a completely honest panel creates a fairness issue that does not exist when you use all the judges' marks. It's possible to have the random panel come up with a different result than the entire panel would have. If the 14 judges are split 8/6 in favor of skater A over skater B, there's a 24% chance that B will end up with a majority of the 9 selected judges. This raises the issue of who "really" won the competition, and it's an unavoidable consequence of using a subset of the judges. If a skater is consistently a little better than the next competitor over the course of a season, that 24% would translate into losing about once a season simply by bad luck of the judging panel draw.

Here's a recent example of a case where one skater almost certainly was placed below another due to such bad luck. These are the marks for two skaters from the 2003 Jr. Worlds:

Skater A

4.6 4.7 4.8 4.8 4.8 4.9 4.9 5.0 5.1 5.1 5.1 5.2 5.2 5.2  
4.9 5.0 5.1 5.1 5.1 5.1 5.1 5.1 5.2 5.3 5.3 5.3 5.3 5.4

finished ahead of

Skater B

4.6 4.6 4.9 4.9 5.0 5.0 5.1 5.1 5.3 5.3 5.3 5.3 5.4 5.5  
5.2 5.2 5.2 5.3 5.4 5.4 5.4 5.5 5.5 5.5 5.5 5.5 5.7 5.7

Although we can't know which four marks were assigned by each judge, we can see that Skater B's technical merit and presentation marks were both substantially higher on average. A simulation study randomly pairing 9 sets of technical merit and presentation marks for each skater produces the better placement for Skater B 99% of the time. A simulation study randomly comparing sets of 9 technical merit marks only produces a better placement for Skater B 90% of the time. While we can't say for sure what happened, the evidence suggests that the random set of judges selected was not at all representative of the panel as a whole as far as these two skaters' placements were concerned. If this particular result had been between the top two finishers at Worlds, the public reaction would not have been pleasant, and it would be due entirely to using a subset of the panel.

If the ISU is going to use 14-judge panels, they should use the entire panel, even if they continue to hide which judge gave which scores. This would eliminate the losing by random selection problem while leaving the threat of corruption essentially unchanged.

Several mathematically proficient skating fans have studied the potential problems of the proposed new scoring system in depth. One place to start reading more about them is George Rossano's excellent material at <http://www.iceskatingintl.com/>

The probabilities above were computed from the appropriate hypergeometric distributions. If you took probability courses in high school or college, these are used for the problems of drawing marbles out of an urn, equivalent to randomly selecting judges off a panel. You can read more about computing these probabilities in most introductory statistics or probability texts or searching for "hypergeometric" on the Web. One on-line discussion is at

<http://www.pitt.edu/~jrclass/e20/notes/OH10.html>