JUDGE BIAS IN AESTHETIC GROUP GYMNASTICS

Seppo Suominen

Haaga-Helia University of Applied Sciences, Finland

Original article *DOI: 10.52165/sgj.15.3.441-452*

Abstract

There are several competitions where the objectivity of judges has raised some questions, in particular the style point evaluation. If part of the actual performance can be objectively measured with an instrument and part of the performance is based on subjective evaluation of judges, an error in ranking is possible. There are some sport disciplines, like ski jumping, where both metrics are used (Krumer et al., 2020); however, in gymnastics, the only criteria are judges' subjective evaluation (Bučar et al., 2012; Leskošek et al., 2012; Rotthoff, 2014).

Our analysis reveals that biased judging in aesthetic group gymnastics is more than probable in domestic competitions in Finland. The local judge that evaluates their own team is not overestimating the performance in any of its three parts: technical value, artistic value, or execution value. However, it seems that judges strategically underestimate the performance of the most important rival. This underscoring is truncated since the highest and lowest scores are truncated in the case of four judges. The local judge's scoring is usually within the two middle scores, which is taken into account in the final score of a performance.

Our analysis used evaluations of 66 different competitions including 585 performances in a period of 22 months. All competitions were domestic, with domestic teams and domestic judges only. Many competitions had 12 judges: 4 evaluating technical value, 4 artistic value, and 4 execution value. All judges were nominated before the actual competition.

Keywords: *Aesthetic group gymnastics, Finland, Judge bias.*

INTRODUCTION

There have been several competitions in which the objectivity of judges has raised some questions. In particular, the style point evaluation – namely, the objectivity of judges – has been challenged. If part of the actual performance can be objectively measured with an instrument and part of the performance is based on the subjective evaluation of judges, an error in ranking is possible. There are some sport specialties, like ski jumping, where both metrics are used (Krumer et al., 2020); however, in most gymnastics the only criteria are the judges' subjective evaluations (Bučar et al., 2012; Leskošek et al., 2012; Rotthoff, 2014).

Reliability in measurement is consistent when, under identical conditions, the same results are achieved. Laboratory testing falls into this category. However, achieving identical performances in sports can be challenging as athletes may not perform identically in repeated attempts, often due to factors such as muscle fatigue.

Objectivity is defined as obtaining the same result from different authorities evaluating the same performance. There are several reasons why judges may fail to be objective. For instance, judges may display favoritism towards their own ethnic group, resulting in fewer fouls being called when their race matches that of the refereeing officials in the NBA (Price & Wolfers, 2010). Similarily, they might call more penalties on soccer players with a different mother tongue (Faltings et al., 2019). Judges may award more style points to their compatriots in ski jumping (Balmer et al., 2010; Krumer et al., 2020; Zitzewitz, 2006) , boxing (Balmer et al., 2007), gymnastics (Balmer et al., 2011) or figure skating (Findlay & Ste-Marie, 2004; Zitzewitz, 2014). Corruption is also a potential issue (Moriconi & de Cima, 2021; Zitzewitz, 2014). Furthermore, experience plays a role in judging. More experienced judges tend to be better than novices at perceptually anticipating upcoming gymnastic elements based on advance information (Ste-Marie, 1999). Soccer referees can also be influenced by home team spectators' noise, leading to fewer penalties being awared to the home team (Boyko et al., 2007).

Ranking in group gymnastics is based on the technical value of both obligatory and non-obligatory movements, the artistic value of the performance, and the execution value of the performance. Consequently, there are three judge groups involved in each competition. The judging process typically divides scores into components: a difficulty score which evaluates the complexity of each movement executed by a gymnast, and an execution score, which assesses the performance of each movement performed by the group during their routine in the competition.

The ranking in aesthetic group gymnastics (AGG) is based solely on the aesthetic criteria of the participating teams. AGG evolved from traditional Finnish women's gymnastics and shares some similarities with rhythmic gymnastics. An AGG competition program is a fusion of artistry and athleticism, set to music, and it encompasses various elements such as body movements, balances, jumps, and combinations. Beyond the obligatory difficulty elements, the entire performance is unified by a continuous, flowing movement and the program's overall atmosphere or theme, which is conveyed through the language of movement, the music selection, and the artistic elements woven into the routine. The entire performance is elevated by the performers' capacity for empathy and expression.

The jury panel is typically selected approximately one to two hours prior to the competition. In most cases, each participating team is obligated to provide one judge for the competition. However, the availability of authorized judges is often limited, and these judges are typically coaches affiliated with one of the competing teams. Consequently, there exists the potential for judges to render biased rulings.

The objective of this study is to evaluate potential bias within judge panels based on a sample of various group gymnastics competitions held in Finland. The range of competition varies from local and rather small competitions to Finnish championship competitions where almost all existing groups are represented. Additionally, there are a few important competitions that sigbnificantly impact the selection of the Finnish national team. The bias exhibited by judges is quantified by either awarding higher scores to the home team compared to other judges' assessments (referred to as over-grading) or assigning lower scores to other teams, typically the primary rivals, in contrast to other judges' evaluations (known as under-grading). The study aims to determine whether over-grading or under-grading bias exists in Finnish group gymnastics competitions. Furthermore, it seeks to identify potential strategies to mitigate or reduce such bias in future competitions.

Our analysis looks at scores assigned to teams by various judges, including the possibility of a coach judging their own team. Specifically, we examine whether the coach serving as a judge tends to award higher scores to their own team compared to the scores given by other judges, who are likely associated with different teams. To investigate this potential bias, we conduct paired-sample t-tests to assess the disparity between the scores provided by the team's own judge and those assigned by the other judges (referred to as over-grading). Additionally, we perform similar paired-sample t-tests to identify any instances of under-grading by judges associated with rival teams.

METHODS

Depending on the competition series, an AGG competition team typically consists of 6 to 12 gymnasts who perform a routine lasting approximately 2 to 3 minutes. The specific requirements for the routine, such as the number of body movements, balances, jumps, and combinations, can vary depending on the series.

The competition takes place in a designated area measuring 13 x 13 meters, and teams are expected to utilize the space creatively throughout their program. AGG competitions are organized into different age and level categories:

Children's series are available for 8 to 10-year-olds, 10 to 12-year-olds, and 12 to 14-year-olds. Gymnasts under the age of 12 compete within their respective age groups, either in the 8- to 10-year-olds or 10- to 12-year-olds series.

AGG gymnasts aged 12 and above have the option to compete at three different series levels: the Finnish Championships, the Racing Series, and the Hobby Series. Teams can choose the level that suits them best.

The national competitions within the Finnish Championship series hold significant prestige. All series feature age groups for children (12- to 14-year-olds), juniors (14- to 16-year-olds), and women (over 16 years old). Additionally, there are competition series for gymnasts in the 12–14, 14–16, 16–20, and over-18 (women) age categories.

The Hobby Series is open to participants in age groups 12–14, 14–16, and over 16.

In all series except the 8–10 series, the jury will be composed of three different jury panels. One panel of judges assesses the technical value of the program (TV = technical value). For example, in the 10- to 12-year-old series, this jury observes whether all required skill parts are included in the competition program (body movements, balances, jumps, hand movement sets, step sets and jumps, acrobatic movements, mobility movements). In all age groups (10–12, 12–14, 14–16, and over 16 years old) of 10–12, the maximum TV score is 6.0 with 0.1 increments. The task of the second jury is to assess the artistic value of the program (AV

= artistic value). AV points are affected, for example, by team gymnastics technique, the synchronicity of gymnasts, the composition and originality of the program, and the unity of the group. The maximum AV score is 4.0 with 0.1 increments. The third panel of judges evaluates the execution (EXE = execution). This jury makes deductions for observed errors, such as incomplete stretches, poor posture, falls, and loss of balance. The maximum performance score is 10.0 with 0.1 increments. In addition, in each criterion (technical, artistic and execution), the judges can award a bonus if the performance is extremely good. A deduction is also possible if any gymnast is outside the designated area, or the costume is in breach with the rules. A typical reduction might take place if the gymnastic slipper falls off the leg. The slipper does not cover the heel. The maximum number of points in all series is 20. In the Finnish Championship series (highest level), the maximum of technical points is 6, of artistic points 4, and of performance 10. In the competition (medium level) and hobby (lowest level) series, the maximum technical and artistic points are 5. The 8-to 10-year-old series has two different judging panels: (1) composition; and (2) performance and expressiveness.

If there are four judges evaluating, for example, technical value, the extreme points, lowest and highest, are truncated, and the average of the two remaining is the score given to a team. In the case of three judges, no truncation is made and the final score is the average of three evaluations. It is important to note that all judges are making their evaluations simultaneously. Below are two examples of evaluations from four judges when a team performed really well. It is assumed that one judge represents their own team (judge #1) and one represents the most important rival team (judge #4). The two remaining judges (#2 and #3) are independent. Suppose that the team's own judge gives the highest possible points, and in scenario A, the rival judge does the same. In this case the final score is 5.95.

Table 1:
Technical value points given by four judges when the rival judge values the performance as extremely good.

Scenario A	Judge #1 (own)	Judge #2 (indep.)	Judge #3 (indep.)	Judge #4 (rival)
	6	5.9	5.9	6
Highest and lowest truncated	truncated	truncated	5.9	6
Final			5.95	

Scenario B shows a case where the team's own judge gives the maximum points, and the rival judge gives a slightly lower point value. Both of these are truncated, and the final score in Scenario B is 5.9, a lower final score than in scenario A.

Table 2:
Technical value points given by four judges when the rival judge values the performance as rather good.

Scenario B	Judge #1 (own)	Judge #2 (indep.)	Judge #3 (indep.)	Judge #4 (rival)
	6	5.9	5.9	5.8
Highest and lowest truncated	truncated	5.9	5.9	truncated
Final			5.9	

In Scenario B, the rival judge uses strategic voting in order to lower the final score of the team (the team with its own judge in the panel). Scenario B in comparison with Scenario A shows that the strategy of Judge #4 is to give the most important rival fewer points (lower enough) since this results in a lower final score for the rival. Since the composition of the judge panels is randomised for each competition, the game is played only once; however, it is possible that the teams will remember the strategic voting bias, and during the next competition the judges may use a tit-for-tat tactics. This strategy is not studied in this paper.

RESULTS

The data used cover all competitions in the Finnish Championship series (highest level) between 2 November 2019 and 25 September 2011, including 66 competitions and 585 performances of 31 different gymnastic associations and 152 judges. Most gymnastic associations have teams in all three age series: 12- to 14-year-old, 14- to 16-year-old and over 16-year-old categories. The data source is public and available from www.kisanet.fi. Some descriptive statistics of the technical values awarded are shown below in Table 3. All judge points have been collected from these public websites. In addition, since the judges names are also public, the affiliations or home teams of the judges have been collected. If the coach of team has been also one of the judges, there is a possibility to over-grade one's own team and under-grade a rival team.

Based on Table 3, the average technical value is higher in the senior (over 16) category, and it seems that only the best teams have continued to the senior level. The number of performances (110) and evaluations (431) is only half the number of performances and evaluations in the juniors' (14- to 16-year-old) and children's (12- to 14-year-old) categories.

Judges gave their own teams average points. There is no significant difference between their own assessment and the middle assessment by the other three judges. The highest points awarded to other teams have been significantly higher than those given to their own team in the children's (t-test -3.582) and juniors' (t-test -3.914) categories. The lowest points given have been significantly lower than those awarded by the team's own judge. In the case of four judges, the highest and lowest points given are truncated in the final evaluation, which is the average of the two middle point values. Thus, the technical value evaluation of the team's own judge is usually included in the final score.

Table 3:
Technical value, descriptive statistics, all competitions at the highest level (Finnish Championship rules) between 2.11.2019 and 25.9.2021

Technical value max 6.0	All age series	12- to 14-year-olds #perf. 244	14- to 16-year-olds #perf. 231	16-year-olds+ #perf. 110
Average (std)	5.19 (0.65) n = 2252	5.18 (0.55) n = 920	4.98 (0.72) n = 901	5.67 (0.37) n = 431
Final score with truncation (std)	5.19 (0.63) n = 585	5.20 (0.50) n = 244	4.95 (0.71) n = 231	5.69 (0.34) n = 110

Table 4:
Technical value awarded by the team's own judge and other judges at the highest level between 2.11.2019 and 25.9.2021

Technical value	All age series	12	14	16
Team's own judge	5.265 (0.683) n = 95	5.165 (0.543) n = 31	5.138 (0.746) n = 45	5.732 (0.541) n = 19
The highest of the other three judges	5.448 (0.561)	5.445 (0.379)	5.298 (0.651)	5.811 (0.416)
Paired sample t-test (own vs. highest)	-5.39***	-3.582***	-3.914***	-1.662
Middle of the other three judges	5.160 (0.994)	5.194 (0.462)	4.916 (1.277)	5.684 (0.614)
Paired sample t-test (own vs. middle)	1.338	-0.387	1.420	1.634
The lowest of the other three judges	4.734 (1.498)	4.439 (1.590)	4.571 (1.578)	5.600 (0.642)
Paired sample t-test (own vs. lowest)	4.277***	2.715*	3.088**	3.371**

Table 5:
Artistic value, descriptive statistics

Artistic value max 4.0	All age series	12- to 14-year-olds #perf. 244	14- to 16-year-olds #perf. 231	16-year-olds+ #perf. 110
Average (std)	3.10 (0.48) n = 2243	2.99 (0.42) n = 920	3.03 (0.49) n = 901	3.52 (0.37) n = 431
Final score with truncation (std)	3.11 (0.45) n = 585	3.11 (0.37) n = 244	3.03 (0.47) n = 231	3.50 (0.36) n = 110

Artistic value is highest in the senior category, which is reasonable considering that seniors have more experience than younger gymnasts. Furthermore, there are fewer teams and performances than in juniors' or children's series. The final score, excluding the highest and lowest points that are truncated, is higher in the children's series.

Table 4:
Artistic value awarded by the team's own judge and other judges at the highest level between 2.11.2019 and 25.9.2021

Artistic value	All age series	12	14	16
Team's own judge	3.258 (0.451) n = 97	3.184 (0.403) n = 45	3.190 (0.511)) n = 29	3.487 (0.396) n = 23
The highest of the other three judges	3.325 (0.429)	3.231 (0.401)	3.283 (0.449)	3.561 (0.382)
Paired sample t-test (own vs. highest)	-3.328***	-1.773($^{*)}$	-1.823$^{(*)}$	-3.364**
Middle of the other three judges	3.139 (0.475)	3.022 (0.417)	3.110 (0.498)	3.404 (0.469)
Paired sample t-test (own vs. middle)	5.450***	4.951***	1.771$^{(*)}$	2.646**
The lowest of the other three judges	2.580 (1.222)	2.344 (1.224)	2.459 (1.341)	3.196 (0.836)
Paired sample t-test (own vs. lowest)	6.478***	5.281***	3.446**	2.034$^{(*)}$

The artistic value awarded by tean's own judges is the second highest evaluation in most cases, and is therefore included in the final score. The highest value given in all age categories is higher than that of the team's own judge. The difference is significant in the senior category. Moreover, there is evidence that the lowest point value awarded is substantially lower than the points given by the team's own judge. The result indicates the possibility of strategic voting by judges from rival associations.

The execution value counts for half of the points given to a performance and, therefore, is the most important. The final scores for each part [t(echnical), a(rtistic) and e(xecution)] are positively correlated: $\rho_{ta} = 0.828$, $\rho_{te}= 0.820$ and $\rho_{ae} = 0.869$. Table 7 presents some descriptive statistics of the execution value given. The points increase with age: seniors have a substantially higher average (8.9) score than juniors (8.0) or children (7.8). The evaluation (Table 8) lies between the highest and the lowest.

The team's own judge evaluated technical value in 95 out of 585 competitions., whereas the number for artistic value is 97 and execution value 107. The share of competitions with a team's own judge is slightly higher than 50%. The highest share is in the senior category, i.e., 58%.

There is some evidence of bias, however. Since the lowest points awarded in each section seem to be significantly lower than those awarded by the team's own judge, the strategic underscoring is plausible.

Table 5:
Execution value, descriptive statistics

Execution value max 10.0	All age series	12- to 14-year-olds #perf. 244	14- to 16-year-olds #perf. 231	16-year-olds+ #perf. 110
Average (std)	8.098 (0.798) n = 2243	7.846 (0.728) n = 920	8.025 (0.776) n = 901	8.876 (0.451) n = 431
Final score with truncation (std)	8.094 (0.759) n = 585	7.850 (0.660) n = 244	8.019 (0.744) n = 231	8.891 (0.428) n = 110

Table 6:
Execution value awarded by teams' own judgse and other judges at the highest level between 2.11.2019 and 25.9.2021

Execution value	All age series	12	14	16
Team's own judge	8.373 (0.708) n = 107	8.020 (0.593) n = 41	8.486 (0.670) n = 44	8.805 (0.691) n = 22
The highest of the other three judges	8.546 (0.652)	8.283 (0.563)	8.609 (0.648)	8.909 (0.633)
Paired sample t-test (own vs. highest)	-5.768***	-4.416***	-3.351**	-1.994$^{(*)}$
Middle of the other three judges	8.328 (0.702)	8.029 (0.571)	8.407 (0.731)	8.727 (0.648)
Paired sample t-test (own vs. middle)	1.723$^{(*)}$	0.089	2.093*	1.882$^{(*)}$
The lowest of the other three judges	7.885 (1.562)	7.549 (1.358)	7.818(1.917)	8.645 (0.663)
Paired sample t-test (own vs. lowest)	3.612***	2.190*	2.587*	3.332**

Recently, in Finnish aesthetic group gymnastics, two teams have consistently outperformed all others. One is based in the Helsinki region (Espoo), and the other hails from Tampere. Both have won several World Cup and World Championships competitions. In international competitions, there are typically two combined events, and the winner is the team with the highest cumulative score from the preliminary competition (usually held on Saturday) and the final competition (held on Sunday). International regulations stipulate that only the top two teams from each country can advance to the finals. Teams from Espoo (E) and Tampere (T) have most often secured their places in the final competition. The following analysis utilizes the points awarded in domestic competitions in comparison to the other leading team in Finland.

The results in Table 9 indicate that judges from the other top team seem to undervalue the rival's performance. This is especially notable in the execution points.

Table 7:
Pairwise comparison of rival judge's evaluation for the top two teams

	TV n = 13	AV n = 24	EXE n = 44		TV n = 18	AV n = 3	EXE n = 4
Points to Espoo awarded by Tampere judge	5.43 (0.65)	3.46 (0.44)	8.55 (0.65)	Points to Tampere awarded by Espoo judge	5.86 (0.17)	3.70 (0.10)	9.12 (0.09)
The highest of the other three judges	5.61 (0.48)	3.53 (0.37)	8.79 (0.58)	The highest of the other three judges	5.93 (0.13)	3.83 (0.05)	9.27 (0.28)
Paired sample t-test (rival vs. highest)	-2.71^{*}	$-2.00^{(*)}$	-5.87^{***}	Paired sample t-test (rival vs. highest)	-2.24^{*}	-2.00	-1.13
Middle of the other three judges	5.49 (0.63)	3.31 (0.49)	8.62 (0.56)	Middle of the other three judges	5.88 (0.16)	3.73 (0.05)	9.17 (0.22)
Paired sample t-test (rival vs. middle)	-0.97	3.71^{***}	-2.13^{*}	Paired sample t-test (rival vs. middle)	-0.83	-0.50	-0.480
The lowest of the other three judges	5.41 (0.70)	3.20 (0.51)	8.50 (0.60)	The lowest of the other three judges	5.81 (0.23)	3.66 (0.05)	9.05 (0.26)
Paired sample t-test (rival vs. lowest)	0.35	5.92^{***}	1.36	Paired sample t-test (rival vs. lowest)	1.37	1.00	0.54

DISCUSSION

The analysis reveals that biased judging in aesthetic group gymnastics is more than probable in domestic competitions in Finland. The team's own judge that evaluates their own team is not overestimating the performance in any of the three parts: technical value, artistic value, or execution value. However, it seems that the judges of the top teams strategically underestimate the performance of the most important rival. This underscoring is truncated from the final score, since the highest and lowest scores are truncated in the case of four judges. The team's own judge scoring usually is within the two most middle scores, which is taken into account in the final score given to a performance. The national gymnastic association should monitor bias in judging and if necessary, impose a fine to the home team of the particular judge and a temporary moratorium.

The analysis used evaluations of 66 different competitions with 585 performances in a period of 22 months. All competitions were domestic, with only domestic teams and domestic judges. Many of the competitions had 12 judges: 4 evaluating technical value, 4 artistic value, and 4 execution value. All judges were drawn prior to the actual competition. Since there is a shortage of judges, all teams must register one judge for each competition in which their team is performing. If the team cannot register any judge, the team must pay a penalty payment to the organizer of the competition.

CONCLUSIONS

Regrettably, biased judging appears to be a prevalent issue in entirely domestic aesthetic group gymnastics competitions. These competitions mandate the presence of over ten judges, with each participating team financially penalized for failing to provide one judge, often selected from the coaching staff. Remarkably, judges do not exhibit a tendency to overestimate their own team's performance. The scores provided by the team's own judge are neither the highest nor the lowest which would be omitted from the final score calculation.

However, there is evidence suggesting the practice of strategic underscoring when evaluating the most prominent rival team. This strategic underscoring potentially allows the judge to lower the final score of the rival team. Addressing this issue warrants immediate attention and action by the judge committee of the gymnastic federation.

REFERENCES

Balmer, N. J., Nevill, A. M., & Lane, A. M. (2007). Do judges enhance home advantage in European championship boxing? *Https://Doi.Org/10.1080/02640410400021583*, *23*(4), 409–416. https://doi.org/10.1080/02640410400021583

Balmer, N. J., Nevill, A. M., & Williams, A. M. (2010). Home advantage in the Winter Olympics (1908-1998). *Https://Doi.Org/10.1080/026404101300036334*, *19*(2), 129–139. https://doi.org/10.1080/026404101300036334

Balmer, N. J., Nevill, A. M., & Williams, A. M. (2011). Modelling home advantage in the Summer Olympic Games. *Https://Doi.Org/10.1080/0264041031000101890*, *21*(6), 469–478. https://doi.org/10.1080/0264041031000101890

Boyko, R. H., Boyko, A. R., & Boyko, M. G. (2007). Referee bias contributes to home advantage in English Premiership football. *Https://Doi.Org/10.1080/02640410601038576*, *25*(11), 1185–1194. https://doi.org/10.1080/02640410601038576

Bučar, M., Čuk, I., Pajek, J., Karacsony, I., & Leskošek, B. (2012). Reliability and validity of judging in women's artistic gymnastics at University Games 2009. *European Journal of Sport Science*, *12*(3), 207–215. https://doi.org/10.1080/17461391.2010.551416

Faltings, R., Krumer, A., & Lechner, M. (2019). *Rot-Jaune-Verde. Language and Favoritism: Evidence from Swiss Soccer*. http://www.seps.unisg.ch

Findlay, L. C., & Ste-Marie, D. M. (2004). A Reputation Bias in Figure Skating Judging. *Journal of Sport and Exercise Psychology*, *26*(1), 154–166. https://doi.org/10.1123/JSEP.26.1.154

Krumer, A., Otto, F., & Pawlowski, T. (2020). *Nationalistic bias among international experts: Evidence from professional ski jumping Sports economics View project Soccer Analytics View project*. https://www.researchgate.net/publication/337192531

Leskošek, B., Cuk, I., Pajek, J., Forbes, W., & Bucar-Pajek, M. (2012). Bias of judging in men's artistic gymnastics at the european championship 2011. *Biology of Sport*, *29*(2), 107–113. https://doi.org/10.5604/20831862.988884

Moriconi, M., & de Cima, C. (2021). Why some football referees engage in match-fixing? A sociological explanation of the influence of social structures [Article]. *International Journal of Sport Policy and Politics*, *13*(4), 545–563. https://doi.org/10.1080/19406940.2021.1928731

Price, J., & Wolfers, J. (2010). Racial Discrimination Among NBA Referees [Article]. *The Quarterly Journal of Economics*, *125*(4), 1859–1887. https://doi.org/10.1162/qjec.2010.125.4.1859

Rotthoff, K. W. (2014). (Not Finding a) Sequential Order Bias in Elite Level Gymnastics. *Southern Economic Journal*, 140827142349001.

https://doi.org/10.4284/0038-4038-2013.052

Ste-Marie, D. M. (1999). *Expert–novice differences in gymnastic judging: an information-processing perspective*. Applied Cognitive Psychology. https://onlinelibrary.wiley.com/doi/abs/10.1002/(SICI)1099-0720(199906)13:3%3C269::AID-ACP567%3E3.0.CO;2-Y

Zitzewitz, E. (2006). Nationalism in Winter Sports Judging and Its Lessons for Organizational Decision Making. *Journal of Economics & Management Strategy*, *15*(1), 67–99. https://doi.org/10.1111/J.1530-9134.2006.00092.X

Zitzewitz, E. (2014). Does Transparency Reduce Favoritism and Corruption? Evidence From the Reform of Figure Skating Judging [Article]. *Journal of Sports Economics*, *15*(1), 3–30. https://doi.org/10.1177/1527002512441479

Appendix: Averages of all scores in Senior's competitions

Technical value	Judge #1 (own)	Judge #2 (highest)	Judge #3 (middle)	Judge #4 (lowest)
Seniors (16+)	5.732	5.811	5.684	5.600
Highest and lowest truncated		truncated		truncated
Final			5.708	

Artistic value	Judge #1 (own)	Judge #2 (highest)	Judge #3 (middle)	Judge #4 (lowest)
Seniors (16+)	3.487	3.561	3.404	3.196
Highest and lowest truncated		truncated		truncated
Final			3.445	

Execution	Judge #1 (own)	Judge #2 (highest)	Judge #3 (middle)	Judge #4 (lowest)
Seniors (16+)	8.805	8.909	8.727	8.645
Highest and lowest truncated		truncated		truncated
Final			8.766	

Execution	Judge #1 (rival)	Judge #2 (highest)	Judge #3 (middle)	Judge #4 (lowest)
Seniors (16+)	8.55	8.79	8.62	8.50
Highest and lowest truncated		truncated	(8.63) = aver. of three others	truncated
Final			8.58	

Corresponding author:

Seppo Suominen
Haaga-Helia University of Applied
Sciences
Hietakummuntie 1 A, 00700 Helsinki,
Finland
e-mail: seppo.suominen@haaga-helia.fi
tel num: +358404887142

Article received: 29. 5. 2023
Article accepted: 27. 9. 2023