Proceedings of the 2nd International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/131/2024.20583

Judging Bias in Olympic Diving: Fairness at Risk Zones During the Tokyo 2021 Games

Kangqi Song[1,5,*,†], Xuanrui Lin[2,6,†], Jiayin Li[3,7,†], Yingxia Yao[4,8,†]

[1]School of Political Science and Public Administration, Shandong University, Qingdao, 266200, China
[2]School of Faculty of Science and Technology, University of Beijing Normal university-Hong Kong Baptist University United International College, Zhuhai, 519000, China
[3]College of Letters and Science, University of Wisconsin-Madison, Madison, WI 53706, United States
[4]School of AI and Advanced Computing, Xian Jiaotong-Liverpool University, Suzhou, 215123, China

[5]songkangqi@mail.sdu.edu.cn
[6]s230033025@mail.uic.edu.cn
[7]jli2789@wisc.edu
[8]Yingxia.Yao23@student.xjtlu.edu.cn
*corresponding author
[†]All the authors contributed equally to this work and should be considered as co-first author.

Abstract. This study aimed to investigate whether judges exhibit bias when scoring divers from the same country, especially at the "risk moment", which are competitively significant moments in the sequence of the event. Using the Tokyo 2021 Olympic Diving dataset (Smith 2021), the study first identified the risk zones and analyzed the judges' scores in these areas in the preliminary and semi-final rounds. Subsequently, the permutation tests and t-tests were employed to examine whether the nationality of the divers affected the judges' scores. The findings suggest that in non-risk zones, judges may support divers from the same country. Nevertheless, in the risk zone, there is no evidence to suggest that judges are biased, which means that judges may consciously maintain fairness during the competition rounds especially at the risk moment. In addition, we explored anti-bias, which means the home judge may give low scores to those divers who are in the risk zone but not from their own country, as these divers may pose a threat to the progress of their own diver at the risk moment. However, the results show that while some judges' behavior is consistent with this assumption, such anti-bias lacks statistical significance with reference to the results of the corresponding permutation test.

Keywords: Risk Zone, Bias Detection, Tokyo 2021 Olympics Diving Competition, Permutation Tests.

DOI: 10.54254/2755-2721/131/2024.20583

1. Introduction

Diving is an aquatic sport that requires a high degree of skill. Diver's leap into the water from a diving platform or springboard accompanied by a series of aerial maneuvers. These maneuvers can range from simple front flips to complex twists and flips. Divers must demonstrate control, grace, and precision in order to earn high scores. This sport is also one of the highly anticipated events in the Olympics. In the Tokyo 2021 Olympic Games, diving competitions are categorized into four events (men's 3-meter springboard, men's 10-meter platform, women's 3-meter springboard and women's 10-meter platform) according to the gender and the height of the diving platform or springboard. Each event is divided into three rounds (preliminaries, semi-finals and finals), with each round allowing athletes to perform a set number of dives. For women, there are five dives in each round, while men have six dives in each round. Besides, there are always seven judges scoring for each dive. The final score for each dive is calculated by removing the highest and lowest scores given by the judges, adding up the remaining scores, and multiplying by the Degree of Difficulty (DD) for that dive.

In diving competitions, the fairness of the judges is crucial to ensure fair results and to preserve the integrity of the sport. However, subjective scoring systems are susceptible to bias, particularly nationalistic bias, whereby judges may favor athletes from their own country. Previous research has highlighted this issue, but there are still research gaps in understanding how this bias manifest itself at different stages of the competition and in specific ranking intervals known as risk zones. This study mainly investigated the presence of nationalistic bias in the men's 10-meter platform (M10mPF) diving event at the Tokyo Olympics, and made some attempts to extend this study to other events. The study first focused on identifying and analyzing specific risk zone ranking intervals that are critical for divers to advance to subsequent competitions. We only consider the preliminary and semifinal rounds in the middle of our study because the final does not find any cases where divers and judges are from the same country, which is irrelevant to the concerns of our study. Then, we used statistical methods including permutation tests to assess whether judges' ratings at the risk moment have any tendency to favoritism based on nationality. Besides, our analysis also compares the behavior of judges at different stages of the competition to gain insight into whether pressure in key rounds affects the fairness of judges' scoring.

The results of this study have important implications for the fairness of diving competitions. They emphasize the need to address potential biases to ensure that all the athletes from all countries are able to compete on a level playing field. Understanding and minimizing these biases can improve the fairness and credibility of subjective scoring systems in sport. This study contributes to the broader field of sport science and judging fairness by identifying specific instances and stages where bias may occur. It provides a structure for future research to further explore and address subjective biases in a variety of sports to promote fairness and equity in competition.

This paper is organized as follows: Chapter 2 briefly reviews relevant research on nationalist bias in international sporting events. Chapter 3 introduces the core variables, concepts, and methods for studying nationalist bias among referees in diving competitions at the 2021 Tokyo Olympics. Specifically, DoAD and risk zones are introduced. Chapter 4 presents the details of the "Bias" and "Anti-bias" analyses and the relevant experimental results. Chapters 5 conclude this paper and discuss the implications of the findings for future research.

2. Literature Review

The issue of nationalistic bias in sports judging has long been recognized as a threat to the legitimacy of international sport competitions [1]. Various studies have consistently highlighted the tendency of judges to favor athletes from their own country. For instance, Emerson and Meredith conducted an analysis of the 2000 Olympic Diving competition and identified significant nationalistic bias in the judges' scoring, particularly in cases where judges awarded higher scores to athletes of their own nationality [2]. Similarly, Emerson et al. provided compelling evidence of favoritism in diving events, emphasizing that such biases, though sometimes subtle, can significantly impact competition outcomes [3]. The magnitude of these biases varies across judges, nations and disciplines [4], and is widespread

DOI: 10.54254/2755-2721/131/2024.20583

across various types of competition [5][6], indicating the complex nature of bias in international sports competitions.

Previous studies have employed various methodologies to detect and analyze nationalistic bias in sports judging. Ansorge et al. examined the gymnastics judges at the 1984 Olympic Games and used sign tests to determine whether there was a significant bias [7]. Emerson et al. described a methodology similar to the discrepancy-based model for the analysis of diving judges' performance [3]. Sandro Heiniger et al. developed a statistical engine for the similar purpose [8]. Related research also used permutation tests, which compare observed discrepancies with those obtained through random data permutations [2].

In further research on nationalist bias in sports judging, Zitzewitz highlighted that judges score athletes from their own countries higher than other judges do, and they appear to vary their biases strategically in response to the stakes, the scrutiny given the event, and the degree of subjectivity of the performance aspect being scored [9]. Similarly, Emerson et al. revealed that the evidence of primary judging bias (i.e., judges in favor of divers from their own countries) is particularly noteworthy, but many other biases are evident as well [3]. Complicating this issue further, research suggests that nationalistic bias may also be driven by broader social and psychological factors. Youngju Kim et al.'s study showed that the Olympics are associated with temporary increases in intergroup biases [10]. Nationalistic bias in sports judging may be influenced not only by the specific conditions of the competition but also by broader social and psychological factors that temporarily heighten intergroup biases during such high-stakes international events.

Unlike previous studies, which have extensively examined nationalistic bias throughout the competition, our study introduces the concept of "risk zones"-critical moments in the competition when a diver's advancement is uncertain and bias is most likely to occur. Using this more targeted approach, we can gain a nuanced understanding of the potential for bias. In addition, our study selects statistical methods based on permutation tests and applies these methods specifically to the risk zone. This approach enables a clearer understanding of judicial behavior at these critical moments and to provide useful research tools and effective research ideas for understanding and reducing bias in sports officiating.

3. Methodology

3.1. DoAD

In cases where the diver and the judge are of the same nationality, the discrepancy between this judge's score and the average of all seven judges' scores is calculated and recorded. We calculate the average of all these types of discrepancies for each judge, which we refer to as avg(diff_match) of one judge. Conversely, when the judge is not of the same nationality as the diver, the difference between the score given by the judge and the average score will also be calculated and recorded. This average of all the differences is calculated, and we refer to it as avg(diff_nonmatch). The discrepancy between the avg(diff_match) and avg(diff_nonmatch) for each judge is defined as the DoAD.

The DoAD setting offers a valuable observation of the level of judges' nationalistic bias, as it removes the influence of judges' individual generalized scoring preferences. In the Tokyo 2021 Olympic Diving dataset, DoAD values were calculated and permutation tests were performed on each judge's scoring, and we were able to obtain similar conclusions as in previous studies. As shown in the Table 1, of the 18 judges who scored their own country's divers, 14 judges exhibited DoAD values greater than 0, while 7 judges demonstrated DoAD values greater than 0.2. Furthermore, the test yielded 6 p-values less than 0.01 (e.g., Mexican judge ROCHA CHAVEZ Sergio, DoAD=0.247, p-value<0.001), indicating that referees still exhibit a preference for their own country when scoring, and that this may not be a mere coincidence.

DOI: 10.54254/2755-2721/131/2024.20583

Table 1. Differences of Average Discrepancies (DoAD) by Judge

No	Judge	DoAD	p_value
1	AXTELIUS Peter	-0.086	0.624
2	HASSAN Mohamed	0.036	0.338
3	PETERSON Gord	0.171	0.005
4	SCHLEPPS Holger	0.217	<0.001
5	WRIGHT Lisa	-0.009	0.508
6	GOLOVAN Anatoliy	0.212	0.007
7	RAO Lang	0.266	<0.001
8	ROCHA CHAVEZ Sergio	0.247	<0.001
9	ZAMPIERI Marco	0.080	0.260
10	AHLERING Julie	0.030	0.300
11	ASADA Masako	-0.172	0.922
12	BROOKER Gillian	-0.004	0.531
13	FEOKTISTOV Eduard	0.078	0.157
14	FRASER Lindsey Ann	0.030	0.335
15	RODRIGUEZ AMADEO Angelique M.	0.319	0.086
16	MIN Suckhong	0.260	0.001
17	REIS Violeta	0.362	0.056
18	van der VOORT Ronald	0.127	0.199

3.2. Risk Zone and Boba Plot

3.3. Risk ID

We chose the men's 10-meter platform (M10mPF) for our analysis. First, we created an empty dataset and stored the ranking information of the diver in it. The ranking lists were recorded after the third to sixth dives, respectively. According to the rules of the competition, from the preliminary round to the semifinal round, the top 18 divers need to be selected to advance; from the semifinal round to the final round, the top 12 divers need to be selected. And the final gold, silver and bronze medals will be won by the top three divers in the final. Therefore, we lock in the rankings as 18, 12 and 3 and focus on the following **risk zone**: 14-22 in the preliminary round, 9-15 in the semifinal round, and 1-6 in the final round. These risk zones help us better understand how divers compete and their probability of advancing at each stage.

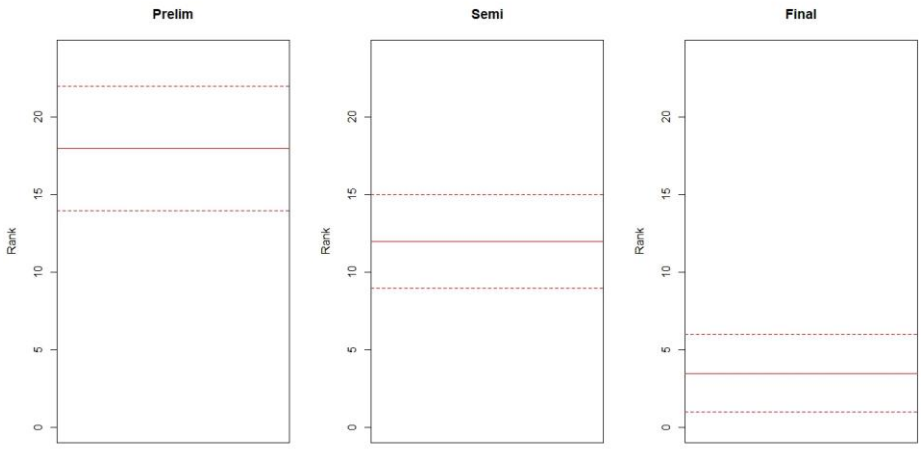

Figure 1. Risk-Zone Plot

DOI: 10.54254/2755-2721/131/2024.20583

Figure 1 shows the risk ranking zone of divers at different stages of the competition (preliminary, semi-final, and final). Each panel represents a separate stage of the competition, the y-axis indicating the ranking of the competitor (Rank). The red dashed ranges indicate the risk ranking zone for each stage where competitors within these ranking ranges are on the verge of advancing or being eliminated. The solid red line indicates the risk zone that we locked in rankings as 3, 12, and 18, showing the key points. In the risk zone, those who are too far ahead do not have to worry about being eliminated, while those who are too far behind are guaranteed to be eliminated, so we target the interval around the advancement rankings to be considered a risk zone. This kind of chart helps us visualize and understand how divers are competing at each stage of the game, making it easier for judges and divers to better strategize their matches.

However, this plot still has room for improvement. In the final round, we could not find any diver and judge matches in the risk zone, so it did not meet our criteria. Consequently, we chose to disregard it. Additionally, we did not consider the fact that the ranking of each jump is constantly changing, which suggests a more efficient way of defining the risk zones. Therefore, a new plot, Boba-Plot, will more accurately represent the risk zone.

3.4. Boba-Plot

In the Men's 10m Platform Diving competition, Boba-plot (Figure 2) focuses on the Preliminary round and Semifinal round. The x-axis is labeled "DiveNum" and represents the different rounds, preliminary round and semifinal round, from "Prelim_1" to "Prelim_6" and "Semi_1" to "Semi_6". y-axis represents the ranking of the divers. Each circle represents a dive, and its position along the y-axis indicates the diver's ranking for that dive. The difference between one judge and the average of all 7 judges from the same dive (Diff shows in the Boba-Plot). The size of the circle represents the absolute difference in scores (AbsDiff), with larger circles indicating larger deviations from the average scores. Circles are colored according to the sign of the difference in scores (Diff): **Red** indicates a positive difference (score higher than average judges score). And, **Blue** indicates a negative difference (score lower than average judges score). **Red dashed lines** indicate the cutoff points for advancing to the next round. **red shaded areas** represent risk zones, lines connect three circles performed the panel of judges changed after three dives. Risk zones show key performance areas where judges' scores will be scrutinized.

The four rectangles in Boba-plot (Figure 2), highlighted by red, show the divers in the risk zone. For example, in the fourth and fifth dives (blue dark circles) and from the preliminary round, the judges did not give high scores to the diver of the same nationality, but in the sixth dive (red darker circle), the judge gave a high score, allowing the diver to advance. That is the reason that we selected this as a risk zone. Within this range, such situations are very likely to occur.

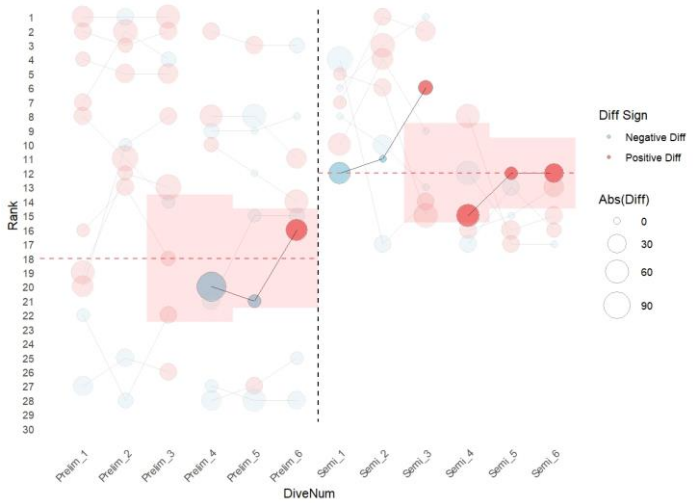

Figure 2. Boba plot for displaying multiple score information

DOI: 10.54254/2755-2721/131/2024.20583

4. Data Analysis and Results

This study focuses on bias and anti-bias within the Risk Zone, using the DoAD to observe the level of nationalistic bias of judges, and the permutation test to determine the bias. In the study of bias, we focus on whether judges will give higher scores to athletes from their own country in the Risk Zone relative to those from their own country but outside the Risk Zone. In the case of a Risk, the DoAD will be named DoADr. In the case of a Non-Risk, it will be named DoADn. The observed difference for each judge (observed_diff) is calculated as the DoADR = (DoADr - DoADn). In addition, the Anti-Bias study will investigate whether judges will give lower scores to divers from other countries in the Risk Zone that may affect the advancement of the divers from their own country, and the DoADo will be introduced to denote the DoAD value of the judges' scores for these athletes.

4.1. Bias Study

Based on an analysis of the Differences in DoAD in the "Olympic_Tokyo_2021" diving competition, we examined whether there was a significant difference between the judges' scores in risky and non-risky diving events. To evaluate whether the difference in DoAD values between "risk" zone and "non-risk" zone dives was statistically significant, a permutation test was performed, utilizing a one-tailed, positive direction approach to assess the significance of the observed difference. The dataset of 10,000 permutations is generated by disrupting the labeling of "match" (i.e., whether the diver's country is the same as the coach's). For each permutation, the DoAD is recalculated, which is the distribution of the permutation statistic. The observed DoAD values are then compared to the permutation distribution, and the p-value is obtained by calculating the position of the observed DoAD values in the permutation distribution. We used a p-value of 0.05 as a criterion, with lower p-values (< 0.05) likely to be biased, indicating no bias (p-value ≥ 0.05).

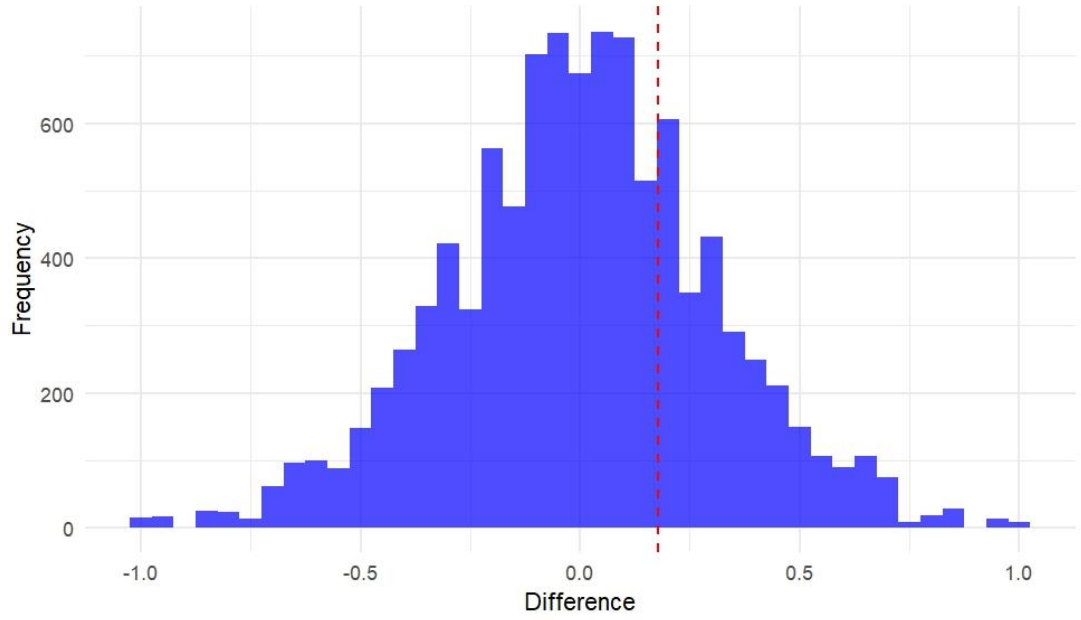

Figure 3. Histogram of Permutation Test Differences for PETERSON Gord

The histogram (Figure 3) shows a roughly normal distribution of differences centered at 0. The observed differences (indicated by the red dashed line) are slightly skewed away from the zero value. If the red dashed line (observed difference) is in the tail of the distribution, the difference is significant. However, since the observed difference is relatively close to the center of the distribution, this indicates that the observed difference is not significantly different from the difference expected by chance. This can be quantified by calculating the proportion of the piezometric difference (p-value) that is greater than or equal to the observed difference.

DOI: 10.54254/2755-2721/131/2024.20583

Table 2. DoADr and DoADn by Judge

(Judges in the risk zone and non-risk zone for the divers from their own country)

No	Judge	DoADr	DoADn	p_value
1	AXTELIUS Peter	-0.352	0.133	0.949
2	HASSAN Mohamed	0.030	0.039	0.514
3	PETERSON Gord	0.329	0.152	0.273
4	SCHLEPPS Holger	0.174	0.229	0.588
5	GOLOVAN Anatoliy	-0.026	0.256	0.843
6	ROCHA CHAVEZ Sergio	0.452	0.202	0.191
7	ZAMPIERI Marco	0.080	0.070	0.492
8	AHLERING Julie	-0.012	0.052	0.600
9	ASADA Masako	-0.254	-0.146	0.660
10	BROOKER Gillian	0.479	-0.035	**0.046**
11	FEOKTISTOV Eduard	-0.199	0.120	0.872
12	FRASER Lindsey Ann	0.013	0.033	0.534
13	RODRIGUEZ AMADEO Angelique M.	0.672	0.141	**0.041**
14	MIN Suckhong	0.041	0.319	0.837
15	van der VOORT Ronald	0.250	0.104	0.304

As demonstrated in Table 2, the calculated DoADr and DoADn values for Peterson Gord are 0.3288 and 0.1524, respectively, with a p-value of 0.2731. The p-values indicate that the observed differences in PETERSON Gord's DoAD values were not statistically significant, suggesting that there is no strong evidence of bias in scoring. According to the permutation test, most of the judges, including Peterson Gord, showed no significant bias in their scoring. However, a few judges, such as BROOKER Gillian and RODRIGUEZ AMADEO Angelique M., showed significant differences in their DoAD values, indicating possible bias in their scoring. In conclusion, the histogram helps to visualize whether the observed difference in DoAD for PETERSON Gord is anomalous compared to the expected difference under the null hypothesis of no bias. In this case, the observed differences do not appear to be significant.

4.2. Anti-Bias Study

In the study about anti-bias, the preliminary round of the men's 10-meter diving competition is still selected as the object of study in this research first. After the first four dives, the divers from 14th to 22nd place were in the "Risk Zone" as we defined earlier, and they came from 8 different countries (first country set). In the corresponding panel of judges, there were 7 judges from 7 different countries (second country set). Observing the two country sets, we can see that USA appears in both sets (from Table 3), meaning that the judge AHLERING Julie from USA not only scores divers from his own country that are in the risk zone, but also scores divers from other countries that are also in the risk zone.

Table 3. An example of selecting countries and judges from two country sets

Country Set 1 (Diver)	MEX	GER	BRA	CAN	KOR	**JPN**	**USA**	GBR
Country Set 2 (Judge)	**USA**	**JPN**	AUS	RUS	CUB	HUN	IND	

There is one new question at this point, whether the judges, in addition to having a tendency to score divers from their own country higher or lower, also have a tendency to score which of the other divers that may affect the advancement of the divers from their own country higher or lower. For the diver from the USA who was in the risk zone in the previous round, the American judge AHLERING Julie gave him a score of 8, which is lower than the average of the scores given by the other judges 0.071. It seems to support the conclusion we drew earlier, that judges may not score their own country's divers on the higher side at risk moments. For divers from other countries who were in the risk zone in the

DOI: 10.54254/2755-2721/131/2024.20583

previous round, the mean difference between the scores given by this American judge and the scores given by the other judges was -0.125. This shows that he did seem to have intentionally given low scores to other countries' competitors in order to enable his own country's competitors to advance. However, considering that this judge may prefer to score the divers lower, we examine the average of all the other scoring diffs of this judge and the result is -0.072, indicating that his scoring level is indeed lower than that of the other judges. However, this average value of the difference is quite small compared to the average value of the difference of the scores he gave to divers from other countries in the danger zone. Referring to the previous calculation similar to DoAD, we can get the discrepancy between these two as DoADo = -0.072 - (-0.125) = 0.053.

Subsequently, we expanded our view to examine the countries and the judges from these countries that appeared in both the "Judges'Country Set" and the " Risk Zone Divers Country Set" after the fourth jump in each competition and each round. We then observed the scores given by judges to non-national divers who were in the risk zone in the previous rounds 5 and 6 (for women, only round 5), respectively, by looking at the average of the difference in the scores given by judges to these divers compared to the scores given to these divers by the other judges (avgo). We will also calculate the average of the difference in scoring between this judge and the other judges scoring all the divers (avg), and derive the difference between these two averages (DoADo = avg - avgo), showing whether this judge tends to score those divers in the risk zone who are not from their own country lower or higher. Similar to the previous section, we assume that the judges will tend to give lower scores to those athletes who are in the danger zone but are not from their own country, and perform a permutation test on the DoAD value of each judge. The results are shown in Table 4.

Table 4. DoADo by Judge

(Judges in the risk zone and non-risk zone for the divers from their own country)

No	Country	Judge	avg	avgo	DoADo	p_value
M10mPF-Prelim-5	JPN	ASADA Masako	0.044	0.188	-0.143	0.803
M10mPF-Prelim-5	USA	AHLERING Julie	-0.072	-0.125	0.053	0.313
M10mPF-Prelim-6	USA	AHLERING Julie	-0.072	-0.143	0.071	0.286
M10mPF-Prelim-6	JPN	ASADA Masako	0.048	0.083	-0.036	0.573
M10mPF-Semi-5	KOR	MIN Suckhong	0.003	-0.167	0.169	0.120
M10mPF-Semi-5	BRA	REIS Violeta	-0.026	0.048	-0.074	0.671
M10mPF-Semi-6	BRA	REIS Violeta	-0.028	0.196	-0.225	0.865
M10mPF-Semi-6	KOR	MIN Suckhong	0.000	0.000	0.000	0.501
M3mSB-Prelim-5	NZL	WRIGHT Lisa	-0.013	-0.125	0.112	0.204
M3mSB-Prelim-5	USA	AHLERING Julie	-0.071	-0.204	0.134	0.127
M3mSB-Prelim-5	GER	SCHLEPPS Holger	-0.006	0.071	-0.078	0.766
M3mSB-Prelim-6	GER	SCHLEPPS Holger	-0.003	-0.107	0.104	0.199
M3mSB-Prelim-6	NZL	WRIGHT Lisa	-0.015	0.012	-0.027	0.568
M3mSB-Prelim-6	USA	AHLERING Julie	-0.071	-0.202	0.132	0.149
M3mSB-Prelim-6	ITA	ZAMPIERI Marco	0.024	0.024	0.000	0.499
M3mSB-Semi-6	MEX	ROCHA CHAVEZ Sergio	0.028	0.000	0.028	0.448
W10mPF-Prelim-5	GER	SCHLEPPS Holger	-0.003	-0.071	0.068	0.262
W10mPF-Prelim-5	EGY	HASSAN Mohamed	-0.055	0.045	-0.099	0.825
W3mSB-Prelim-5	UKR	GOLOVAN Anatoliy	-0.031	-0.205	0.174	0.062
W3mSB-Semi-5	KOR	MIN Suckhong	0.000	-0.024	0.024	0.434

Based on the results of the DoAD and permutation tests in the table, in about half of the cases, the judges will give lower scores (positive DoAD) to athletes who are in the risk zone but not from their own country. However, even if these were positive values, we could not statistically claim the existence

DOI: 10.54254/2755-2721/131/2024.20583

of judges' scoring bias, since all p-values in all DoAD permutation tests are greater than 0.05, suggesting that the observed bias is likely to be due to chance. As a result, our hypothesis that "the judges will tend to give lower scores to those athletes who are in the risk zone but are not from their own country" is rejected. Judges do not give lower scores to divers who are not from their own country in the risk zone, thus making it easier for divers from their own country in the risk zone to advance, thus demonstrating the fairness and professionalism of the judges in the competition. They will strictly follow the rules of the competition in the risk zone, judging on the criteria of technique, performance and artistry, without being influenced by nationality, region or other non-competitive factors, thus ensuring the fairness and authority of the competition.

5. Conclusion

5.1. Summary

This study builds on previous research on nationalistic bias in sports scoring, particularly with regard to diving. By focusing on the diving competition at the Tokyo 2021 Olympic Games, we aim to provide a more nuanced understanding of when and how bias manifests itself. According to previous studies, the presence of nationalistic bias in subjective scoring systems has been noted. Our study goes a step further by introducing the concept of "risk zones". By focusing on these risky moments, we provide a deeper analysis of scoring behavior. The study began by identifying and examining the risk zones where such bias is most likely to occur. Through detailed statistical analysis using permutation tests, the results of the study revealed that while there was some evidence of nationalistic "Bias" in the non-risk zones, this bias was not more pronounced in the risk moments. In addition, we explored the issue of "Anti-Bias" that judges may work against divers from other countries at risk zones and in favor of their own. As with the same results obtained from the "Bias" study, while individual cases were observed that met the hypothesis, the lack of statistical significance in these cases suggests that there is no certain or pervasive bias. These results highlight that although nationalist sentiments are associated with judges' judgments, the judges did not appear to be more unfair in the risk moments, both to their own athletes and to those from other countries who posed a threat to their own athletes.

5.2. Future Work

One of the major strengths of our study lies in the methodology used in defining and analyzing risk zones. By identifying moments in specific ranking zones that are critical to a diver's advancement, we are able to focus on those moments where bias has a significant impact. This targeted approach provides deeper insights into scoring behavior. In addition, our statistical analyses using permutation tests provided a comprehensive understanding of potential biases. Future research could focus on how the same judges score divers from various countries in different risk zones at different competitions to analyze judge bias more objectively and comprehensively. Besides, understanding the psychological and social factors that influence scoring decisions, such as exploring judges' motivations and pressures, can provide a more thorough understanding of bias.

Continuous monitoring and analysis of judging patterns is important to maintaining fairness in diving competitions. This study provides the stage for future research that will examine and eliminate bias in a variety of competitive situations. It also provides insightful information about the subject of sport science and evaluating fairness. Furthermore, the results of this study have significant implications for athletes and judges. On the one hand, judges might use the knowledge gained from this research to consider and enhance how they score, particularly in terms of preserving fairness during important moments in the competition. On the other hand, athletes stand to gain from being aware of efforts aimed at guaranteeing them fair scoring during crucial moments in their path to success. In conclusion, this study emphasizes the significance of preserving fairness at these important moments by providing thorough insights into nationalistic bias in Olympic diving competitions and by concentrating on the risky moments where bias is susceptible to arise.

DOI: 10.54254/2755-2721/131/2024.20583

Acknowledgement

Kangqi Song, Xuanrui Lin, Jiayin Li and Yingxia Yao contributed equally to this work and should be considered co-first authors. The article was prepared by Jiayin Li and Yingxia Yao and subsequently revised by Kangqi Song and Xuanrui Lin. The code for data processing and analysis was developed by Kangqi Song and Xuanrui Lin, and Jiayin Li and Yingxia Yao provided assistance in revising the plots and tables.

References

[1] Lyngstad, T. H., Härkönen, J. and Rønneberg, L. T. S. (2020) 'Nationalistic bias in sport performance evaluations: An example from the ski jumping world cup', *European Journal for Sport and Society*, 17(3), pp. 250–264. doi: 10.1080/16138171.2020.1792628.

[2] Emerson, J. W. and Meredith, S. (2011) 'Nationalistic Judging Bias in the 2000 Olympic Diving Competition', *Math Horizons*, 18(3), pp. 8–11. doi: 10.4169/194762111X12954578042812.

[3] Emerson, J. W., Seltzer, M. and Lin, D. (2009) 'Assessing Judging Bias: An Example From the 2000 Olympic Games', *The American Statistician*, 63(2), pp. 124–131. doi: 10.1198/tast.2009.0026.

[4] Heiniger, S. and Mercier, H. (2019) 'National Bias of International Gymnastics Judges during the 2013-2016 Olympic Cycle', *arXiv.org*, 16 August. Available at: https://arxiv.org/abs/1807.10033 (Accessed: 19 August 2024).

[5] Sandberg, A. (2017) 'Competing Identities: A Field Study of In-Group Bias Among Professional Evaluators', *The Economic Journal*, 128(613), pp. 2131–2159. doi:10.1111/ecoj.12513.

[6] Waguespack, D.M. and Salomon, R. (2016) 'Quality, Subjectivity, and Sustained Superior Performance at the Olympic Games', *Management Science*, 62(1), pp. 286–300. doi:10.1287/mnsc.2014.2144.

[7] Ansorge, C. J. and Scheer, J. K. (1988) 'International Bias Detected in Judging Gymnastic Competition at the 1984 Olympic Games', *Research Quarterly for Exercise and Sport*, 59(2), pp. 103–107. doi: 10.1080/02701367.1988.10605486.

[8] Heiniger, S. and Mercier, H. (2021) 'Judging the judges: evaluating the accuracy and national bias of international gymnastics judges', *Journal of Quantitative Analysis in Sports*, 17(4), pp. 289–305. doi:10.1515/jqas-2019-0113.

[9] Zitzewitz, E. (2006) 'Nationalism in Winter Sports Judging and Its Lessons for Organizational Decision Making', *Journal of Economics & Management Strategy*, 15(1), pp. 67–99. doi: 10.1111/j.1530-9134.2006.00092.x.

[10] Kim, Y. and Na, J. (2020) 'The Olympic paradox: The Olympics and intergroup biases', *Group Processes & Intergroup Relations*, 25(1), pp. 123-145. doi: 10.1177/1368430220931160.