© *Idōkan Poland Association*
“IDO MOVEMENT FOR CULTURE. Journal of Martial Arts Anthropology”,
Vol. 22, no. 4 (2022), pp. 33–40
DOI: 10.14589/ido.22.4.5

KINESIOLOGY & COACHING

Andrew Clark[1]
1 ORCID: 0000-0003-1273-9630
Department of Biomedical Engineering, The Ohio State University, Columbus (USA)
Contact: Department of Biomedical Engineering, The Ohio State University, 270 Bevis Hall, 1080 Carmack Rd., Columbus, OH 43210; e-mail: clark.1759@buckeyemail.osu.edu, Tel.: +1 740-649-7922

A Statistical Analysis of the Kata Scoring System in Sport Karate

Submission: 23.01.2020; acceptance: 1.02.2021

Key words: karate, kata, judging, World Karate Federation, Karate1

Abstract

Background. A year and a half before *karate* was scheduled to make its debut appearance in the Olympics in 2020, the World Karate Federation (WKF) elected to change its judging system for *kata* to a point-based system reminiscent of other Olympic sports.

Problem and aim. This paper analyzes data collected from WKF tournaments over the first year of the adoption of the new karate judging system to evaluate their implementation and factors that affected scores and inter-judge agreement.

Methods. The result books from the WKF’s 7 Karate1 Premier League and 4 Karate1 Series A League tournament were analyzed (3445 total kata performances). The distribution of scores and the standard deviation of judge scores per kata performance were calculated. Technical and athletic scores of each judge were plotted and Pearson’s correlation coefficient was found. Singular value decomposition (SVD) was used to cluster kata that were often run together by competitors and ANOVA and Tukey’s post hoc test were run on scores from the kata clusters.

Results. The frequency of scores follows a normal distribution with a slight negative skew. Judge agreement was affected by the quality of performance where better kata performances had a higher agreement. While judges give two independent scores for each technique and athleticism, there is a strong correlation between the two scores (r=0.932). Clustering and analysis of scores based on kata chosen to run showed that certain kata styles score higher.

Conclusions. This paper is the first attempt to evaluate the new kata scoring system and to identify extrinsic factors that affect the scoring.

Introduction

Karate is an international sport and martial art with origins in Japan. It is estimated that around 100 million people practice karate in 192 countries [Aina 2017]. As a sport, karate has two main events: kata and kumite. Kata involves the competitor performing a routine containing predefined sequences of karate techniques upon which they are scored. Kumite consists of standup sparring where successfully landed attacks are scored as points. While one can compete in both events (and often do train in both), elite competitors usually specialize in only one due to the unique technical and athletic demands of each [Doria *et al.* 2009; Koropanovski *et al.* 2011]. Many different karate organizations exist at all levels—local, national, international—each with their own different rule sets. However, the only organization recognized by the International Olympic Committee is the World Karate Federation (WKF). In 2016, the WKF secured karate in the 2020 Summer Olympics in Japan.

Prior to 2019, the WKF scored kata on a flag system. In this system, two competitors performed a kata of their choice decided before the round began [World Karate Federation 2018]. After both competitors completed their kata, a panel of five judges voted for the winner using a flag. The competitor receiving the most flags was the winner and moved onto the next round. In 2019, the WKF changed the judging system for kata to a point-based system at their tournaments and will be used in the 2020 Summer Olympics [World Karate Federation 2019]. In this system, all competitors are divided into one of multiple groups. Each competitor in the group performs his or her kata and is immediately scored by a panel of seven judges. At the end of the round, a predefined number of the highest scoring competitors move on to the next round where they must perform a different kata. This continues until the overall winner is decided.

Each judge gives two scores to the competitor upon completion of the kata, one for technical performance and one for athletic performance. The scores range from 5.0 to 10.0 with intervals of 0.2. The official criteria for evaluation for the technical performance are: stances, techniques, transitional movements, timing, correct breathing, focus (*kime*), and conformance of technique to the style of the kata. The athletic performance criteria are: strength, speed, and balance. Once all seven judges submit their scores electronically, the two highest and the two lowest scores for each technical performance and athletic performance are dropped. The remaining scores for the technical performance are then summed and multiplied by 0.7 to become the final technical score. The final athletic score is calculated similarly except multiplied by 0.3 instead of 0.7. The final total score is then the sum of both the final technical and final athletic scores.

There are 102 kata that a competitor can run [World Karate Federation 2019]. Traditionally the WKF has used kata from four main styles of Japanese karate: Goju Ryu, Shito Ryu, Shotokan, and Wado Ryu. However, in the last few years, the WKF has included a few kata from other styles such as Ryuei Ryu and Gensei Ryu. Some kata exist in multiple styles and thus have different variations. While a competitor may perform kata from different styles, each style has unique principles and technical considerations that the competitor must adhere to for the performance of their kata. Generally speaking, Goju Ryu is characterized by its combination of “soft” grappling techniques and strong strikes. Shito Ryu is characterized by short rapid techniques. Shotokan emphasizes strong linear long-range techniques. Wado Ryu uses principles of Jujitsu (Japanese grappling martial art) to emphasize softness and evasion. The official scoring system does not intrinsically favor any style, only that competitors perform kata from a style in conformity to the technique of the style.

Some articles have been published looking at factors that affect karate athletes’ performance [Camomilla *et al.* 2009; Augustovicova *et al.* 2018; Tabben *et al.* 2018; Tabben *et al.* 2019] and how to increase karate performance [Weinberg *et al.* 1981; Chaabene *et al.* 2019; Rezaei *et al.* 2019]. However, no article has ever investigated judging agreement, biases, or the influence of extrinsic factors in karate scoring. Judging agreement is the distribution of scores different judges would score a specific performance. At the ideal judge agreement, all judges would always give the same performance the same score. This also ensures that different performances can be accurately compared and that one performance does not score better due to which judges were scoring. Biases are the tendencies of a judge, or judges, to increase or decrease their score based on things that are not being scored (such as the competitor’s nationality). Extrinsic factors are everything that affects a performance score that is not dictated by the rules. Judge agreement and biases are considered extrinsic factors. By identifying and minimizing these factors, competitions become fairer, and competitors can better understand what to train and how to perform in order to maximize their scores.

Different analytical methods for analyzing sports judging have been made [Looney 2004; Scoppa 2008; Diaz-Pereira *et al.* 2014; Heiniger, Mercier pre-print]. These methods have been used to investigate many different sports [Emerson, Arnold 2011; Pajek *et al.* 2013; Gift 2018] as well as to evaluate changes in judging systems [Zitzewitz 2014; Premelc *et al.* 2019]. Through different studies, people have found evidence of extrinsic factors affecting scores of performances. In Olympic synchronized diving, it was found that performances were at a disadvantage when following a high scoring performance [Kramer 2017]. There was also shown to be evidence of difficulty bias in gymnastics, in which the difficulty and execution scores are judged separately however there was evidence of more difficult routines receiving higher execution scores [Morgan 2012]. By understanding these biases and extrinsic factors, organizations can modify the rules to avoid them, judges can try to avoid them in their judging, and athletes can try to best navigate them to maximize their score. The goals of this paper are, in the setting of elite WKF Karate1 tournaments, to identify extrinsic factors (overall quality of performance, corresponding technical/athletic performance, kata choice) that influence kata scoring and its inter-judge agreement.

Methods

Data collection

Result books for the WKF’s 7 Karate1 Premier League and 4 Karate1 Series A League tournaments held in 2019 were downloaded from sportdata.org. Result books detailed competitor’s nationality, individual judge’s scores, final score, kata name, gender, and turn order. A total of 3445 kata performances were conducted and analyzed from these 11 tournaments. All statistical tests and singular value decomposition (SVD) were done in JMP Pro 14.

Intrinsic judging error variability

Intrinsic judging error variability was calculated as previously done by Mercier and Heiniger [Mercier, Heiniger pre-print]. For each kata the performance median was treated as the control score as the median is less affected by an outlier score than the mean. The overall standard deviation for each kata performance with a given median score was calculated separately for each technical score and athletic score. This standard deviation by median score was plotted to visualize the judging error variability by performance quality.

Score correlation

All technical and athletic scores from all judges were pooled together and analyzed using Pearson's correlation coefficient. This correlation was compared with any 2 judges' scores by finding the Pearson's correlation coefficient between all combinations of 2 judges for each technical and athletic score.

Kata clustering and analysis

For each combination of 2 kata, the number of competitors that performed both of them (looking through all 11 tournaments) was recorded. This frequency of kata run together by a competitor then underwent singular value decomposition (SVD) for data reduction. A graph was made plotting SVD1 and SVD2 (the 2 SVD that explain the data the most). Kata were clustered based on proximity to other kata. Once clusters were established, an ANOVA was run on the dependent variable of total score with the main effects of kata cluster, gender, tournament and interactions of kata cluster with gender and tournament each. Tukey's post hoc test was run after significance was found in the ANOVA.

Results

Distribution of scores

For a point-based scoring system with a pre-defined range, it is important to see how this range was utilized. For both technical and athletic scores in this dataset, the distribution of all scores given by the judges followed normal distributions with a slight negative skew. However, both the average and median were 7.8 for both scores (Figure 1). Scores spanned almost the whole range of 5.0-10.0. The presented data includes scores from all rounds, analyzing scores from only the first round yielded similar distributions with less than 5% change in dispersion, but with an average and median of 7.6 (data not shown).

Effect of quality of performance on judge agreement

One metric that quantifies judge agreement is the standard deviation of the judges' scores for each performance. For each performance, the average standard deviation for each median score was plotted against the median score for both technical and athletic scores (Figure 2). Both categories show that there is less deviation between judge scores for better performances. An exception to this trend exists for the lowest of scores (5.0-5.6 for technical, 5.0 for athletic) in which judge deviation lowered with lower scores. These low-scoring outlier data points represent one performance each (for a total of 3).

Technical vs. athletic scores

The Pearson correlation coefficient for athletic score and technical score given by the same judge for the same kata performance was 0.932 indicating a strong correlation between the two. This correlation was much higher than that seen between any two judges for the same category for the same performance (Table 1). Furthermore, the slope of the trend line through a plot of the technical and athletic score data is 0.9431 with an intercept of 0.411 (Figure 3). This makes the athletic score nearly identical to the technical score with only the very extreme scores having up to a 0.1 difference. The lowest extreme technical scores have around 0.1 higher athletic scores than technical and the highest extreme technical scores have 0.1 lower athletic scores than technical scores. This makes the technical score slightly more sensitive to performance quality than the athletic score.

Kata selection and effect on scores

Lastly, how the scoring was affected by which kata was run was analyzed, especially with reference to the style the kata belongs to. While potentially 102 kata can be chosen, only 35 different kata were performed throughout all the tournaments. Out of the 3445 kata performances, 88% of them were one of 10 different kata. Singular value decomposition (SVD) data reduction on data recording each kata a competitor performed from all competitions was used to create a cluster graph (Figure 4A). This graph of SVDs places kata that were often run together close to each other on the graph. Two main clusters can be observed. Cluster 1 exclusively includes Shotokan kata and 1 kata from Gensei Ryu. Shotokan and Gensei ryu kata are absent from Cluster 2. Cluster 2 includes mostly kata from Shito Ryu, have a version in Shito Ryu, or have been adopted into Shito Ryu from other styles.

In total, kata contained in cluster 1 were run 1123 times and kata contained in cluster 2 were run 2321 times. An ANOVA showed a statistical significance for total score based on kata cluster as well as interactions with gender and the specific tournament. Cluster 2 scored better than cluster 1 for both genders and at all tournaments. However, for some tournaments the differences in scores between kata clusters were not statistically significant. Table 2 shows the differences between total scores based on kata cluster and statistical significance based on Tukey's post hoc test. Figure 4B shows the average total score for each individual kata for each cluster, which visualizes that only 1 of the highest 18 scoring kata was from cluster 1. The breakdown of which kata belongs to which cluster and their average score is in Table 3.

Discussion

The goal of this paper is to analyze potential extrinsic factors in the scoring of WKF Karate1 tournaments in 2019, the first year of the implementation of the new kata scoring system. Switching from a flag system to a point-based system also allowed for the opening up of the analyses performed in this paper. Both technical and athletic scores given by judges followed a normal distribution with a slight negative skew and an average of 7.8.

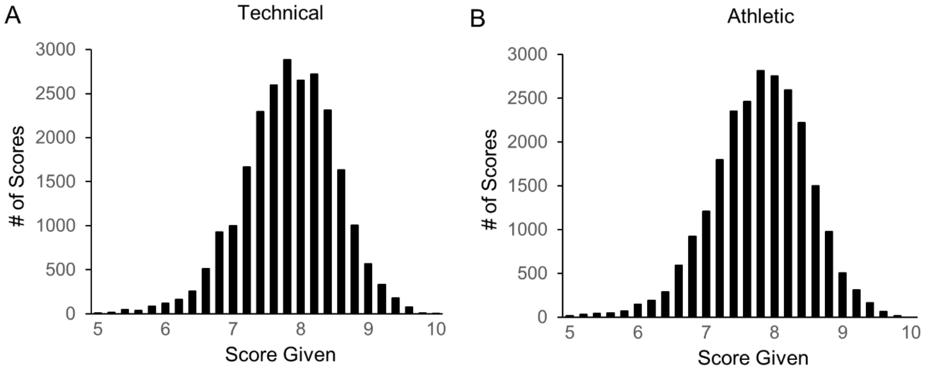

Figure 1. Distribution of scores given by judges for technical performance (A) and athletic performance (B)

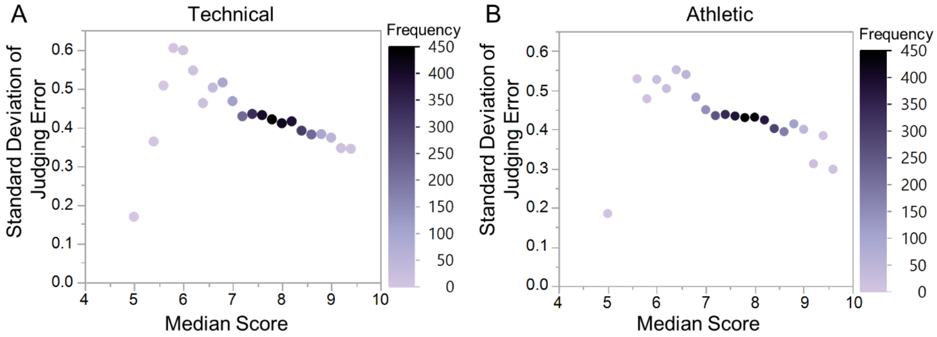

Figure 2. Standard deviation of judging marks versus median score for technical performance (A) and athletic performance (B)

This distribution may be indicative of central tendency bias — the difficulty of differentiating between average level performances causing an over utilization of the middle scores [Myford, Wolfe 2004]. However, it could also indicate that the performance level of athletes truly approximates a normal distribution. This latter argument could be supported by the fact that scores in the middle range had less inter-judge agreement than the lowest scores. Regardless, due to the bulk of the scores being in this middle range, it is important for judges to be adequately trained to accurately differentiate between performances scoring in this range. In addition, it may be beneficial to increase the available score range (use 0-10, or use 0.1 increments) to help differentiate average performances. Having a median score of 7.8 slightly decreases the available range of scores for good performances. While this might be a disadvantage, it provides a broader range of scores for the lowest scoring performances, which have the lowest inter-judge agreement.

Previous analyses of inter-judge agreement of similar sports events typically show one of two different trends [Heiniger *et al.* 2018]. One trend is that the better the performance, the greater the inter-judge agreement. The other common trend is a quadratic relationship in which both the worst and the best performances have the highest inter-judge agreement and the scores in the middle have the least inter-judge agreement. The data found for this data set is that it exhibits the former. That inter-judge agreement increases with better performance. This suggests that it is easier for judges to accurately assess the quality of good performances, but there is less agreement on how poor or mediocre performances should be rated. The exception to this trend of inter-judge agreement increasing with better performances is at the very lowest few scores, in which case there is high inter-judge agreement. This could be due to the low number of performances on these scores. These performances could also have scored low due to obvious performance errors, such as a momentary loss of balance, which could lead to increased inter-judge agreement on these performances.

With the change in the kata judging system from a flag-based system to a point-based system, the WKF changed the weight of athletic versus technical performance. Previously, both athletic and technical performances were to have equal consideration [World Karate Federation 2018]. With the adoption of the current point-based system, the technical performance was weighted to be 70% of the final score and the athletic performance to be weighted 30%. While the intent may be to score technical and athletic performances independently, in practice they have a high correlation. In fact, the technical score given by a judge is more predictive of the athletic score the judge gives than it is of a technical score given by another judge. The trend line through athletic versus technical scores is nearly 1. It is possible that this data could be explained by better competitors training both athletic and technical components and thus are truly better at both. However, the high correlation throughout all scores makes this seem unlikely. It seems more plausible that the judges are unable to distinguish between technical and athletic performances and thus assign similar scores for both. This might be able to be rectified through better training of judges on how to score each aspect. At an organizational level, it may be useful to better define the two components and how to distinguish the two. As it remains, it appears for this elite level dataset that having both technical and athletic scores are redundant and adds little value. However, perhaps having the two scores is beneficial when having a participant that is especially lacking in either athletic ability or technical skill. These two scores may also be more independent for other demographics such as novice level or youth divisions.

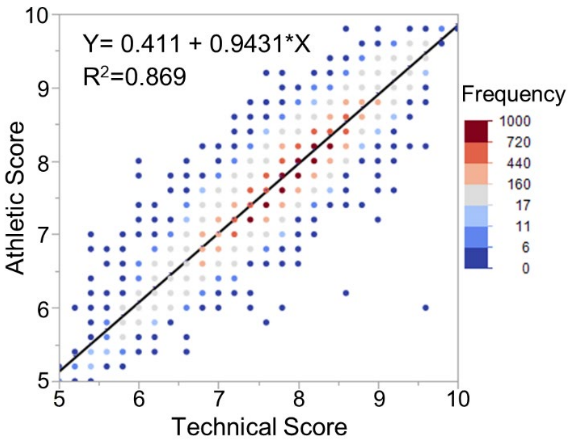

Figure 3. Plot of the relationship between athletic and technical scores given by the same judge.

While potentially 102 kata can be chosen and some competitors might have to run up to 5 different kata at a single tournament, there was clear favoritism for which kata were run by competitors with 88% of the ran kata being one of 10 different kata. There is no official degree of difficulty score but some kata showcase the judged attributes better than others and thus can contribute to a higher score. Also of consideration is that these kata were originally created to be a collection of self-defense moves for an individual to be able to practice on their own with no intention of tournament performance in mind.

Table 1. Pearson Correlation Coefficients for Judge Scores

	Pearson's Correlation Coefficient	
	Average	Range
Between any 2 judges		
Technical vs Technical	0.62	0.59 – 0.67
Athletic vs Athletic	0.62	0.58 – 0.66
For the same Judge		
Technical vs Athletic	0.932	

Historically, kata in WKF had been limited to four styles—Goju Ryu, Shito Ryu, Shotokan, and Wado Ryu. However, starting in 2013 the WKF rulesets stopped segregating kata into one of these four styles and since than it has included a few kata from other styles. Most of these newly included kata have been adopted by Shito Ryu stylists. Cluster analysis of kata run by each competitor reveals a sharp distinction in kata choice. Competitors can be broken into running kata from one of these two clusters. Cluster 1 contains 8 kata from the Shotokan style as well as 1 kata from Gensei Ryu. Cluster 2 contains a larger number of kata with a total of 25 kata. Most of these kata are kata from Shito Ryu, have a version in Shito Ryu, or have been adopted into some Shito Ryu schools from other styles. On average, kata from cluster 2 scored 0.58 higher in overall score than those from cluster 1. Cluster 2 had over twice as many performances as cluster 1 (2321 compared to 1123). There were extremely few competitors that did kata from both cluster 1 and cluster 2. This perhaps highlights the stylistic difference between the Shotokan style from the rest of the styles in that they do not perform kata from cluster 2. This is more likely due to stylistic differences in how movements are performed rather than the actual sequence of movements in the kata as Shito Ryu variants of Shotokan kata cluster into cluster 2 (ex. Shotokan's Unsu in cluster 1 and Shito Ryu's variant Unshu in cluster 2).

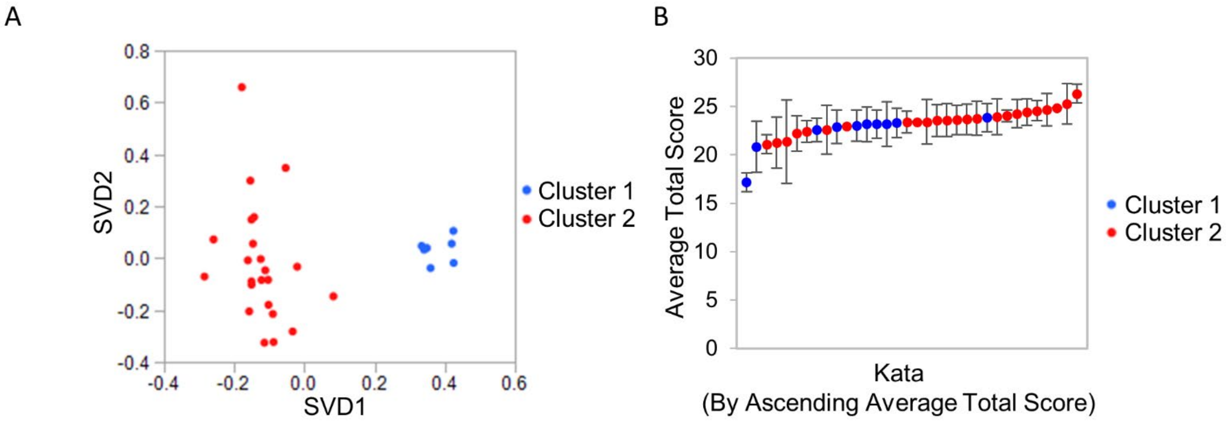

Figure 4. Results of spatial clustering based on SVD of kata performed together by a competitor (A). Graph showing the average total score of kata in ascending order and denoting kata cluster (B).

Table 2. Total Scores for Kata Clusters (mean ± SD)

	Total Score		
	Kata Cluster 1	Kata Cluster 2	Difference
Overall	23.12 ± 1.67	23.68 ± 1.68	0.56*
Tournament			
Dubai	22.91 ± 0.98	23.26 ± 1.3	0.35
Istanbul	23.23 ± 1.64	23.87 ± 1.8	0.64*
Tokyo	23.81 ± 1.28	23.85 ± 1.69	0.04
Madrid	23.8 ± 1.2	23.89 ± 1.46	0.09
Montreal	22.55 ± 2.55	22.9 ± 2.04	0.35
Moscow	23.83 ± 1.48	24.02 ± 1.58	0.19
Paris	22.76 ± 1.08	23.17 ± 1.3	0.41
Rabat	23.9 ± 1.49	24.11 ± 1.5	0.21
Salzburg	22.58 ± 1.68	23.51 ± 1.73	0.93*
Santiago	23.33 ± 1.55	23.86 ± 1.48	0.53
Shanghai	23.93 ± 1.52	24.38 ± 1.55	0.45
Sex			
Male	23.26 ± 1.66	23.74 ± 1.67	0.48*
Female	22.76 ± 1.62	23.64 ± 1.69	0.88*

* – denotes statistical significance between kata cluster 1 and 2 (Tukey’s Post Hoc Test, P < 0.05)

Table 3. List of scores and cluster category of kata ran

			Average Score			Standard Deviation		
Kata	Cluster	Times Ran	Tech score	Athl Score	Total Score	Tech score	Athl Score	Total Score
Bassai Dai	1	2	11.9	5.3	17.2	0.6	0.4	1.0
Enpi	1	10	14.6	6.3	20.8	1.9	0.8	2.6
Nijushiho	N/A	1	14.7	6.3	21.0	N/A	N/A	N/A
Heiku	2	2	14.8	6.3	21.1	0.6	0.4	1.0
Gojushiho	2	5	15.0	6.3	21.3	1.8	0.9	2.7
Chinto	2	2	15.0	6.4	21.4	3.2	1.1	4.3
Nipaipo	2	16	15.6	6.7	22.2	1.3	0.5	1.8
Pachu	2	3	15.7	6.7	22.4	0.8	0.3	1.1
Paiku	2	24	15.8	6.8	22.6	0.9	0.4	1.2
Kosukun Dai	2	6	15.9	6.7	22.6	1.8	0.7	2.5
Kanku Sho	1	123	16.0	6.9	22.9	1.2	0.5	1.7
Sanseiru	2	1	16.1	6.8	22.9	N/A	N/A	N/A
Unsu	1	303	16.1	6.9	23.0	1.1	0.5	1.6
Gojushiho Sho	1	237	16.2	6.9	23.2	1.3	0.6	1.8
Gojushiho Dai	1	199	16.2	6.9	23.2	1.0	0.4	1.4
Sochin	1	11	16.3	6.9	23.2	1.6	0.7	2.3
Gankaku	1	186	16.3	7.0	23.3	1.1	0.5	1.5
Kusanku	2	3	16.3	7.1	23.4	0.8	0.3	1.1
Matsumura Bassai	2	1	16.4	7.0	23.4	N/A	N/A	N/A
Tomari Bassai	2	16	16.4	7.0	23.4	1.6	0.7	2.3
Anan	2	285	16.5	7.0	23.6	1.1	0.5	1.6
Suparinpei	2	493	16.6	7.0	23.6	1.2	0.5	1.7
Anan Dai	2	458	16.6	7.1	23.6	1.1	0.5	1.5
Papuren	2	435	16.6	7.1	23.7	1.1	0.5	1.6
Chatanyara Kusanku	2	318	16.7	7.1	23.8	1.3	0.5	1.8
Sansai	1	54	16.7	7.2	23.9	1.0	0.4	1.5
Chibana No Kushanku	2	81	16.8	7.2	24.0	1.3	0.6	1.9
Kyan No Chinto	2	2	16.9	7.1	24.1	0.4	0.2	0.6
Ohan	2	42	17.0	7.3	24.3	1.0	0.5	1.5
Kururunfa	2	60	17.1	7.3	24.4	1.0	0.4	1.4
Shisochin	2	6	17.2	7.4	24.6	0.7	0.3	1.0
Ohan Dai	2	53	17.3	7.4	24.7	1.1	0.5	1.7
Seisan	2	1	17.5	7.4	24.9	N/A	N/A	N/A
Unshu	2	2	17.6	7.6	25.3	1.6	0.5	2.1
Oyadomari No Passai	2	4	18.5	7.8	26.3	0.6	0.3	1.0

Previous data had been collected when the WKF was using a flag-based judging system by Augustovicova et al [2015]. They found similar trends in style success in Karate1 kata competitions. They found Shotokan kata had a 43.3% win rate (compared to Shito Ryu’s 54.8%) and also cited that of the top 20 for both men and women, 6 men and 3 women did Shotokan whereas 13 men and 17 women performed Shito Ryu kata that year. However, in 2015 they found Shotokan kata made up 44.7% of all kata performed [Augustovicova *et al.* 2018]. The data presented here from 2019 shows a dramatic change within these 4 years, where Shotokan kata went from being performed 44.7% of the time to 33% (which includes the Gensai Ryu kata Sansei). This decrease could be due to fewer Shotokan practitioners doing well enough to qualify for these elite international tournaments. However, it could also be that competitors are switching away from Shotokan to Shito Ryu which has a larger repertoire of competition-quality kata and score higher. Augustovicova et al. [2019] has also done a study looking at kata selection at different age groups and concluded there were no significant differences found in kata selection between adults and younger age groups. This indicates that these kata may be higher scoring in younger divisions as well or that younger age groups are adopting what is being done in the adult division.

Gymnastics has found artistic scoring to have a low judge agreement [Bucar *et al.* 2014]. They suspected that the poorly defined rules for the artistry evaluation contributed to an overall lack of consistency and increased effect of individual judges' tastes. They also proposed a repetitive examination of artistry judging. While karate kata do not explicitly have an artistic score, some of the criteria are similar to the artistic score, such as timing and focus. Karate kata judging may also benefit from better-defined criteria and a repetitive examination of the current state of kata judging.

Conclusion

The adoption of WKF's new scoring system allows for more nuanced scoring for a kata performance than the previous flag scoring system which only provided how a kata performance did relative to one competitor's performance. This provides better feedback for coaches and athletes about the quality of the performance as well as potentially help spectators understand the decision. The new system also makes athlete drawing much less of an extrinsic factor in determining the final rank placement of competitors. At the current state in elite adult WKF Karate1 tournaments, average kata performances are hard to differentiate from each other and placement for these performances may depend on the judging panel due to the variation in judge scores. However, high judge agreement for the top performances indicate more accurate placement of their performances. Future judge training could focus on how to best distinguish between these average performances and increase their judge agreement. Having two different scores (technical and athletic) for these performances appeared to not be effectively utilized as the athletic score correlated more to the judge's technical score than judge agreement for technical score, rather than the two scores being independent. Lastly, kata choice has evidence to be an extrinsic factor for the score with a bias towards specific styles. Athletes may choose to adopt a higher scoring kata to improve their tournament rank placement.

Acknowledgements

I would like to acknowledge and thank Alan Burgess for his constructive suggestions during the planning and development of this research work.

Source

http://www.sportdata.org (access: Dec. 2019)

References

1. Aina K. (2017), *The Global Allure of Karate*. Available at: https://www.nippon.com/en/views/b06601/the-global-allure-of-karate.html.
2. Augustovicova D., Argajova J., Saavedra M., Matabuena M., Arriaza R. (2018), *Top-level karate: Analysis of frequency and successfulness of katas in K1 Premiere League*, "Ido Movement for Culture. Journal of Martial Arts Anthropology", vol. 18, no. 4, pp. 46-53; doi: 10.14589/ido.18.4.6.
3. Augustovicova D., Stefanovsky M., Argajova J., Kampmiller T. (2019), *The issue of early specialization in karate: the same pool of katas in all top-level WKF competition age categories*, "Archives of Budo", vol. 15, pp. 241-248.
4. Bucar Pajek M., Kovac M., Pajek J., Leskosek B. (2014), *The judging of artistry components in female gymnastics: A cause for concern?* "Science of Gymnastics Journal", vol. 6, no. 3, pp. 5–12.
5. Camomilla V., Sbriccoli P., Mario A.D., Arpante A., Felici F. (2009), *Comparison of two variants of a kata technique (unsu): the neuromechanical point of view*, "Journal of Sports Science & Medicine", vol. 8, no. CSSI3, pp. 29-35.
6. Chaabene H., Negra Y., Capranica L., Prieske O., Granacher U. (2019), *A Needs Analysis of Karate Kumite With Recommendations for Performance Testing and Training*, " Strength and Conditioning Journal.", vol. 41, no. 3, pp. 35-46; doi: 10.1519/ssc.0000000000000445.
7. Diaz-Pereira M.P., Gomez-Conde I., Escalona M., Olivieri D.N. (2014), *Automatic recognition and scoring of olympic rhythmic gymnastic movements*, "Human Movement Science", vol. 34, pp. 63-80; doi: 10.1016/j.humov.2014.01.001.
8. Doria C., Veicsteinas A., Limonta E., Maggioni M.A., Aschieri P., Eusebi F., Fano G., Pietrangelo T. (2009), *Energetics of karate (kata and kumite techniques) in top-level athletes*, "European Journal of Applied Physiology", vol. 107, no. 5, pp. 603-610; doi: 10.1007/s00421-009-1154-y.
9. Emerson J.W., Arnold T.B. (2011), *Statistical Sleuthing by Leveraging Human Nature: A Study of Olympic Figure Skating*, "The American Statistician", vol. 65, no. 3, pp. 143-148.
10. Gift P. (2018), *Performance Evaluation and Favoritism:Evidence From Mixed Martial Arts*, " Journal of Sports Economics", vol. 19, no. 8, pp. 1147-1173; doi: 10.1198/tast.2011.10165.
11. Heiniger S., Mercier H., *Judging the Judges: A General Framework for Evaluating the Performance of International Sports Judges*. Pre-print available at: https://arxiv.org/pdf/1807.10055.pdf

12. Koropanovski N., Berjan B., Bozic P.R., Pazin N., Sanader A., Jovanovic S., Jaric S. (2011), *Anthropometric and physical performance profiles of elite karate kumite and kata competitors*, "Journal of Human Kinetics", vol. 30, pp. 107-114; doi: 10.2478/v10078-011-0078-x.
13. Looney M.A. (2004), *Evaluating judge performance in sport*, "Journal of Applied Measures", vol. 5, no. 1, pp. 31-47.
14. Mercier H., Heiniger S., *Judging the Judges: Evaluating the Performance of International Gymnastics Judges*. Pre-print available at: https://arxiv.org/pdf/1807.10021.pdf
15. Morgan H., Rotthoff K. (2012), *The Harder the Task, the Higher the Score: Findings of a Difficulty Bias*, "Economic Inquiry", vol. 52; doi: 10.2139/ssrn.1555094.
16. Myford C.M., Wolfe E.W. (2004), *Detecting and measuring rater effects using many-facet Rasch measurement: Part II*, "Journal of Applied Measures", vol. 5, no. 2, pp. 189-227.
17. Pajek M.B., Cuk I., Pajek J., Kovac M., Leskosek B. (2013), *Is the quality of judging in women artistic gymnastics equivalent at major competitions of different levels?*, "Journal of Human Kinetics", vol. 37, pp. 173-181; doi: 10.2478/hukin-2013-0038.
18. Premelc J., Vuckovic G., James N., Leskosek B. (2019), *Reliability of Judging in DanceSport*, "Frontiers in psychology", vol. 10, pp. 1001-1001; doi: 10.3389/fpsyg.2019.01001.
19. Rezaei S., Akbari K., Gahreman D.E., Sarshin A., Tabben M., Kaviani M., Sadeghinikoo A., Koozehchian M.S., Naderi A. (2019), *Caffeine and sodium bicarbonate supplementation alone or together improve karate performance*, "Journal of the International Society of Sports Nutrition", vol. 16, no. 1, p. 44; doi: 10.1186/s12970-019-0313-8.
20. Scoppa V. (2008), *Are subjective evaluations biased by social factors or connections? An econometric analysis of soccer referee decisions*, "Empirical Economics", vol. 35, no. 1, pp. 123-140; doi: 10.1007/s00181-007-0146-1.
21. Tabben M., Conte D., Haddad M., Chamari K. (2019), *Technical and Tactical Discriminatory Factors Between Winners and Defeated Elite Karate Athletes*, "International Journal of Sports Physiology and Performance", vol. 14, no. 5, p. 563; doi: 10.1123/ijspp.2018-0478
22. Tabben M., Miarka B., Chamari K., Beneke R. (2018), *Decisive Moment: A Metric to Determine Success in Elite Karate Bouts*, "International Journal of Sports Physiology and Performance", vol. 13, no. 8, p. 1000; doi: 10.1123/ijspp.2017-0526
23. Weinberg R.S., Seabourne T.G., Jackson A. (1981), *Effects of Visuo-motor Behavior Rehearsal, Relaxation, and Imagery on Karate Performance*, "Journal of Sport and Exercise Psychology", vol. 3, no. 3, p. 228; doi: 10.1123/jsp.3.3.228.
24. World Karate Federation (2018), *Kata and Kumite Competition Rules for 2018.*
25. World Karate Federation (2019), *Kata and Kumite Competition Rules for 2019.*
26. Zitzewitz E. (2014), *Does Transparency Reduce Favoritism and Corruption? Evidence From the Reform of Figure Skating Judging*, "Journal of Sports Economics", vol. 15, no. 1, pp. 3-30; doi: 10.1177/1527002512441479.

Analiza statystyczna systemu punktacji *kata* w karate sportowym

Słowa kluczowe: karate, kata, sędziowanie, Światowa Federacja Karate, Karate1

Streszczenie

Tło. Na półtora roku przed planowanym debiutem karate na Olimpiadzie w 2020 roku, Światowa Federacja Karate (WKF) zdecydowała się zmienić system sędziowania *kata* na system punktowy, przypominający ten w innych sportach olimpijskich.

Problem i cel. W niniejszej pracy przeanalizowano dane zebrane z turniejów WKF w ciągu pierwszego roku od przyjęcia nowego systemu sędziowania karate, aby ocenić ich wdrożenie oraz czynniki, które miały wpływ na punktację i porozumienie między sędziami.

Metody. Analizie poddano książki z wynikami z 7 turniejów *Karate1 Premier League* i 4 *Karate1 Series A League* WKF (łącznie 3445 wykonań *kata*). Obliczono rozkład punktów i odchylenie standardowe ocen sędziowskich przypadających na 1 *kata*. Wyniki techniczne i sportowe każdego sędziego zostały sporządzone na podstawie wykresu i współczynnika korelacji Pearsona. Dekompozycja wartości pojedynczych (SVD) została użyta do grupowania *kata*, które były często wykonywane razem przez zawodników, a na wynikach z grup *kata* zostały przeprowadzone testy ANOVA i *post hoc* Tukeya.

Wyniki. Częstotliwość ocen miała rozkład normalny z lekkim ujemnym odchyleniem. Zgodność sędziów była zależna od jakości wykonania, gdzie lepsze wykonania *kata* miały wyższą zgodność. Chociaż sędziowie przyznali dwie niezależne oceny dla każdej techniki i wysportowania, istnieje silna korelacja między tymi dwoma ocenami (r=0,932). Analiza wyników w oparciu o wybrane *kata* pokazała, że niektóre style *kata* są wyżej punktowane.

Wnioski. Niniejsza praca jest pierwszą próbą oceny nowego systemu punktacji *kata* oraz określenia czynników zewnątrzpochodnych, które mają wpływ na punktację.