

Reliability in the evaluation of international and national judges in an artistic swimming routine

Confiabilidade na avaliação de juízes internacionais e nacionais em uma rotina de nado artístico

Katia Regina Ponciano¹

<https://orcid.org/0000-0002-7274-9606>

Meico Fugita¹

<https://orcid.org/0000-0001-6301-6868>

Cláudia Borim da Silva²

<https://orcid.org/0000-0002-3234-7581>

Roberta Luksevicius Rica³

<https://orcid.org/0000-0002-6145-1337>

Ana Carolina Gomes¹

<https://orcid.org/0000-0002-5917-5679>

Cássio de Miranda Meira Júnior¹

<https://orcid.org/0000-0002-9775-935X>

Danilo Sales Bocalini⁴

<https://orcid.org/0000-0003-3993-8277>

Abstract – Artistic swimming (AS) is a sport evaluated by fifteen judges in routine sessions. The athletes' goal is to achieve proficient motor patterns according to pre-established criteria. The present study analyzed whether there is difference between the two groups of AS judges with different levels of experience. The group of the International Swimming Federation - FINA (IG) evaluates AS at national and international level and the non-FINA group (NG) evaluates AS only at national level. Twenty experienced judges were divided into groups, 10 IG judges and 10 NG judges. Thus, judges evaluated the item execution of three routines with five required elements. Cronbach's alpha coefficient showed high internal consistency in IG ($\alpha = 0.85$ in T1 and 0.83 in T2). In NG, internal consistency was observed in T1 and low consistency in T2 ($\alpha = 0.82$ in T1 and 0.39 in T2). Evaluation analysis between IG and NG was significant ($p > 0.0330$) and reliability analysis (bias: -0.1266 95% agreement limit: -1.642 to 1.388) showed consistency and high degree of confidence in results. The findings suggest that the item execution of required elements showed high objectivity regarding judges with different levels of experience, IG and NG, regardless of categorization and time of practice. FINA has changed the number of judges and the number of items evaluated in routine sessions. It is suggested that the reduction of items has contributed in a positive way so that judges can focus more on evaluation itself.

Key words: Attention; Judgment; Performance evaluation; Swimming.

Resumo – Nado artístico (NA) é um esporte avaliado por quinze juízes em provas de rotinas. A meta dos atletas é a realização de padrões motores proficientes de acordo com critérios pré-estabelecidos. O presente estudo analisou se havia diferença entre os dois grupos de juízes de NA com diferentes níveis de experiência. O grupo da Federação Internacional de Natação - FINA (GI) avalia a nível nacional e internacional e o grupo não FINA (GN) avalia apenas a nível nacional. Participaram 20 juízes experientes separados em grupos, 10 juízes GI e 10 juízes GN. Assim, os juízes avaliaram o item execução de três rotinas com cinco elementos obrigatórios. O coeficiente alfa de Cronbach apresentou elevada consistência interna no GI ($\alpha = 0,85$ em T1 e $0,83$ em T2). No GN observou-se consistência interna em T1 e baixa em T2 ($\alpha = 0,82$ em T1 e $0,39$ em T2). A análise de avaliação entre GI e GN, foi significativa ($p > 0,0330$) e análise de confiabilidade (viés: $-0,1266$ de limite de concordância 95%: $-1,642$ a $1,388$) mostrando consistência e alto grau de confiança nos resultados. Os achados sugerem que o item execução dos elementos requeridos apresentou alta objetividade no que diz respeito aos juízes com diferentes níveis de experiência, GI e GN, independentemente da categorização e do tempo de prática. A FINA alterou o número de juízes e o número de itens que são avaliados nas provas de rotina. Sugere-se que a redução dos itens tenha contribuído para que os juízes se concentrassem mais na avaliação.

Palavras-chave: Atenção; Avaliação de desempenho; Julgamento; Natação.

1 University of Sao Paulo. Department of Physical Activity Sciences School of Arts, Science and Humanities. Sao Paulo, SP, Brazil.

2 Sao Judas Tadeu University. Graduate Program in Physical Education. Sao Paulo, SP, Brazil.

3 Estacio de Sa University. Department of Physical Education. Vitoria, ES, Brazil.

4 Federal University of Espirito Santo. Experimental Physiology and Biochemistry - Laboratory of Physical Education and Sport Center. Vitoria, ES, Brazil.

Received: August 19, 2020

Accepted: December 08, 2020

How to cite this article

Ponciano K, Fugita M, Silva CB, Rica RL, Gomes AC, Meira Júnior CM, Bocalini DS. Reliability in the evaluation of international and national judges in an artistic swimming routine. Reliability in the evaluation of international and national judges in an artistic swimming routine. Rev Bras Cineantropom Desempenho Hum 2021;23:e76587. DOI: <http://doi.org/10.1590/1980-0037.2021v23e76587>

Copyright: This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).



INTRODUCTION

Artistic swimming (AS) is an Olympic sport characterized by different events: solo (1 woman), duet (2 women), mixed duet (1 man and 1 woman), team (8 women), combined free and highlight (10 women). In this sport, athletes' performance consists of a sequence of movements with the accompaniment of a song in routine tests or without music in figure tests.

As in other acrobatic and rhythmic sports, in AS, the objective is the performance itself and the score that reflects the quality of movements is attributed by judges¹.

The performance of AS athletes is evaluated by seven judges who give scores in figure tests and fifteen judges who give scores in routine tests. In AS, there are two types of routines, technical (TR - with required elements) and free (FR - with free content). Fifteen judges divided into two tables, eight on one side of the pool and seven on the other, evaluate athletes in the following evaluation components: (1) execution / synchronization; (2) artistic impression (FR) or general impression (TR); and (3) difficulty (FR) or execution of elements / synchronization (TR)². Each judge must study the totality of evaluation components for both tests, as they must be able to evaluate all components, for example, evaluation in the routine test occurs alternately so that judge 1 evaluates component 1, the judge 2 evaluates component 2, repeating this alternation until judge 15.

According to the old rules³, five judges evaluated figure tests and 10 judges evaluated FR and TR tests. Scores were attributed to each component by each judge. This format resulted in many items to be considered by judges before attributing the final score. Therefore, in the last quadrennium (2013-2017) and in the current one (2017-2021), FINA believed that it was easier to increase the number of judges and hand out some components, as well as isolating the difficulty in the FR.

Consequently, there are fewer components to be assessed so that each judge can focus on the specificity of each. This is very important, considering that there are 100 units between scores 0 and 10 (0.1, 0.2, ... 9.9, 10) to differentiate a complete failure from a perfect performance and although scores 4.8 and 5.2 differ by four units, they are in different categories - Deficient and Satisfactory, respectively. The explanatory scoring scale contained in the AS manual² was constructed to guide the judgment, thus serving as basis for the judges' assessment. Therefore, the challenge is to train judges and to promote constant updates.

Even though judges are part of a population that has been neglected in studies on the development of their knowledge, coaches and athletes depend on scores given by judges to have feedback on the movements of their routines. Although the physical demands of AS judges during the trial are minimal, perceptual and cognitive demands are very high.

One of the stages of the learning process in the formation of an AS judge is visual training, one of the most discussed topics in study groups, courses and meeting in competitions. The preparation of judges is carried out through studies, training, and opinion evaluation (national and

international), especially when there are courses with judges of various nationalities and with different experience levels. This exchange of information is necessary in order to pursue desired homogeneity among judges.

They are trained to observe performance and apply specific judgment criteria for assigning scores. Although often referred to as subjective, these judgments are intersubjective, as they do not depend on the purely peculiar perspective of an individual judge, but on the possibility of a consensus by a group of trained individuals⁴. An efficient visual system on the part of judges is essential for them to judge athletes' movements in an objective and reliable way⁵, especially with regard to sports with artistic evaluations.

Furthermore, it is through the process of practice, repetition, training and experience that individuals improve their ability to process information and perform tasks until reaching high proficiency levels. In this context, there are several studies that have investigated the development of sports specialists, most of them in the visual search field^{6,7}. Based on literature, it could be suggested that the behavior of looking at AS judges translates into perceptual and cognitive capacities, which demand from memory a search for the image of the ideal movement for comparison with the observed movement. This process involves short and long-term memory, divided and selective attention and detection and identification of complex movement patterns to identify the degrees of execution difficulty⁸, which can become a skill / expertise in the judgment of a specific sport.

In this sense, it is worth mentioning that some studies that analyzed the visual search of judges in artistic gymnastics have indicated that skilled ones fix their gaze on different relevant areas of the athlete's body, who performs the movements and detect errors more quickly compared to novice judges⁸⁻¹¹.

Attributing scores in artistic gymnastics is very similar to AS, studies on the visual search strategies of AS judges can provide important information in order to facilitate the pedagogical guidelines for directing attention and to serve as a parameter for the standardization of positioning judges and assessment equipment. An efficient visual system on the part of judges is essential to enable them to judge athletes' movements in an objective and reliable way. Improving internal consistency among judges is a constant goal and many aspects can influence it.

According to the above, the hypotheses of the present study were: 1- international judges would have high and positive consistency and internal agreement; 2- national judges would not have consistency and internal agreement; and 3- there would be no positive correlation in scores attributed between groups due to the different experience times. The objective was to verify if there is significant difference between the two groups of AS judges with different experience levels.

METHODS

Participants

After approval by the Ethics and Research Committee of USJT (1.266.821) and signing the Consent Form, subjects voluntarily participated in this study.

Three AS athletes aged 17-18 years (8 years of experience in the sport), 10 international judges listed on the FINA committee (IG) with at least 25 years of experience and 10 judges participating only in national competitions unlisted on the FINA committee (NG) with at least 5 years of experience.

Video Shooting

Athletes watched the video of a technical team routine with the following required elements: 1st) Starting in a submerged back pike position with legs in vertical position, 301 – a Barracuda is executed, 2nd) 435 – a Nova is executed to the bent knee Surface Arch Position. A 360° rotation is executed as legs are lifted to a vertical position followed by continuous 720° spin (2 rotations), 3rd) Starting in a front pike position, legs are lifted to a vertical position. A full twist is executed; legs are lowered to a split position. A walkout front is executed, 4th) Starting in a submerged back pike position with legs in vertical position, 308 – a Barracuda Airborne Split is executed and 5th) Travelling ballet leg sequence. Starting in a back-layout position travelling headfirst, a Ballet Leg is executed, the horizontal leg bends to a Flamingo Position and is then lifted to a Ballet Leg Double Position². Two days were allocated for athletes to become familiar with TR elements, none of the athletes watched the execution of the other. Before attempts, each athlete watched the TR video to find out what the technical elements would be without any instructions from the researcher. After this familiarization period, athletes performed four attempts on a single day, all attempts were filmed with an iPad installed on a tripod with a retina display with resolution of 2048 x 1536 pixels, 326 pixels per inch.

Procedure

Among the four filmed attempts, only one attempt by each athlete was selected so that judges could evaluate the execution of technical elements. Videos were distributed in a randomized, double-blind manner and sent by email to each of the judges. Each judge evaluated the three videos, which was considered the first evaluation (test / T1) and, 7 days after the first evaluation, the second evaluation (retest / T2) was performed. All procedures for standardization and evaluation week were sent to each of the judges so that they could evaluate test and retest in the same week.

Statistical analyses

The average score of all judges, for each element of each athlete, was used in the test to compare the average scores of judges (test and retest, by group and as a whole). All values were expressed as average \pm standard deviation. Internal consistency of each evaluator was expressed for each element and for each athlete at T1 and one week later at T2. Agreement between evaluators was expressed for each element and for each athlete at T1 and T2. Concordance between IG vs NG was expressed for each element and for each athlete at T1 and T2. Cronbach's alpha¹² was used assuming 0.70 as the lower limit¹³ and the Bland-Altman¹⁴ scatter to analyze the degree

of agreement of assigned scores as well as among judges. All analyses were performed with the SPSS software (v 15.0, IBM, Armonk, NY, USA).

RESULTS

There was no significant difference in the total score between IG (6.89 ± 0.28) and NG judges (6.84 ± 0.22), Figure 1.

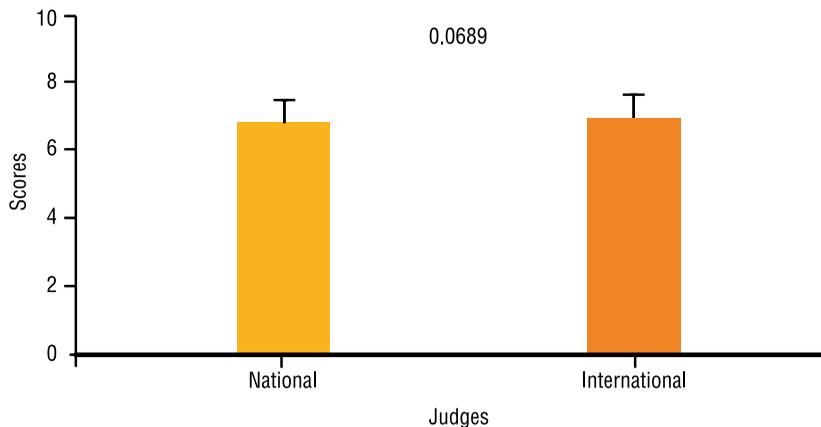


Figure 1. Score attributed by both groups of judges. Values expressed as mean \pm standard error.

Cronbach's alpha coefficient indicated high internal consistency at T1, value of 0.82 and low consistency at T2, value of 0.39 for NG judges. In IG, both moments indicated high internal consistency, T1 – 0.85 and T2 – 0.83 shown in Tables 1 and 2.

Table 1. Pearson correlation between T1 and T2 NG judges on 5 required elements.

T1	Judge 1	Judge 2	Judge 3	Judge 4	Judge 5	Judge 6	Judge 7	Judge 8	Judge 9	Judge 10	a
Judge 1	1.000										
Judge 2	.434	1.000									
Judge 3	.671**	.823**	1.000								
Judge 4	.256	.040	.056	1.000							
Judge 5	.459	.676**	.415	.239	1.000						
Judge 6	.581*	.766**	.744	-.001	.652**	1.000					
Judge 7	.434	1.000*	.823**	.040	.676**	.766**	1.000				
Judge 8	.438	.213	.239	-.215	.138	.426	.213	1.000			
Judge 9	.421	.108	.270	-.186	.051	.398	.108	.879**	1.000		
Judge 10	.076	.650**	.358	-.169	.649**	.460	.650**	.268	.133	1.000	.828
T2	Judge 1	Judge 2	Judge 3	Judge 4	Judge 5	Judge 6	Judge 7	Judge 8	Judge 9	Judge 10	a
Judge 1	1.000										
Judge 2	.421	1.000									
Judge 3	.048	.719**	1.000								
Judge 4	-.149	.284	.600**	1.000							
Judge 5	-.022	.182	.239	.171	1.000						
Judge 6	.013	.179	.253	.258	-.550*	1.000					
Judge 7	-.014	.524*	.621**	.424	.710**	-.370	1.000				
Judge 8	-.157	-.760**	-.788**	-.230	-.171	.005	-.525	1.000			
Judge 9	.292	-.341	-.336	-.218	.388	-.218	-.165	.346	1.000		
Judge 10	.128	.235	.420	.174	.472	.051	.495	-.146	.415	1.000	.399

Note. * $p < 0.05$; ** $p < 0.01$

Table 2. Pearson correlation between T1 and T2 IG judges on 5 required elements.

T1	Judge 1	Judge 2	Judge 3	Judge 4	Judge 5	Judge 6	Judge 7	Judge 8	Judge 9	Judge 10	a
Judge 1	.1000										
Judge 2	.718**										
Judge 3	.661**	.507									
Judge 4	.597*	.565*	.138								
Judge 5	.584*	.556*	.258	.255							
Judge 6	.498	.693**	.494*	.178	.725**						
Judge 7	-.665**	-.599*	-.078	-.801**	-.201	-.132					
Judge 8	.800**	.642**	.326	.875**	.366	.299	.798**				
Judge 9	.862**	.877**	.515**	.677**	.665**	.694**	.707**	.753**			
Judge 10	.703**	.627**	.455*	.628*	.389	.154	-.687**	.733**	.717**	1.000	.856
T2	Judge 1	Judge 2	Judge 3	Judge 4	Judge 5	Judge 6	Judge 7	Judge 8	Judge 9	Judge 10	a
Judge 1	1.000										
Judge 2	.539*										
Judge 3	.508	.704**									
Judge 4	-.205	-.168	-.576*								
Judge 5	.391	.518*	.244	.247							
Judge 6	.674**	.853**	.710**	-.150	.579*						
Judge 7	.262	.229	.174	.260	.439	.577*					
Judge 8	-.276	.011	-.239	.655**	-.228	-.066	.182				
Judge 9	.680**	.818**	.566*	-.240	.371	.779**	.202	-.088			
Judge 10	.490	.511	.496	-.032	.411	.705**	.763**	.113	.481	1.000	.831

Note. *p< 0.05; **p< 0.01

Regarding evaluation analysis between IG and NG at moments T1 and T2, they are equivalent, revealing significant correlation ($r^2: 0.01516$ $p>0.0330$) and confirmed after Bland-Altman reliability analysis (bias: -0.1266 of limit agreement 95%: -1.642 to 1.388) according to Figure 2 A and B.

DISCUSSION

Initial visual search studies were carried out in laboratory environment using pro-

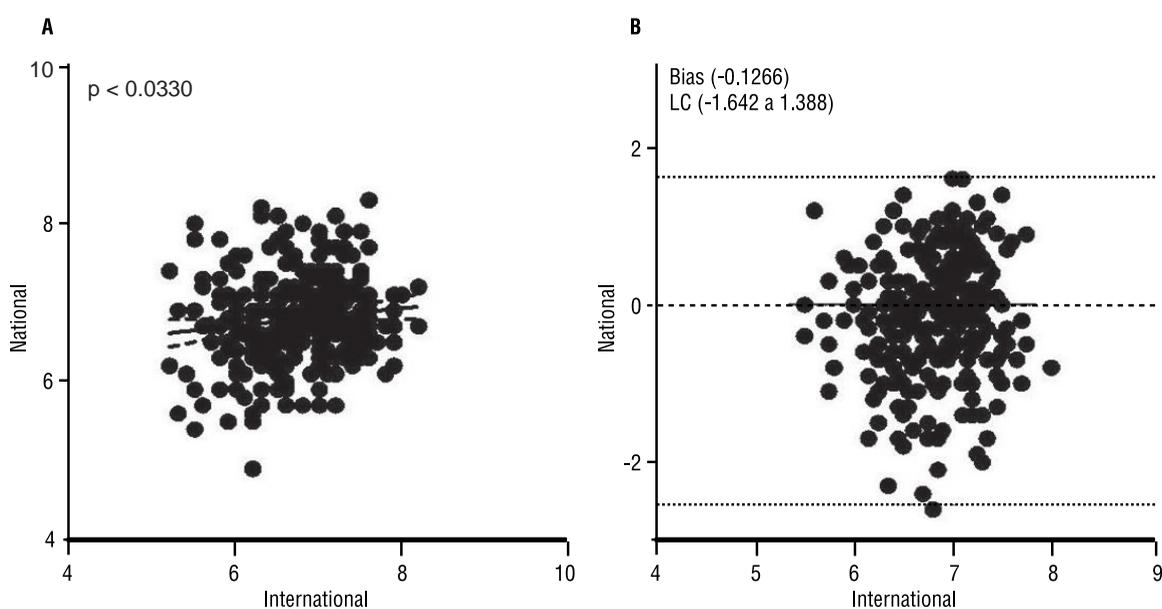


Figure 2. Linear Pearson correlation (Panel A) and reliability analysis (Panel B) of scores attributed between groups, IG vs NG, at moments T1 and T2.

jected videos that represented the sports scene⁶. With technological advancement, many authors started to explore scenarios in real or close to real situations^{15,16}.

These studies brought explanations about the types of visual strategies used by skilled and non-skilled judges in the dynamics of performing tasks during situations similar to competition¹⁶.

The motor performances of skilled judges are generally more proficient due to structured and systematic patterns of visual search instead of random, meaningless visual strategies¹⁷. These knowledge structures direct the visual search strategy of skilled judges to more important areas of the scene based on past experience and contextual information. The visual search seems to be controlled by this knowledge, which has been developed over years of observation, training and competition⁶.

A recent systematic review¹⁸ carried out from 2006 to 2016 in national and international journals described that most AS studies have investigated the physiological mechanisms of athletes, especially in routine tests. This review covered only studies whose variables of interest were physiological responses. Therefore, there is lack of studies on the role of AS judges using the visual search in actions of those who evaluate movements in this sport. It is worth mentioning that there are investigations of visual perception in judges of other modalities such as football and gymnastics^{4,7,11,19-24}.

A study⁴ with rhythmic gymnastics judges investigated the gaze behavior of 30 judges with different experience levels: international, national and beginner. The objective was to investigate whether judges made visual fixation of errors in an efficient manner to assist in the decision-making process regarding the performance of gymnasts. The study considered the fixation of judges' gaze on specific errors and, to differentiate between location of the eye fixation and location of the error, the gymnast's hands on the apparatus and the location by a single spatial area, the right-left corner, were considered of the video projection screen. The error analysis showed that international level judges are more efficient at detecting errors (40%) compared to the other groups (23%). The capture of visual fixation at a specific location in the process can be explained by the lack of experience of novice judges in the evaluation task, resulting in an excessive amount of time to visually process certain errors. This analysis showed that novice judges, although spending long time correcting detected errors, were not efficient in using fixation to detect them; however, national-level judges were also efficient in using visual fixation to detect errors. Although it was expected that judges with international experience would make more efficient use of visual fixation to detect errors, the findings do not allow differentiating national from international judges. However, the findings suggest a different strategy used by international judges to detect errors, which does not depend on visual fixation, but on more complex cognitive strategies, based on extensive experience and larger knowledge base. Such strategies that may not be based on specific visual perception mechanisms can help them to detect greater number of errors in general. Perceptual anticipation is one of these strategies that have been reported. It is likely that the anticipation of a next gymnastic element is based on advanced

visual cues that experienced judges can identify in advance when compared to novice judges in each sequence of movements in gymnastics^{10,11}.

Another study⁷ with gymnastics judges investigated visual search strategies and use of tips in artistic gymnastics judges, who were instructed to evaluate a film by two national gymnasts performing two mandatory barbell routines and two optional solo routines. Differences were evident in the distribution of fixations for the screen areas : experienced judges fixed more on gymnasts' upper body (head and upper limbs), while novices concentrated on lower limbs, and may have been influenced by the fact that, at the beginning of their careers and consequently being able to participate only in less important competitions or in smaller categories, they had experience only in judging gymnasts of lower technical level, which generally perform more errors in the placement of feet and lower limbs than experienced gymnasts. In addition, novice judges detected only half the number of errors when compared to skilled judges.

Studies with gymnastics judges helped in our pioneering study in relation to AS judges, being our research object¹, which was to verify the evaluation reliability and internal consistency among 10 international-level judges through video. Two tests were performed, T1 test and T2 retest, adopting a seven-day pause between one test and another. The study assumed 0.70 as the lower limit of Cronbach's alpha coefficient¹² and results indicated correlation values of 0.85 for T1 and 0.83 for T2, which means high reliability between judges. Regarding the limit of agreement between scores attributed at T1 and T2, Pearson's correlation coefficient and the Bland-Altman technique¹⁴ showed that dispersion diagrams indicated mean differences between T1 and T2 close to zero, with minimal confidence intervals. The average value of bias and limits of agreement was 95%. Thus, based on these findings¹, it could be safely inferred that the video analysis for the assessment of AS judges is reliable.

The analyses of the present study did not confirm two of the hypotheses raised. Initially, there was no significant difference in the total score between groups. The coefficient analysis showed high consistency in both groups at two times T1 and T2. Only in T2, the value of 0.39 for NG judges was low. However, the analysis of scores between IG vs NG groups at moments T1 and T2 were equivalent, revealing significant correlation.

The results of the present paper and study¹ may suggest that changes in AS rules in the last eight years helped making evaluations less subjective in the component execution of TR elements. The fact that the FINA committee divided the number of components to be evaluated and separated the execution difficulty may have been an important factor in the internal consistency in the evaluations of judges with different experiences.

As a result of studies on visual search presented with gymnastics judges and results with AS judges, another way of preparation for these individuals is associated with the relationship of the visual search with the evaluative actions of assigning scores and preparing dissertation feedbacks^{10,11}. Through the study of the gaze behavior of this population, it is possible to produce increasingly relevant knowledge to improve the performance of judges

in the evaluation of AS athletes. Knowledge about more adequate visual search rates, time percentages when gaze is directed to important areas of the scene and variations in pupil diameters for deductions of cognitive effort are examples of variables that can be further studied in this line of investigation. Such specific knowledge could encourage instructional programs in the training of AS judges, which is expected in future investigations.

CONCLUSION

AS judges lack investigations on visual behavior and its relationship with the assessment of evidence. Our group is exploring new possibilities for understanding visual behavior in different tests. Thus, it is possible to consider that progress in other visual training techniques is important, as it can improve efficiency in evaluations and in an attempt to minimize subjectivity in this sport in relation to some components. Finally, we believe that visual search studies by AS judges can further contribute to the evolution of the sport.

Acknowledgments

The authors thank FAPES (Espírito Santo Research and Innovation Support Foundation) (590/19 – No. 84417625/2018). Fund providers had no role decision on the paper.

COMPLIANCE WITH ETHICAL STANDARDS

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or non-profit sectors.

Ethical approval

Ethical approval was obtained from the local Human Research Ethics Committee of São Judas Tadeu University, process No. 1.266.821. The protocol was written according to standards established by the Declaration of Helsinki.

Conflict of interest statement

The authors have no conflict of interests to declare.

Author Contributions

Conceived and designed experiments: KP, DSB, CMMJ. Performed experiments: KP, MF, ACG. Analyzed data: KP, CBS, DSB, RLR, ACG. Contributed with reagents/materials/analysis tools: CBS, DSB, CMMJ. Wrote the paper: KP, ACG, DSB, RLR.

REFERENCES

1. Ponciano KR, Fugita M, Figueira Jr AJ, Silva CB, Meira Jr CM, Bocalini DS. Reliability of judge's evaluation of the synchronized swimming technical elements by video. *Rev Bras Med Esporte* 2018;24(3):182-185.

2. Fédération Internationale De Natation. Manual for synchronised swimming Judges, coaches & referees. FINA Office. Lausanne: Suíça; 2017.
3. Fédération Internationale De Natation. Manual for synchronised swimming Judges, coaches & referees. FINA Office. Lausanne: Suíça; 2009.
4. Flessas K, Mylonas D, Panagiotaropoulou G, Tsopani D, Korda A, Siettos C, et al. Judging the judges' performance in rhythmic gymnastics. *Med Sci Sports Exerc* 2015;47(3):640-648.
5. Annett, J. Subjective rating scales: science or art? *Ergonomics* 2002;45(14):966-987.
6. Williams AM, Davids K, Williams JGP. Visual percept action in sport. London: Routledge; 1999.
7. Bard C, Fleury M, Carrière L, Hallé M. Analysis of gymnastics judges' visual search. *Res Q Exerc Sport* 1980;51(2):267-273.
8. Pizzera A, Möller C, Plessner H. Gaze behavior of gymnastics judges: Where do experienced judges and gymnasts look while judging? *Res Q Exerc Sport* 2017;89(1):112-119.
9. Plessner H, Haar T. Sports performance judgments from a social cognitive perspective. *Psychol Sport Exerc* 2006;7(6):555-575.
10. Ste-Marie DM, Lee TD. Prior processing effects on gymnastic judging. *J Exp Psychol Learn Mem Cogn* 1991;17(1):126.
11. Ste-Marie DM. Expert–novice differences in gymnastic judging: an information-processing perspective. *Appl Cog Psychol* 1999;13(3):269-281.
12. Cronbach L. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;16(3):297-334.
13. Hopkins WG. Measures of reliability in sports medicine and Science. *Sports Med* 2000;30(1):1-15.
14. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;327(8476):307-310.
15. Appelbaum LG, Erickson G. Sports vision training: A review of the state-of-the-art in digital training techniques. *Int Rev Sport Exerc Psychol* 2018;11(1):160-189.
16. Kredel R, Vater C, Klostermann A, Hossner EJ. Eye-tracking technology and the dynamics of natural gaze behavior in sports: A systematic review of 40 years of research. *Front Psychol* 2017; 8:1845.
17. Mann DL, Causer J, Nakamoto H, Runswick OR. Visual search behaviours in expert perceptual judgements. In: Williams AM, Jackson RC. *Anticipation and decision making in sport*. New York: Routledge 2019. p. 59-78.
18. Ponciano K, Miranda MLDJ, Homma M, Miranda JMQ, Figueira Jr AJ, Meira Jr CM, et al. Physiological responses during the practice of synchronized swimming: a systematic review. *Clin Physiol Funct Imaging* 2017;38(2):163-175.
19. Gilis B, Weston M, Helsen WF, Junge A, Dvorak J. Interpretation and application of the laws of the game in football incidents leading to player injuries. *Int J Sport Psychol* 2006;37(2/3):121.
20. Helsen W, Gilis B, Weston M. Errors in judging “offside” in association football: Test of the optical error versus the perceptual flash-lag hypothesis. *J Sports Sci* 2006;24(05):521-528.
21. Ferreirinha J, Carvalho J. Tendencies and deviations of judging in gymnastics. *Rev Engym* 2012; 1:2-3.
22. Leandro C, Ávila-Carvalho L, Sierra-Palmeiro E, Bobo-Arce M. Judging in Rhythmic Gymnastics at different levels of performance. *J Hum Kinet* 2017;60(1):159-165.
23. Mascarenhas D, O'Hare D, Plessner H. The psychological and performance demands of association football refereeing. *Int J Sport Psychol* 2006;37(2/3):99.

Corresponding author

Kátia Ponciano
Rua Granjas, 342, Cotia, Chacara Eliana, SP.
Apartamento 08. Brazil.
Postal code: 06708-620
Email: kcapristi@gmail.com