

# RELIABILITY OF JUDGE'S EVALUATION OF THE SYNCHRONIZED SWIMMING TECHNICAL ELEMENTS BY VIDEO



ORIGINAL ARTICLE  
ARTIGO ORIGINAL  
ARTÍCULO ORIGINAL

CONFIABILIDADE DA AVALIAÇÃO DE JUÍZES POR VÍDEO DOS ELEMENTOS TÉCNICOS DE NADO SINCRONIZADO

CONFIABILIDAD DE LA EVALUACIÓN POR JUECES DEL VÍDEO DE LOS ELEMENTOS TÉCNICOS DE NADO SINCRONIZADO

Katia Regina Ponciano<sup>1</sup>  
(Physical Education Professional)  
Meico Fugita<sup>1</sup>  
(Physical Education Professional)  
Aylton José Figueira Júnior<sup>1</sup>  
(Physical Education Professional)  
Cláudia Borim da Silva<sup>1</sup>  
(Mathematician)  
Cássio de Miranda Meira Júnior<sup>2</sup>  
(Physical Education Professional)  
Danilo Sales Bocalini<sup>3</sup>  
(Physical Education Professional)

1. Universidade São Judas Tadeu,  
Department of Post-Graduation in  
Physical Education, São Paulo,  
SP, Brazil.

2. Universidade de São Paulo, Escola  
de Artes, Ciências e Humanidades.  
São Paulo, SP, Brazil.

3. Universidade Federal do Espírito  
Santo, Laboratório de Fisiologia e  
Bioquímica Experimental do Centro  
de Educação Física e Desporto,  
Vitoria, ES, Brazil.

## Correspondence:

Av. do Oratório 2635, apto. 101,  
Torre 1, Parque São Lucas,  
São Paulo, SP, Brazil. 03221-100.  
kcapristi@gmail.com

## ABSTRACT

**Introduction:** In acrobatic, rhythmic and expressive gymnastics, the goal is performance and the score is given by judges. In synchronized swimming the panel is composed of seven judges who assess figures and fifteen judges who assess technical and free routines. Accordingly, the aim of this study was to verify the reliability of this strategy - the evaluation of technical elements in a synchronized swimming routine via video. **Method:** The study included three synchronized swimming athletes aged 17 to 18 and ten level A and B judges on the FINA list with at least ten years of experience in national and international events. **Results:** Cronbach's Alpha coefficient was 0.85 for T1 (test) and 0.83 for T2 (retest), indicating high internal consistency above 0.70. As regards agreement between scores awarded at both T1 and T2, significant correlation ( $r: 0.530 p > 0.0005$ ) was found between them, confirmed after Bland-Altman reliability analysis (bias: 0.0553334 95% of limit of agreement -1.25043 to 1.36110). **Conclusion:** The results of this study suggest that video is a reliable tool for training synchronized swimming judges. **Level of Evidence II; Diagnostic studies - Investigating a diagnostic test.**

**Keywords:** Evaluation; Videos; View..

## RESUMO

**Introdução:** Na ginástica acrobática, rítmica e expressiva, o objetivo consiste no desempenho e a pontuação é atribuída pelos juízes. No nado sincronizado (NS), a banca examinadora é composta por sete juízes que avaliam as figuras e quinze juízes que avaliam rotinas técnica e livre. Sendo assim, o objetivo deste estudo consiste em verificar a confiabilidade dessa estratégia - a avaliação dos elementos técnicos em uma rotina de natação sincronizada via vídeo. **Método:** O estudo incluiu três atletas de NS com idade entre 17 e 18 anos e dez juízes de nível A e B listados no comitê da FINA com, no mínimo, dez anos de experiência em eventos nacionais e internacionais. **Resultados:** O coeficiente Alfa de Cronbach foi 0,85 para o T1 (teste) e 0,83 para o T2 (reteste) indicando uma alta consistência interna acima de 0,70. Em relação à concordância entre as pontuações atribuídas em ambos T1 e T2: encontrou-se uma correlação significativa ( $r: 0,530 p > 0,0005$ ) entre eles, confirmada após a análise de confiabilidade de Bland-Altman (viés: 0,0553334, limite de concordância de 95%: -1,25043 a 1,36110). **Conclusão:** Os resultados desse estudo sugerem que o vídeo é uma ferramenta confiável para o treinamento dos juízes de NS. **Nível de Evidência II; Estudos diagnósticos-Investigação de um exame para diagnóstico.**

**Descritores:** Avaliação; Vídeos; Visão..

## RESUMEN

**Introducción:** En la gimnasia acrobática, rítmica y expresiva, el objetivo consiste en el desempeño y la puntuación es dada por los jueces. En el nado sincronizado (NS) la banca examinadora está compuesta por siete jueces que evalúan las figuras e quince que evalúan las rutinas técnica e libre. Siendo así, el objetivo de este estudio consiste en verificar la confiabilidad de esa estrategia - la evaluación de los elementos técnicos en una rutina de natación sincronizada a través de vídeo. **Método:** El estudio incluyó tres atletas de NS con edad entre 17 y 18 años y diez jueces de nivel A y B listados en el comité de Fina con al menos diez años de experiencia en eventos nacionales e internacionales. **Resultados:** El coeficiente alfa de Cronbach fue de 0,85 para T1 (test) y 0,83 para T2 (retest) indicando una alta consistencia interna superior a 0,70. Con relación a la concordancia entre las puntuaciones atribuidas en T1 y T2: se encontró una correlación significativa ( $r: 0,530 p > 0,0005$ ) entre ellos, confirmada después del análisis de Bland-Altman (sesgo: 0,0553334; límite de concordancia de 95% -1,25043 a 1,36110). **Conclusión:** Los resultados de este estudio sugieren que el video es una herramienta confiable para el entrenamiento de los jueces de NS. **Nivel de Evidencia II; Estudios de diagnósticos-Investigación de un examen para diagnóstico.**

**Descriptorios:** Evaluación; Vídeos; Visión.



## INTRODUCTION

In acrobatic, rhythmic and expressive sports as in Artistic Gymnastics, Rhythmic Gymnastics, Diving and Synchronized Swimming, the goal is the performance itself and the score is given by judges.

In Synchronized Swimming seven judges award scores to figure competitions and fifteen (15) judges to competitions routines.<sup>1</sup>

The explanation score scale and the Synchronized Swimming Manual for Judges, Coaches and Referees were built to guide the judgment, they are the basis for the judge. As the sport evolves, the amount of specifications increases demanding from judges more knowledge and experience.<sup>1</sup> So, training judges and promoting updates to ensure their efficiency are great challenges.

Nowadays there are two types of routines, the technical routine (RT), with compulsory elements and the free routine (RL), with free content. Fifteen (15) judges sit on the pool side to judge the following components: (1) execution/synchronization; (2) artistic impression (RL) or general impression (RT); (3) difficulty (RL) or execution of the elements (RT).<sup>1</sup> Thus judge 1 evaluates component (1); judge 2 evaluates component (2); judge 3 evaluates component (3), repeating this alternation up to the 15<sup>th</sup> judge.

According to the old rules (2009-2013), five (5) judges would award scores for technical merit and other five (5) for artistic impression, resulting in large numbers of items to consider before awarding the final score. On the new rules (2013-2017) FINA increased the number of judges and distributed the contents of the technical merit in components (1) and (3). Therefore, there are less items to be evaluated in order to focus on the specificity of each component. This is very important, considering that there are 100 units between score 0 and 10 to differentiate between a perfect performance and a complete failure. For example, in the "Good" score categories, from 7.0 to 7.9, there is a range of 10 decimal intervals. It is important that each group of five judges uses the same evaluation criteria. Not only shouldn't the scores overpass 10 units difference but they also should be at the same category, regardless the competition level or location. That is, although the scores 4.8 and 5.2 have 5 units difference, they are in different categories - Deficient and Satisfactory, respectively. This means that the judges are probably using different criteria for the evaluation.

Attentive view to the rules and homogeneity among the judges are developed through study, training, national and international judging experience and information exchange. The visual training is extremely important for a synchronized swimming judge to organize and diversify his view, from the simplest to the most complex repertoire of actions, empowering the judge with greater objectivity and reliability.

FINA uses routine videos in judges training, aiming at exemplifying every aspect of judgment described in its manual. However, although this tool is widely used to homogenize the judge's view, it is still not clear in the literature how efficient it is. Therefore, the purpose of this study was to verify the reliability of this strategy - the evaluation of technical elements in a synchronized swimming routine via video.

## MATERIALS AND METHODS

After the approval of the ethics and research committee of the USJT (nº 1.266.821) and signing of the consent form and free and informed consent, they voluntarily participated in this study: three NS female athletes, aged between 17 and 18 years, 10 judges listed in FINA, level A and B committee with at least 10 years of experience in national and international events.

All three NS athletes were submitted to biometric evaluation and height measure in estadiometer Cardiomed (WCS model) with accuracy of 115/220 cm. Body weight was measured in Filizola electronic scale (Personal Line 150 model) with a resolution of 100g and a maximum capacity of 150 kg. The body mass index (BMI) was used under the following

equation: BMI: body mass/height.<sup>2</sup>The body composition was determined by density skinfold technique as envisaged on our group publication.<sup>2</sup>

All routines were continuously recorded using iPad Mini (Apple Inc., 1 Infinite Loop, Cupertino, CA 95014 USA) with retinal display by 2048 x 1536 pixels, with 326 pixels per inch. A synchronized swimming judge was responsible for shooting in order to ensure that the recording was consistent to a routine reality in a competitive situation. The videos were recorded with the judge's lateral movement, following the routine, ensuring that each movement performed by the athletes was visually adequate for the judges to assign their scores. In this way, it was assured that the judges (volunteers on this research) evaluated the same competitive reality.

All three NS athletes watched a technical routine with the required elements: 1<sup>st</sup>) Starting in a submerged back pike position with legs vertical, a barracuda is executed; 2<sup>nd</sup>) A nova is executed to the bent knee surface arch position. A rotation of 360° is executed as the legs are lifted to a vertical position followed by a continuous spin of 720° (2 rotations); 3<sup>rd</sup>) Starting in a front pike position, the legs are lifted to a vertical position. A full twist is executed; the legs are lowered to a split position. A walkout front is executed; 4<sup>th</sup>) Starting in a submerged back pike position with legs vertical, a barracuda airborne Split is executed and the last element 5<sup>th</sup>) Travelling ballet leg sequence. Starting in a back layout position travelling head first, a ballet leg is assumed, the horizontal leg bends to a flamingo position and is lifted to a ballet leg double position. All official elements according to the rules by FINA (2013).

All three NS athletes were submitted to four trials in order to process their knowledge of the five elements into the routine. After this period of knowledge all athletes were submitted to four trials to shoot the required elements 1 to 5 into de routine, according to the official rules.

Therefore, 3 videos will be used in this study, one of each athlete, to verify that the video analysis is a reliable evaluation strategy of technical elements on synchronized swimming routine. The videos were randomized to double-blind way and sent by email to the judges. Each of the judges reviewed the 3 videos at the first evaluation (test) and after 7 days, in the second evaluation (retest).

## Statistical analysis

The verification of intra-rater reliability was performed by test-retest method within an interval of one (1) week between applications. The inter-rater reliability grade refers to the degree of agreement among the 10 judges and among the analyzed elements. The agreement between the values of the scores was analyzed according to the following procedures: Cronbach's Alpha comparison between the judges' evaluation of all the participants and all elements in the test and retest. Admitting to 0.70 as the lowest limit for Cronbach's Alpha coefficient.<sup>3</sup>The dispersion was used by Bland-Altman to analyze the degree of agreement of the awarded scores. All values were expressed as mean, standard deviation ( $\pm$ SD). All analyzes were performed with SPSS software (v 15.0; IBM Armonk NY USA).<sup>4</sup>

## RESULTS

The physical characteristics of NS athletes are described in Table 1. The experience of athletes corresponded to  $8 \pm 0.0$  years practicing the sport.

**Table 1.** Physical characteristics of the athletes.

Parameters	Mean $\pm$ SD
Age (years)	17.67 $\pm$ 0.47
Weight (kg)	60.33 $\pm$ 4.78
Height (kg)	1.68 $\pm$ 0.06
BMI (kg/m <sup>2</sup> )	13.19 $\pm$ 2.98
Body fat (%)	21.63 $\pm$ 3.10

The Cronbach's Alpha coefficient was 0.85 for T1 (test) and 0.83 for T2 (retest) indicating high internal consistency once they are above 0.70. Tables 2 and 3 show the correlation between judges two to two.

Table 2 shows that the judge 3 presented indices below the critical (RC: 0514) compared to the judges 5 to 10. Note that the judge 7 correlates negatively with all the judges although some correlations are not significant.

Table 3 shows in T2 no correlation between the scores assigned by the judge 8 and all others (except judge 9). The same happens to judge 9 who is positively correlated to judge 8 and negatively to judge 3.

In relation to the agreement between the scores awarded at both time points T1 and T2: significant correlation was found ( $r: 0.530$   $p > 0.0005$ ) between the scores assigned between T1 and T2 (Figure 1) confirmed after Bland-Altman reliability analysis (bias: 0.0553334 95% limits of agreement: -1.25043 to 1.36110).

## DISCUSSION

Reliability means stability, predictability, consistency or lack of "error" in a set of measures, which may have variability and fluctuations in measurements. As an example, one can cite the following situation: if judge A gives the person 1 a high score and so do the judges B, C and D there is, in other words, a high concordance among the judges, the same is valid for the four judges when awarding lower grades.<sup>5</sup> To the extent that the classifications agree, ratings are reliable. Thus the use of agreement is a measure used between two or more sets of ratings.<sup>6</sup>

There are many studies that analyze and investigate reliable visual methods of measurement. Photogrammetry and photometry as reliable measurement have been frequently used both in sports and in rehabilitation.<sup>7-13</sup>

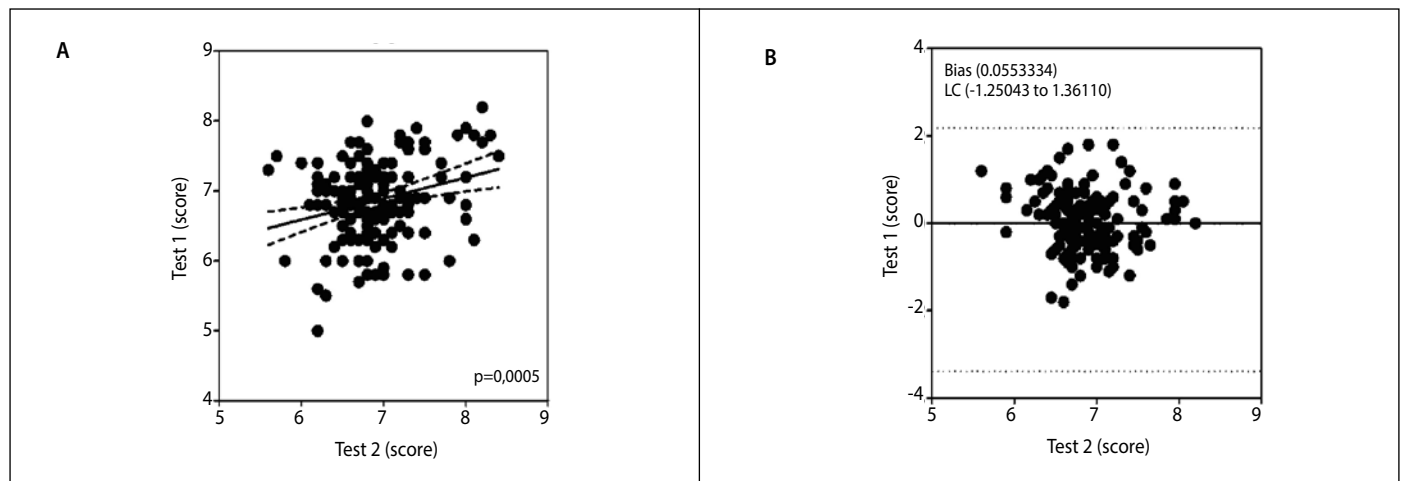
Some studies<sup>7,9-12</sup> used of photometry and photogrammetry to investigated the reliability on issues related to the improvement of

**Table 2.** Analysis of the scores assigned by the judges in the 5 elements analyzed, time T1.

T1	Judge 1	Judge 2	Judge 3	Judge 4	Judge 5	Judge 6	Judge 7	Judge 8	Judge 9	Judge 10	$\alpha$ de Cronbach
Judge 1	1.000										
Judge 2	.718										
Judge 3	.661	.507									
Judge 4	.862	.877	.623								
Judge 5	.584	.556	.445	.665							
Judge 6	.498	.693	.468	.694	.725						
Judge 7	-.665	-.599	-.403	-.707	-.201	-.132					
Judge 8	.800	.642	.485	.753	.366	.299	-.798				
Judge 9	.597	.565	.354	.677	.255	.178	-.801	.875			
Judge 10	.703	.627	.430	.717	.389	.154	-.687	.733	.628	1.000	*856

**Table 3.** Analysis of the scores assigned by the judges in the 5 elements analyzed, T2 time.

T2	Judge 1	Judge 2	Judge 3	Judge 4	Judge 5	Judge 6	Judge 7	Judge 8	Judge 9	Judge 10	$\alpha$ de Cronbach
Judge 1	1.000										
Judge 2	.539										
Judge 3	.508	.704									
Judge 4	.680	.818	.566								
Judge 5	.391	.518	.244	.371							
Judge 6	.674	.853	.710	.779	.579						
Judge 7	.262	.299	.174	.202	.439	.577					
Judge 8	-.276	.011	-.239	-.088	-.228	-.066	.182				
Judge 9	-.205	-.168	-.576	-.240	.247	-.150	.260	.655			
Judge 10	.490	.511	.496	.481	.411	.705	.763	.113	-.032	1.000	*831



**Figure 1.** Linear correlation (Panel A) and reliability analysis (Panel B) of the grades given between judges in the T1 and T2 times.

posture in different joints of the body in rehabilitation programs and prevention of sport injuries<sup>9,11,12</sup> and to improve athletic performance.<sup>8,13</sup>

Researches based on measure evaluation must be attentive to the reliability because it enables the reproducibility of a finding.<sup>14</sup> The intra-subject variation is a kind of important reliability measure because it precisely infers in estimation of changes and in tests often used among coaches and other professionals who monitor the athletes performance.<sup>15</sup> In addition, retesting correlations are excellent ways to verify reliability because if the correlation shows that the classification of the participants in a trial is replicated in the second trial and obtains value .1 this result has high significance.<sup>15</sup>

Videos of routines performed in competitions are used in the synchronized swimming FINA judges formation to standardize the criteria used by them. However, the reliability of this strategy is not clear in the literature. The physical demands of the judges during their judging are minimal, but the mental demand is intense and requires a lot of concentration. In the evaluation of technical elements of a given routine, the focus is only on the execution, with emphasis on technical efficiency. Differently from soccer judge who must follow the players from the beginning to the end of the game, running all the time. A wrong decision of a soccer judge may be associated with his positioning.<sup>16</sup> For synchronized swimming judges the wrong assignment of whole scores or entire scores tenths may be related to factors such as, for example, not visualizing one or more routine moves.

In this study, assuming 0.70 as the lower limit of the Cronbach's Alpha coefficient,<sup>3</sup> it was found 0.85 for T1 and 0.83 for T2, which means high reliability among the judges. As there wasn't variability between the given grades, the results of this study indicated high reliability.

In relation to the limit of agreement between the scores awarded in T1 and T2, the Pearson correlation coefficient and the Bland-Altman technique showed that scatter diagrams indicated average differences between T1 and T2 close to zero, with minimal confidence intervals.

The average value of bias and agreement limits was 95%. We can say that the video analysis can be considered reliable.

The reliability among judges has wide use in the academic field and the focus is on the evaluators decision consistency.<sup>6</sup> The results of this study showing consistency is a synonym of reliability and that high degree of confidence is associated to the stability of the observed parameter. Thus the evaluations and ratings made by 10 judges showed high stability and confidence level.

In short, the correlations between the scores of each element in technical routine assigned by the judges and the consistency between them in the present study, using the Alpha Cronbach, Pearson's correlation and graphic dispersion by Bland-Altman, legitimize the analysis through video.

## CONCLUSION

Although there are criticisms regarding the subjectivity of judgment the values found showed the opposite, that is, objectivity, because of the consistency intra and inter evaluators. Even though the video analysis situation allows longer analysis, not being necessary to award immediately after the execution, as on the actual situation of competition, one can say that it is an effective training strategy. First, of course, because it is reliable. The content of the video was interpreted almost the same way by the 10 evaluators and allowed evaluation consistency after 7 days. Further studies may occur with an increased number of videos and evaluators. In addition, specific items of analysis could be strengthened if differences in the assessments of certain actions or specific elements are noticed.

The results of this study suggest that the use of videos is a reliable tool for training the synchronized swimming judges. Thus, it is possible to consider that progress in other visual training techniques is important because it can further improve the objectivity and efficiency in the ratings. This is useful in training programs for judges.

---

All authors declare no potential conflict of interest related to this article.

---

---

**AUTHORS' CONTRIBUTIONS:** Each author made significant individual contributions to this manuscript. KRP (0000-0002-7274-9606)\*, DSB (0000-0003-3993-8277)\* and CBS (0000-0002-3234-7581)\*: were the main contributors to the writing of the manuscript; MF (0000-0001-6301-6868)\*: evaluated the data from the statistical analysis; AFJ (0000-0001-8069-2366)\* and CMMJ (0000-0002-9775-935X)\*: conducted the literature review and revision of the English text. All authors contributed to the intellectual concept of the study and approved the final version of the manuscript. \*ORCID (Open Researcher and Contributor ID).

---

## REFERENCES

1. FINA: Synchronized swimming manual for judges, coaches & referees; 2013-2017. Suíça; 2013.
2. Serra AJ, Amaral AM, Rica RL, Barbieri NP, Reis Júnior D, Silva Júnior JA, et al. Determinação da densidade corporal por equações generalizadas: facilidade e simplificação no método. *ConScientiae Saúde*. 2009;(8):19-24.
3. Hair Jr JF, Black WC, Babin JB, Anderson RE, Tatham RL. Análise multivariada de dados. 6.ed. Porto Alegre: Bookman; 2009.
4. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;327(8476):307-10.
5. Kerlinger FN. Metodologia da pesquisa em Ciências Sociais: um tratamento conceitual. São Paulo (SP): Ed. Pedagógica e Universitária; 2003.
6. Matos DA. Confiabilidade e concordância entre juízes: aplicações na área educacional. *Est. Aval. Educ*. 2015;25(59):298-324.
7. Rocha EA, Baroni MP, Pereira AL, Assis SJ, Dantas DS. Confiabilidade inter e intraexaminador da fotogrametria computadorizada por meio do software AutoCAD® R12. *ConScientiae Saúde*. 2015;14(4):617-26.
8. Guerreiro RC, César EP, Périllier R, Assis CA, Santos TM. Confiabilidade de fotogrametria na medida do deslocamento vertical da alçada de egg no nado sincronizado. *R Bras Ci e Mov*. 2013;21(3):80-7.
9. Luna NM, Nogueira GB, Saccol MF, Leme L, Garcia MC, Cohen M. Amplitude de movimento rotacional glenoumeral por fotogrametria computadorizada em atletas de seleção brasileira de handebol masculino. *Fisioter Mov*. 2009;22(4):527-35.
10. Sacco IC, Alibert S, Queiroz BW, Pripas D, Kieling I, Kimura AA, et al. Confiabilidade da fotogrametria em relação a goniometria para avaliação postural de membros inferiores. *Rev Bras Fisiot*. 2007;11(5):411-7.
11. Cardoso JR, Boer MC, Oliveira BI, Kawano MM, Carregaro RL. Confiabilidade intra e interobservador da mensuração do ângulo de flexão anterior do tronco pelo método de Whistance. *Fisioter. Pesq*. 2007;14(3):44-9.
12. Sato TO, Vieira ER, Gil Coury HJ. Análise da confiabilidade de técnicas fotométricas para medir a flexão anterior do tronco. *Rev Bras Fisiot*. 2003;7(1):53-99.
13. Perin A, Ulbricht L, Ricieri DV, Neves EB. Use of biophotogrammetry for assessment of trunk flexibility. *Rev Bra Med Esporte*. 2012;18(3):176-80.
14. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951;16(3):297-334.
15. Hopkins WG. Measures of reliability in sports medicine and science. *Sports Med*. 2000;30(1):1-15.
16. Silva AI, Oliveira MC. Fatores que podem interferir na tomada de decisão do árbitro de futebol. *RBPFE*. 2012;6(32):113-27.