JUDGES' EVALUATION OF ROUTINES IN MEN ARTISTIC GYMNASTICS

George Dallas [1], Paschalis Kirialanis[2]

[1]Kapostriakon University of Athens, Department of Physical Education and Sport Science, Greece

[2]Dimokrition University of Thrake, Department of Physical Education and Sport Science, Greece

Original research article

Abstract

For competition judging, the practice of assigning gymnastics judges into one of two groups (D-Jury and E-Jury) is internationally accepted. International judges (the highest level) are placed in the D-Jury and national judges are allocated to the E-Jury. Performance evaluations are the jurisdiction of the E-Judges who record the deductions in the exercises, determining the exercise final score. The purpose of this study was to examine if there were significant differences between D-Jury and E-Jury judges (international vs. national), based on their evaluations of gymnastics performances; allowing for an assessment of the necessity to split judges into these two groups. Twenty experienced judges, who volunteered to participate in the study, were divided into two groups (National vs International). The judges evaluated, via videotape, nine gymnastics routines performed on the rings. Points were deducted (in tenths of units) based on the severity of errors in the routines. According to the results, for the judges level effect the results approached significance and significant differences were found across the 9 separate programs. The observed differences raise questions concerning the existing placement system of judges (international vs. national) in Greece.

Keywords: artistic gymnastic, judges, evaluation, level of judging.

INTRODUCTION

In various events in artistic gymnastics (floor exercises, side horse, rings, etc.), it is at the level of the judges' knowledge and experience that a "winner" is decided. For that decision to be made, the judges are engaged in an extensive process related to information concerning the movement patterns observed (Ste-Marie, 1999). For this reason, they record the difficulty values of the elements that are performed (according to the Code of Points that is valid for every Olympic cycle), the connections of these elements (D-Jury) and the technical aspects of these elements (performance, composition) (FIG, 2009). In international competitions, all members of the Juries (D- and E-Panels, Assistants and Secretaries) must possess exact, applicable and thorough knowledge of the F.I.G. Code of Points for men and the F.I.G. rules for judges. They must have successfully participated in an international or intercontinental judges course and possess the corresponding FIG category . Prior to the competition, they participate in the Judges' Review Session (instruction) and the final draw of the judges to their functions.

Literature in this area states that the cognitive and perceptual differences that exist between expert and novice athletes can also be applied to judges, because they can also be classed as "performers", since they evaluate gymnasts' performances (Abernethy and Russell, 1984; Allard,

Graham, and Paarsalu, 1980; Allard and Starkes, 1980; Bard and Fleury, 1981). It can be stated that perceptual differences are related to the elements of the display that are selectively attended to (Allard and Starkes, 1980), the way and the speed at which the visual display is searched (Bard and Fleury, 1981) and how quickly the important information is extracted from the visual display prior to movement (Abernethy and Russell, 1984). On the contrary, the cognitive differences between expert and novice athletes refers to the interpretation and organisation of the skill-related information in memory, so as to facilitate superior recall of that knowledge (Allard, Graham, and Paarsalu, 1980).

Previous studies state that expert judges (more than 10 years experience) are superior to novice judges (up to 3 years experience) because they are more effective at interpreting biomechanical information available from the gymnast's body (Abernethy, 1997), they have greater breadth and depth of knowledge (Ste-Marie, 1999) and they can focus on different areas of the body better than novice judges (Bard et al, 1980). In addition, expert judges are more accurate when recognising form errors (correct body positions) than novice judges (Ste-Marie and Lee, 1991). This is because they are more able to predict what elements follow up during performance of one or more combinations of elements (Ste-Marie and Lee, 1991) and can better adhere to the speed of performances in various apparatus (Salmela, 1978).

A gymnasts' final score is calculated as follows: D-Score (from the D-Jury) + E-Score (from the E-Jury) = final score for each apparatus.

The D-Score is concerned with difficulty, element groups and connection values, while the E-Score is concerned with execution and composition. The E-Score is calculated by averaging the middle two of four (or four of six) scores (deductions).

Internationally and nationally, the level of the athlete's performance is evaluated by the judges and there is a common agreement about the final score that the gymnast receives. However, it is often unclear whether the final sum of deductions comes from the same number and kind of faults that receive deductions (small, medium, large, very large).

In Greece, judges are divided into three categories (novice, national and international). International judges have successfully participated in an international or intercontinental judges course. National and novice judges have only participated in national judges' courses. For them, the results of the examination of these courses serve as the main criteria for further categorisation (i.e. from novice to national, from national to international). However, it is the opinion of specialists that experience is of greater value than judging courses.

Although there are no differences in the total number of deductions (sum of deductions) that judges give whilst evaluating athletes' routines, it is unclear whether the sum of deductions comes from the same number of faults or the same technical error. This is even more evident in routines of lower technical level than in routines executed by elite athletes. It is therefore questionable whether differences in scores between experienced judges result from the judges' different category (national, international); whether differences in the final score result from the same technical faults in the same elements; or if they have come from different elements. It is possible that result accuracy would improve in national competitions if international level judges also judged in the E-Jury, allowing for more accurate and objective evaluations. The purpose of this study was to examine if there are significant differences between national and international judges in: a) the total amount of deductions in all the routines performed, b) the total amount of deductions in each routine, c) the deductions for every element separately and d) the deductions between competition and video evaluation.

METHODS

Participants

Twenty experienced national and international judges from the Hellenic Federation Gymnastics volunteered to participate in the study. They were divided into two groups: a) international judges (n=8) with 14.47 ± 4.35 years of judging experience and had judged 80.43 ± 28.43 competitions and b) National judges (n=12) with 6.25 ± 1.55 years of judging experience and had judged 18.50 ± 6.54 competitions. The differences for these two parameters (years of judging and number of competitions) were statistically significant ($p < .05$).

Instruments

Competition routines were recorded using a video camera (JVC GR-Ax2) during an international meeting of artistic gymnastics. The video camera was placed so that the optical axis of the camera was perpendicular with the transverse axis of the performance of the routines on the rings. The distance of the camera from the nearest ring was 3.00 ± 0.20m and the camera's height from the floor was 1.00 ± 0.12m. This placement of the camera is identical to the corresponding position of judges (E-jury) that evaluate the technical execution according to the Code of Points.

Procedure

To evaluate the gymnastics routine, the judges watched the routines via a video link on a monitor. Judges sat one meter from the monitor. Judges independently evaluated the same nine rings routines; each routine contained ten elements resulting in a total of ninety elements. The sum of deductions of every element that was performed was the total score of these deductions in every routine. After the end of each performance element, a black screen appeared for 5 seconds on the monitor, allowing the judges enough time to record the deductions on a record sheet and to prepare for the next performance.

Two expert international judges also evaluated all routines to provide a more objective evaluation and reference point (gold standard) for comparison. The evaluated routines in the preset study represented a broad range of technical gymnastics abilities, thus providing routines with many errors, as well as routines with few errors. The dependent variable, which was the score of each gymnast in the nine routines—as well in each routine separately—was used for statistical analysis (Student's t-test). Statistical significance was set at the 0.05 level.

RESULTS

The scores of the National and International judges across the 9 separate programs and 10 separate exercises are presented in tables 1 and 2 respectively.

Multivariate Analysis of Variance (MANOVA) was used to examine the differences between National and International judges in the deductions across the separate exercises. The multivariate and univariate post hoc results were not significant ($\Lambda= .926$, $F= 1.349$, $p= .208$, $n^2= .074$), indicating that the two groups of national and international judges were not significantly different when evaluating the deductions across the 10 separate exercises. The overall univariate post hoc findings are presented in table 3.

We examined the interaction between judge's level (International vs National) and programs (9 separate programs), with respect to the judges' evaluation score. The interaction effect of the 2 X 9 independent groups ANOVA was not significant ($F= .588$, $p=.786$, $\eta^2=.028$). Accordingly, we examined the main effects for judge's level and programs. For the judge's level effect, the results approached significance ($F= 3.881$, $p= .051$, $\eta^2= .023$) and significant differences were found across the 9 separate programs ($F= 11.633$, $p= .000$, $\eta^2= .365$). The post hoc LSD test was used to detect the sources of significance across the 9 separate programs. The overall findings are presented in table 4.

Table 1. *Mean scores (deductions) of judges, across separate programs.*

Variable	M	SD	N
Program 1			
National	178.33	57.18	12
International	**183.75**	**65.88**	**8**
Program 2			
National	266.67	72.15	12
International	237.50	69.02	8
Program 3			
National	127.50	61.07	12
International	118.75	64.01	8
Program 4			
National	166.67	78.66	12
International	143.75	79.45	8
Program 5			
National	189.17	62.88	12
International	131.25	57.18	8
Program 6			
National	135.00	48.71	12
International	108.75	24.74	8
Program 7			
National	156.67	41.63	12
International	140.00	27.25	8
Program 8			
National	92.50	41.14	12
International	102.50	62.28	8
Program 9			
National	132.50	45.15	12
International	122.50	46.83	8

Table 2. *Mean scores (deductions) of judges, across separate exercises.*

Variable	M	SD	N
Exercise 1			
National	1.62	1.44	108
International	1.36	1.20	72
Exercise 2			
National	2.07	1.14	108
International	1.90	1.22	72
Exercise 3			
National	1.48	1.27	108
International	1.55	1.34	72
Exercise 4			
National	2.75	1.98	108
International	2.29	1.75	72
Exercise 5			
National	1.38	1.16	108
International	1.33	1.09	72
Exercise 6			
National	1.44	1.31	108
International	1.14	1.06	72
Exercise 7			
National	1.64	1.14	108
International	1.33	0.98	72
Exercise 8			
National	1.15	1.12	108
International	1.35	1.43	72
Exercise 9			
National	0.68	0.99	108
International	0.79	1.10	72
Exercise 10			
National	1.79	1.97	108
International	1.26	1.92	72

Table 3. *ANOVA table, examining the differences between international vs national level judges, across the 10 separate exercises.*

Effect	SS	df	MS	F	p	n^2
Exercice 1						
BG	2.904	1	2.904	1.595	.208	.009
WG	324.046	178	1.820			
Exercice 2						
BG	1.268	1	1.268	.918	.339	.005
WG	245.727	178	1.380			
Exercice 3						
BG	.237	1	.237	.140	.708	.001
WG	300.741	178	1.690			
Exercise 4						
BG	9.445	1	9.445	2.649	.105	.015
WG	634.616	178	3.565			
Exercise 5						
BG	.093	1	.093	.072	.788	.000
WG	227.435	178	1.278			
Exercise 6						
BG	4.033	1	4.033	2.727	.100	.015
WG	263.278	178	1.479			
Exercise 7						
BG	4.033	1	4.033	3.470	.064	.019
WG	206.917	178	1.162			
Exercise 8						
BG	1.712	1	1.712	1.081	.300	.006
WG	281.949	178	1.584			
Exercise 9						
BG	.490	1	.490	.471	.493	.003
WG	185.171	178	1.040			
Exercise 10						
BG	12.245	1	12.245	3.208	.075	.018
WG	679.505	178	3.817			

BG. Between groups
WG. Within groups

Table 4. *Post hoc LSD test examining significance across the 9 separate programs.*

					Mean Difference			
program	2	3	4	5	6	7	8	9
1	-74.50*	56.50*	23.00	14.50	56.00*	30.50	84.00*	52.00*
2		131.00*	97.50*	89.00*	130.50*	105.00*	158.50*	126.50*
3			-33.50	-42.00*	-0.50	-26.00	27.50	-4.50
4				-8.50	33.00	7.50	61.00*	29.00
5					41.50*	16.00	69.50*	37.50*
6						-25.50	28.00	-4.00
7							53.50*	21.50
8								-32.00
9								

*: p < .05

DISCUSSION

It should be noted that the interaction effect between judges category and programs in the present study—with respect to the evaluation score—was not significant. Although international judges have considerably more years and competitions of judging experience, the years of judging experience of national judges provides sufficient knowledge to identify errors in gymnastics routines. However these results should be interpreted with caution, because the sum of deductions between these two categories of judges weren't from the same errors and the same severity (degree of error). This means that judges in these two categories may differ in declarative knowledge (Ste-Marie, 1999), meaning they "record" errors in a different way. It is clear that an attempt has been made internationally to minimise the subjectivity in the judging process and although judges aim to evaluate in an objective way, we should mention that the judging procedure is based on judges' own perceptions of what constitutes the 'perfect performance'.

Based on the findings from this study, judges evaluated different deductions for errors in separate programs. These findings are similar to findings from previous studies in this area (Ste-Marie, 1999; Ste-Marie and Lee, 1991). Possibly these statistically-significant differences in the deductions in each element are a result of insufficient comparison of the deviation of the technique of the element performed with the perfect technique. Bard et al., (1980) supported that novice and experienced judges focus their attention in "different areas" of the body of the athletes, agreeing with the results of Tenenbaum and his colleagues (1996), who also supported that judges gain experience and become better at their work in every competition.

According to Ste-Marie (1999) and Thomas (1994), the amount of declarative and procedural knowledge is different between national and international judges. Knowledge concerning real information based on specific rules and decision for movement (exercise) possibly differs between national and international judges. The present findings are in conflict with Ste-Marie and Thomas, since no significant differences were evident between national and international level judges. The results, however, approached significance and replication study may be necessary in the future to confirm the present findings. Further, the nine separate programs were associated with athlete's high speed performance, a factor that according to Salmela (1978) is associated with judging

errors and is a decisive factor for the accuracy of evaluation. We suggest with caution that previous athletic experience of the judges— if they have been athletes in gymnastics or not—may also affect their evaluations. It could also be supported that the complexity of the "multi-joint" system of the human body is also affecting the evaluation from the judges.

Finally, the differences that were revealed between real marks given during competition and those that were given through video evaluation agree with the theory of Puhl (1980), who stated that isolated presentation of the elements through video is giving the judges the possibility to evaluate with more precision. It is possible that the speed of performance and the different connections of the elements during competition also influence the accuracy of the evaluation. This is supported by the results of previous studies (Salmela, 1978).

The results of the present study agree with other findings (Abernethy, 1997, Ste-Marie, 1999) which support that experienced judges better interpret the biomechanical information coming from the athlete's body and need to focus their attention less on the performance, allowing them to concentrate more on the analysis of the element. Additionally—as already observed in the present research—based on the differences of the deductions in isolated routines there is a difference in the capacity of "anticipation through perception" between national and international judges. This fact is in agreement with findings of previous research (Ste-Marie and Lee, 1991; Tenenbaum et al, 1996). In conclusion, we can say that the accuracy of judging between national and international judges is satisfactory based on the very small percentage of statistically-significant differences in the total amount of deductions in all the routines.

Though all judges that participated in this study have a sufficiently long practicing experience, there are some differences in the evaluation between national and international judges. Probably these differences result from different opinions and knowledge about the performance of the elements or from personal experience, or from differing ability to recognise the nature of mistakes. For the elimination of these differences the presence of international judges in the E-panel is recommended. Special judges' courses to present and analyse the mistakes of performance of the elements and the resulting deductions from the judges will contribute to a fair result.

REFERENCES

Abernethy, B. (1997) Movement expertise: A juncture between psychology, theory, and practice. Paper presented at the meeting of the Association for the Advancement of Applied Sport Psychology, San Diego, CA, June.

Abernethy, B. and Russell, D.G. (1984) Advance cue utilization by skilled cricket batsmen. *The Australian Journal of Science and Medicine in Sport*, 16, 2, 2-10.

Allard, F., Graham, S. and Paaraslu, M.E. (1980) Perception in sport: Basketball. *Journal of Sport Psychology,* 2, 14-21.

Allard, F. and Starkes, J.L. (1980) Perception in sport: Volleyball. *Journal of Sport Psychology*, 2, 22-33.

Bard, C. and Fleury, M. (1981) Considering eye movement as a prediction of attainment. In I.M. Cockerill and W.W. MacGillivray (eds.), *Vision and Sport* (pp.28-41). Chetenham, UK: Stanley Thornes Publishers.

Bard, C., Fleury, M., Carriere, L. and Halle, M. (1980) Analysis of gymnastic judges' visual search. *Research Quarterly for Exercise and Sport*, 51, 267-273.

Federation International Gymnastics (2009) The Code of Points. Lucerne, Switzerland: Raeber.

Puhl, J. (1980) Use of video replay in judging gymnastics' vaults. *Perceptual and Motor Skills*, 51, 51-54.

Salmela, J.H. (1978) GYMNASTIC JUDGING: A complex information processing task, or (Who's putting one over

on who?). *International Gymnast,* 20, 54-57, 62-63.

Ste-Marie, D.M. and Lee, T.D. (1991) Prior processing effects on gymnastic judging, *Journal of Experimental Psychology*, 17, (1), 126-136.

Ste-Marie, D.M. (1999) Expert-Novice Differences in Gymnastic Judging: An Information-processing Perspective. *Applied Cognitive Psychology*, 13: 269-281.

Tenenbaum, G., Levy-Kolker, N., Sade, S., Lieberman, D.G. and Lidor, R. (1996) Anticipation and confidence of decisions related to skill performance. *International Journal of Sport Psychology,* 27, 293-307.

Thomas, K. T. (1994) The development of sport expertise: From Leeds to MVP legend. *Quest,* 1994, 46, 199-210.