COBISS: 1.08
Agris category code: /

STATISTICAL EVALUATION OF JUDGING AT THE CDI-W DRESSAGE COMPETITION IN KAPOSVÁR 2011

Csaba SZABÓ [1, 2], Orsolya VANCZER [1]

ABSTRACT

The relationship between humans and horses has a long history. Now days the horses used mainly for sport purposes. The highest level of cooperation and harmony is needed in the discipline of dressage. Due to its nature the scoring has some subjectivity which can result in inconsistent judging even on Olympic Games. To improve the consistency the statistical evaluation of judging can be helpful. Therefore, the aim of our investigations was to statistically evaluate the judging on a lower level international dressage competition. Data were collected on the 2011 CDI-W competition held in Kaposvár, Hungary. Those competitions were included in the study, where at least ten riders started. Five out of eight judges evaluated each competition. The scores given by judges and the index of disagreement were evaluated with variance analyses. The judging position had no effect of on the average score given. However, in higher level tests the judges tended to give higher points. Some of the judges give significantly different average scores to others. This underlies the necessity of the participation on regular refresher seminars. The index of disagreement detected significant difference at judging position 'B' and in Grand Prix frees style test. In conclusion judges tended to give higher scores and have higher level of disagreement in higher level of competitions. However, it seems that in lower level international competitions the magnitude of disagreement is less compared to top ranked events. The index of disagreement is more robust to detect slight differences in inconsistent judging compared to variance analyses.

Key words: horses /sport / dressage / judging / index of disagreement

1 INTRODUCTION

Equestrian sports are the only ones among Olympic-game sports where animals involved and women and men competing in the same competition. Dressage is the highest level of the inward cooperation between the rider and the horse, where the rider communicates with the horse with 'invisible' aids to inexperienced observers. The completion of the dressage test requires full cooperation from the horse. The beauty and strenuousness of the movements, the precision of implementation and the harmony between the rider and horse raise a program to the level of art (Mays, 1937). The evaluation of the dressage programs – similarly to other artistic performances – can be quite subjective. Thus even at Olympic Games the result of independent judges can have a significant variance (Hawson *et al.*, 2010). Inconsistent judging in Beijing 2008 contributed to the dismissal of the Dressage Committee in Fédération Equestre International (FEI) (Stachurska and Bartyzel, 2011). The breeding value based on competition results suitable for dressage horse lines selection, therefore the consistent judging even more important (Stewart *et al.*, 2010). However, judges used to compare their given scores after the competitions, but statistical analyses based evaluation rarely can be found in the literature and the few found evaluates Olympic Games (Stachurska *et al.*, 2006; Stachurska and Bartyzel, 2011). Therefore, our aim was to statistically evaluate the judging on a lower level international dressage competition.

[1] Kaposvár Univ., Guba Sándor út 40, 7400 Kaposvár, Hungary
[2] Corresponding author, e-mail: szabo.csaba@ke.hu

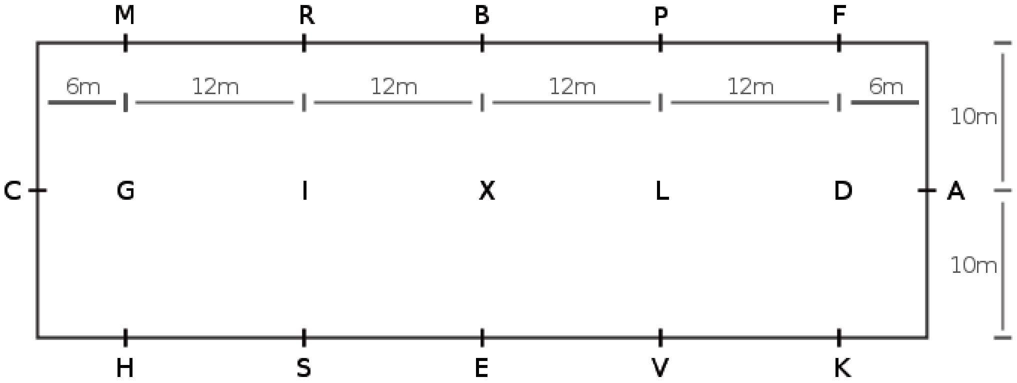

***Figure 1:** The setup of the dressage arena*

2 MATERIALS AND METHODS

Data were collected at the 2011 CDI-W dressage competition organized by the EC Pannon Equestrian Academy, Kaposvár. Eighty-one horse, seventy competitors from fifteen nations participated on the event. The competitors presented dressage programs with variable difficulty, according to the knowledge level of the horses. For the statistical analyses we used data from only those competitions where at least ten riders started, namely: St. George (StG), Intermediare II (Int2), Intermediare I (Int1), Grand Prix (GP), Intermediare I free style (Int1f) and Grand Prix free style (GPf). Dressage tests were presented in a 20 m × 60 m arena. Certain points of the arena marked with letters (Fig. 1.). Five out of eight judges judged each competition, situated at the following letters: E, H, C M and B. The tests consist of a series of movements, which must be followed (except the free style test). These movements or series of movements judged separately from 0 to ten points and summarized for each judge and expressed as a percentage to the maximum obtainable points. The result of the rider is the average of the five judge score.

The results of the rider were recorded in MS Excel. The fixed effects (competition, judge position and judge) were evaluated with the GLM procedure of SAS (SAS Institute Inc, Cary, NC, USA). In case of significant treatment effect differences were tested with Duncan range test. Since the correct ranking is more important than the actual difference in a given score the Index of Disagreement (ID) were calculated according to the method of Stachurska and Bartyzel (2011). The ID evaluates as a percentage the disagreement of ranking by a particular judge relative to the general ranking based on the scores of five judges.

3 RESULTS AND DISCUSSION

The average scores given by the judges at the judging positions did not differ significantly (Table 1). The average 63–70% results show that the pairs completed the test on the average international level. At the Dressage

***Table 1:** Evaluating the scores (%) according to the position of the judge and the competition* [1]

	Judge position					
Competition	B	C	E	H	M	Mean
St. Georges	63.5	62.8	62.3	62.0	62.7	62.7^a
Intermediaire II	62.4	63.8	63.1	62.4	62.8	62.9^a
Intermediaire I	63.8	63.0	64.1	62.9	63.6	63.5^a
Grand Prix	65.6	64.4	65.0	64.4	64.4	64.8^b
Intermediaire I free style	65.5	64.2	65.2	64.0	65.5	64.9^b
Grand Prix free style	70.6	69.8	69.4	70.1	68.1	69.6^c
Mean	65.3	64.6	64.9	64.3	64.5	

[1] Model: Position P = 0.543, Competition P < 0.001, Position * Competition P = 0.997

[a, b] Means in a column with the same superscript do not differ significantly (P > 0.05)

***Table 2:** The effect of judges and its position on the average scores given* [1]

	Position (n = 103)							
Judge	B	E	C	M	H	P	n	Mean
J1	-	65.1^{b}(GP)	63.8^{b}(Int2)	-	70.1^{a}(GPf)	0.000	52	66.3^{c}
J2	65.7(GP)	63.1(Int2)	66.0(StG,GPf)	65.6(Int1f)	-	0.311	84	65.4^{cd}
J3	63.5^{b}(StG)	66.5^{a}(Int1,GPf)	-	-	64.1^{ab}(Int1f)	0.027	66	65.1^{cd}
J4	-	65.2(Int1f)	-	64.6(StG, Int1,GPf)	63.8(Int2,GP)	0.517	103	64.4^{d}
J5	67.2^{a}(Int2,GPf)	-	63.6^{b}(Int1,Int1f)	-	62.0^{b}(StG)	0.005	77	64.4^{d}
J6	-	-	64.5(GP)	62.9(Int2)	-	0.267	37	64.0^{d}
J7	63.9(Int1)	-	-	-	-		19	63.9^{d}
J8	65.5(Int1f)	62.4(StG)	-	64.4(GP)	62.9(Int1)	0.066	77	63.7^{d}

[1] Model: Position P = 0.482; Judge P = 0.008; Position × Judge P = 0.001

$^{a, b}$ Means in a row with the same superscript do not differ significantly (P > 0.05)

$^{c, d}$ Means in a column with the same superscript do not differ significantly (P > 0.05)

World Cup final 70.31% was the average score which is quite similar to the average result of the Kaposvár CDI-W competition. Interesting result that the competitors average score was significantly lower in lower level competitions. This can be explained partly with the younger age of the horses in lower class tests. However, in these tests riders has to perform easier movements. Based on that we could expect at least similar, if not even higher results. It is also interesting result that there is a significant difference between the average score given in the Grand Prix and Grand Prix free style competition, despite that horses has to perform the same movements. The only difference is that in the freestyle test the riders determine the order of the movements, which best fits to the music.

According to our results – the position of the judge do not affect the scores given (Table 2). We found significant differences in the judge's average score. Since all judges judged lower and higher class tests, this difference can not be explained by that. This result confirms the necessity to improve the common view of the judges in course of refresher seminars (Janson and Olsson, 2004). Since there was a significant position*judge interaction, we evaluated the scores by judges separately as well. Some judge gave significantly higher scores in certain positions. After reviewing the evaluated competitions (notes in parenthesis) the reason is that judges give higher scores in Grand Prix freestyle test.

***Table 3:** The effect of the competition and the position of the judges on the Index of Disagreement (%)* [1]

	Position					
Competition	E	H	C	M	B	Mean
StG	6.0	1.2	3.4	3.5	11.7	5.16^{yz}
Int2	3.7	5.6	0.6	1.3	3.7	2.98^{z}
Int1	5.0	2.2	3.6	1.8	8.7	4.26^{yz}
GP	5.0	5.4	4.7	5.7	14.5	7.06^{xy}
Int1f	2.7	4.0	2.6	7.8	2.2	3.86^{yz}
GPf	10.2	11.1	3.5	8.4	12.3	9.10^{x}
Average	5.43^{a}	4.92^{a}	3.07^{a}	4.75^{a}	8.85^{b}	5.40

[1] Model: Position P = 0.019; Competition P = 0.017; $^{a, b}$ Means in a row having similar superscript do not differ significantly (P > 0.05);

$^{x, y, z}$ Means in a column having similar superscript do not differ significantly (P > 0.05)

The correct placing is more important from the rider's point of view. To the measurement of that the Index of Disagreement developed by Stachurska and Bartyzel (2011) is applicable. This index shows that the placing of a judge, judges in a certain position or judges in different competitions in how many percentages differ. If this value is 0% it means that the placing is in complete agreement, while if it has a value of 100% the placing is reversed. We can conclude that the judges in this competition judged with similar view, since the average disagreement was about 5% (Table 3.). Interesting result that in the judging position of 'B' resulted significantly higher disagreement. Stachurska and Bartyzel (2011) could not detect such a difference in the case of the Athen and Hong Kong Olympic Games. In the Grand Prix free style competition the index of disagreement had significantly higher rate. One reason could be that in this competition highly educated riders and highly trained horses started, resulting very similar performances. Due to the subjective nature of the judgment, it is unavoidable to having more disagreement in the placing. This theory is supported by the result of Stachurska and Bartyzel (2011) who found considerable higher indexes in the case of Olympic Games competitions.

4 CONCLUSIONS

Judges tend to give higher scores and have higher level of disagreement in higher level of competitions. However, it seems that in lower level international competitions the magnitude of disagreement is less compared to top ranked events. The index of disagreement is more robust to detect slight differences in inconsistent judging compared to variance analyses.

5 REFERENCES

Hawson L.A., Mclean A.N., McGreevy P.D. 2010. Variability of scores in the 2008 Olympic dressage competition and implications for horse training and welfare. Journal of Veterinary Behavior – Clinical Applications and research, 5: 170–176

Janson H., Olsson U. 2004. A measure of agreement for interval or nominal multivariate observations by different sets of judges. Educational and Psychological Measurement, 64: 62–70

Mays E. 1937. The Piaffe (in Hungarian). Magyar Katonai Szemle, 6, 4: 224

Stachurska A., Bartyzel K. 2011. Judging dressage competitions int he view of improving horse performance assessment. Acta Agriculturae Scandinavica, Section A, 61: 92–102

Stachurska A., Pieta M., Niewczas J., Markowski W. 2006. The freestyle dressage competition as a test of the horse's performance. Equine and Comparative Excercise Physiology. 3: 93–100

Stewart I.D., Woolliams J.A., Brotherstone S. 2010. Genetic evaluation of horses for performance in dressage competitions in Great Britain. Livestock Science, 128: 36–45