

*Original study*

## Effect of the judge and definition of the trait for horse free jumping evaluation

Dorota Lewczuk

Polish Academy of Science, Institute of Genetics and Animal Breeding Jastrzębiec, Magdalena, Poland

### Abstract

The presented research investigates the effect of the judge on scores for horse free jumping skills, the agreement of judge's scores and relations of these scores with measured jumping parameters received by video image analysis in order to recognise judging preferences and trait definition. The investigation was based on a group of 32 warm-blooded stallions that were judged in free jumping by six experienced judges in five routinely evaluated jumping traits. Simultaneously horses were filmed during jumping and linear jumping parameters were measured. Additional jumping parameters were calculated to describe the jumping style in a more detailed way. The influence of the effect of the judge was estimated by analysis of variance, the relationships between judges' notes and jumping parameters by analysis of correlations. The effect of the judge was statistically significant for all traits. The correlations between individual judges' notes were not equal. The notes for particular traits were in some cases more correlated with other traits like with notes for the same trait. Mean notes for evaluated traits were correlated above 0.6 between each other. Correlations between judges' notes and measured jumping parameters were low and medium. Some jumping parameters were correlated with all traits whereas some parameters were not correlated at all. Received results showed that the definition of the traits evaluated by judges is not the same for all of them. However, for all judges the distance of landing, elevation of the body and lifting of front limbs were the most important parameters of the jump.

**Keywords:** judgement, performance test, free jumping, kinematics, horse

Archiv Tierzucht 56 (2013) 64, 638-649

doi: 10.7482/0003-9438-56-064

**Corresponding author:**

Dorota Lewczuk; email: d.lewczuk@ighz.pl

Institute of Genetics and Animal Breeding of the Polish Academy of Sciences Jastrzebiec, ul. Postępu 36A, 05-552 Magdalena, Poland

Received: 23 May 2012

Accepted: 08 May 2013

Online: 24 May 2013

© 2013 by the authors; licensee Leibniz Institute for Farm Animal Biology (FBN), Dummerstorf, Germany.

This is an Open Access article distributed under the terms and conditions of the Creative Commons Attribution 3.0 License (<http://creativecommons.org/licenses/by/3.0/>).

## Introduction

The quality of breeding is always based on conducting correct selection, which should be founded on the proper and adequate evaluation of selected traits. That is of special importance for traits scored subjectively like it is widely used in sport horse breeding. The same or compared definitions are also of main importance for every step of selection. Comparison of the data definition is the basic requirement for further national or international horse breeding evaluation (Koenen & Aldrige 2002). The decision-making process is a very complicated task, especially in case of biological objects like moving horses connected with real-time evaluation. The judges' visual cognition may be affected by unforeseen effects. Psychometric tests for static discrimination between two opposing stimuli – dots with parts placed »up« or »down« – may lead to only 65 % of positive judgement (Braddick 1997). So the evaluation of objects in movement might be more difficult and affected by judges' expectations or bias in observations. The practice of judging requires certainty and understanding of the task, a stabile performance level and a learned ability to activate specific parts of the brain (Braddick 1997). According to Funder (1999 by Morris *et al.* 2002) good judging is connected with the trait (some traits are judged more easily than others), the judge (some persons are better judges than others) and the target (some individuals are judged more easily than others).

In recent research the role of motivation and emotion in judging and the decision-making process is widely underlined (Maner *et al.* 2007). Agreement between judges is treated as an indicator for good judging (Morris *et al.* 2002, Fuller *et al.* 2006, Lloyd *et al.* 2007). The subjective evaluation of traits provides basic information on the skills of a young horse's performance. From the point of view of genetic research evaluations made by judges seem to be a good prediction of the horse's future performance (Thorén-Hellsten *et al.* 2006). However, individual horse evaluations across countries are not always comparable (Koenen & Aldrige 2002, Koenen *et al.* 2004). Biomechanical studies underline the possible bias on horse jumping skills evaluation (Santamaria *et al.* 2006). Other authors wrote about possible systematic manipulation by owners and breeders as well as about insufficient and overlapping definitions (König von Borstel *et al.* 2012). Further investigations could be helpful for a better understanding of judging systems. The tools to study horse biomechanics are advanced (Colborne 2004) and ready as the methods for calibration of judgements. The aim of the present paper was to investigate the effect of the judge on horse free jumping evaluation, the agreement of judges' scores and their correlations with measured jumping parameters received using video image analysis. The description of evaluated skills and judges' preferences will be ascertained by an analysis of relations between intra- and inter-judges scores for jumping and relationships between these scores and kinematic measurements of the jump.

## Material and methods

### *Horses and traits*

The material was collected at a performance test station for young horses. The group of horses in investigation consisted of 32 warm-blooded stallions at the age of three years. The horses were previously trained for a period of 100 days and the jumping test was conducted at the end of that period. Six experienced judges were asked to evaluate horse jumping skills

according to the rules used for young horses' free jumping evaluation. The following traits of jumping description were marked: »willingness to jump«, »ease of the jump«, »work of the front«, »work of the back« as well as »work of the trunk, head and neck«. The traits were not described more precisely. Horses were evaluated in accordance with the scale designed for performance test stations, ranging from 0 to 10 points for every separate trait.

### *Video analysis*

The horses were simultaneously filmed during jumping by a digital Panasonic AG-EZ 35 camera (25 FPS) (Panasonic Corp., Osaka, Japan) with the set up standing at a distance of 10 m from the main obstacle, perpendicular to the horse's movement. Horses jumped in the riding hall and the line of obstacles was placed in the following order:

- 1) indicator pole on the ground at a distance of 2.5 m before the first vertical obstacle,
- 2) the first vertical obstacle measuring 0.6 m at a distance of 6.4 m,
- 3) the second vertical obstacle of 0.6 m at a distance of 6.8 m and
- 4) a spread obstacle (*doublebarre*) with a fixed width of 0.8 m and height ranging from 0.9 m to 1.2 m.

The warm-up of all horses consisted of 20 min on the lunge (mostly trot). Horses jumped every height of the obstacle (0.9 m, 1 m, 1.1 m and 1.2 m) once or twice depending on the decision of the trainer. Almost all repetitions were observed on the fourth height of the obstacle. The investigated group of 32 horses performed 156 jumps in total. Filmed material was analysed by a manual programme for video image analysis (Cytowski & Sakowski 1998). The scale for measurements was achieved by measuring real distances on the wall behind the path of the horses' movement. The following linear kinematic parameters were considered: taking off, landing, lifting of the limbs and elevation of specific points of the trunk (bascule points) above the obstacle. In order to describe the silhouette of the horse during the jump in more detail some additional parameters were calculated based on the basic measurements: symmetry of limbs lifting over the obstacle, »work« of the front and hind limbs, »work« of the head, croup and curve of the upper body line. All measured and calculated parameters are defined in Table 1.

### *Statistical analysis*

The analysis of the data was performed by statistical package SAS v9.1 (SAS Institute Inc., Cary, NC, USA) using different procedures. The aim of the study was realised by following analysis: the effect of the judge on the horse evaluation, the correlations between individual judges' score and the means of them as well as the correlations between judges' scores and jumping parameters.

The effect of the judge was estimated by analysis of variance MIXED procedure with the following statistical model:

$$y_{ijk} = \mu + a_i + J_j + e_{ijk} \quad (1)$$

where  $y_{ijk}$  is the judges' score,  $\mu$  is the mean,  $a_i$  is the random effect of the horse ( $i=1, \dots, 32$ ),  $J_j$  is the fixed effect of the judge ( $j=1, \dots, 6$ ) and  $e_{ijk}$  is the error.

Table 1  
Definitions of the measured and calculated linear parameters

Trait	Definition
1. Taking off	distance between the first point of hoof of hind limb placed closer to obstacle before jump flying phase and the second stand of obstacle
2. Landing	distance between the last point of hoof of front limb placed closer to obstacle after jump flying phase and the second stand of obstacle
3. Lifting of FL	distance between the highest pole of second stand of obstacle and the lowest point of front limb above that pole
4. Lifting of FR	distance between the highest pole of second stand of obstacle and the lowest point of front right limb above that pole
5. Lifting of HL	distance between the highest pole of second stand of obstacle and the lowest point of hind left limb above that pole
6. Lifting of HR	distance between the highest pole of second stand of obstacle and the lowest point of hind right limb above that pole
7. Elevation of head	distance between the highest pole of second stand of obstacle and the highest point of head (occiput) on bascule frame
8. Elevation of withers	distance between the highest pole of second stand of obstacle and the highest point of withers on bascule frame
9. Elevation of croup	distance between the highest pole of second stand of obstacle and the highest point of croup (sacrum) on bascule frame
10. Symmetry of the jump	ratio of the taking off distance (1) to the landing distance (2)
11. Symmetry of front limbs	difference between the lifting of front left (3) and the lifting of front right (4)
12. Symmetry of hind limbs	difference between the lifting of hind left (5) and the lifting of hind right (6)
13. »Work« of head	difference between the elevation of withers (8) and the elevation of head (7)
14. »Work« of croup	difference between the elevation of withers (8) and the elevation of croup (9)
15. Curve of the upper line	difference between the elevation of croup (9) and the elevation of head (7)
16. »Work« of front limbs	difference between the elevation of withers (8) and the mean of lifting of front limbs (3,4)
17. »Work« of hind limbs	difference between the elevation of withers (8) and the mean of lifting of hind limbs (5, 6)

The analysis of relationships between the judges' scores for every trait was calculated as Pearson's correlations using the CORR procedure of the programme. The same procedure was performed to find out the relationships of the individual judge's score with the mean of all judges. Results were presented as the range of correlations between individual judges and correlations between individual judge and the mean of all judges as well as correlations between means for all judges.

The relationship between measured jumping parameters and scores for single traits evaluated for jumping skills were calculated as correlations corrected for effects that influence jumping parameters (height of the obstacle, successive number of the jump) using the Manova option in GLM procedure. This analysis was calculated for every judge's score as well as for the mean score separately.

## Results and discussion

### *Variance of judges' scores and jumping parameters*

The mean scores of six judges and their basic characteristics are presented in Table 2. Most of the scores were between 6 and 7 with a standard deviation of less than 1. Most scores were given within the range of 3 points, in some cases they were within a wider range of 4 points (13 %) and one judge used for only one trait half of the scale (5 points). The investigated judging is comparable with the results obtained by other authors and countries. The cited values were between 6 and 7 with a standard deviation between 0.7 and 1.7 (Bruns *et al.* 2001). Some values of standard deviation in the presented study were below the level given by the cited authors working on the international horse breeding evaluation (Interstallion). This was mainly observed for the trait »willingness to jump«. The effect of the judge was statistically significant for all traits ( $P < 0.001$ ). In the case of »willingness to jump« and »work

Table 2  
Statistical characteristics of the jumping measurements and traits evaluated by judges

Traits	Mean	SD	Min	Max
Measured linear parameters				
1. Taking off	262.7	40.3	163.3	330.3
2. Landing	155.9	50.9	75.5	300.0
3. Lifting FL	16.1	9.5	1.1	48.8
4. Lifting FR	17.4	10.1	1.1	55.8
5. Lifting HL	19.0	10.8	1.1	59.3
6. Lifting HR	20.9	11.6	1.1	60.7
7. Elevation – head	113.1	19.2	77.0	158.3
8. Elevation – withers	114.2	13.8	89.9	158.3
9. Elevation – croup	105.8	14.3	77.6	143.7
10. Symmetry of the jump	1.9	0.60	0.56	3.81
11. Symmetry of front limbs	-1.3	6.52	-20.7	20.0
12. Symmetry of hind limbs	-1.8	4.9	-18.6	15.8
13. »Work« of head	1.1	12.3	-28.1	39.8
14. »Work« of croup	8.4	4.0	-6.9	23.6
15. Curve of the upper line	-7.3	10.7	-32.8	32.8
16. »Work« of front limbs	97.5	11.3	76.4	147.3
17. »Work« of hind limbs	85.8	14.5	52.0	132.1
Traits				
18. Willingness to jump	7.9	0.4	6	8
19. Easy to jump	7.3	0.6	6	9
20. Work of front	6.8	0.6	4	9
21. Work of back	7.0	0.5	5	9
22. Work of trunk, head and neck	6.8	0.5	5	8

of front« one judge differs in a spectacular manner from the others. All differences are presented in Table 3.

Table 3  
The effect of the judge on the jumping traits

Trait	SE	Score LSM*					
		Judge 1	Judge 2	Judge 3	Judge 4	Judge 5	Judge 6
Willingness to jump	0.11	7.84 <sup>A</sup>	8.06	7.84 <sup>B</sup>	8.09 <sup>C</sup>	8.06 <sup>D</sup>	7.50 <sup>ABCD</sup>
Ease to jump	0.15	7.59 <sup>AB</sup>	7.47 <sup>cD</sup>	7.16 <sup>Aceh</sup>	7.47 <sup>Dfh</sup>	7.47 <sup>eG</sup>	6.90 <sup>BFG</sup>
Work of front	0.16	6.81 <sup>ABc</sup>	7.56 <sup>ADEFG</sup>	6.34 <sup>BDHi</sup>	6.97 <sup>EHJ</sup>	6.71 <sup>Fi</sup>	6.37 <sup>GJ</sup>
Work of back	0.13	6.59 <sup>AbdE</sup>	7.03 <sup>A</sup>	6.96 <sup>b</sup>	6.88 <sup>cf</sup>	7.18 <sup>df</sup>	7.21 <sup>eE</sup>
Work of trunk, head and back	0.14	6.66 <sup>AbC</sup>	7.06 <sup>ADe</sup>	6.34 <sup>bDFg</sup>	7.06 <sup>Ch</sup>	6.88 <sup>F</sup>	6.68 <sup>egh</sup>

\*means marked with the same letter differ significantly, small letter  $P \leq 0.05$ , capitals  $P \leq 0.001$

The jumping parameters achieved in this study are presented in Table 2. The results obtained were similar to the values measured on other groups of young horses in free jumping (Lewczuk 2007). Only lifting of the hind limbs seems to be measured higher in the group of horses presented in this paper. Comparable values were obtained for adult sport horses (Puchała 2005, Pietrzak *et al.* 2006). However, the standard deviations for measured parameters of young horses were higher. Similarly, the coefficient of variance reached higher values.

#### *Correlations between individual judges' scores and means for different traits and within a trait*

The range of correlations between individual judges' scores for each trait are presented in Table 4. The results obtained were on a medium level. The highest values of correlations between individual judge's scores for the same trait were calculated for »ease of the jump« and »work of the front«. The lowest values were achieved for the traits »work of the back« and »work of trunk, head and neck«. In some cases correlations between individual judge's scores for different traits were higher than correlations between individual judge's scores for the same trait. This was observed for the traits »ease of the jump« and »work of the front« (judge 3), as well as for the traits »willingness to jump« and »ease of the jump« (judges 3 and 4). Some high values of correlations were calculated also between scores of the judges for different traits on the level of 0.8-0.9 (judges 4 and 5). The lack of correlation between some individual judge's scores for the same traits was rather surprising. This was observed for »work of the trunk, head and neck«, »willingness« and »work of the front«.

The correlations between the means of the judges' scores for every trait and the correlations between every individual judge's note with the mean score for each trait are also presented in Table 4. The correlations between the mean score for different traits were high and ranged from 0.62 for the traits »willingness to jump« and »work of the trunk, head and neck« to the highest value of 0.88 for the traits »ease of the jump« and »work of the front«. Some differences in style of judging could be noticed on the basis of the correlation between the individual judge's score and the mean for the traits. It seems that two tendencies could be observed for the traits »ease of the jump« and »willingness to jump«. The correlations of

Table 4  
Correlations between notes of six judges for every evaluated trait

Trait	Judge	Work of trunk, head and neck		Work of back		Work of front		Ease of jump		Willingness to jump	
		Corr	Mean	Corr	Mean	Corr	Mean	Corr	Mean	Corr	Mean
Willingness to jump	1	0-0.48	0.33	0-0.52	0.37	0-0.50	0.42	0-0.46	0.51	0.36-0.57	0.81
	2	0-0.38	0.34	0-0.55	0.32	0-0.66	0.50	0.19-0.62	0.53	0-0.58	0.63
	3	0.31-0.53	0.58	0-0.54	0.45	0.24-0.47	0.51	0.29-0.53	0.52	0-0.36	0.51
	4	0-0.63	0.39	0.22-0.64	0.66	0.33-0.56	0.60	0.34-0.69	0.72	0.34-0.49	0.71
	5	0.20-0.63	0.46	0-0.53	0.51	0.30-0.62	0.60	0.38-0.77	0.69	0.30-0.58	0.78
	6	0.17-0.65	0.56	0.27-0.50	0.58	0.24-0.54	0.53	0-0.66	0.61	0.30-0.58	0.68
Mean			0.62		0.68		0.75		0.85		1.00
Ease of jump	1	0-0.73	0.56	0.40-0.73	0.82	0.32-0.77	0.67	0.45-0.73	0.86		
	2	0-0.52	0.47	0-0.62	0.58	0.32-0.70	0.62	0.37-0.56	0.69		
	3	0.33-0.65	0.71	0.29-0.56	0.66	0.24-0.77	0.80	0.34-0.53	0.74		
	4	0.19-0.78	0.65	0.18-0.63	0.66	0.36-0.66	0.66	0.56-0.71	0.84		
	5	0-0.66	0.49	0.19-0.61	0.55	0.29-0.71	0.66	0.37-0.63	0.72		
	6	0-0.76	0.61	0.38-0.63	0.76	0.31-0.54	0.60	0.34-0.73	0.75		
Mean			0.77		0.88		0.88		1.00		
Work of front	1	0.26-0.62	0.60	0.30-0.59	0.68	0.38-0.69	0.85				
	2	0-0.56	0.47	0-0.63	0.46	0-0.49	0.51				
	3	0.32-0.58	0.70	0.21-0.52	0.55	0.35-0.55	0.71				
	4	0.15-0.62	0.48	0.15-0.91	0.63	0.35-0.72	0.79				
	5	0-0.81	0.54	0.35-0.69	0.68	0-0.72	0.84				
	6	0-0.63	0.51	0.16-0.49	0.50	0-0.61	0.72				
Mean			0.77		0.79		1.00				
Work of back	1	0-0.36	0.31	0.30-0.42	0.63						
	2	0-0.44	0.40	0.21-0.58	0.62						
	3	0.31-0.50	0.63	0.21-0.38	0.64						
	4	0-0.63	0.53	0.34-0.58	0.75						
	5	0-0.61	0.42	0.21-0.68	0.75						
	6	0.25-0.68	0.49	0.20-0.68	0.70						
Mean			0.70		1.00						
Work of trunk, head and neck	1	0.24-0.56	0.66								
	2	0.31-0.61	0.73								
	3	0-0.56	0.70								
	4	0.32-0.61	0.75								
	5	0.35-0.63	0.75								
	6	0-0.63	0.63								
Mean			1.00								

Corr: range of correlations

three judges (1, 2, 3) were on an equal level of 0.5, while the correlations for judges (4, 5, 6) reached higher values, about 0.6-0.7. Another trend in judging could be observed on the basis of calculations for the traits »ease of the jump« and »work of the front«. Except for two extreme values (0.6, 0.8), all others were around 0.65.

The correlations between individual judge's scores and the mean of all judges ranged from 0.51 to 0.86. The most even values were calculated for the traits »work of the back« and »work of the trunk, head and neck«. The relationship between individual judge's scores and the mean of all judges was more differentiated for »work of the front« and »willingness to jump« than for other traits.

### *Correlations between judges' scores and measured jumping parameters*

The results obtained are presented in Table 5. All correlations are on a low and medium level ranging from -0.29 to 0.42. Some jumping parameters were not correlated with the judges' scores at all like lifting of the hind limb and symmetry of the hind limbs or were correlated only with individual judge's scores like »work« of the croup or the taking off distance. The jumping parameters landing distance and lifting of the front limbs were correlated with all traits evaluated by the judges. Elevation of the trunk (bascule points) was correlated with the traits »willingness«, »ease of the jump« and »work of the back«.

Table 5  
Correlations between the individual judges' notes and measured jumping parameters

Trait	Willingness to jump		Easy of jump		Work of front		Work of back		Work of trunk, head and neck	
	J 1-6	Mean	J 1-6	Mean	J 1-6	Mean	J 1-6	Mean	J 1-6	Mean
1. Taking off			0.0-0.19		-0.15-0.30		0-0.29		0-0.28	
2. Landing	0.27-0.31	0.40	0-0.45	0.37	0-0.20	0.18	0-0.41	0.30	0-0.40	0.36
3. Lifting FL	0-0.20	0.15	0-0.23	0.18	0-0.24	0.18	0-0.24		0-0.24	0.21
4. Lifting FR	0.14-0.24	0.24	0.14-0.37	0.36	0-0.32	0.31	0-0.30	0.27	0.20-0.38	0.37
5. Lifting HL	-0.16-0		-0.11-0				0-0.23	0.14		
6. Lifting HR	-0.20-0		-0.17-0				0-0.16			
7. Elevation – head	0-0.33	0.25	0-0.25	0.17	-0.24-0.32		0-0.24	0.14	-0.29-0	
8. Elevation – withers	0-0.25	0.21	0-0.27	0.21	-0.17-0.31		0-0.19	0.20	0-0.21	
9. Elevation – croup	0-0.24	0.24	0-0.28	0.24	-0.14-0.35		0-0.27	0.23	0-0.23	
10. Symmetry – jump	-0.36-0	-0.31	-0.16-0							
11. Symmetry – front	-0.17-0		-0.26-0	-0.21	-0.21-0		-0.20-0	-0.22	-0.29-0.17	-0.19
12. Symmetry – hind	0-0.15		0-0.18		0-0.20				0-0.18	
13. »Work« of head	-0.18-0				0-0.21		0-0.16		0-0.33	0.22
14. »Work« of croup	-0.17-0		-0.18-0		-0.23-0		-0.20-0		0-0.16	
15. Curve – upper line	-0.15-0				0-0.23		0-0.20		0-0.31	0.24
16. »Work« of front limbs	0.15-0		-0.18-0		-0.29-0	-0.16	-0.16-0		-0.19-0	
17. »Work« of hind limbs	0-0.42	0.28	0-0.32		0-0.39		0-0.22		0-0.18	

J 1-6: Judge 1-6 range



The highest correlation for mean of all judges was achieved for the distance of landing and the trait »willingness to jump« (0.4) as well as the trait »ease of the jump« (0.37). Some correlations between lifting of the front limbs and the individual judge's scores reached values above 0.3. The highest values of correlations (0.45) were calculated for individual judge's scores between the parameter landing and »ease of the jump«. The most even correlations between individual judge's scores were noted for lifting of the front limb and »work of the trunk, head and neck« as well as for the parameters distance of landing, the trait »willingness to jump« and lifting of the front limb and the trait »willingness to jump«. The same tendencies in judging were observed for four out of six judges in relation to the parameters of elevation of the trunk and limbs and »willingness to jump« and »ease of the jump«. The most differentiated correlations between individual judge's scores for specific traits were calculated for the parameter elevation of the head and the trait »work of front«. Received results ranged from  $-0.24$  to  $0.32$ . The additional parameters »work« of the head and curve of the upper body line were correlated with the trait »work of the trunk, head and neck« on the low level about  $0.22$  and  $0.24$ , respectively. The parameter which was in negative correlation with the judges' scores was symmetry of the front limbs. The parameters which were related only with some individual judge's scores were »work« of front limbs, lifting of the hind limbs and symmetry of the hind limbs. Symmetry of the jump was the only parameter that was negatively correlated with all judges' scores. The height of received correlations between traits and jumping parameters are comparable with the French results. The kinematics measurements were correlated with the judges' notes in the range between  $-0.36$  and  $0.18$  (Dufosset & Langlois 1984).

The consistency of the judges' opinion is widely used as a predictor of good judgement. However, the classification of the meaning of received heights of values was not always the same. Landis & Koch (1977) suggested that values above 0.8 were classified as »excellent« and values above 0.4 as »moderate«. Martin & Bonnett (1987) qualified relationships between 0.3-0.5 as »acceptable«, between 0.5-0.7 as »good« and above 0.7 as »excellent«. According to Fleiss (1981 by Fuller *et al.* 2006), values above 0.4 are »acceptable«. The same level of importance was accepted by Keegan (Keegan *et al.* 1998). Most of the relationships are presented as kappa ( $\kappa$ ) values or Kendall's ( $W$ ) statistic. However, as the scale of all of these coefficients is the same, ranging from 0 to 1, relationships calculated in different ways seem to be comparable in the meaning. Most correlations received in the presented study for the same traits were above 0.6, therefore reaching the common acceptable level.

Correlations between individual judge's scores were not always on an acceptable level, especially if there was a lack of correlations between some of them. On the basis of relations between the judge's scores also the character of the traits could be discussed (Morris *et al.* 2002). The traits »willingness to jump«, »ease of the jump« and »work of front« could be considered to be easier to judge because of judges' agreement on their evaluation. Higher correlations between individual judges' scores for different traits than correlations between individual judges' scores for the same trait implied that the definition of traits may not be the same for all judges. The same could be concluded on the basis of connections between judge's scores for evaluated traits and jumping parameters. The discrepancy of the definition for the investigated traits is also observed on the basis of the correlations between an individual judge's scores and the mean, which showed in some cases rather two different ways of judging.

The need for a global system of scoring has been emphasized in veterinary medicine (Fuller *et al.* 2006) and in judging for breeding purposes (Thorén-Hellsten *et al.* 2006). Also the fact that using the scale of judging may involve the results was known (Fuller *et al.* 2006). However, it seemed that even the same scale does not guarantee the same results. The way of using different scales may influence results. The judge's personal point of reference in her/his mind, that is the basis for judgement, seemed to affect the results even more significantly. The veterinary scale of judging lameness could be a very good example for this problem. Except for the scales 0-5 points or 0-10 points, also the correction on the last observations of judgement is used (Fuller *et al.* 2006). This provides for the situation where the earlier observation is the reference point for the next judgement. The role of experience in judgement is obvious. It was suggested that people who handled horses regularly are much more likely to agree on scores (Lloyd *et al.* 2007). Subjective judgement and problems connected with such kind of evaluation were underlined in many papers on horses' evaluation. Some authors suggested a new methodology of calculations of breeding value estimation based on official performance tests judged in the traditional way just because of the narrow scale of evaluation used (Borowska *et al.* 2011). New weights for traits in indexes of horses were also suggested because of differences in heritabilities of almost the same traits (observed in different time period) judged by different experts (Dietl *et al.* 2005). Another possibility of correction of horse evaluation was discussed in the paper that studied benefits of descriptive (linear) evaluation (Rustin *et al.* 2009). The usefulness of biokinematics in research on dressage horse judgement was underlined as an explanation for introduction of new traits for gait evaluation (Becker *et al.* 2012, Becker *et al.* 2013). Some new possibilities were also seen by using behavioural tests (König von Borstel *et al.* 2012). That suggested that the existing lack of definitions and irregular judgement is a long lasting international problem and should be solved by the introduction of new technologies and ideas.

In conclusions of the paper, it seems that the style of judging could be predicted on the basis of correlations between measured jumping parameters and the judges' scores. The correlations between these two sources of data received in the study are low and middle high. Diversified results of individual judges let recognise some different tendencies in judging. Comparison of judges' notes within and between the evaluated partial traits indicates that free jumping note is not precisely defined. Detailed traits' scores could be helpful by establishing the definition of overall trait. Imprecise trait definition is hidden if only the mean score of judges is analysed. Linear kinematical measurements launch trait definition for horse free jumping. However, it could be expected that the correlations between the judges' scores and the measured jumping parameters would be on a more equal level for every individual judge. Judges' preferences are evident due to kinematical and detailed traits analysis. For all judges, distance of landing, elevation of the body and lifting of front limbs were the most important parameters of the jump.

## References

- Becker AC, Stock KF, Distl O (2011) Genetic analyses of new movement traits using detailed evaluations of warmblood foals and mares. *J Anim Breed Genet* 129, 390-401

- Becker AC, Stock KF, Distl O (2013) Correlations of unfavorable movement characteristics in warmblood foals and mares with routinely assessed conformation and performance traits. *Animal* 7, 11-21
- Borowska A, Wolc A, Szwaczkowski T (2011) Genetic variability of traits recorded during 100-day stationary performance test and inbreeding level in Polish warmblood stallions. *Arch Tierz* 54, 327-337
- Braddick O (1997) Visual psychophysics. *Curr Biol* 7, R209-R211
- Bruns E, Schober M, Fredricson I (2001) Interstallion – a cross-country evaluation of testing methods and data availability/suitability. 52nd Ann Meet EAAP, Budapest, Hungary, 26-29 September 2001
- Colborne GR (2004) Gait analysis: technology looking for a place to happen? *Vet J* 168, 112-113
- Cytowski J, Sakowski T (1998) [Selection of traits in the digital analysis of biological images]. *ICS PAS Reports* 871, 1-32 [in Polish]
- Dietl G, Hoffmann S, Reinsch N (2005) Impact of the trainer and judges in the mare performance test of Warmblood Horses. *Arch Tierz* 48, 113-120
- Dufosset JM, Langlois B (1984) [Analysis of jumps over obstacles of 122 horses Selle Français and evaluation of free jumps]. *Compte rendu de la journée d'étude sur le cheval, CEREDPA (Ed.) Paris 7 mars 1984*, 2-26 [in French]
- Fuller CJ, Bladon BM, Driver AJ, Barr ARS (2006) The intra- and inter-assessor reliability of measurement of functional outcome by lameness scoring in horses. *Vet J* 171, 281-286
- Keegan KG, Wilson DA, Wilson DJ, Smith B, Gaughan EM, Pleasant RS, Lillich JD, Kramer J, Howard RD, Bacon-Miller C, Davis EG, May KA, Cheramie HS, Valentino WL, van Harreveld PD (1998) Evaluation of mild lameness in horses trotting on a treadmill by clinicians and interns or residents and correlation of their assessments with kinematic gait analysis. *Am J Vet Res* 59, 1370-1377
- Koenen EPC, Aldridge LI, Philipsson J (2004) An overview of breeding objectives for warmblood sport horses. *Livest Prod Sci* 88, 77-84
- Koenen EPC, Aldridge LI (2002) Testing and genetic evaluation of sport horses in an international perspective. *Proc 7th World Congr Appl Genet Livest Prod* 30, 367-374
- König von Borstel U, Pirsich W, Gauly M, Bruns E (2012) Repeatability and reliability of scores from ridden temperament tests conducted during performance tests. *Appl Anim Behav Sci* 139, 251-263
- Landis JR, Koch GG (1977) The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33, 159-174
- Lewczuk D (2007) The effect of training on linear jumping parameters in young stallions. *Equine Comp Exerc Physiol* 4, 159-165
- Lloyd AS, Martin JE, Bornett-Gauci HLI, Wilkinson RG (2007) Evaluation of a novel method of horse personality assessment: Rater-agreement and links to behaviour. *Appl Anim Behav Sci* 105, 205-222
- Martin SW, Bonnett B (1987) Clinical Epidemiology. *Can Vet J* 28, 318-325
- Maner JK, Gerend MA (2007) Motivationally selective risk judgments: Do fear and curiosity boost the boons or the banes? *Organ Behav Hum Decis Process* 103, 256-267
- Morris PH, Gale A, Duffy K (2002) Can judges agree on the personality of horses? *Pers Individ Dif* 33, 67-81
- Pietrzak S, Bocian K, Strzelec K (2006) [Determination of horse jump parameters over up-right and oxer fences with use of computer picture analysis]. *Rocz Nauk PTZ* 2, 23-33 [in Polish]
- Puchała J (2005) [Analysis of jumping parameters over selected obstacles during competitions]. PhD thesis. Agriculture Academy Lublin [in Polish]
- Rustin M, Janssens S, Buys N, Gengler N (2009) Multi-trait animal model estimation of genetic parameters for linear type and gait traits in the Belgian warmblood horse. *J Anim Breed Genet* 126, 378-386
- Santamaria S, Bobbert MF, Back W, Barneveld A, van Weeren PR (2006) Can early training of show jumpers bias outcome of selection events? *Livest Sci* 102, 163-170
- Thorén-Hellsten E, Viklund Å, Koenen EPC, Ricard A, Bruns E, Philipsson J (2006) Review of genetic parameters estimated at stallion and young horse performance tests and their correlations with later results in

dressage and show-jumping competition. *Livest Sci* 103, 1-12