INTERNATIONAL JOURNAL OF AGRICULTURE & BIOLOGY
ISSN Print: 1560–8530; ISSN Online: 1814–9596
13–450/2014/16–3–557–563
http://www.fspublishers.org

Full Length Article

Assessment Scores in Morphological Competitions of *Pura Raza Español* Horse

Sánchez María José[1*], Gómez María Dolores[1], Molina Antonio[2] and Valera Mercedes[1]
[1]*Department of Agroforestry Sciences, University of Seville, Ctra, Utrera km 1, 41013 Seville, Spain*
[2]*Department of Genetics, University of Cordoba, C.U. Rabanales, Ctra, Madrid-Cadiz, km 396a, 14071 Cordoba, Spain*
*For correspondence: v32sagum@gmail.com

Abstract

The aim of this work was to establish the external factors affecting morphological scores in competitions of *Pura Raza Español* (PRE – Spanish Purebred) horses. The dataset included 8633 participations, in 69 different morphological competitions of 5097 horses, belonging to 1113 studs. In total, the dataset had 21,760 records (each with eight partial morphological scores) with a balanced frequency of males and females. Our results showed that there are several external factors that influence the judges' scores. Morphological competition (15.70% of the total variance), judge (9.65%), judge*morphological competition (7.45%) and judge*type (2.41%) were the most important external factors in the analyzed traits. The reliability of judgements was evaluated by the index of disagreement, and results showed that it might be necessary to train judges to get more agreement in these scores. Moreover, for a breeding program it is very important that judges use all the range of scores in order to collect the maximum level of variation. The analyses showed that competitions with three judges were the most appropriate.

Keywords: Andalusian horse; Morphological evaluation; Index of disagreement; Judge; Spanish Purebred

Introduction

In horse breeding conformation, as beauty, appropriate body proportions and breed standard, still has great importance in many breeds. Moreover, relative economic values of selection criteria, indicate that conformation is still the most important thing (Bruns *et al*., 1978; Schwark *et al*., 1988) to achieve high prices for horses (Preisinger *et al*., 1991). In fact studs use morphological scores obtained in morphological competitions to select horses for future breeding.

A linear scoring system was first introduced in dairy cattle, although a linear assessment of conformation traits was applied to most livestock species. The linear system has been developed in horses to collect suitable data for genetic evaluation of conformation traits mainly to improve functionality, but unable to assess the beauty and breed standard of the animals. The morphology assessment by classical methods is carried out in a subjective way based on the experience of the judge, which explains most of the phenotypic variance through the morphological scores obtained by the animals (Holmström *et al.*, 1990). Since objectively measuring beauty and breed standards of animals seems impossible, morphology assessment remains the only valid model.

The Pura Raza Español (PRE – Spanish Pure Breed) horse is a breed with ancestors dating back thousands of years in the Iberian Peninsula. They have been recognized as an individual breed since the 15th century, are strongly built, yet extremely elegant horses. The PRE horse is held for its beauty and fine temperament and therefore, its selection has always been based on these qualities (Muñoz *et al.*, 1997). Their morphological evaluation, from a phenotypic point of view (beauty according to the breed standard), up to now, has been made by means of specific morphological competitions. The scores recorded in the PRE morphological competitions are an important source of information to improve the morphological assessment of this breed because they remain an important reference for the studs. For example, in 2012, international morphological final championship of PRE (called SICAB) had a total turnover of 30 millions Euros (ANCCE, 2012). Only PRE horses registered in the official stud-book can participate in these competitions, where horses are grouped in different sections by their sex and age. The animals are scored numerically on a "desirability point" scale as a distance from the "ideal" using a score ranged from 1 to 10 points: It is a subjective judging defined as judging through the use of an individual feeling as the ultimate criterion for what is deemed good and correct (Magnusson and Thafvelin, 1985).

Collective scores for each region are designed to summarize the characteristics of each anatomical area of the animal, which must reflect the qualities of the entire performance and the relation with the breed standard. The

To cite this paper: Sánchez, M.J., M.D. Gómez, A. Molina and M. Valera, 2014. Assessment scores in morphological competitions of *Pura Raza Español* horse. *Int. J. Agric. Biol*., 16: 557–563

final classification is decided according to the average score awarded by all of the judges. But, even in this case, the judging is sometimes controversial because morphological beauty is difficult to assess, since it cannot be objectively measured. So, many times, the score given by a judge, is far from the average of all participating judges.

In fact judging needs to be transparent, since all cases of protective or unjust scores may meet with protest and result in the wrong breeding; furthermore, trends in judging behaviour can shape the future of breed selection. Nowadays the reliability of morphological scores is not taken into account and scores are not verified. Therefore, this system, based on breed standard and beauty, needs to be improved since it is a criterion actually used by studs to choose breeding animals. And therefore, because of the high influence the judge has on the scores, the reliability of judging has to be assessed.

The aim of this work is to establish the external factors affecting morphological scores in PRE competitions to contribute to a better design of the assessment methodologies for this breed.

Material and Methods

Dataset

Data for the analysis were collected from 69 different competitions held between 2006 and 2010. It included 5097 different horses, belonging to 1113 studs. A dataset with 21,760 records (each one have 8 morphological partial scores described in Table 1), corresponding to 8633 participations. The dataset had a balanced frequency of males and females (49.7% of records– 49.2% of horses- from males) with an average of 4.3 available records per horse. The number of judges that participated in a competition varied with the level of the competition, with the highest number of judges present for competitions with a large number of horses. The lower level of morphological competition (with 1 judge), called type A, had a total of 838 records from 8 different morphological competitions. Next level, type B (2 judges), had 3766 records from 18 morphological-competitions. And the last two levels were competitions with 3 or 5 judges and the highest number of horses 16,561 and 595 including all the final-championships, called type C and D, respectively were collected from 43 different competitions.

Assessment records included 8 morphological traits, related with anatomical regions; head and neck, shoulders and withers, chest and thorax, back and loin, croup and tail, forelimbs, hindlimbs and overall form. Abbreviations and broad definitions of the analysed traits are given in Table 1.

The assessed traits were scored numerically on a desirability scale from 1 (not at all similar to the conformation characteristic of breed standard) to 10 points (completely identical to the conformation completely identical to the most perfect morphology). All 8 traits were evaluated by each judge. And the final score of the animal in the competition corresponded with the average of the scores of the different judges, which defined the ranking. Each judge appraises horses independently to avoid influences on the scores.

Statistical Analysis

Basic statistics were estimated for all the analyzed traits grouped by the level of competition, according to the type of competition (number of participant judges: A, B, C and D) and total data (0). The 95% reference intervals were calculated by removing the upper and lower 2.5% of the range for each morphological parameter (2.5 and 97.5 percentiles) and then the confidence limits of these reference values were estimated.

To analyze the possible influence of various factors on the classification of different regions, a MANOVA was fitted including the following fixed effects: section (16 levels, where the animals were grouped by their sex and age as shown in Table 2), judge (43 levels); stud (grouped in 49 classes by the number of animals that have morphological controlled, since it was assumed that the studs with more animals carry a similar handling of horses); coat (7 levels: chestnut, buckskin, white, bay, black, roan and grey); competition (69 levels) and type of competition (4 levels according to the maximum number of judges in the competitions: with 1, 2, 3 and 5 judges); and the combination between: judge × section (431), judge × breeder-stud (1486), judge × coat (113), judge × competition (267) and judge × type (162). All these factors were included in the model, because they can produce errors of perception, categorization and memory process, because of the previous experiences of the judges (Plessner and Haar, 2006). Additionally, the percentage of variance of the different factors for each trait was also calculated. Finally, the Principal Components Analysis (PCA) was estimated to analyze the relationship between the different traits and types of competitions.

The statistical analyses were carried out using the program SAS package v. 6.11 (SAS, 2001).

Index of Disagreement

To evaluate the agreement between the evaluations of the different judges in the PRE competitions, we estimated the Index of Disagreement (ID) as described by Stachurska and Bartyzel (2011). The freely available program takes into account the ranking of the animals by each final score. These authors consider that when judges agree this would result in all the horses having the same similar ranking order thus making the similarities between scores less important. The ID is an estimation of the judging quality, which showed, in percentage, how much a judge ranked the horses differently from the total ranking in a trait based on the sum of the scores given by the other judges that participate in the same competition. In this sense, the

Table 1: Abbreviations and descriptions of the eight morphological traits analyzed in the PRE morphological competitions

Abbreviations	Variables	Descriptions
HN	Head-neck	Evaluation of head and neck region
SW	Shoulder-withers	Evaluation of shoulder and withers region
CT	Chest- thorax	Evaluation of chest and thorax region
BL	Back-loin	Evaluation of back and loin region
CrT	Croup and tail	Evaluation of croup and tail region
FL	Forelimb	Evaluation of forelimb region
HL	Hindlimb	Evaluation of hindlimb region
OF	Overall form	Evaluation of overall form

Table 2: Description of the age and the sex of the different sections established in PRE morphological competitions

Section	Age (years)	Sex
1	1	Female
2	1	Male
3	2	Female
4	2	Male
5	3	Female
6	3	Male
7A	4	Female
7C	5	Female
8A	4	Male
8C	5	Male
9A	6	Female
9C	7	Female
10A	6	Male
10C	7	Male
11	8 or more	Female
12	8 or more	Male

lower ID, the more consistent a judge was with the total ranking of the animal in a given event. As Stachurska and Bartyzel (2011) recommended, the particular traits were considered in this analysis, instead of the total final ranking, because the consistency of the final ranking of the animal in the competition could sometimes be coincidental. The ID was modified to obtain a more clearly estimation of the difference between each judge and the other members of the jury, since scores awarded by an evaluated judge were not included from the ranking that was being compared Stachurska and Bartyzel (2011). This parameter is evaluated as a percentage of the disagreement of ranking in a single score by a particular judge relative to the general ranking based on the scores of the other judges excluding the evaluated judge. The ID was calculated 3 times separately, for the different morphological type of competition with more than 1 judge evaluating (2, 3 and 5 judges). Therefore, the records belonging to the competitions with only one judge were excluded in this analysis giving a total of 20,922 records, from a total of 7288 participations, held in 61 different competitions. These data belong to 4524 animals from 1042 studs.

Results

Descriptive statistics of the eight analysed traits are given in Table 3. In the PRE competitions, the average values (regardless of the type of competition) ranged between 6.9 scores (forelimb and hindlimb) and 7.9 scores (shoulder-withers and chest-thorax). In all the traits, except for the croup and tail, the competitions with 1 judge had the highest coefficients of variation (CV). The competitions with 2 judges (B) had the lowest score (3.5) and never achieved the maximum (10.0). And finally, the competitions with 5 judges (D) had the lowest range of the scale with 3 traits having only 2.5 points of variation: chest-thorax, forelimb and hind limb.

The relationships between the eight analyzed traits and the four type of competition are shown as a graphical representation of a principal components analysis in figure 1. Factor 2 absorbed 26.3% of total variance and separates the fore-region of the animal, the back-loin and morphological type of competition B and D of the rest of traits and type of competition. Whereas, factor 1 absorbed 30.9% of total variance and included type of competition B.

The influence of several external factors and their combinations on the scores given in the competitions for PRE horses were analyzed by a MANOVA and summarized in Table 4. Competition (15.7% of the total variance), judge (9.7%), judge*competition (7.5%) and judge*type of competition (2.4%) had the highest values for the percentage of variance as explained by the MANOVA analysis.

Finally, the average ID of each PRE type of competition is shown in Table 5. Competitions with 2, 3 and 5 judges had an average ID of 36.18%, 14.52% and 22.30%, respectively. Competitions with 3 judges (C) had the highest use of the scale (range of points) and the lowest ID.

Discussion

In general, it is difficult to compare evaluations of conformation traits derived from several breeds and scoring systems, due to serious differences in breeding goals and the way of scoring (Druml *et al.,* 2008). According to the criteria of the judges, the fore and hind limbs in PRE horses are the regions with less breed-quality, whereas shoulder-withers and chest-thorax are those areas with more quality related to the breed standard.

In general, this breed could be defined as a homogeneous population because of the low CV (lower than 10.00% for all analyzed traits); the highest values were shown for the hindlimb (9.44%) and the back-loin region (8.71%). As expected, the overall form was a homogeneous score, which is a global score for the whole animal. The CV were similar to those reported by Dobek *et al.* (2012) for the conformation traits in Polish Warmblood stallions; but higher than those reported by Molina *et al.* (1999) that ranged from 2.3 to 5.1 in the same breed, and slightly lower than those shown by Dietl *et al.* (2005) in Warmblood Horses.

Table 3: Descriptive statistics for the eight traits analyzed in PRE morphological competitions and grouped by the type of competitions

Trait	Type	Mean± se	Range	Reference limits*	95% CI for lower reference limit	95% CI for upper reference limit	CV %
HN	0	7.64±0.004	4.0-10.0	6.5-9.0	5.3-7.7	7.8-10.0	8.04
	A	7.50±0.020	5.0-10.0	6.0-9.0	4.6-7.4	7.6-10.0	9.63
	B	7.63±0.011	4.5-9.5	6.0-9.0	4.7-7.3	7.7-10.0	8.71
	C	7.64±0.005	4.0-9.5	6.5-9.0	5.3-7.7	7.8-10.0	7.70
	D	8.02±0.025	6.0-9.5	7.0-9.5	5.8-8.2	8.3-10.0	7.71
SW	0	7.93±0.004	4.9-10.0	6.5-9.0	5.3-7.7	7.8-10.0	7.45
	A	7.69±0.020	4.9-9.5	6.0-9.0	4.6-7.4	7.6-10.0	9.30
	B	7.95±0.010	5.0-9.5	6.7-9.0	5.6-7.8	7.9-10.0	7.09
	C	7.94±0.005	5.0-10.0	7.0-9.0	5.9-8.1	7.9-10.0	7.28
	D	8.30±0.023	6.5-9.5	7.0-9.3	5.9-8.1	8.2-10.0	6.74
CT	0	7.92±0.003	4.9-10.0	7.0-9.0	6.0-8.0	8.0-10.0	6.38
	A	7.89±0.017	5.5-10.0	6.5-9.0	5.3-7.7	7.8-10.0	7.72
	B	7.89±0.009	5.0-9.5	7.0-9.0	6.0-8.0	8.0-10.0	6.42
	C	7.92±0.004	4.9-9.5	7.0-9.0	6.0-8.0	8.0-10.0	6.20
	D	8.29±0.021	7.0-9.5	7.5-9.0	6.5-8.5	8.0-10.0	6.15
BL	0	7.23±0.004	4.0-9.5	6.0-8.5	4.8-7.2	7.3-9.7	8.71
	A	7.45±0.020	5.0-9.5	6.0-9.0	4.6-7.4	7.6-10.0	9.72
	B	7.12±0.011	4.9-9.0	6.0-8.4	4.8-7.2	7.2-9.6	8.59
	C	7.24±0.005	4.0-9.5	6.0-8.5	4.8-7.2	7.3-9.7	8.62
	D	7.22±0.021	4.0-9.0	6.0-8.0	5.0-7.0	7.0-9.0	7.21
CrT	0	7.75±0.003	4.0-10.0	6.5-9.0	5.5-7.5	7.8-9.8	6.78
	A	7.90±0.015	6.0-9.5	7.0-9.0	5.9-8.1	7.9-10.0	7.11
	B	7.69±0.009	4.9-9.5	6.5-8.6	5.4-7.6	7.5-9.7	7.17
	C	7.75±0.004	4.0-10.0	6.5-8.7	5.5-7.5	7.8-9.8	6.60
	D	8.07±0.021	6.5-9.5	7-0-9.0	6.0-8.0	8.0-10.0	6.49
FL	0	6.94±0.004	3.5-9.0	5.5-8.0	4.4-6.6	6.9-9.1	8.29
	A	6.97±0.019	4.5-8.5	5.5-8.0	4.1-6.9	6.6-9.4	10.13
	B	6.87±0.010	3.5-9.0	5.5-7.8	4.4-6.6	6.7-8.9	8.31
	C	6.96±0.004	4.0-9.0	5.5-8.0	4.4-6.6	6.9-9.1	8.15
	D	6.78±0.017	5.5-8.0	6.0-7.5	5.2-6.8	6.7-8.3	6.29
HL	0	6.91±0.004	3.5-9.5	5.5-8.0	4.2-6.8	6.7-9.3	9.44
	A	7.04±0.021	5.0-9.0	5.5-8.5	4.0-7.0	7.0-10.0	10.93
	B	6.77±0.011	3.5-9.0	5.5-8.0	4.2-6.8	6.7-9.3	9.64
	C	6.95±0.005	4.0-9.5	5.5-8.0	4.2-6.8	6.7-9.3	9.24
	D	6.65±0.018	5.5-8.0	5.5-7.5	4.6-6.4	6.6-8.4	6.72
OF	0	7.80±0.004	4.5-10.0	7.0-9.0	6.0-8.0	8.0-10.0	6.62
	A	7.90±0.016	6.0-10.0	7.0-9.0	5.9-8.1	7.9-10.0	7.19
	B	7.78±0.009	5.5-9.5	6.5-8.5	5.5-7.5	7.5-9.5	6.92
	C	7.79±0.004	4.6-10.0	7.0-9.0	6.0-8.0	8.0-10.0	6.35
	D	8.37±0.022	7.0-10.0	7.5-9.5	6.6-8.4	8.6-10.0	6.29

* Values between 2.5 and 97.5 percentile. Where: 0 overall is the value including all the competitions; A including only the competitions with 1 judge; B with 2 judges; C with 3 judges and D with 5 judges

For the abbreviations of the analyzed traits see Table 1

Table 4: Percentage of variance and significance level for the different factors on the eight morphological traits analyzed in the PRE competitions

Factors	HN	SW	CT	BL	CrT	FL	HL	OF
Section	5.019 (***)	1.669 (***)	7.004 (***)	0.892 (***)	0.819 (***)	0.617 (**)	0.584 (***)	1.329 (***)
Judge	0.210 (**)	8.394 (***)	13.247 (***)	3.542 (***)	3.560 (**)	26.151 (***)	15.809 (***)	6.282 (***)
Stud	2.133 (***)	0.996 (***)	1.873 (***)	2.147 (***)	2.044 (***)	1.234 (***)	0.801 (***)	1.822 (***)
Coat	0.917 (***)	0.181 (n.s.)	0.138 (n.s.)	0.181 (n.s.)	0.072 (n.s.)	0.509 (*)	0.196 (n.s.)	0.349 (*)
Competition	15.563 (***)	32.365 (***)	6.441 (***)	24.846 (***)	7.109 (***)	8.830 (***)	28.051 (***)	2.420 (***)
Type	0.561 (**)	0.102 (***.)	0.876 (***)	0.508 (**)	0.319 (*)	0.124 (**)	0.097 (***)	8.403 (***)
Judge*section	2.214 (***)	2.565 (***)	3.062 (***)	1.489 (***)	2.635 (***)	2.542 (***)	1.608 (***)	1.800 (***)
Judge*stud	1.271 (***)	0.982 (***)	1.604 (***)	0.702 (**)	1.057 (***)	0.726 (***)	0.286 (n.s.)	2.881 (***)
Judge*coat	0.966 (***)	0.186 (n.s.)	0.290 (n.s.)	0.208 (*)	0.398 (**)	0.161 (n.s.)	0.076 (n.s.)	0.648 (***)
Judge*competition	4.605 (***)	10.290 (***)	4.767 (***)	7.022 (***)	4.843 (***)	7.207 (***)	11.529 (***)	9.357 (***)
Judge*type	2.934 (***)	1.966 (***)	0.666 (n.s.)	1.819 (*)	2.595 (***)	2.873 (***)	1.386 (n.s.)	5.002 (***)

Where: significance level is * $p<0.05$, ** $p<0.01$, *** $p<0.001$, and ns is not significant

For the abbreviations of the analyzed traits see Table 1

When analyzed by type of competition, a high CV was shown in competitions with one judge (A) and almost a progressive decrease in CV was observed with increased number of judges. It might be the case that animals that participate in competitions with one judge are less homogeneous, because they are in low level competitions

where all kinds of animals can participate. Conversely in the highest level competitions despite having more judges (3-5 judges), therefore more likely to be more variability in assessments, the CV was the lowest. In these competitions, the judges feel free to use a bigger scale, because the scores of one judge are not going to be compared directly with the score of other judges. When there were three different judges (C) in the competition, they used the lowest range of the scale more than the highest one. The judges demonstrated high differences amongst the horses, but they were homogeneous in their appraisers when compared between them amongst themselves. This is consistent with this result, in the competitions with five judges (D), the lowest CV and the highest mean scores. This was because in this competition there were highly selected horses that go to the final championship (classified in previous competitions with lower level), and therefore they are expected to be more homogeneous and with a higher breed quality.

As in previous analysis, every trait analyzed individually, and independently of the number of judges, also showed that judges didn't use all the range of the scores and the CV was usually lower than 10%. Reference limits (95%) ranged from 2.0 to 2.5. So, it might be necessary to train judges to use all the scale in order to collect the maximum level of variation from the population. Differences between minimum and maximum scores for the different traits varied from 6.0 points for head-neck, croup-tail and hind limb to 5.1 points for shoulder-withers and chest-thorax. Similar to these results were observed by Preisinger *et al.* (1991) in Trakehner judges, Dietl *et al.* (2005) in Mecklenburger Warmblood horse judges and Schroderus and Ojala (2010) in Finnhorse and Standardbred horse judges. The use of the whole scoring range would bring up better the differences between horses (Schroderus and Ojala, 2010) and it could be described as a resource for more exactness. The lowest scores for each trait are never given, with the lowest score used in this breed being 3.5 points for limbs and the upper one of 10 points in all traits excluding limbs and back-loins region. This could be explained by the fact that judges avoid extreme categories at the beginning of a sequence to maintain their degrees of freedom, that is, to have room for further fine tune their scoring (Fasold *et al.*, 2012). In this sense, only when judges have calibrated an internal scale, that is, a transformational rule that maps external input onto an available category system, they can use extreme judgments. Since the external input is heavily context dependent, this mapping process needs a certain number of observations from them to be-judges-series (Fasold *et al.*, 2012).

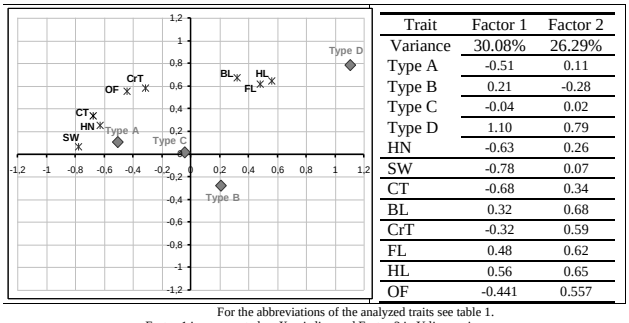
In the factor analyses, Factor 1 includes separated morphological type of competition B of the rest of body traits and type of competition; whereas, right upper quartile show a clear relationship between the scores for the conformation of the limbs (forelimb and hindlimbs) and back-loins region. Limb conformation is clearly the most serious morphological problem of the PRE horse (Molina *et al.*, 1999). Morphological type of competition D has had the highest relation with those "problematic" traits. This can be explained with the more horses participating in a morphological competition, the more relevance is given by the judges to the correct conformation of the limbs. In the left upper quartile type of competition A and C were more correlated with traits that were closely related to the breed quality of PRE horses: overall forms, head-neck, shoulder-withers and chest-thorax and croup-tail. Besides the impact on the breed quality, these regions are important for the capacity of performance of the horse. The head and neck determine athletic ability (Lawrence, 2001), back movement and stride characteristics at trot, as well as the stride length (Rhodin *et al.*, 2005). Indeed, Holmström (2001) suggested that good head-neck and neck-body insertion are more important than neck length for dressage ability. Lawrence (2001) also affirmed that the head-neck connection must be favourable to achieve free movement and flexion. Two conformation variables were analyzed in PRE to illustrate these two claims: *head-neck perimeter* and *neck-body perimeter*, both of which are correlated with biokinematic variables at trot (Sánchez *et al*., 2013).

Another important issue for study of scores in morphological competitions is the analysis of possible external factors that influence the judge's score; different from the morphology of the animal evaluated (Table 4). Due to the structure of data, sex and age were not included because the available scores were collected in different sections according to sex and age (2.2% of the total variance). The percentage of variance (8.4%) for type of competition for overall forms of the horse in this study is remarkable. It seems that the broader and subjective the trait is, the more influential is the type of competition where the horses participated for its evaluation. Besides, the system of judging, based on the simultaneous assessment of all the individuals in a section, leads judges to decide after intuitively adjusting for these external factors. Suontama *et al.* (2009) also reported the importance of age and sex because of probable differences in selection intensity between the sexes and the stronger influence of the environmental factors (nutrition, training and overall management) on conformation scores when judging adult horses or foals. The influence of sex and age on this kind of data for PRE horses was also seen in the section. It is also important to remark that the competition had the highest percentage of variance for most of the analyzed traits and judges had higher importance in the variance than the type of competition by the number of judges. The importance of judge × type and judge × competition would suggest that scores have a high subjective component due to the individual interpretation of the scoring for a given judge and also according to the competition and to the level of the competition (type of competition).

Table 5: Estimations of the indexes of disagreement between the judges in two different levels (3 types) of PRE morphological competitions, grouped by the number of judges

Trait	Type		
	B (2 Judges)	C (3 Judges)	D (5 Judges)
HN	37.76	12.29	15.44
SW	53.68	19.86	28.02
CT	42.68	16.08	20.66
BL	32.57	9.49	24.06
CrT	35.44	14.52	18.43
FL	39.91	22.82	34.78
HL	31.14	7.85	37.64
H	33.59	12.61	16.19
Mean±s.e.	36.18±8.94	14.52±4.67	22.30±9.12

For the abbreviations of the analyzed traits see table 1



Trait	Factor 1	Factor 2
Variance	30.08%	26.29%
Type A	-0.51	0.11
Type B	0.21	-0.28
Type C	-0.04	0.02
Type D	1.10	0.79
HN	-0.63	0.26
SW	-0.78	0.07
CT	-0.68	0.34
BL	0.32	0.68
CrT	-0.32	0.59
FL	0.48	0.62
HL	0.56	0.65
OF	-0.441	0.557

For the abbreviations of the analyzed traits see table 1.
Factor 1 is represented en X axis line and Factor 2 in Y lines axis

Fig. 1: Graphical representation of the principal components analysis for the eight morphological traits analyzed in the PRE competitions

According to Stefani (1998), some sports have a performance rating system in which judging plays a major role. So, the performance judgments are an inherent part of competitive sports behavior where judges perceived a stimulus, categorized it, store it in memory, and finally put it together with the retrieved memories and other available information to be integrated into a judgment and expressed as a decision in the scores (Plessner and Haar, 2006). The same psychological procedure is expected with the judges that participate in the PRE competitions, where "judge" was also one of the three main factors that influence the scores obtained by the animals (87.50%). The scores in these competitions drive subsequent re-orientation programs for individual participants but, perhaps more importantly, trends in judge behaviour shape the future of the breeding in the population (Hawson *et al.*, 2010). And therefore, because of the high influence of the judge on the scores, the reliability of judging has to be assessed.

Conformation of horses according to breed quality cannot be objectively measured. In this sense, although judges pass a lot of courses to widen their knowledge and to standardize basic principles, the scores on how to judge in a particular case are often not sufficient and difficult to apply (Stachurska and Bartyzel, 2011). Therefore, the reliability was evaluated using the methodology proposed by these authors. Although ID cannot be fully accurate, the general ranking was assumed as a proper indicator of an ideal horse, since it was decided by all the judges and no better indicators are available (Stachurska *et al.*, 2006). So, the lower ID, the more consistent a judge was with the others. An ID of 0% means that the judge was entirely consisted with the others, whereas an ID of 100% means that it is totally inconsistent. It seems that even now, the best way to enhance the quality of judging does not consist-in increasing the number of judges who participate in each event, but of estimating and verifying their scores, as suggested by Stachurska and Bartyzel (2011). It is important to point out that Stachurska *et al.* (2005) proposed the elimination of the scores from judges with an ID higher than 20%, because they assumed that these judges did not agree with the total ranking. The increasing in the number of judges who participate in high level competition (3 to 5) is not recommended, because it raises the cost of the competition and gives disagreement results. So, on average, only the competitions with 3 judges (C) have adequate reliability according to this criterion. Therefore, in order to minimize the effects of the judge and to obtain the most reliable results, verifying the competition scores by checking the quality of judging and excluding the scores of the judges who considerably disagree with the others could be the best solution. Itemize the scores to describe conformation traits in more details and widening the judging scale to cover a large range of scores would also be very useful. Therefore, the training of the judges would be very important to ensure the validity of the records collected.

Acknowledgements

The authors wish to thank the Asociación Nacional de Criadores de Caballos de Pura Raza Española (ANCCE) for making available the data used in this research. We are also most grateful to Stachurska, A. and Bartyzel, K. on University of Life Sciences in Lublin for the freely available program for the estimation of the index of disagreement.

References

Asociación Nacional de Criadores de Caballos Españoles (ANCCE), 2012. *SICAB Closes his XXII Edition Meeting their Expectations in Terms of Number of Visitors; [cited 25 November 2012]*. Available from URL: www.ancce.es/mostrar_noticia.php?orden=861

Bruns, E., M. Bierbaum, D. Frese and H.J. Haring, 1978. Die Entwicklung von Selektionskriterien für die Reitpferdezucht. IV. Schätzung relativer o¨konomischer Gewichte anhand von Auktionsergebnissen. *Züchtungskunde*, 50: 93–100

Dietl, G., S. Hoffmann and N. Reinsch, 2005. Impact of trainer and judges in the mare performance test of Warmblood horses. *Arch. Tierz.*, 48: 113–120

Dobek, A., A. Borowska, K. Molinski and T.Szwaczkowski, 2012. Entropy analysis of performance test results of warmblood stallions. *J. Anim. Feed Sci*., 21: 77–88

Druml, T., R. Baumung and J. Sölkner, 2008. Morphological analysis and effect of selection for conformation in the Noriker draught horse population. *Lives. Sci*., 115: 118–128

Fasold, F., D. Memmert and C. Unkelbach, 2012. Extreme judgments depend on the expectation of following judgments: a calibration system. *Phychol. Sport Exercise,* 13: 197–200

Hawson, L.A., A.N. McLean and P.D. McGreevy, 2010. Variability of scores in the 2008 Olympic dressage competition and implications for horse training and welfare. *J. Vet. Behav.*, 5: 170–176

Holmström, M., L. E. Magnusson and J. Philipsson, 1990. Variation in conformation of Swedish warmblood horses and conformational characteristics of elite sport horses. *Equine Vet. J.*, 22: 186–193

Holmström, M., 2001. Effects of conformation, editors. Equine Locomotion, United Kingdom: (ed), Saunders: 281-295

Lawrence, L.A., 2001. *Horse Conformation Analysis*. Cooperative Extension; Washington State University

Magnusson, L.E. and B. Thafvelin, 1985. Studies of the conformation and related traits of standardbred trotters in Sweden, *Ph. D. Thesis*, Swedish University of Agricultural Sciences, Skara

Molina, A., M. Valera, R. Dos Santos and A. Rodero, 1999. Genetic parameters of morphofunctional traits in Andalusian horse. *Livest. Prod. Sci.*, 60: 295–303

Muñoz, A., R. Santisteban, M.D. Rubio, R. Vivo, E.I. Agüera, B.M. Escribano and F.M. Castejón, 1997. Training as an influential factor on the locomotor pattern in Andalusian horses. *Zentralblatt für Veterinärmedizin. Reihe A.*, 44: 473–480

Plessner, H. and T. Haar, 2006. Sports performance judgements from a social cognitive perspective. *Psychol. Sport Exercise,* 7: 555–575

Preisinger, R., J. Wilkens and E. Kalm, 1991. Estimation of genetic parameters and breeding values for conformation traits for foals and mares in the Trakehner population and their practical implications. *Livest. Prod. Sci*., 29: 77–86

Rhodin, M., C. Johnston, KR. Holm, J. Wennerstrand and S. Drevemo, 2005. The influence of head and neck position on kinematics of the back in riding horses at the walk and trot. *Equine Vet. J*., 37: 7–11

SAS, 2001. *Sas/Statistical Users Guide for Personal Computer*. release 6.18.SAS. Institute, Inc, cary, N, C, USA

Sánchez, M.J., M.D. Gómez, F. Peña, J.L. Morales A. Molina and M. Valera, 2013. Relationship between conformation trait and gait characteristics in Spanish Purebred horses, *Arch. Tierz.,* 13: 137–148

Schroderus, E. and M. Ojala, 2010. Estimates of genetic parameters for conformation measures and scores in Finnhorse and Standardbred foals. *J. Anim. Breed. Genet*., 127: 395–403

Schwark, H.J., P. Petzold and I. Nörengerg, 1988. Untersuchungen zur Auswahl von Selektionskriterien bei der Weiterentwicklung der Pferdezucht der DDR. *Arch. Tierz*., 31: 279–289

Stachurska, A. and K. Bartyzel, 2011. Judging dressage competitions in the view of improving horse performance assessment. *Acta Agric. Scan. Section A – Anim. Sci.*, 61: 92–102

Stachurska, A., J. Niewczas and M. Markowski, 2005. *An Estimation of Reliability of Judging the Horse Dressage Competitions*. 56th Annual Meeting of the European Association for Animal Production June 5-8 2005, Uppsala, Sweden

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

Stefani, R., 1998. Predicting outcomes. *In*: *Statistics in Sport*, pp: 249–275. Bennet, J. (ed.). Arnold, London

Suontama, M., M.T. Saastamoinen and M. Ojala, 2009. Estimates of non-genetic effects and genetic parameters for body measures and subjectively scored traits in Finnhorse trotters. *Livest., Sci.,* 124: 205–209

(Received 04 April 2013; Accepted 28 May 2013)