The stability of conformation and movement traits evaluation tested in cold-blooded horses of different endangerment status

Grażyna Maria Polak and Dorota Lewczuk

ABSTRACT
Proper assessment of the horse conformation is fundamental for the proper breeding and progress in the breeds undergoing improvement and/or maintaining the right level of traits in conservation breeds, as it is the main and the first-achieved element of selection. The objective of the study was to analyse the stability of traits assessment at the show of cold-blooded horses by individual judges by analysing the factors that influenced the results. The analysis of variance was conducted on scores of 93 horses of different endangerment status, judged at the same horse show by 6 judges. The fixed effects of sex, breed of the sire and dam, type of breeder (state, national) and age class were taken into account. The Pearson correlations were calculated between scores for individual judges and the mean score. The obtained results showed significant effects of the breed of parents and type of breeder on the scores of individual judges. The trait ‘body condition’ was the most difficult trait to evaluate and the ‘trot’ the easiest one. The ‘trot’ was the trait most dependent on genetic endangerment status of the horse pedigree. New definitions for these traits should be established for the needs of conservational programmes.

1. Introduction
Proper assessment of the horse conformation being the first possible horse evaluation is fundamental for making proper breeding evaluation, achieving further progress in the breeds undergoing improvement or conservation programmes. Selection on horse conformation is done at different stages of breeding – the assessment of foals, horses entering into the stud book or the evaluation at breeding shows and exhibitions seems essential in both kinds of breeding programmes (developmental and conservational). Every assessment should be as objective as possible; so many studies have explored the possibility of improving the assessment methods and analysed the correctness of scoring, the relationship between trait measurements (Kapron et al. 2003, 2005; Lewczuk 2008), or the introduction of new, more detailed scoring systems (Becker et al. 2013; Folla and Mantovani 2013). However, more objective systems also have their limitations, such as the risk of omitting an important detail, which, even if unnamed as the separate trait, may subconsciously determine the subjective scores of judges and be positively related to the future productive results of the horses. Some publications point to the inevitable subjectivity of assessment and the fact that evaluators unintentionally give in to old frames of thinking (Caspar et al. 2015). The accuracy of scores is particularly relevant when restoring old breeds or types that are phenotypically similar. This situation occurs in the endangered populations of Sztumski and Sokolski horses, which show considerable similarities because they had been intensively crossed with Ardennes stallions in the past (Polak 2013). Therefore, the improvement of judges’ scores is essential for improving horses. The experience of the judges and breeding commissions allows scoring the present and future conditions and potential of individual horses, even if individual differences in scoring are noted in almost every study that investigated this aspect (Weaver and Stewart 2012). The first step in improving the quality of scoring is to know the differences in scoring and the tendencies in judging the horses (Lewczuk 2013). The objective of the study was to analyse the stability of traits assessment at the show of cold-blooded horses by individual judges by analysing the factors that influenced the results. The effect of different environmental (type of breeder, class and age) and genetic (breed of sire, breed of dam) factors on the show scores of cold-blooded horses representing four different breed groups, including those covered by the genetic resources conservation programme, was analysed. The hypothesis that there are no differences between evaluation of different conformation traits in horses of endangerment status and improved pedigree is tested. All horses were evaluated according to the official Polish system of evaluation expressed in points.

2. Materials and methods
2.1. Horses and traits
The conformation evaluations of 93 cold-blooded horses scored during championship show in four different age and sex classes (23 one-year-old colts, 29 two-year-old colts, 20 one-year-old
fillies, and 21 two-year-old fillies) were analysed. Most of the horses came from private breeders (86%) and the others came from state-owned herds and studs. A large majority of the horses descended from stallions registered in the Polish Horse Breeders Association database as Polish cold-blooded horses (54 head), followed by Ardennes horses (15 head), other foreign cold-blooded breeds (12 head), and endangered Sztumski and Sokólski horses under genetic resources conservation (7 head). A similar pattern was found for the proportion of dams of the studied horses, with 68 horses from Polish cold-blooded mares, 10 horses from Ardennes mares, 6 horses from other foreign cold-blooded horses, and 4 horses from endangered cold-blooded horses. The scores were collected independently from six judges, who evaluated the horses in the different age and sex classes. Because one of the judges always served as a reserve judge, the results were not always given for each judge in each scoring class, but judges were identified between classes. The following traits were scored: ‘type’ – breed and sex type, according to the breeding programme; the quality of the tissues and frames of the horse; ‘conformation’ – the quality of horse body build, correctness of different body parts, especially head with the neck, trunk, croup, legs, and feet; ‘walk’ – quality of walk; conspicuous and symmetrical steps, energy and elasticity of movement; ‘trot’ – quality of trot; elastic, rhythmical, energy of movement; ‘condition and management’ – quality of body condition and preparation for the exhibition; ‘final result’ – sum of points from other traits. Traits were evaluated on the scale of 1–10, where 10 is excellent and 1 very bad. The accuracy of scoring was 0.5 point. Such kinds of the data are treated statistically as normally distributed.

2.2. Statistical methods

In order to determine the influence of different effects on judges’ scores, analysis of variance was performed for the mean official score and for single scores of individual judges, with consideration of the following effects: sex, age class, breed of sire, breed of dam, and type of breeder. Such method allowed us to study the influence of effects on individual judges’ scores and their style of judging. Differences between certain levels of factors were analysed with t-test for least squares means. GLM procedure of the SAS package was used with five fixed factors. The following model was applied:

\[ Y_{\text{abcd}} = \mu + P_a + K_b + O_c + M_d + H_e + e_{\text{abcd}}, \]

where \( Y_{\text{abcd}} \) – judges’ score; \( \mu \) – population mean; \( P_a \) – sex (\( a = 1, 2 \)); \( K_b \) – age class (\( b = 1, 2 \)); \( O_c \) – breed of sire (\( c = 1, \ldots, 4 \)); \( M_d \) – breed of dam (\( d = 1, \ldots, 4 \)); \( H_e \) – type of breeder (\( e = 1, 2 \)); \( e \) – error.

Additional calculations were done with the regression on the age on days within classes. In order to determine the relationship of individual scores among the judges, a Pearson correlations were performed between judges’ scores for every individual trait using the CORR procedure of the SAS package, as well as the correlation of every judge with the mean of all judges for that trait.

3. Results

Analysis of variance of the mean scores of the horses conformation evaluation (Table 1) showed that the most important effect that influenced the results was sex, which was highly significant for the ‘conformation’ (\( p < .0001 \)) and ‘type’ scores (\( p = .0019 \)), as well as significant for the final result (\( p = .0189 \)). In almost all cases, stallion scores were higher than mare scores (Table 2). Type of breeder (private, state) had an effect on two traits of the horses. For both traits (‘type’ (\( p = .02 \)) and ‘trot’ (\( p = .04 \))), higher scores were given to state-bred horses. As regards the effect of breed, statistically significant differences for both the breed of sire and the breed of dam were obtained for the ‘trot’ score (0.019 and 0.020, respectively). In both cases, horses born of foreign parents (other cold-blooded foreign breed for the breed of sire and Ardennes breed for the breed of dam) obtained higher scores for ‘trot’. The highest values of the ‘trot’ were received by horses sired by foreign stallions (8.87 points) that were significantly different for values received for horses sired by Arden stallions (8.23), Polish cold-bloods (8.39), and endangered sires (8.28). Almost the same differences were noted for the breed of dams, as horses with foreign and Arden dams were evaluated higher (8.79–8.80) than horses with Polish cold-bloods or endangered dams (7.85). Age class had no effect on the scores given to horses, whereas regression on age within class was not significant.

Analysis of the effects (\( p \)-values) of the studied influences on scores given by individual judges and the mean of their notes is shown in Table 1. The highest agreement between the significance of effects on judging (\( p \)-values) was obtained for the influence of sex on ‘conformation’ and ‘type’ scores, as well as for the effect of parents’ breed on ‘trot’ score. The lowest agreement for the \( p \)-values of significant effects on different judges’ notes was obtained for the type of breeder on ‘type’ and ‘trot’ scores. The effect of type of breeder was significant for ‘trot’ score in evaluation of three judges, and non-significant based on the scores of the other three judges. Also, the ‘body condition’ score was sex dependent for two judges and sex independent for the other judges. The obtained results suggest that judges took different approaches to different traits of the horses, and their preferences were recognizable.

Additional information about the compatibility between judges’ scores is provided by phenotypic correlations between the scores of individual judges and the correlations between judges’ scores and the mean score of all of them. The highest agreement for the scores was observed for the ‘type’ scores, where correlations range from 0.59 to 0.90 between the judges, and from 0.72 to 0.95 between the judges and the mean score (Table 3). The lowest agreement was estimated for the ‘body condition’ score, where the correlations between judges ranged from 0.04 to 0.77 and the correlations for the scores of individual judges with the mean score varied between 0.26 and 0.88.

4. Discussion

Results obtained for the effect of sex are in line with expectations. The higher scores for stallions are fully justified considering the higher level of male selection that has been used for
years in the horse breeding all over the world. Differences between evaluations for different sexes should be seen for all traits as a result of different degree of selection in both sexes. Much more rigorous selection of stallions should be performed for movement traits especially in the case of breeds of genetic resources. The lack of significant differences between the age classes of investigated horses seems to be the proper direction in the horse evaluation. This indicates the same tendency of assessment regardless of the horses’ age, which is not always applied in equine breeding (Wejer and Lewczuk 2016). The effect of type of breeder for the traits differing in this regard – the ‘type’ and ‘trot’ score – varied considerably for individual judges.

The effect of breed of the parents proved to be significant and should confound Polish breeders, because the significant differences concerned the ‘trot’ quality, which is a trait associated with exercise performance ability of the horses. The fact that currently this is not the main use of cold-blooded horses does not entitle breeders to waste the quality of this trait in the population. The trait should be particularly improved in the breeding of protected breeds because the changes introduced in 2015 to the conservation programmes of Sztumski and Sokólski horses indicate that the main goal of breeding the native types of cold-blooded horses is their draft use in agrotourism and organic farming (IZ PIB 2015a, 2015b). Despite its relatively low level, evaluation of the trot in horses sired by stallions under the conservation programme is at a higher level than in horses born from dams under the conservation programme. This fact indicates the high potential and selection potential for the protected horses and allows for rapid progress.

The study on the evaluation of type, which is the most important for restoring the Sztumski and Sokólski horses, shows the significant effect of the breed of sire on the detailed scores of individual judges. The interesting result was obtained for the effect of type of breeder. The cold-blooded horses were always characterized by a disproportion between animals kept in state-owned herds and studs, and private breeders (Chrzanowski et al. 1989; Chrzanowski 2008), where, in order to improve the population, stallions were imported from countries with a high level of cold-blooded horse breeding (Germany, France, and Sweden). Our results support this statement. The results obtained are dominated by state horses. This may be due to two reasons – either the studied population constituted only one champion year group and the results are characteristic of a given year, or the judges commission had its own (perhaps subconscious) preferences. The second reason seems more probable because the effect of type of breeder of the traits differing in this regard – the ‘type’ and ‘trot’ score – varied considerably for individual judges, as described above. Most of the traits were judged in harmony, but it seems that the effect of breeder should be investigated more thoroughly because the existence of judges’ preferences has been shown. Also the effect of the breed of the horse should be investigated on other populations more detailed as it can mean not only horse quality, but also judges’ preferences of different breeds, as in most shows the horse catalogues are presented with pedigrees or horses grouped by the sire lines. The possibility of identification of judges’ preferences obtained in

### Table 1. The influences of effects on individual assessment of judges and mean scores (p-values).

| Trait             | Breeder | Breed of sire | Breed of dam | Sex | Class |
|-------------------|---------|---------------|--------------|-----|-------|
| Type 1st judge    | 0.3550  | 0.6369        | 0.8084       | 0.0187 | 0.1814 |
| Type 2nd judge    | 0.0815  | 0.3381        | 0.1934       | 0.0004 | 0.7715 |
| Type 3rd judge    | 0.0198  | 0.9785        | 0.5440       | 0.0104 | 0.5962 |
| Type 4th judge    | 0.1300  | 0.6700        | 0.0829       | 0.0023 | 0.3639 |
| Type 5th judge    | 0.0678  | 0.4829        | 0.3229       | 0.1820 | 0.6751 |
| Type 6th judge    | 0.6526  | 0.6783        | 0.4660       | 0.1701 | 0.1499 |
| Type, All         | 0.0474  | 0.7917        | 0.5390       | 0.0019 | 0.2790 |
| Conformation 1st judge | 0.2170  | 0.9777        | 0.4793       | 0.0016 | 0.0564 |
| Conformation 2nd judge | 0.9345  | 0.2753        | 0.0417       | 0.0002 | 0.5001 |
| Conformation 3rd judge | 0.1143  | 0.2680        | 0.5368       | 0.0002 | 0.9478 |
| Conformation 4th judge | 0.9119  | 0.3413        | 0.2322       | 0.0010 | 0.8926 |
| Conformation 5th judge | 0.9081  | 0.3526        | 0.6274       | 0.0002 | 0.2870 |
| Conformation 6th judge | 0.2098  | 0.3086        | 0.1945       | Not available | 0.1412 |
| Conformation, All | 0.6933  | 0.2825        | 0.2138       | 0.0001 | 0.4236 |
| Walk 1st judge    | 0.3700  | 0.5815        | 0.5220       | 0.2580 | 0.0251 |
| Walk 2nd judge    | 0.4101  | 0.1095        | 0.5472       | 0.1655 | 0.3875 |
| Walk 3rd judge    | 0.2257  | 0.2286        | 0.1625       | 0.0180 | 0.4191 |
| Walk 4th judge    | 0.7147  | 0.3011        | 0.5621       | 0.7126 | 0.0412 |
| Walk 5th judge    | 0.2520  | 0.4302        | 0.6836       | Not available | Not available |
| Walk 6th judge    | 0.5101  | 0.6648        | 0.6712       | Not available | Not available |
| Walk, All         | 0.8064  | 0.1613        | 0.3306       | 0.3914 | 0.1184 |
| Trot 1st judge    | 0.0429  | 0.0356        | 0.0086       | 0.4180 | 0.0798 |
| Trot 2nd judge    | 0.1502  | 0.0596        | 0.0077       | 0.1194 | 0.6122 |
| Trot 3rd judge    | 0.0102  | 0.0854        | 0.0123       | 0.6900 | 0.9980 |
| Trot 4th judge    | 0.3039  | 0.0190        | 0.0130       | 0.7825 | 0.3826 |
| Trot 5th judge    | 0.0112  | 0.0280        | 0.0119       | 0.9483 | 0.8712 |
| Trot 6th judge    | 0.4792  | 0.5531        | 0.0665       | Not available | Not available |
| Trot, All         | 0.0225  | 0.0199        | 0.0020       | 0.8325 | 0.4438 |
| Body condition 1st judge | 0.8554  | 0.9351        | 0.1391       | 0.3996 | 0.1499 |
| Body condition 2nd judge | 0.3275  | 0.8300        | 0.6961       | 0.3059 | 0.6209 |
| Body condition 3rd judge | 0.1378  | 0.8227        | 0.4536       | 0.1516 | 0.7110 |
| Body condition 4th judge | 0.4460  | 0.8464        | 0.3232       | 0.0030 | 0.6878 |
| Body condition 5th judge | 0.5917  | 0.2366        | 0.0212       | 0.0454 | 0.1701 |
| Body condition 6th judge | 0.2263  | 0.0266        | 0.7240       | Not available | Not available |
| Body condition, All | 0.3201  | 0.7988        | 0.3734       | 0.0744 | 0.2854 |
the same letters show significant differences between groups. Statistically significant differences in columns within effect are marked with small letters, p ≤ 0.05; capital letters, p ≤ 0.01.

LSM – least squares means and SE – standard errors.

our study could help in the schooling process for official people involved in the evaluation of horses.

The well-known problems in the assessment of animals include the different perceptual abilities of the judges and the differences between the ease of evaluating individual traits (Weaver and Stewart 2012). According to the breeding rules, two traits can be considered the same, if the genetic correlations between them reached the value above 0.8. If we had

Table 3. Correlation coefficients between the scores of individual judges and the composite score (above diagonal), and significance of these correlations (below diagonal).

Score for the trait 1st judge 2nd judge 3rd judge 4th judge 5th judge 6th judge Mean

Type 1st judge x 0.59 0.65 0.62 0.64 0.73 0.72
Type 2nd judge 0.001 x 0.80 0.83 0.68 0.73 0.86
Type 3rd judge 0.001 0.001 x 0.88 0.81 0.90 0.95
Type 4th judge 0.001 0.001 0.001 x 0.80 0.88 0.94
Type 5th judge 0.001 0.001 0.001 0.001 x not available 0.87
Type 6th judge 0.001 0.001 0.001 0.001 x not available 0.92
Conformation 1st judge x 0.57 0.58 0.67 0.70 0.32 0.74
Conformation 2nd judge 0.001 x 0.79 0.83 0.70 0.53 0.85
Conformation 3rd judge 0.001 0.001 x 0.83 0.77 0.60 0.89
Conformation 4th judge 0.001 0.001 0.001 x not available 0.94
Conformation 5th judge 0.001 0.001 0.001 0.001 x 0.53 0.90
Conformation 6th judge 0.006 0.001 0.001 0.001 not available 0.001 x 0.73
Walk 1st judge x 0.62 0.47 0.67 0.81 0.43 0.76
Walk 2nd judge 0.001 x 0.74 0.80 0.86 0.64 0.93
Walk 3rd judge 0.001 0.001 x 0.62 0.81 0.34 0.79
Walk 4th judge 0.001 0.001 0.001 x 0.84 0.75 0.89
Walk 5th judge 0.001 0.001 0.001 0.001 x not available 0.91
Walk 6th judge 0.001 0.001 0.001 0.001 x not available 0.74
Trot 1st judge x 0.72 0.76 0.82 0.79 0.70 0.88
Trot 2nd judge 0.001 x 0.81 0.95 0.80 0.55 0.91
Trot 3rd judge 0.001 0.001 x 0.84 0.76 0.55 0.88
Trot 4th judge 0.001 0.001 0.001 x not available 0.96
Trot 5th judge 0.001 0.001 0.001 0.001 x 0.70 0.89
Trot 6th judge 0.001 0.001 0.001 not available 0.001 x 0.84
Body condition 1st judge x 0.48 0.67 0.54 0.04 0.27 0.75
Body condition 2nd judge 0.004 x 0.66 0.77 0.31 0.49 0.77
Body condition 3rd judge 0.127 0.001 x 0.69 0.14 0.38 0.88
Body condition 4th judge 0.001 0.001 0.001 0.001 x 0.50 0.88
Body condition 5th judge 0.001 0.001 0.001 0.001 x 0.42 0.26
Body condition 6th judge 0.001 0.001 0.001 0.001 0.001 x 0.58
used this rule for the phenotypic correlation between the scores of individual judges for the same trait – no one of investigated traits could be considered as uniform. However, according to the scale used in other research on agreement of assessment, the values above 0.4 (in the scale of 0–1 of weighted kappa value) are good and above 0.75 very good (Fuller et al. 2006). Therefore, most of the values between individual judges are positive in accordance to such scaling.

Based on our study of the agreement of correlations between judges’ scores, the ‘body condition’ is the most difficult trait to assess. Such differences can result from different approaches to horse breeding directions; however, the element of such importance – the basic trait of slaughter horses (Mantovani et al. 2014) – should be evaluated more equally. The body condition is very important because it largely reflects the health status of animals (Martinson et al. 2014). The optimum assessment of young horses seems particularly important because improvements in the breeding efficiency of cold-blooded horses are seen in shortening the generation interval (Mantovani et al. 2013), especially as the expression of some traits in cold-blooded horses changes with age, which means that it is more heritable in young horses (Mantovani et al. 2010; Suontama et al. 2011). The type is a trait characterized by greater scoring conventionality both in our study and in other studies (Druml et al. 2015). The scores and correlations obtained are largely dependent on within-breed variation in the studied traits (Komosa et al. 2013), which was rather limited in the present study.

5. Conclusions
In conclusion, the preferences of judges’ evaluation were observed; so the method can be used to determine and control the objective evaluation of horses. The trait ‘body condition’ seems the most difficult to assess and the trait ‘trot’ score is the easiest one in coldblood horse analysed in this study. It seems that clarification of definitions for different evaluated traits should be established.

Disclosure statement
No potential conflict of interest was reported by the authors.

References
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(1):11–21.

Caspar GL, Dhand NK, McGreevy PD, Pritchett-Corning KR. 2015. Human preferences for conformation attributes and head-and-neck positions in horses. PLoS ONE. 10(6):e0131880.

Chrzansowski Sz. 2008. Rody męski w hodowli koni zimnokrwistych. Hodowca i Jeździec VI nr. 2(17):12–16.

Chrzansowski Sz, Chachula J, Szalgowska-Wąsik U, Oleksiak S, Wilczak J. 1989. Konie zimnokwiste w Polsce środkowej, środkowo-wschodniej i południowej. Warszawa: PWN.

Druml T, Dobretzberger M, Brem G. 2015. The use of novel phenotyping methods for validation of equine conformation scoring results. Animal. 9(6):928–937.

Folla F, Mantovani R. 2013. Genetic parameters of linear type traits scored at adult age in Italian heavy draught horse. Ital J Anim Sci. 12 s.1:C-015.

Fuller CJ, Bladon BM, Driver AJ, Barr ARS. 2006. The intra- and inter-assessor reliability of measurement of functional outcome by lameess scoring in horses. Vet J. 171:281–286.

Instytut Zootechniki PIB. 2015a. Program ochrony zasobów genetycznych koni sztumskich.

Instytut Zootechniki PIB. 2015b. Program ochrony zasobów genetycznych koni sokólskich.

Kapron M, Janczarek I, Suska A 2005. Próba oceny współzależności między dwoma systemami bonitacji pokroju ogierów półkrwi i ilościowymi parametrami ich skoków pod jeźdźcem. Roczn Nauk PTZ. 11(1):45–56.

Kapron M, Kapron H, Litwinczuk A, Florek M, Sochacki S, Scalecki P. 2003. Wpływ wielkości różnic między wymiarami wybranych partiń ciała na wskaźniki rżennej przydatności koni różnych typów pochodzeniowych. Roczn Zauf Nauk Zool., Supl. 18:75–78.

Komosa M, Frackowiak H, Purzyck H, Wojnowska M, Gramacki A, Gromacki J. 2013. Differences in exterior conformation between primitive, Half-bred, and Thoroughbred horses: anatomic-breeding approach. J Anim Sci. 91(4):1660–1668.

Lewczuk D. 2008. Analiza systemu sędziowania zdolności skokowych koni w skokach luzem za pomocą komputerowej analizy obrazu. Monografie Rozprawy IGHZ PAN Jastrzębiec Nr.21.

Lewczuk D. 2013. Effect of the judge and definition of the trait for horse free jumping evaluation. Archiv Tierz. 56(6):638–649.

Mantovani R, Guzzo N, Sartori C, Bailoni L. 2014. In vivo performance of Italian heavy draft horse weanlings fed two protein levels and slaughtered at two ages. J Anim Sci. 92:4998–5008.

Mantovani R, Sartori C, Pigozzi G. 2010. Genetcs of temperament and productive traits in the Italian heavy draught horse breed. www.kongressband.de/wcgalp2010/assets.

Mantovani R, Sartori C, Pigozzi G. 2013. Retrospective and statistical analysis of breeding management on the Italian Heavy Draught Horse breed. Animal. 7(7):1053–1059.

Martinson KL, Coleman RC, Rendahl AK, Fang Z, McCue ME. 2014. Estimation of body weight and development of a body weight score for adult equids using morphometric measurements. J Anim Sci. 92:2230–2238.

Polak G. 2013. Problemy związane z prowadzeniem programów ochrony zasobów genetycznych koni w typie sokólskim i sztumskim, w latach 2008-2011. Roczn Nauk Zoot. 39(1):47–59.

Suontama M, van der Werf JH, Juga J, Ojal M. 2011. The use of foal and studbook traits in the breeding programmes of Finnhorse and Standardbred trotters. J Anim Breed Genet. 128:114–123.

Weaver EA, Stewart TR. 2012. Dimensions of judgment: Factor analysis of individual differences. J Behav Dec Making. 25:402–413.

Wejer J, Lewczuk D. 2016. Effect of the age on the evaluation of horse conformation and movement. Ann Anim Sci. 16:36.