Applied Animal Behaviour Science 234 (2021) 105184

Contents lists available at ScienceDirect

Applied Animal Behaviour Science

journal homepage: www.elsevier.com/locate/applanim

Rider effects on horses' conflict behaviour, rein tension, physiological measures and rideability scores

Janne Winther Christensen [a], Rikke Munk [b], Lesley Hawson [c], Rupert Palme [d], Torben Larsen [a], Agneta Egenvall [e], Uta U. König von Borstel [f], Maria Vilain Rørvang [g,*]

[a] Aarhus University, Department of Animal Science, Blichers Allé 20, 8830 Tjele, Denmark
[b] Højgaard Hestehospital Odense, Rugårdsvej 696, 5462 Morud, Denmark
[c] "Safer Horses", Kilmore, Victoria 3764, Australia
[d] University of Veterinary Medicine, Department of Biomedical Sciences, Veterinärplatz 1, 1210 Vienna, Austria
[e] Swedish University of Agricultural Sciences, Department of Clinical Sciences, Veterinary Epidemiology Unit, Ulls väg 26, 750 07 Uppsala, Sweden
[f] Justus-Liebig University, Institute of Animal Breeding and Genetics, Section Animal Husbandry, Behaviour and Welfare, Leihgesterner Weg 52, 35392 Giessen, Germany
[g] Swedish University of Agricultural Sciences, Department of Biosystems and Technology, Sundsvägen 16, 230 53 Alnarp, Sweden

A R T I C L E I N F O

Keywords:
Equitation science
Heart rate
Horse riding
Sport horse
Temperament
Training cues

A B S T R A C T

Many breeding organisations include a subjective scoring of rideability by a professional rider into their evaluation of sports horses, but the consistency and reliability of the scoring system is debateable. The aim of this study was to investigate (i) whether professional riders agree in their scoring of rideability, and (ii) whether rideability scores are affected by rein tension, horse conflict behaviour, heart rate, and salivary cortisol, and (iii) whether riders induce different levels of conflict behaviour and physiological responses in the horses. Ten professional, female riders each rode 10 dressage horses (level M German scale; $n = 100$ combinations) through a standardised dressage test (10 min warm-up followed by a 4-min test) and subsequently scored the horses for rideability on the official 1–10 scale (1 = poor to 10 = excellent) from the Danish Riding Federation. Rein tension, horse heart rate, saliva cortisol and conflict behaviour were measured for each rider-horse pair. The riders were inconsistent in their scoring of rideability to the individual horses, e.g. scores for one of the horses ranged from 1 to 8. There was a significant effect of rider ($P = 0.003$) and the frequency of conflict behaviour (undesired head movements: $P < 0.001$, breaking the gait: $P = 0.013$, and other evasive behaviour: $P = 0.032$) on rideability scores, i.e. the more conflict behaviour the lower the score. There was no significant effect of rein tension and the physiological measures on rideability scores. However, there was a significant effect of rider on rein tension, horses' heart rate and increases in saliva cortisol concentrations and a tendency for some types of conflict behaviour, suggesting that some riders induced more discomfort in the horses. Future studies could help shed light on which elements of riding style are particularly important for sports horse welfare. In conclusion, this study found a large variation in rideability scores assigned to ten sports horses by ten professional riders. Rideability scores were dependent on the level of horse conflict behaviour, but not rein tension and physiological measures. Further studies are needed to improve the objectivity, consistency and reliability of rideability assessment of sports horses.

1. Introduction

Behavioural traits of horses are important for safe horse-human interactions (Hawson et al., 2010) as well as animal welfare (e.g. Schork et al., 2018). Accordingly, the majority of riders consider the horse's behaviour during riding and training to be of major importance (Górecka-Bruzda et al., 2011; Graf et al., 2013). Rideability is described as the degree of comfort a rider feels when riding a horse and the ease with which a horse can be ridden (König von Borstel et al., 2013; König Von Borstel and Glißman, 2014). This particular trait has been suggested to be one of the most important traits when assessing a horse, by both riders (Górecka-Bruzda et al., 2015) and breeders (Teegen et al., 2008). It is common practice to assign a rideability score to sports horses during performance testing. In a study on 234, 3-yr old Danish Warmblood

* Corresponding author.
E-mail address: mariav.rorvang@slu.se (M.V. Rørvang).

https://doi.org/10.1016/j.applanim.2020.105184
Received 8 June 2020; Received in revised form 15 November 2020; Accepted 22 November 2020
Available online 8 December 2020

horses, weak negative correlations were found between rideability scores and reactivity, measured as evasive behaviour during an official conformation evaluation, suggesting less reactive horses were easier to ride (Rothmann et al., 2014). Although rideability is aimed to be a standardised evaluation, streamlined among professional riders and judges, the measurement is inevitably affected by the level and quality of prior training and individual preferences. The performance test commonly involves the horse being ridden for a brief period by a professional test rider who then assigns a score for rideability for that horse. The test rider is appointed by the national equestrian federation conducting the tests. There is no calibration between riders prior to the horse assessment process.

Ground judges may also score rideability based on observed communication between rider and horse, e.g. through rein contact and the riders' legs and seat (König Von Borstel and Glißman, 2014). However, there appears to be limited agreement among test riders and judges when applying existing scoring guidelines for rideability (König von Borstel et al., 2013), and it is debatable, if rideability can be considered a stable temperamental trait and as such, it should be re-considered whether it is appropriate to include it into the evaluation of sports horses. In a survey with 1087 riders (competition 49 %, leisure 38 %, and professional 9%), the conclusion was a need for more objective assessment methods of horses' temperament, and the survey participants supported a restructuring of the current assessment (Graf et al., 2013). Thus, there is a need to further investigate to which extent professional riders agree in their scoring of rideability to the same horses as well as to implement more objective measures to assess this particular trait.

This study aimed to investigate (i) whether professional riders agree in their scoring of rideability, and (ii) whether rein tension, horse conflict behaviour, heart rate and saliva cortisol reflect the scores given by riders, and (iii) whether riders induce different levels of conflict behaviour and physiological responses in the horses.

2. Materials and methods

2.1. Funding, ethics and consent

The study was conducted on a private facility in Denmark in December 2013 and was partly funded by the Danish Horse Levy Foundation. It conformed with national legislation on animal experimentation (Danish Ministry of Justice, Act. no. 253 and §12 in Act. no. 1459), the ARRIVE guidelines (Kilkenny et al., 2010) and the guidelines by the Ethical Committee of the ISAE (International Society of Applied Ethology) (Duncan et al., 2013). All riders and horse owners gave written consent for inclusion in the study.

2.2. Riders and horses

Ten professional, female riders (age: mean ± s.e.: 39.3 ± 1.6 (range: 29–45 years), weight: 63.7 ± 1.9 (range: 53–70 kg)) took part in this study. They had 30 ± 2.0 years of riding experience (range: 20–40 years). All riders were right handed. The ten riders each rode ten dressage horses through a standardised dressage test, resulting in 100 rider-horse combinations. The ten privately owned warmblood horses (age: mean ± s.e.: 9.0 ± 0.8 (range: 6–13 years), four mares and six geldings; height at withers: 166 ± 2.9 cm (range 150–175)) were all educated to the second highest national level in dressage (Level M). They were either stabled at the facility (n = 4) or used to being trained at the facility (n = 6). Two riders participated on each of five test days, i.e. each horse was ridden by two riders per day (Test day 1: Rider 1 and 2; Test day 2: Rider 3 and 4 etc.). One horse unfortunately showed dangerous behaviour during Test day 2, apparently as a reaction to the test equipment, and the professional rider advised this horse should not continue in the experiment to ensure rider safety. Thus, this particular horse was excluded from the data analysis as data were only obtained during the first test day. A replacement horse was recruited for Test days 3–5 (i.e. for riders 5–10), resulting in a total of n = 96 rider-horse combinations to be analysed.

The entire experimental period lasted 10 days to ensure a maximum of two subsequent test days for the horses. The horses wore their usual saddle and bridle (with single- or double-jointed snaffle bit) with all riders.

2.3. Test protocol

The first saliva sample was collected prior to saddling (Sample A, baseline). Horses arriving by trailer (n = 6) were stabled in a box for at least 30 min before the sample was collected. The horse was then saddled by the usual rider and heart rate equipment was attached to the saddle and girth. The horse was led to the indoor riding arena, and warmed up by its usual rider for 15 min after which a rein tension meter was attached. The test was then initiated in the same indoor riding arena. The professional rider was allowed a 10 min warm-up period before riding the horse through a standard dressage test (approx. 4 min duration, just below M-level; containing exercises such as shoulder-in, leg-yield and extended gait in trot, canter and walk). According to the results in Christensen et al. (2014), saliva samples were collected 0 and 5 min after the test (Sample B and C). After the first rider finished her test, the horse was walked by hand for 5 min after which the second rider mounted and had a 10 min warm-up period before her test. Saliva samples were collected again at 0 and 5 min after the second rider finished the test (Sample D and E). After the second rider, the test day for that particular horse was terminated and the usual rider walked their horse by hand during a cool-down phase before leading it back to the stable to be groomed. In total, the horses were worked for approx. 50 min/day.

2.4. Recordings

2.4.1. Rideability scoring by riders

After each test ride, the professional riders scored the horses for rideability, quality of gaits, and capacity on the official 1–10 scale (from 1 = poor to 10 = excellent) from the Danish Riding Federation. All the professional riders were familiar with scoring horses for these traits according to this scale.

2.4.2. Rein tension

Rein tension (RT) was measured using the equipment and software from SignalScribe (Version 2.6; Crafted Technology). The equipment consisted of 2 lightweight load cells (3 × 4 cm, weight 60 g with measuring range up to 5 kg, i.e. 49 N (N)), one for each rein, connected via a custom interface to a data logger, positioned under the chin and attached to the bridle and noseband by Velcro straps. Rein tension data were sampled at a rate of 100 Hz and recorded to a mini SD card (Transcend, 1 Gb). After each test, the data were downloaded to a PC, using the SignalScribe data capture software. The sensors were calibrated each test day by hanging known weights (0.5–5 kg) vertically from the reins and measuring the tension (SignalScribe, Help Manual, Version 2.04.01, see further description in Christensen et al. (2011).

Rein tension was calculated as the mean load (N) on the left and right rein, for each rider-horse pair. Approx. 10 % of the data were above the upper detection limit of the load cells (49 N), i.e. the tension was higher than approx. 5 kg, and the calculated means are therefore underestimates. Data from ten pairs were lost due to technical issues with the equipment, and for four pairs data were only available for the left rein. These data were included in the analysis resulting in n = 86 for RT data.

In addition to the measured rein tension, riders were also asked to assess their impression of the rein tension on a scale from 1 to 10 (1 = loose/light on the reins, to 10 = heavy on the reins), for each horse.

2.4.3. Behavioural recordings

All tests were recorded on video for later analysis of horse behaviour (Table 1) by one trained observer (international student) who was unfamiliar with all horses and riders and blind regarding the aim of the study.

2.4.4. Heart rate

Horse heart rate was recorded using Polar Equine RS800cx (RR recordings; Polar Electro OY, Kempele, Finland), consisting of two sensors, a W.I.N.D. transmitter and a wristwatch receiver. Before fitting the electrodes, water and ultrasound gel were used to optimise the contact between electrode and skin. Data were downloaded using the software Polar ProTrainer, Equine Edition 5™. Artefacts in the data were corrected using the error correction function in this program (default settings, i.e. moderate Watt and minimum zone: 6 bpm). The average heart rate during each test ride was extracted (in beats per minute, bpm). Heart rate variability was not evaluated due to the short test duration (< 5 min) (von Borell et al., 2007).

Some of the heart rate files were lost due to errors caused by e.g. low conductance between skin and electrodes (n = 14), leaving a total of 82 rider-horse pairs to be included in the analysis of heart rate data.

2.4.5. Salivary cortisol

Saliva was collected with Salivette® cotton rolls (Sarstedt, Nümbrecht-Rommelsdorf, Germany) placed onto the tongue of the horse for 1 min with forceps, until the swab was well soaked. The Salivettes were frozen at −18 °C immediately and were later defrosted and centrifuged for 10 min at 1000 g. Concentrations of cortisol were determined using a direct enzyme immunoassay without extraction (Palme and Möstl, 1997) validated for equine saliva (Schmidt et al., 2009). The intra-assay coefficient of variation was 5.0 %, the inter-assay variation was 6.7 % and the minimal detectable concentration was 0.3 pg/well. All samples were analysed at the Inst. Animal Science, Aarhus University, Denmark.

There was large variation in the cortisol concentrations and some apparent outliers, particularly in the baseline samples, which may relate to prior feed intake. Although horse owners were advised not to feed their horse for 30 min prior to baseline sampling, feed remains were often encountered during sampling. Since cortisol data should be calculated as a difference from baseline to account for individual differences, this variation made the analysis challenging. To avoid the effect of outliers in the baseline samples, we used the median of the five baseline values (one from each test day) as the standard baseline value for each horse. For one of the resident horses, however, all five baseline samples were unrealistically high (i.e. higher than all post-exercise values) and data from this horse could not be included in the analysis. Since we only had two post-test samples (at 0 and 5 min) per rider, we used the mean of these samples to calculate the difference from baseline, i.e. test response = sample (B + C)/2 – median baseline', or 'sample (D + E)/2 – median baseline' for the first and second rider on each test day, respectively. One sample E was lost and a few were clearly outliers (e.g. when a single post-test value was lower than the median baseline and all other post-exercise samples were higher), resulting in n = 72 for the cortisol analysis. Consequently, the cortisol results should be interpreted with caution.

Table 1
Ethogram of recorded behaviour.

Behavioural parameter	Description[1]
Tail swishing (freq)	Lateral, dorsoventral or circular motion of the tail that interrupts the rhythmical waving motion of the tail corresponding to the gait[1]
Head movements (freq)	Head raising, lowering and tossing (i.e. a clear movement of the head away from the desired frame)[1]
Breaking the gait (freq)	The horse switches from the gait specified in the program to any other unintended gait[2]
Other evasive behaviour (freq)	Shying, i.e. the horse shows a startle reaction with a subsequent attempt to flee, or bolting, i.e. the horse suddenly runs (usually galloping) out of control by the rider[2]

[1] Adapted from (Christensen et al., 2014).
[2] Modified from (König von Borstel et al., 2011).

2.5. Data analysis

Prior to the analyses, skewed variables were log-transformed to comply with normality (tail swishing, undesired head movement, breaking the gait and other evasive behaviour). Initially, the rideability scores were analysed using a One-way Anova to investigate if horses differed in rideability score. To investigate the aims of the study (i and ii), the effect of rider (1–10), rein tension (mean N), frequency of conflict behaviour (tail swishing, undesired head movement, breaking the gait, and other evasive behaviour, separately), HR (mean bpm), and cortisol concentrations (difference from baseline, ng/mL) on rideability score (1–10) was investigated in a linear mixed-effects model, accounting for repeated measures per horse. The full model included all explanatory variables as listed above as main fixed effects, and interactions were not evaluated. This model was reduced using step-wise backwards reduction with P < 0.10 as threshold, removing the variable with the highest P-value in each step. The final model included the variables: rider, undesired head movement, breaking the gait and other evasive behaviour. Results of the final model are shown as likelihood ratio F-tests and p-values.

The same model type (linear mixed-effects model accounting for repeated measures per horse) was used to analyse the effect of rider on each measured variable, separately: rein tension, heart rate, cortisol, and expression of conflict behaviour (tail swishing, undesired head movements, breaking the gait and other evasive behaviour, separately).

The association between the measured rein tension and the riders' subjective impression of the rein tension was analysed using Repeated Measures Correlation Analysis.

In all analyses, significance was evaluated using 5% as the significance level, and all analyses were carried out using R version 3.5.2 "Egshell Igloo" (The R Foundation for Statistical Computing, 2019) and packages "nlme" (Pinheiro et al., 2019), "lme4" (Bates et al., 2016), "matrix" (Bates et al., 2019), "lmerTest" (Kuznetsova et al., 2017) and "rmcorr" (Bakdash and Marusich, 2018). Plots were also made using R and package "ggplot2" (Wickham, 2009).

3. Results

3.1. Do riders agree on rideability scores?

There was a significant difference among the horses in their assigned rideability scores (One-way Anova: F = 8.4, df = 9, P < 0.001, Fig. 1), i.e. some horses were generally considered easier to ride. However, there was a large variation among riders in their scoring of rideability of the individual horses. The horse with the highest variation in rideability scores received scores from 1 to 8 (Horse 4, Fig. 1) and there was a significant effect of rider on rideability scores (Table 2).

3.2. Rein tension, conflict behaviour, heart rate and cortisol

Mean levels of rein tension, conflict behaviour, heart rate and saliva cortisol are presented in Table 2. In addition to the rider, undesired head movements, breaking the gait and "other evasive behaviour" had a significant effect on the rideability scores (Table 2), i.e. rideability score dropped with increasing conflict behaviour. Rein tension, heart rate, saliva cortisol, and tail swishing had no effect on the rideability scores. Test results from the final model are presented in Table 2.

3.2.1. Rein tension

The overall level of rein tension for the 96 rider-horse pairs was (mean ± se) 26.9 ± 0.6 and 27.4 ± 0.6 N for the right and left rein,

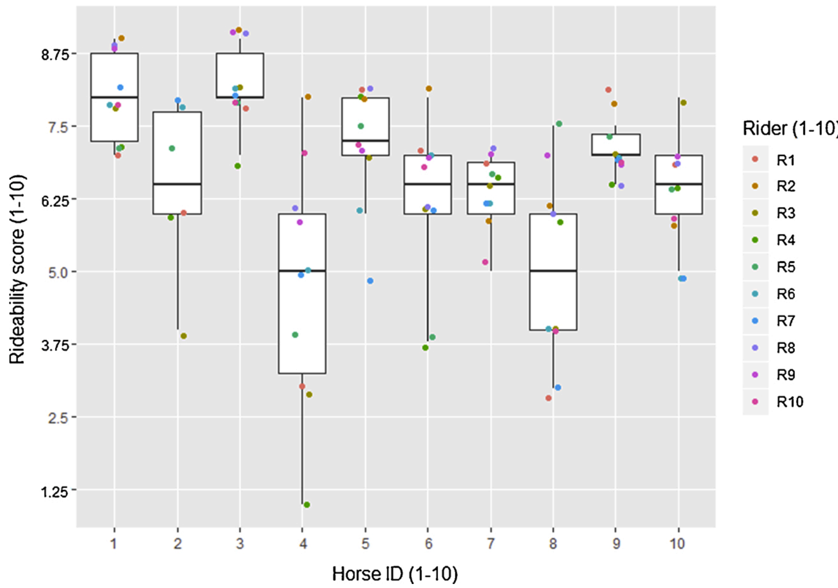

Fig. 1. Graphical illustration of rideability scores (scale 1; very poor - 10; excellent) assigned to each of the 10 horses by the 10 professional riders. To avoid scores being super-imposed (i.e. when riders gave the same score), points are plotted with 0.12 jitter distance. White boxplots illustrate the median, 25 % and 75 % quartiles and the range of the scores. Horses differed significantly with regard to their median rideability score (One-way Anova: F = 8.4, df = 9, P < 0.001).

Table 2
Summary statistics of all variables (Rider, Rein tension, Conflict behaviour, Heart rate and Salivary cortisol), presented as mean ± se, range and median (with 25 and 75 % quartiles). Results (F_{df} and P-values) from the final model fit of rideability scores are also shown.

Variables	Type	No. of rider-horse combinations	Mean ± se	Range	Median [25 %;75 % quartiles]	F_{df}	P-value
Rider	Factorial: 10 levels	96				3.18_{74}	0.003
Rein tension (N)[1]	Continuous	86	27.2 ± 1.0	17.7–36.8	27.0 [24.5;29.4]	–	–
Conflict behaviour[2]		96	12.2 ± 0.7	0–42	9.3 [8.0;16.0]		
Tail swishing	Continuous		4.7 ± 0.6	0–34	2.0 [0.0;8.0]	–	–
Head movements	Continuous		0.6 ± 0.2	0–11	1.0[0.0;11.0]	24.87_{74}	<0.001
Breaking the gait	Continuous		1.1 ± 0.1	0–7	2.0[0.0; 1.3]	6.51_{74}	0.013
Other evasive behaviour	Continuous		0.6 ± 0.2	0–14	0.0 [0.0;0.0]	4.80_{74}	0.032
Heart rate (bpm)	Continuous	82	116.3 ± 1.0	96–152	116.5 [110.0;122.0]	–	–
Saliva cortisol[3] (ng/mL)	Continuous	72	2.2 ± 0.2	0.18–7.09	1.8 [1.0;3.1]	–	–

[1] Mean Newton (avg. left and right) per rider-horse pair.
[2] Mean frequency per rider-horse pair.
[3] Calculated as increase from baseline.

respectively. There was a significant effect of rider on the level of rein tension ($F_{9,67}$ = 16.63; P < 0.001; Fig. 2), i.e. some riders applied significantly more rein tension than others when riding the same 10 horses through the same dressage test. Riders' subjective impression of the rein tension did not correlate with the actual rein tension (Fig. 3).

3.2.2. Conflict behaviour

Tail swishing was the most commonly observed conflict behaviour and was observed in all horses (Table 2). Undesired gait changes were also observed in all horses although the overall frequency was low and the variance high (Table 2). The other types of conflict behaviour (undesired head movements and other evasive behaviour) occurred less frequently (Table 2), and three horses expressed neither of them. There was no effect of rider on the frequency of undesired head movements, or breaking the gait, whereas there was a tendency for an effect of rider on other evasive behaviour ($F_{9,77}$ = 1.8; P = 0.07) and frequency of tail swishing ($F_{9,77}$ = 1.9; P = 0.07).

3.2.3. Mean heart rates and saliva cortisol

Mean heart rates of the horses during the dressage tests and increases in saliva cortisol concentrations are shown in Table 2. Both heart rates ($F_{9,63}$ = 67.5, P < 0.001) and saliva cortisol concentrations ($F_{9,53}$ = 3.9, P < 0.001) were significantly affected by rider.

4. Discussion

This study found large within-horse variation in rideability scores assigned to 10 sports horses by 10 professional riders, and most of the objective measures did not reflect the scores. Hence, individual riders differ in their interpretation of rideability and each rider emphasise different aspects of rideability, such as responsiveness to leg, seat and rein cues suppleness, quality of rein contact with the bit in the horse's mouth, comfort and security when sitting on the horse etc. Perceived rideability could also reflect the horse-rider match, which is influenced by personality traits of both rider and horse (Visser et al., 2008) as well as laterality (horse) and handedness (rider) (Kuhnke et al., 2010). In addition, several studies have highlighted the profound effects of individual rider style on horses' movement symmetry (e.g. Clayton and Hobbs, 2017; Persson-Sjodin et al., 2018). Thus, even if the test riders agree on the definitions of how to score rideability, their own riding style may affect the horse and consequently the rideability score. These findings suggest that the individual rider affects the horse during riding evaluations, which inevitably influences the horse's rideability.

In addition to the significant effect of the rider on rideability scores, we also found a significant effect of some types of conflict behaviour. These were undesired head movements, breaking the gait and other evasive behaviour. Horses displaying more of these conflict behaviours

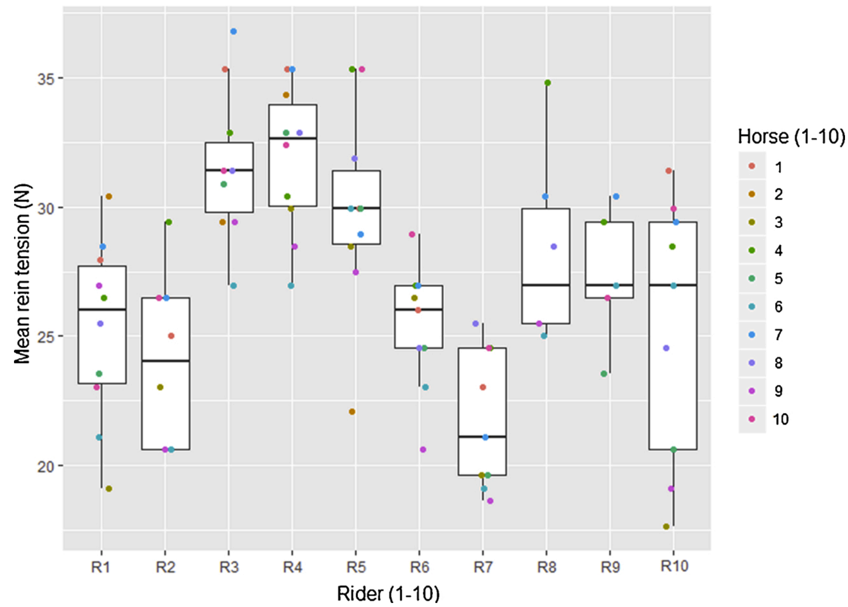

Fig. 2. Graphical illustration of the mean rein tension (N) for each rider (1-10), with each point representing a rider-horse pair. Points are plotted with 0.12 jitter distance to avoid cases where points are super-imposed. White boxplots illustrate the median, 25 % and 75 % quartiles and the range of the scores. There was a significant effect of rider on the level of rein tension (Linear mixed-effects model considering repeated measures: $F_{9,67} = 16.63$; $P < 0.001$).

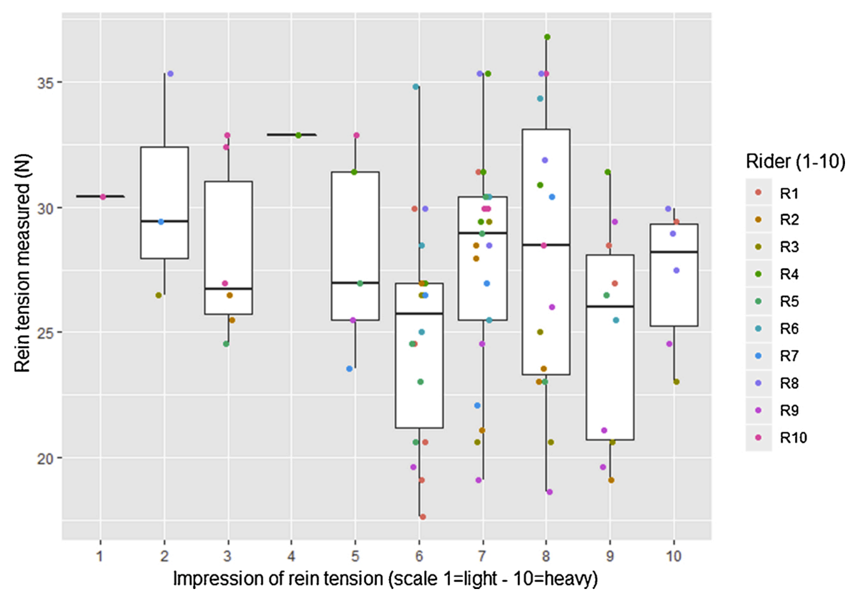

Fig. 3. Graphical illustration of measured rein tension (N) and rider's impression of the rein tension during the test (scale from 1-10, 1=loose, 10=heavy). Each point represents a rider marked by a corresponding colour. To avoid points being super-imposed (i.e. when riders had the same measured and impression of rein tension), points are plotted with 0.12 jitter distance. White boxplots illustrate the median, 25 % and 75 % quartiles and the range of the scores. Correlation analysis: $P > 0.05$.

were scored negatively by the riders and found to be more difficult to ride by individual riders. This result is in agreement with previous studies, e.g. Visser et al. (2008) where riders' assessment of horse cooperation during riding was more negative as frequency of evasive behaviour increased. König von Borstel and Glißman (2014) found that increased tail-swishing, shying and changes in gait reduced rideability scores. In this study, tail swishing did not affect rideability scores. The riders in this experiment might not have factored this into their rideability assessment as tail swishing is common in high-level dressage (Kienapfel et al., 2014) and it may be difficult to sense tail swishing from the saddle.

We did not find a significant effect of mean rein tension on rideability scores. This is in contrast to previous findings by König von Borstel and Glißman (2014) where rideability scores dropped significantly with increasing mean, maximum and variability in rein tension, and similarly, variance in rideability scores could be explained by maximum rein tension measured during a dressage performance test in another study (König von Borstel et al., 2012). In both of these studies, rideability was scored by both judges on the ground and a test rider and the final score was a mean of all scores, whereas in our study rideability was scored by riders only, which could explain this discrepancy. Also, the mean rein tension at two different testing stations were lower in their study (9.1 ± 1.6 and 21.7 ± 1.3 N, respectively) compared to the mean tension in the current study (27.2 ± 1.0 N). In our study, the upper detection limit of 49 N was exceeded in all tests (approx. 10 % of all recordings), i.e. the means should be considered underestimates and the maximum tension and variability could not be calculated. Similarly, in a previous study on dressage horses which were ridden by their usual rider in a 10 min dressage test, an average 15 % of the data exceeded the upper detection limit, using the same equipment (Christensen et al., 2014). König von Borstel and Glißman (2014) suggested not only the quantity but also the stability of rein tension is an important factor in relation to rideability, and future studies should ideally use equipment with a higher upper detection limit. It is also interesting to note the notably poor correlation among the riders' subjective impression of rein tension and the measured tension, as also reported by Randle and Abbey (2013). This may suggest riders' perception of rein tension could be more related to the quality (e.g. stability) rather than the quantity or variability of tension.

Mean heart rates (116.3 ± 1.0 bpm) were in agreement with previous studies, e.g. Christensen et al. (2014) where horses were ridden in a dressage test by their usual rider (mean heart rates: 115.9 ± 0.7 bpm), but heart rates did not significantly affect rideability scores, which agrees with findings from a study by König von Borstel et al. (2012). Similarly, the mean increase in salivary cortisol concentrations from baseline (2.2 ± 0.2 ng/mL) were similar to the increase in Christensen et al. (2014) (also approx. 2.2 ng/mL), but did not significantly affect the rideability scores. However, due to the large variation in cortisol concentrations and a number of apparent outliers, this result should be interpreted with caution. Further studies are required to investigate whether physiological parameters are unrelated to riders' perception of rideability.

We found a significant effect of the rider on the level of rein tension, horses' heart rate and increases in saliva cortisol concentrations, and a tendency for the frequency of tail swishing and other evasive behaviour. Thus, it appears that individual riders induce different levels of discomfort in the horses, which likely relates to riding style, incl. experience, use of aids, balance, handedness and personality (as discussed above). Similarly, König von Borstel and Glißman (2014) found significant effects of rider on rein tension and behavioural parameters and suggested that some rider-horse combinations may be a more or less good match. Future studies could help shed light on which elements of riding style are particularly important in relation to horse welfare. Collectively, the studies highlight the inevitable effect of individual riding style, which influences rideability scores and may explain the reported very low heritability of rideability (Dietl et al., 2005).

It should be noted that although our study included n = 100 horse-rider combinations, our results are based on only 10 horses and 10 female riders. Ideally, future studies should include a larger number of horses and riders to account for the many factors discussed above.

5. Conclusions

In conclusion, this study found a substantial variation among rideability scores assigned to 10 sports horses by 10 professional riders. The results suggest that these scores are rider-dependent and thus of little validity in identifying actual differences among horses. Rideability scores further depended on the occurrence of some types of conflict behaviour, which may be indicative of a rider-horse mismatch. Riders significantly affected the horses' physiological responses. Further studies are needed to identify which elements of riding style that are particular important for sports horse welfare, as well as to improve the objectivity, consistency and reliability of rideability assessment in sports horses.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgements

Thank you to the voluntary riders, horse owners and helpers.

References

Bakdash, J., Marusich, L., 2018. rmcorr: Repeated Measures Correlation.

Bates, D., Martin, Maechler, Walker, S., 2016. Package "lme4" [WWW Document]. CRAN Repos. https://doi.org/10.18637/jss.v067.i01.

Bates, D., Maechler, M., Davis, T.A., 2019. Matrix: Sparse and Dense Matrix Classes and Methods.

Christensen, J.W., Zharkikh, T.L., Antoine, A., Malmkvist, J., 2011. Rein tension acceptance in young horses in a voluntary test situation. Equine Vet. J. 43, 223–228. https://doi.org/10.1111/j.2042-3306.2010.00151.x.

Christensen, J.W., Beekmans, M., van Dalum, M., VanDierendonck, M., 2014. Effects of hyperflexion on acute stress responses in ridden dressage horses. Physiol. Behav. 128, 39–45. https://doi.org/10.1016/j.physbeh.2014.01.024.

Clayton, H.M., Hobbs, S.J., 2017. The role of biomechanical analysis of horse and rider in equitation science. Appl. Anim. Behav. Sci. 190, 123–132. https://doi.org/10.1016/j.applanim.2017.02.011.

Dietl, G., Hoffmann, S., Reinsch, N., 2005. Impact of trainer and judges in the mare performance test of Warmblood Horses. Arch. Anim. Breed. 48, 113–120. https://doi.org/10.5194/aab-48-113-2005.

Duncan, I., Hagan, K., Jose Hotzel, M., Lay, D., Olsson, A., Sherwin, C., Whittaker, A., Martin, F., 2013. Ethical treatment of animals in applied animal behaviour research: ethical stand-point & decision models. Int. Soc. Appl. Ethol.

Górecka-Bruzda, A., Chruszczewski, M.H., Jezierski, T., Murphy, J., 2011. Behaviour-related traits of Polish sport horse stallions participating in 100-day performance tests. Livest. Sci. 136, 207–215. https://doi.org/10.1016/j.livsci.2010.09.014.

Górecka-Bruzda, A., Kosińska, I., Jaworski, Z., Jezierski, T., Murphy, J., 2015. Conflict behavior in elite show jumping and dressage horses. J. Vet. Behav. 10, 137–146. https://doi.org/10.1016/j.jveb.2014.10.004.

Graf, P., König von Borstel, U., Gauly, M., 2013. Importance of personality traits in horses to breeders and riders. J. Vet. Behav. Clin. Appl. Res. 8, 316–325. https://doi.org/10.1016/j.jveb.2013.05.006.

Hawson, L.A., McLean, A.N., McGreevy, P.D., 2010. The roles of equine ethology and applied learning theory in horse-related human injuries. J. Vet. Behav. Clin. Appl. Res. 5, 324–338. https://doi.org/10.1016/j.jveb.2010.06.001.

Kienapfel, K., Link, Y., König V Borstel, U., 2014. Prevalence of different head-neck positions in horses shown at dressage competitions and their relation to conflict behaviour and performance marks. PLoS One 9, e103140. https://doi.org/10.1371/journal.pone.0103140.

Kilkenny, C., Browne, W.J., Cuthill, I.C., Emerson, M., Altman, D.G., 2010. Improving bioscience research reporting: the arrive guidelines for reporting animal research. PLoS Biol. 8, 6–10. https://doi.org/10.1371/journal.pbio.1000412.

König von Borstel, U., Pasing, S., Gauly, M., 2011. Towards a more objective assessment of equine personality using behavioural and physiological observations from performance test training. Appl. Anim. Behav. Sci. 135, 277–285. https://doi.org/10.1016/j.applanim.2011.10.007.

König von Borstel, U., Peinemann, V., Glißman, C., Euent, S., 2012. Willing to work? Suitability of different riding situations in evaluation of equine personality. In: Proceedings of the 8th International Equitation Science Conference in Edinburgh, UK, July 18-20th. Edinburgh, p. 129.

König von Borstel, U., Pasing, S., Gauly, M., Christmann, L., 2013. Status quo of the personality trait evaluation in horse breeding: judges' assessment of the situation and strategies for improvement. J. Vet. Behav. Clin. Appl. Res. 8, 326–334. https://doi.org/10.1016/j.jveb.2013.04.066.

König Von Borstel, U., Glißman, C., 2014. Alternatives to conventional evaluation of rideability in horse performance tests: suitability of rein tension and behavioural parameters. PLoS One 9, e87285. https://doi.org/10.1371/journal.pone.0087285.

Kuhnke, S., Dumbell, L., Gauly, M., Johnson, J.L., McDonald, K., König Von Borstel, U., 2010. A comparison of rein tension of the rider's dominant and non-dominant hand and the influence of the horse's laterality. Comp. Exerc. Physiol. https://doi.org/10.1017/S1755254010000243.

Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B., 2017. lmerTest package: tests in linear mixed effects models. J. Stat. Softw. 82 https://doi.org/10.18637/jss.v082.i13.

Palme, R., Möstl, E., 1997. Measurement of cortisol metabolites in faeces of sheep as a parameter of cortisol concentration in blood. Int. J. Mammal Biol. 62, 192–197.

Persson-Sjodin, E., Hernlund, E., Pfau, T., Haubro Andersen, P., Rhodin, M., 2018. Influence of seating styles on head and pelvic vertical movement symmetry in horses ridden at trot. PLoS One 13, e0195341. https://doi.org/10.1371/journal.pone.0195341.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., Team, R.C., 2019. nlme: Linear and Nonlinear Mixed Effects Models.

Randle, H., Abbey, A.K.S., 2013. Qualitative (perceived) versus quantitative (actual) assessment of rein tension: what lessons can be learnt?. In: 9th International Equitation Science Conference, 2013, July 17th-20th. Delaware, USA, p. 37.

Rothmann, J., Christensen, O.F., Søndergaard, E., Ladewig, J., 2014. Behavior observation during conformation evaluation at a field test for Danish warmblood horses and associations with rideability and performance traits. J. Equine Vet. Sci. 34, 288–293. https://doi.org/10.1016/j.jevs.2013.06.007.

Schmidt, A., Aurich, C., Neuhauser, S., Aurich, J., Möstl, E., 2009. Comparison of cortisol levels in blood plasma, saliva and faeces of horses submitted to different stressors or treated with ACTH. In: 5th International Symposium on Equitation Science, 2009. Sydney, Australia, p. 53.

Schork, I.G., de Azevedo, C.S., Young, R.J., 2018. Personality, abnormal behaviour, and health: an evaluation of the welfare of police horses. PLoS One 13, e0202750. https://doi.org/10.1371/journal.pone.0202750.

Teegen, R., Edel, C., Thaller, G., 2008. Evaluation of the Breeding Objectives for the Trakehner Horse Breed by a Contingent Valuation Method (CV). Züchtungskunde.

The R Foundation for Statistical Computing, 2019. R Core Team.

Visser, E.K., Van Reenen, C.G., Blokhuis, M.Z., Morgan, E.K.M., Hassmen, P., Rundgren, T.M.M., Blokhuis, H.J., 2008. Does horse temperament influence horse-rider cooperation? J. Appl. Anim. Welf. Sci. 11, 267–284. https://doi.org/10.1080/10888700802101254.

von Borell, E., Langbein, J., Després, G., Hansen, S., Leterrier, C., Marchant-Forde, J., Marchant-Forde, R., Minero, M., Mohr, E., Prunier, A., Valance, D., Veissier, I., 2007. Heart rate variability as a measure of autonomic regulation of cardiac activity for assessing stress and welfare in farm animals - A review. Physiol. Behav. 92, 293–316. https://doi.org/10.1016/j.physbeh.2007.01.007.

Wickham, H., 2009. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York. https://doi.org/10.1007/978-0-387-98141-3.