Conflict behaviors displayed by horses during dressage tests and their relationship to performance evaluation

Digital Object Identifier (DOI):
10.1016/j.jveb.2022.07.011

Published In:
Journal of Veterinary Behavior: Clinical Applications and Research

Journal of Veterinary Behavior 55–56 (2022) 48–57

Contents lists available at ScienceDirect

Journal of Veterinary Behavior

journal homepage: www.journalvetbehavior.com

Equine conflict behaviors in dressage and their relationship to performance evaluation

Kathryn L. Hamilton[a,*], Bryony E. Lancaster[a], Carol Hall[b,c]

[a] Royal (Dick) School of Veterinary Studies and Roslin Institute, University of Edinburgh, Easter Bush, Edinburgh, United Kingdom
[b] School of Animal Rural and Environmental Sciences, Nottingham Trent University, Brackenhurst Campus, Southwell, Nottinghamshire, NG25 0QF, United Kingdom
[c] National Equine Welfare Council, Slad Lane, Princes Risborough, Buckinghamshire, HP27 0PP, United Kingdom

A R T I C L E I N F O

Article history:
Received 21 October 2021
Revised 15 July 2022
Accepted 30 July 2022
Available online 5 August 2022

Keywords:
Equine
Welfare
Conflict behavior
Dressage
Competition performance
Equitation

A B S T R A C T

In the equestrian discipline of dressage, the behavior encouraged through judging should be based on correct and welfare-centered training techniques. Certain behaviors in the ridden horse result from unclear or conflicting cues from the rider and can be referred to as conflict behaviors. This study aimed to investigate the occurrence of these behaviors during Preliminary, Novice and Elementary level British Dressage (BD) tests, and to examine their relationship with performance evaluation by the judge. Data were collected from 75 dressage tests in November and December 2019. Each test was filmed, and the judges' scores were collected. Between five and seven movements (i.e., small numbered sections into which dressage tests are divided) within each test were analyzed and the frequency of conflict behaviors displayed used to derive a behavior score for each movement. These behaviors were recorded in six subsections: head, ears, mouth, tail, auditory and whole body. Conflict behaviors were seen in 97.6% of the movements analyzed, with horses displaying two or more such behaviors in 83% of movements. There was no significant association found between judge score and overall behavior score but there was a negative correlation between whole-body scores and judge score (Spearman's rank correlation: $P<0.001$). Horses with their nasal plane in front of the vertical were awarded lower judge scores than those with their nasal plane either vertical (Wilcoxon rank sum test: $P<0.01$), or less than 30° behind the vertical ($P<0.001$). Judge scores were significantly higher for movements in which horses had their ears forward compared to those in which ears were held back (Wilcoxon rank sum test: $P<0.05$) or to the side ($P<0.05$). No association was found between judge score and mouth or tail behavior. Significantly higher mouth behavior scores were seen within downwards transitions (e.g., canter to trot) compared to movements that involved changing the rein (Wilcoxon rank sum test: $P<0.05$) or circling to the right ($P<0.05$). Conflict behaviors occurred in almost all the dressage movements analyzed, but the only association with performance score was when the behavior involved the horse's whole body and/or the head and neck. Behavioral signs of conflict are indicative of compromised welfare in ridden horses and the results of this study suggest that a greater focus on such behavior should be included in dressage judge training and performance evaluation.

Introduction

The inclusion of animals in sport has become a topic for debate, with the potential for public opinion to put the continuation of equestrian sport at risk (Duncan et al., 2018). In the training and management of equine athletes and during competition, there is a need to ensure that the welfare of the animal is paramount. The aim embedded into the guidelines produced by the Inter-

* Address for reprint requests and correspondence: Miss Kathryn Louise Hamilton. The University of Edinburgh, 34 Alligan Road Crieff, Perthshire PH7 3JS, United Kingdom

E-mail address: kham958@live.co.uk (K.L. Hamilton).

https://doi.org/10.1016/j.jveb.2022.07.011

national Equestrian Federation (FEI), is that training in dressage should entail 'the development of the Horse into a happy Athlete through harmonious education' (FEI, 2020). The horse should be 'calm, supple, loose, flexible, confident, attentive and keen' and should work in 'perfect understanding' with their rider (FEI, 2020). However, time and financial pressures often cause young horses to be rushed through their training (Ödberg and Bouissou, 1999; Heuschmann, 2018). Many training methods used today are derived from traditional methods used when the horse was first domesticated and fail to incorporate consideration of equine welfare or learning theory (Waran et al., 2007). Pain may be an underlying factor that goes undetected. For example, low grade lameness was found in 47% of 506 sport horses tested by Greve and Dyson (2014) suggesting that problems can exist and develop without the trainer/rider/handler noticing. It has been suggested that inappropriate training and poor riding may result in the development of equine behavioral problems that can be linked to large numbers of young horses being slaughtered (Ödberg and Bouissou, 1999). Some approaches used to resolve behavioral problems may worsen the situation. For example, the use of certain training aids (such as strong bits and tight nosebands) has the potential to cause nerve damage and pain (Casey et al., 2013; Uldahl and Clayton, 2019) and can significantly increase the number of problem behaviors displayed by the ridden horse (Hockenhull and Creighton, 2012). Facial expressions of the horse can be used to detect pain in ridden horses (Gleerup et al., 2015; Mullard et al., 2017; Dyson et al., 2017; Gleerup et al., 2018; Dyson et al., 2018a; Dyson et al., 2018b) which is likely to contribute to the development of unwanted / problem behavior.

Certain behaviors in the ridden horse result from unclear, confusing or conflicting cues from the rider and may indicate a negative mental state in the horse (Christensen et al., 2014; Kienapfel et al., 2014; Smiet et al., 2014). Such behavior is often termed conflict behavior, caused by the presence of two contrasting motivations simultaneously, as for example the simultaneous use of hand and leg pressure sometimes encouraged in dressage riding (McGreevy, 2012). Whether pressure is needed and if so, what level of pressure should be exerted, is a matter for debate within the equestrian world (Ödberg and Bouissou, 1999; Clayton et al., 2005; Warren-Smith et al., 2007; Kuhnke et al., 2010). Additionally, Condon et al. (2021) found associations between commonly used equipment (e.g., harsh bits, spurs) and behaviors indicative of conflict. Specific behaviors have been identified as signs of conflict. Previous research has found tail swishing to be common in dressage horses (Kienapfel et al., 2014; Kienapfel, 2011; Górecka-Bruzda et al., 2015) which may represent an agonistic behavioral response to the rider (Górecka-Bruzda et al., 2015) or higher concentration levels in the horse (Hall and Heleski, 2017). Tension in the muscles surrounding the mouth, head movements (for example, shaking, lifting) and head position in relation to withers are suggestive of discomfort and/or conflict (Kienapfel et al., 2014; McGreevy, 2012). The visibility of these signs to the observer is variable, with those associated with the mouth and facial expression being hard to see from any distance (Hall et al., 2014). The movement and position of the head and neck are generally more clearly visible. Having a head position / nasal plane behind the vertical can be indicative of conflict (Kienapfel et al., 2014) while a nasal plane position which exceeds 30° in front of the vertical can indicate natural head carriage, spooking (sudden movement displayed by horse due to shying) or resistance to the bit (McGreevy, 2012).

Despite the FEI (2020) guidelines stating that *'the head should remain in a steady position, as a rule slightly in front of the vertical'*, there has been an increase within competitive dressage in the number of horses being ridden with the nasal plane behind the vertical (Lashley et al., 2014). Additionally, a significant relationship has been found between horses having their nasal plane behind the vertical and higher scores awarded by the judges (Lashley et al., 2014). At the same time, it was found that horses with their nasal plane behind the vertical showed significantly more conflict behaviors than those who had their nasal plane in front of the vertical (Kienapfel et al., 2014). The combined findings of these two studies suggest that the extent to which conflict behaviors are accounted for during performance evaluation requires further investigation.

The evaluation of performance in dressage is currently subjective and hard to quantify. At Olympic level there is high variability within collective scores (the marks awarded for general performance features rather than for specific movements) (Hawson et al., 2010), and low reliability with scores affected by type of movement (i.e. small numbered sections into which dressage tests are divided), rank of horse, judge and location (Stachurska and Bartyzel, 2011). In dressage judging there is more disagreement between judges in relation to the highest scoring horse-rider combinations, which may be caused by a lack of precise and objective guidelines (Heiniger and Mercier, 2018). The FEI are developing a more objective system (based on a code of points) for scoring dressage performance to improve judging quality and consistency across the sport (Dressage Judging Working Group, 2018). Similar judging tools in artistic gymnastics have produced high reliability and consistency within judging (Leskosek et al., 2010; Bucar et al., 2012). To ensure that dressage judging rewards performance that reflects optimum training (as outlined in the FEI guidelines, 2020), further consideration of behavior indicative of conflict and discomfort must be incorporated in future judging criteria.

There is currently evidence that the assessment of elite level dressage performance rewards incorrect head and neck positions which have also been found to be associated with an increased frequency of conflict behaviors (Kienapfel et al., 2014; Lashley et al., 2014). The aim of this study was to determine whether the evaluation of dressage performance at sub-elite levels (Elementary, Novice and Preliminary levels) includes consideration of such conflict behaviors in terms of the scores attributed by judges. If the scores allocated take account of the occurrence of conflict behaviors, the future training of these horses is more likely to be optimised and their ongoing welfare improved.

Materials and methods

Ethical permission for this project was granted by the University of Edinburgh Human Ethics Review Committee (HERC – 378:19) and the University of Edinburgh Veterinary Ethics Review Committee (VERC – 101:19). Consent was obtained from all participants before data collection.

Data collection - recording video footage and photographs

Dressage tests at British Dressage (BD) affiliated competitions at two different venues in Scotland were videoed over three days in November and December 2019. Arenas used were indoor with Cushion Track^{TM} or Andrews Bowen Ltd. surface. A video camera (Panasonic DMC-FZ200, 75mm focal lens) on a tripod was set up about 6m from the edge of the arena (between C and H – Figure 1). Ideally the camera would have been set up next to the judge, but this was not possible at these venues. The tripod was fixed parallel to the ground on the lateral and longitudinal axes of rotation using a spirit level. The camera on the tripod was free to rotate around the vertical axis and therefore could be panned around the arena by the experimenter using a handle. Each movement is allocated a score from 0 (not executed) to 10 (excellent) by a judge according to the judge's interpretation of how well the movement was

Table 1
Test movement numbers[a] analyzed when equine behavior was compared to judge score in 75 dressage tests. Movements are small numbered sections into which dressage tests are divided.

Test level and number	Test movement number (and descriptor) used for behavioral analysis					
	Left-hand circle	Right-hand circle	Change of rein	Upwards transition	Downwards transition	Halt
Preliminary 12 (n = 83)	3 (20 metres diameter in working trot)	5 (20 metres diameter in working trot)	2 (MXK –working trot)	9 (working trot to working canter left)	14 (working canter right to working trot)	16 (working trot to halt)
Preliminary 13 (n = 99)	4 (20 metres diameter in working trot)	2 (20 metres diameter in working trot)	14 (MXK –working trot)	11 (working trot to working canter right)	13 (working canter right to working trot)	15 (working trot to halt)
Preliminary 14 (n = 56)	2 (20 metres diameter in working trot)	6 (20 metres diameter in working trot)	10 (KXM – free walk on a long rein)	7 (working trot to working canter right)	9 (working canter right to working trot to medium walk)	15 (working trot to halt)
Novice 27 (n = 46)	2 (20 metres diameter in working trot)	9 (20 metres diameter in working trot)	15 (KXM – medium trot)	12 (working trot to working canter left)	7 (working canter right to working trot)	19 (working trot to halt)
Novice 28 (n = 117)	9 (20 metres diameter in working canter)	4 (20 metres diameter in working canter)	13 (KXM – medium trot)	n/a	7 (working canter right to working trot)	16 (medium walk to halt)
Elementary 42 (n = 31)	5 (10 metres diameter in working trot)	4 (10 metres diameter in working trot)	9 (KXM – medium trot)	13 (medium walk to working canter right)	23 (working canter left to working trot)	24 (working trot to halt)
Elementary 44 (n = 15)	5 (10 metres diameter in collected trot)	3 (10 metres diameter in collected trot)	2 (EXB – collected trot)	10 (medium walk to working canter left)	6 (medium trot to collected trot)	16 (medium walk to halt)

[a] Movements are small, numbered sections into which dressage tests are divided. Movement numbers correspond to the official British Dressage (BD) test sheets for each test. Preliminary level is the first level of BD which includes canter movements. Novice level tests include stride lengthening. Lateral work comes in at Elementary level.

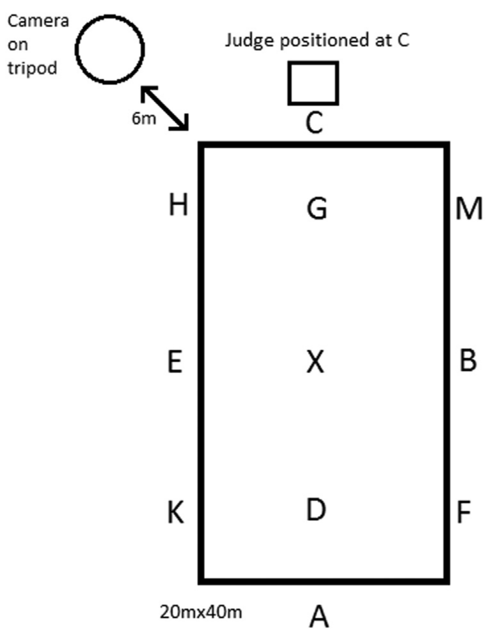

Figure 1. Plan of camera, judge and dressage arena set up for collection of video footage of 75 British Dressage tests at two venues in Scotland.

performed. Judges' scores were obtained from test sheets which were photographed. The time of capture displayed on the camera for each video and photograph was recorded to ensure videos were matched up to the correct test sheet. Videos and photographs were recorded by KLH and two assistants. Videos and photographs were stored on a password-protected, encrypted drive within 24 hours of data collection. Each horse-rider combination was given an individual ID code ('subject ID') and videos and photographs were named accordingly. Within the data, there were four horse-rider combinations who completed two tests (at different levels).

Data collection – behavior analysis

Behavior analysis was carried out by KLH using the video footage. Table 1 outlines movements chosen from each test. To improve accuracy and reliability of the study the movements chosen were the same for each test. Since tests are not identical the movements chosen were those that would occur in all tests included in the study: right-hand circle and left-hand circle, upwards transition, downwards transition, change of rein and halt. One test (Novice 28, see Table 1) had an upwards transition and a circle within the same movement, however, at one of the venues the transitions were not visible and therefore only the circle was used in the behavioral analysis of these movements.

There were 447 movements analyzed for the 75 horse and rider combinations. During statistical analysis, judge scores were compared to behavior scores to investigate effect of horse behavior on marks awarded by the judge. Comments written on the test sheet were noted in case of any outliers in the study (for example, if the horse stepped out of the arena and therefore gained a lower mark than the performed movement otherwise would have scored). Behavioral analysis of movements began at the letter at the beginning of the movement and finished at the last letter in the movement. For any movement that contained wording such as 'before A' or 'after X', behavioral analysis began or ended 2-3 seconds before or after the stated letter. Microsoft® Films and TV (version 10.20022.1101.0) and Microsoft® Excel® for Office 365 MSO (16.0.12527.20170) were used for data collection. RStudio (version 1.2.1335) was used for data analysis.

For the study to be comparable with real-life judging situations, the movement clip was analyzed by KLH (researcher) without the use of behavioral analysis software. Each movement clip was scored twice consecutively by KLH. There were no differences recorded between the first observation and the second observation. Behavioral analysis was completed without the knowledge of final placings or judge scores (blind testing). To determine inter-observer reliability, 12 movements were watched by a second independent observer (veterinarian) and behavior scores were compared to behavior scores from first observer (KLH). Intraclass correlation coefficient showed a substantial (0.62, CI 0.124 – 0.874) inter-observer reliability according to definitions used by Landis and Koch (1977).

Behaviors were divided into body sections and scores were allocated using an ethogram (see Table 2): the behavior scores were based on previously published descriptions of conflict behavior (McGreevy, 2012; Górecka-Bruzda et al., 2015) or pain-related behavior (Hall et al., 2014; Dyson et al., 2018a). For each movement individual body section behavior scores were summed to give a total behavior score. Information on venue, test number, gait (i.e. the pattern of movement in which the horse moves e.g. walk, trot - FEI, 2015), level of competition (i.e. Prelim, Novice or Elementary), judge list (i.e. level of qualification of judge within British Dressage - BD, 2021) and movement type were also collected. Number of head movements (0–2) were recorded. A diagram was used to improve accuracy when recording 'nasal plane angle' data (supplementary information). Presence or absence of spurs and type of noseband (cavesson, drop, flash, grakle or Micklem) was noted for each movement.

Statistical analysis

Initial descriptive analyses were carried out to investigate the mean and range seen in behavior score and score awarded by judges. Occurrence of each behavior was investigated for each body section across the entire dataset and for each movement type. The data set varied significantly from a normal distribution (Shapiro-Wilk normality test: $P < 2.2 \times 10^{-16}$) and therefore non-parametric tests were chosen for the analysis. Spearman's rank correlation coefficient, Wilcoxon rank sum test and Kruskal-Wallis test were used to investigate variation within judge score and behavior score. Within the data, there were four horse-rider combinations who completed two tests (at different levels).

Variation within score awarded by judge was investigated using level of judge, venue, level of test, gait and movement type as predictors. Individual body section behaviors were then used as predictors to investigate their effects on judge score. Variation within behavior score was investigated using venue, level of test, gait, movement type, presence of spurs and noseband type as predictors. Each body section behavior was investigated using predictors of movement type, noseband type and presence of spurs.

Results

Occurrence of conflict behaviors

A total of 75 horse-rider combinations were included in the study giving 447 movements. No conflict behavior was seen in 2.4% (n = 11) of movements. Two or more conflict behaviors were seen in 83% of movements. Table 3 shows percentage of movements in which no conflict behaviors were seen and Table 4 illustrates occurrence of all behavior types recorded.

Variance in score awarded by judge

Minimum score awarded by judge for a single movement was 3 and maximum was 9. Mean score awarded by judge for a single movement was 6.4. No association was found between behavior score and score awarded by the judge (Spearman's rank correlation: $P = 0.188$) (Figure 2).

Movements in which the horse displayed a whole body movement indicative of conflict were awarded lower scores by the judge than those in which horses did not display these behaviors (Spearman's rank correlation: $P = 5.495 \times 10^{-5}$; r = -0.193). <30 behind (Wilcoxon rank sum test: $P = 3.1 \times 10^{-6}$) and vertical (Wilcoxon rank sum test: $P = 0.0013$) head positions both gained significantly higher judge scores than <30 above. >30 above position was significantly associated with lower judge scores compared to vertical (Wilcoxon rank sum test: $P = 2.0 \times 10^{-5}$), <30 above (Wilcoxon rank sum test: $P = 0.0048$) and <30 behind (Wilcoxon rank sum test: $P = 3.1 \times 10^{-6}$). The judge score decreased as the number of head movements increased (Spearman's rank correlation: $P = 2.322 \times 10^{-5}$; r = -0.202). Forward ears gained significantly higher marks than ears held back (Wilcoxon rank sum test: $P = 0.045$) or ears held to the side (Wilcoxon rank sum test: $P = 0.045$). No association between judge score and poll to withers, tail or mouth behavior was found.

No association between judge score and level of judge, venue, level of competition, movement type, gait or test number was found.

Variance in behavior score

Minimum behavior score was 0 and maximum was 12. Mean behavior score was 4.3. Behavior score was significantly lower for change of rein movements when compared to upwards transition movements (Wilcoxon rank sum test: $P = 0.0022$). In upwards transitions, there was a significant increase in number of head movements compared to movements involving circling to the left (Wilcoxon rank sum test: $P = 0.009$), changing the rein (Wilcoxon rank sum test: $P = 3.4 \times 10^{-5}$), downwards transitions (Wilcoxon rank sum test: $P = 0.023$) or halts (Wilcoxon rank sum test: $P = 0.009$). Higher tail behavior scores were seen in upwards transitions compared to change of rein movements (Wilcoxon rank sum test: $P = 9.9 \times 10^{-4}$), downwards transitions (Wilcoxon rank sum test: $P = 0.012$) and halts (Wilcoxon sum rank test: $P = 9.9 \times 10^{-4}$). Mouth behavior score was significantly higher in downwards transitions than in movements involving changing the rein (Wilcoxon rank sum test: $P = 0.034$) or circling to the right (Wilcoxon rank sum test: $P = 0.034$). No association between movement type and nasal plane position or ears behavior was found.

Nasal plane angle behavior score was significantly higher when spurs were present (Kruskal-Wallis test: $H(1) = 19.715$; $P = 8.987 \times 10^{-6}$). Presence of spurs was significantly associated with less head movement (Kruskal-Wallis test: $H(1) = 7.973$; $P = 0.005$) and lower ear behavior scores (Kruskal-Wallis test: $H(1) = 11.963$; $P = 5.426 \times 10^{-4}$). Type of noseband was significantly associated with tail behavior (Kruskal-Wallis test: $H(4) = 38.087$; $P = 1.075 \times 10^{-7}$), mouth behavior (Kruskal-Wallis test: $H(4) = 17.43$; $p = 0.002$) and nasal plane position (Kruskal-Wallis test: $H(4) = 12.349$; $p = 0.015$). Flash nosebands were associated with higher tail behavior scores than cavesson (Wilcoxon rank sum test: $P = 8.7 \times 10^{-5}$), drop (Wilcoxon rank sum test: $P = 8.7 \times 10^{-5}$) or Micklem (Wilcoxon rank sum test: $P = 8.7 \times 10^{-5}$) nosebands. Drop nosebands were significantly associated with lower tail behavior scores than cavesson (Wilcoxon rank sum test: $P = 0.013$) or grakle (Wilcoxon rank sum test: $P = 0.002$) nosebands, while Micklem nosebands were also associated with lower tail behavior scores than grakle nosebands (Wilcoxon rank sum test: $P = 0.035$). Lower mouth behavior scores were seen for cavesson or flash nosebands compared to grakle (Wilcoxon rank sum test: $P = 0.034$) or Micklem (Wilcoxon rank sum test: $P = 0.005$) nosebands. Grakle nosebands were significantly associated with higher nasal plane angle behavior score than cavesson (Wilcoxon rank sum test: $P = 0.024$), drop (Wilcoxon rank sum test: $P = 0.024$) or micklem (Wilcoxon rank sum test: $P = 0.026$) nosebands.

Table 2
Behavior indicative of conflict descriptions and scoring. Behavior score developed from Dyson et al. (2018a), Hall et al. (2014) and McGreevy (2012).

Body section		Behavior	Behavior description	Behavior score[a]
HEAD	Nasal plane angle[b]	>30° above vertical	Nasal plane is more than 30° in front of the vertical	1
		<30° above vertical	Nasal plane is between 30° in front of the vertical and vertical	0
		Vertical	Nasal plane is at the vertical	0
		<30° behind vertical	Nasal plane is between vertical and 30° behind the vertical	2
		>30° behind vertical	Nasal plane is more than 30° behind the vertical	2
	Poll to withers[b]	Above	Poll is higher than the withers	0
		Level	Poll is level with the withers	1
		Below	Poll is below the withers	2
	Head movement[c]	Tipped head	Median plane of the head is not perpendicular to ground but instead deviates to left or right by 15° or more	2
		Nodding	Vertical head movement (out with normal movement for gait) at a speed of one (or less) movement per stride (i.e. a cycle of limb pattern which is completed when all limbs have returned to their initial relative position)	2
		Tossing	Vertical head movement (out with normal movement for gait) at a speed of more than one movement per stride	2
		Head lift	One sudden vertical head movement	2
		Nose side to side	Nose moves medially and laterally	1
		Headshaking	Movement of the head around the longitudinal axis of the neck	2
		Steady head[d]	Head held steady with only natural movements for gait	0
		Turned head	Median plane of head and neck turned more than necessary for movement (i.e. overbent)	1
EARS[b]		Both forward	Pinnae point cranially	0
		Back (not pinned)	Pinnae point caudally but are not flat against neck	1
		Both back (pinned)	Pinnae are laid flat against neck	2
		One forward, one back	One pinna points caudally and one pinna points cranially	0
		Scanning	Pinnae continually move throughout movement; if they stop moving they are only held in one position for less than 2 seconds	1
		Both held to side	Pinnae point laterally	0
MOUTH		Still and shut[d]	Mouth is closed with no lip movement	0
		Still and open (no tongue)	Mouth is open for at least 1 second; tongue is not visible	2
		Still and open (tongue)	Mouth is open for at least 1 second; tongue is visible	2
		Moving (licking and chewing)	Relaxed movement of mouth; tongue may be seen, muscles around mouth and nose relaxed	0
		Moving (tense)	Tense movement of mouth and/or lips; tongue not seen, tension in muscles around mouth and nose	2
TAIL		Swinging[d]	Lateral movement of tail natural for gait	0
		Swish	One quick lateral or vertical movement of tail	1
		Swish multiple	More than one quick lateral or vertical movement of tail	2
		Clamped	Tail clamped to body and not moving with natural movement of gait	2
		Crooked	Tail held to one side of median plane of body	2
WHOLE BODY[c]		Stopping or napping	Horse slowing down/stopping/going sideways	2
		Bucking	Horse lifting hindquarters off the ground more than necessary for gait	2
		Un-cued behavior	Behavior displayed by horse which is not part of test (e.g. breaking gait)	2
		Rearing	Horse lifting head, neck, shoulders and forelimbs off the ground more than necessary for gait	2
		Spooking	Sudden movement displayed by horse due to shying	1
		Stumbling	Tripping displayed by horse (front or hind limb)	1

[a] 0 = not contributing; 1 = possible discomfort from conflicting motivations or pain; 2 = likelihood of discomfort from conflicting motivations or pain.
[b] behavior is recorded as displayed if they are present for the majority of the movement.
[c] behavior is recorded as displayed if they are present at any point in the movement.
[d] behavior is recorded as displayed if no other behaviors for this body section are displayed.

Table 3
The percentage of movements at three levels of competition (Preliminary, Novice and Elementary) in which no equine conflict behaviors were seen for five different types of behavior when 447 movements were analyzed in 75 British Dressage tests.

Type of behavior (italics indicates behavioral descriptor which represents no conflict behavior[a])	Percentage of movements in which behavior was displayed by horse		
	Preliminary level (n = 238)	Novice level (n = 163)	Elementary level (n = 46)
Whole body *(none)*	95% (n=226)	91% (n=149)	93% (n=43)
Tail *(swinging)*	67% (n=159)	55% (n=90)	52% (n=24)
Height of poll in relation to withers *(above)*	83% (n=198)	75% (n=123)[b]	98% (n=45)
Mouth *(still, shut)*	25% (n=60)	26% (n=42)	13% (n=6)
Head movement *(steady)*	56% (n=134)[b]	61% (n=100)[b]	91% (n=42)

[a] these behaviors were not recorded for a trial if seen in conjunction with conflict behaviors (see Table 2), therefore any trial in which these were recorded had no conflict behavior for that type of behavior.
[b] significant difference ($p<0.05$) compared to Elementary level.

Table 4
Occurrence of behavior types in different body sections of the horse when 447 dressage movements were analyzed from 75 dressage tests at Preliminary, Novice and Elementary levels of British Dressage competitions. See Table 2 for ethogram.

Body section		Behavior	Number of movements in which behavior was seen
HEAD	Nasal plane angle	>30° above vertical	15
		<30° above vertical	148
		Vertical	110
		<30° behind vertical	153
		>30° behind vertical	3
		Not visible	18
	Poll to withers	Above	366
		Level	66
		Below	14
		Not visible	1
	Number head movement	0	276
		1	123
		2	48
	Head movement	Tipped head	40
		Nodding	31
		Tossing	12
		Head lift	42
		Nose side to side	31
		Headshaking	4
		Steady head	276
		Turned head	11
EARS		Both forward	82
		Back (not pinned)	126
		Both back (pinned)	1
		One forward, one back	11
		Scanning	78
		Both held to side	137
		Not visible	12
MOUTH		Still and shut	108
		Still and open (no tongue)	48
		Still and open (tongue)	7
		Moving (licking and chewing)	20
		Moving (tense)	232
		Not visible	32
TAIL		Swinging	273
		Swish	72
		Swish multiple	90
		Clamped	0
		Crooked	11
		Not visible	1
WHOLE BODY		None	418
		Stopping or napping	1
		Bucking	1
		Un-cued behavior	6
		Rearing	0
		Spooking	17
		Stumbling	4

Behavior score was significantly higher for Novice level movements compared to Prelim level movements (Wilcoxon rank sum test: $P= 0.017$). Movements in working canter had higher behavior scores than movements performed in medium trot (Wilcoxon rank sum test: $P= 0.036$) or working trot (Wilcoxon rank sum test: $P= 0.018$). No association between behavior score and level of judge, venue or test number was found.

Discussion

In this study there was no significant association found between the overall occurrence of conflict behaviors (behavior score) and the performance score awarded by the judge. However, when the association between judge scores and the individual categories of behavior type was assessed, variation according to body area

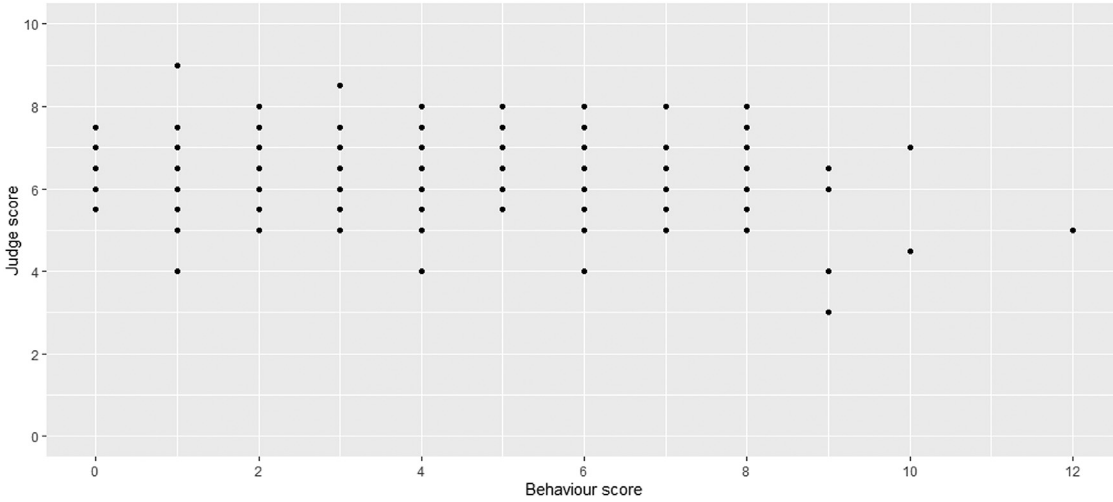

Figure 2. Score awarded by judge and behavior score seen in 447 movements analyzed in 75 dressage tests at Preliminary, Novice and Elementary levels of British Dressage competitions. Behavior score shows presence of conflict behaviors displayed by the horse (behavior score of 0 = no conflict behavior). Score awarded by judge maximum = 9 and minimum = 3. Behavior score maximum = 12 and minimum = 0.

was found. This suggests that judges are accounting for some but not all types of conflict behavior when scoring performance. Whole-body behavior, head position and movement, and ear behavior showed significant correlations with judge scores, while mouth, tail, and poll to withers behavior did not. This suggests that judges will give lower marks when some, but not all, types of conflict behavior are displayed. It may also suggest that some behaviors are more readily noticed by judges than others.

Whole-body conflict behaviors were significantly associated with judge score. Movements in which no whole-body conflict behaviors were displayed resulted in a higher judge score than movements in which bucking, stumbling, spooking, napping or un-cued behavior was seen. These behaviors are clearly visible which may explain why they are used by the judge when determining scores. This finding agrees with previous studies in which the frequency of such whole-body evasive behaviors (including rearing, bucking and refusing to go forward) negatively correlated with dressage scores (De Cartier d'Yves and Ödberg, 2005). In a dressage test, judges are looking for harmony, submission, and acceptance of the bit without tension (Niggli, 2003; FEI, 2020) so overtly evasive behaviors, or those indicative of a lack of relaxation, for example spooking, will undoubtedly accrue lower marks.

A nasal plane angle of vertical to 30° behind the vertical was significantly associated with higher judge score compared with an angle of between vertical and 30° in front of vertical. This finding suggests that judges are not adhering to FEI dressage rules which state that the horse's nasal plane should be 'slightly in front of the vertical' (FEI, 2020). Furthermore, this study agrees with a study in elite dressage horses where a behind vertical head position was significantly associated with higher scores from the judges and this head position has become increasingly prevalent in dressage (Lashley et al., 2014). These results suggest that judges are encouraging incorrect vertical and behind vertical head positions.

A head position of more than 30° in front of the vertical received lower marks than on or 30° either side of the vertical. McGreevy et al. (2010) found that the nasal plane of a horse at liberty is typically 30.7±11.5° in walk, 27.3±12.0° in trot and 25.5±11.0° in canter. A horse's head and neck position should reflect the level of training and should progress from the horse's natural head carriage to a shorter, higher head carriage to achieve collection (Niggli, 2003; Karl, 2012; Heuschmann, 2018; FEI, 2020). The nasal plane will become closer to vertical (but not on or behind the vertical). Therefore, in contrast with the results of this study, arguably horses with natural head carriage should be penalized less than those with a nasal plane on or behind the vertical; the former position may be an indication of lower training level in the horse (especially in low level competitions) and the latter position is likely to be an indication of incorrect riding and training techniques (Heuschmann, 2018). Equine nasal plane positions which exceed 30° in front of the vertical could indicate relaxed movement of the horse, presence of an alarming stimulus, play fighting or resistance to the bit (McGreevy, 2012; Condon et al., 2021). These possibilities highlight the need for using other behavioral signs (for example, ears and mouth) to determine the reason for this type of head carriage.

Judge score was significantly higher for horses with ears forward than for those with ears back. This could indicate that judges deem horses to be more relaxed with ears forward than back. This idea may have stemmed from natural behaviors of the horse: horses put their ears back during displays of aggression, during resistance and when suffering from physical exhaustion or discomfort (McGreevy, 2012).

Judge score in this study did not show any association with mouth behavior suggesting that judges do not consider tension or movement of the mouth when awarding marks. This result was unexpected as 'acceptance of the bit' is important in the ridden horse according to rulebooks (for example, FEI, 2020), dressage test sheets (for example, BD tests: British Dressage, 2020) and theoretical texts (for example, Niggli, 2003). Therefore, any resistance to the bit should be penalized by the judge. Since the mouth is such a small part of the horse's body, judges may choose to focus on larger, more obvious parts (e.g. the head and neck or the limbs). For each movement, the judge has a limited amount of time to assess the horse-rider combination and therefore small components may be missed. During the behavioral analysis, the mouth was not visible in 7% of the movements (see Table 4). Since there was only a small difference between the view of the arena from the camera position and the judge's position (Figure 1) it is likely that the mouth will be visible to the judge for a similar amount of time. In 52% of movements observed, tense/moving mouth behavior was seen and in 13% of movements an open mouth was seen. In only 28% of movements mouth behavior was ideal (either still and shut or relaxed moving). These results suggest a lack of attention to unrestricted mouth behavior by judges, but further research is needed to examine mouth behaviors displayed during training. Training should help the horse to accept the bit and perform the dressage test with the lightest contact possible. The results in this study, alongside other studies (Ödberg and Bouissou, 1999; Kienapfel et al., 2014; Górecka-Bruzda et al., 2015), indicate that there may be core problems in training and riding techniques whereby the horse's head is pulled into the 'correct' outline and time is not taken to properly train the horse to correctly carry itself and establish a relaxed contact. Additionally, Ödberg and Bouissou (1999) suggest time pressures and peer pressure may cause riders and trainers to introduce movements to a horse for which it is not prepared.

Lack of association between tail behavior and judge score suggests that judges do not consider position or movement of the tail when awarding marks. Previous research has found high frequencies of tail swishing in dressage competitions (Górecka-Bruzda et al., 2015) and an association of increased tail swishing with both higher-level competition and more complex movements (Williams and Warren-Smith, 2010; Górecka-Bruzda et al., 2015). Additionally, tail swishing is known to occur as a sign of stress (Young et al., 2012), pain (Malmkvist et al., 2012; Christensen et al., 2014) or irritation (for example, insect presence) (McGreevy, 2012), and to increase with high ambient temperatures (Kasper and Beck, 1997). Since there can be a wide variety of reasons for tail swishing in the horse it does not provide a useful behavioral indicator of conflict on its own and there may be difficulty in distinguishing between natural movement of the tail and swishing. However, tail swishing may have value when assessed alongside other behavioral parameters to evaluate conflict in the horse.

Downwards transitions showed higher behavior scores for mouth than other movement types. Horses may display conflict in different ways depending on what they have been asked to do. Head movements (for example, lifting) and tail swishing were more common in upwards transitions while tension in the mouth was more common in downwards transitions. This information may help judges to acknowledge conflict behavior in their performance assessment by allowing them to focus on specific body sections depending on the movement type. It is impractical for judges to focus on all parts of the horse's body to search for conflict behavior as well as assessing rider position, horse movement and accuracy. In addition, it may also help during training to identify the source of any problems causing conflict behaviors.

The presence of spurs was associated with the horse's nasal plane being behind the vertical which may indicate an equine coping mechanism for increased or prolonged pressure or pain on their sides. However, the presence of spurs was also associated with less head movement and lower ear behavior scores. This may indicate that spurs have useful application as an aid when riding but that they should be used with caution.

Horses who wore flash nosebands showed increased tail swishing and horses who wore Micklem bridles or grakle nosebands showed increased mouth conflict behaviors. Some pieces of tack which are designed to gain better control (for example, by closing the horse's mouth) may be causing conflict in the horse. Although the effect of tack and spurs on the occurrence of conflict behavior was not one of the main aims of this study, these findings agree with other studies concluding that some items of tack may have detrimental effects (Heleski et al., 2009; Fenner et al., 2016; Condon et al., 2021).

This study was conducted at BD affiliated competitions with horse-rider combinations who agreed to participate and were over the age of 18. Since real-life conditions were used there were several variables which could not be controlled. Within the data, there were four horse-rider combinations who completed two tests (at different levels). At both venues, competition took place indoors and measures were implemented (by event organizers) which were designed to keep distractions to a minimum. However, factors such as spectator movement and surroundings could still have affected horse behavior.

The tests used were different to each other, however, no significant effect of test was found. Camera set up ideally would have been closer to the judge, but this was not possible at these venues. This meant that the view of the horse-rider combination in the competition arena was sometimes obscured by another horse or a spectator. However, this only affected behavior recording for the head, mouth, tail and ears and occurred infrequently (see Table 4). Some behaviors in this study had low frequency (for example, headshaking, n=4; bucking, n=1; napping, n=1) and associations may have been seen in a larger dataset. Due to time and resource restrictions only one observer assessed all the movements. A second observer watched 12 movements and intraclass correlation coefficient showed an inter-observer reliability of 0.62. However, despite clear descriptors of the behaviors being recorded there was a degree of variation between observers and therefore by implication there is likely to be variation between judges scoring dressage. This suggests the importance of clear descriptors in dressage scoring in order to reach a point of consistent scoring between judges.

Future research would be valuable to examine training techniques, management style, coaching techniques, or other factors as possible contributors for these behaviors. This study agrees with others (Lashley et al., 2014; Hawson et al., 2010; Heiniger and Mercier, 2018) in highlighting the need for refining and reassessing dressage judging systems. The FEI are currently developing a more objective system (based on a code of points) for scoring Grand Prix dressage movements to improve fairness and consistency in judging and to aid the cognitive processes of judges (Dressage Judging Working Group, 2018). This is an excellent development to improve the sport of dressage, but currently it provides little to no focus on improving horse welfare. One of the requirements for this system is to '*identify and include only measurable observations which are easily visible to each judge*' (Dressage Judging Working Group, 2018). By including more detailed descriptions of behaviors indicative of conflict and embedding these in judge training, this system would serve as a means of rewarding improved approaches to training. Welfare of the horse is emphasized as being of utmost importance within competition rule books (FEI, 2020; British Dressage, 2020), however, this high frequency of conflict behaviors found by multiple studies suggests that the governing bodies need to do more to put their words into action.

Conclusion

Conflict behaviors were found to occur during the movements recorded in the BD dressage competitions assessed in this study. Variation in the relationship between specific conflict behaviors and performance evaluation was found. While certain behaviors, such as bucking and napping, were associated with reduced judge scores, others were associated with higher judge scores (for example, nasal plane on or behind the vertical). No association between more subtle behavioral signs of conflict, including tense mouth movements, and judge scores was found. The results indicate that currently dressage performance is not consistently being evaluated based on the FEI guidelines (FEI, 2020), with some behavioral signs of conflict being disregarded or, even more worryingly, rewarded. In agreement with previous research, the results of this study emphasize the need for governing bodies such as the FEI and BD to reassess their judge training and continuing professional development to better incorporate these behavioral criteria. More focus on accurately interpreting behavior and embedding this into performance evaluation would result in improved training and welfare for the horse, as well as promoting a sustainable future for equestrian sport.

Policy and Ethics Statement

The work described in this article was carried out in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki) and EU Directive 2010/63/EU for animal experiments. Ethical permission for this project was granted by the University of Edinburgh Human Ethics Review Committee (HERC 378:19) and the University of Edinburgh Veterinary Ethics Review Committee (VERC 101:19). Consent was obtained from all participants before data collection.

Authorship Statement

The authors confirm that there was no conflict of interest in relation to this study. The idea for the study was conceived by KLH, BEL, CH. The experiment was designed by KLH, BEL, CH. The experiments were performed by KLH. The data were analyzed by KLH. The paper was written by KLH, BEL, CH.

Declaration of Competing Interest

None.

Acknowledgements

We would like to thank Charlotte Woolley for her very helpful statistical support and Fiona Hamilton and Lesley Clayton for their invaluable assistance in data collection. Lastly, we wish to thank all competitors and their horses who took part in this study.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jveb.2022.07.011.

References

British Dressage (BD). 2020. British Dressage Members' Handbook: Incorporating the 2020 Rules. Available at: https://www.britishdressage.co.uk/media/2842/2020-rule-book-searchable.pdf. Accessed May 2020.
British Dressage (BD). 2021. Becoming a Judge. Available at: https://www.britishdressage.co.uk/get-involved/judges-and-writers/becoming-a-judge/. Accessed January 2022.
Bučar, M., Čuk, I., Pajek, J., Karacsony, I., Leskošek, B., 2012. Reliability and validity of judging in women's artistic gymnastics at University Games 2009. Eur. J. Sport. Sci. 12 (3), 207–215.
Casey, V., McGreevy, P.D., O'Muiris, E., Doherty, O., 2013. A preliminary report on estimating the pressures exerted by a crank noseband in the horse. J. Vet. Behav. 8 (6), 479–484.
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.
Clayton, H.M., Singleton, W.H., Lanovaz, J.L., Cloud, G.L., 2005. Strain gauge measurement of rein tension during riding: a pilot study. Equine Comp. Exerc. Physiol. 2 (3), 203–205.
De Cartier d'Yves, A., Ödberg, F., 2005. A preliminary study on the relation between subjectively assessing dressage performance and objective welfare parameters. In: Proceedings of the 1st International Equitation Science Symposium. Melbourne, pp. 89–110 26-27 AugustAvailable at.
Dressage Judging Working Group (DJWG). 2018. Dressage Judging Working Group report. Available at: https://inside.fei.org/system/files/FinalReportDJWG.pdf. Accessed May 2020.
Duncan, E., Graham, R., McManus, P., 2018. 'No one has even seen… smelt… or sensed a social licence': Animal geographies and social licence to operate. Geoforum 96, 318–327.
Dyson, S., Berger, J.M., Ellis, A.D., Mullard, J., 2017. Can the presence of musculoskeletal pain be determined from the facial expressions of ridden horses (FEReq)? J. Vet. Behav. 19, 78–89.
Dyson, S., Berger, J., Ellis, A.D., Mullard, J., 2018a. Development of an ethogram for a pain scoring system in ridden horses and its application to determine the presence of musculoskeletal pain. J. Vet. Behav. 23, 47–57.
Dyson, S., Ellis, A., Mullard, J., Berger, J., 2018b. Response to Gleerup: understanding signals that indicate pain in ridden horses. J. Vet. Behav. 23, 87–90.
Fédération Equestre Internationale (FEI), 2015. FEI Campus Glossary. Available at: https://campus.fei.org/local/globalglossary/view.php?mode=letter&hook=G. Accessed January 2022.
Fédération Equestre Internationale (FEI), 2020. Fédération Equestre Internationale Dressage Rules 2020. 25th ed. Available at: https://inside.fei.org/sites/default/files/FEI_Dressage_Rules_2020_Clean_Version.pdf. Accessed May 2020.
Fenner, K., Yoon, S., White, P., Starling, M., McGreevy, P., 2016. The effect of noseband tightening on horses' behavior, eye temperature, and cardiac responses. PLoS One 11 (5).
Gleerup, K.B., Andersen, P.H., Wathan, J., 2018. What information might be in the facial expressions of ridden horses? Adaptation of behavioral research methodologies in a new field. J. Vet. Behav. 23, 101–103.
Gleerup, K.B., Forkman, B., Lindegaard, C., Andersen, P.H., 2015. An equine pain face. Vet. Anaesth. Analg. 42 (1), 103–114.
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 (2), 137–146.
Greve, L., Dyson, S.J., 2014. The interrelationship of lameness, saddle slip and back shape in the general sports horse population. Equine Vet. J. 46 (6), 687–694.
Hall, C., Heleski, C., 2017. The role of the ethogram in equitation science. Appl. Anim. Behav. Sci. 190, 102–110.
Hall, C., Kay, R., Yarnell, K., 2014. Assessing ridden horse behavior: Professional judgment and physiological measures. J. Vet. Behav. 9 (1), 22–29.
Hawson, L.A., McLean, A.N., McGreevy, P.D., 2010. Variability of scores in the 2008 Olympic dressage competition and implications for horse training and welfare. J. Vet. Behav. 5 (4), 170–176.
Heiniger, S., Mercier, H., 2018. Judging the Judges: A General Framework for Evaluating the Performance of International Sports Judges. ArXiv e-print, arXiv:1807.10055. Available at: https://arxiv.org/pdf/1807.10055.pdf. Accessed May 2020.
Heleski, C.R., McGreevy, P.D., Kaiser, L.J., Lavagnino, M., Tans, E., Bello, N., Clayton, H.M., 2009. Effects on behavior and rein tension on horses ridden with or without martingales and rein inserts. Vet. J. 181 (1), 56–62.
Heuschmann, G., 2018. Tug of war: Classical versus "modern" dressage. JA Allen, London Rev. edn..
Hockenhull, J., Creighton, E., 2012. Equipment and training risk factors associated with ridden behavior problems in UK leisure horses. Appl. Anim. Behav. Sci. 137 (1-2), 36–42.
Karl, P., 2012. Twisted truths of modern dressage: a search for a classical alternative. Cadmos Books, Schwarzenbek.
Kasper, M., Beck, A.M., 1997. Effect of environmental temperature on the behavior of Clydesdales during preparation time before athletic performances. Equine practice (USA) 19 (6), 25–28.
Kienapfel, K., 2011. Und was meinen die Pferde dazu? Über das Ausdrucksverhalten von Pferden bei verschiedenen Halsstellungen. Pferdeheilkunde 27 (4), 372–380.
Kienapfel, K., Link, Y., Borstel, U.K.V., 2014. Prevalence of different head-neck positions in horses shown at dressage competitions and their relation to conflict behavior and performance marks. PloS one 9 (8).
Kuhnke, S., Dumbell, L., Gauly, M., Johnson, J.L., McDonald, K., Von Borstel, U.K., 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. 7 (2), 57–63.
Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33 (1), 159–174.
Lashley, M.J., Nauwelaerts, S., Vernooij, J.C.M., Back, W., Clayton, H.M., 2014. Comparison of the head and neck position of elite dressage horses during top-level competitions in 1992 versus 2008. Vet. J. 202 (3), 462–465.
Leskošek, B., Čuk, I., Karácsony, I., Pajek, J., Bučar, M., 2010. Reliability and validity of judging in men's artistic gymnastics at the 2009 university games. Sci. Gymnastics J. 2 (1), 25–34.
Malmkvist, J., Poulsen, J.M., Luthersson, N., Palme, R., Christensen, J.W., Søndergaard, E., 2012. Behavior and stress responses in horses with gastric ulceration. Appl. Anim. Behav. Sci. 142 (3-4), 160–167.
McGreevy, P.D., 2012. Equine behavior: a guide for veterinarians and equine scientists. Saunders, 2nd ed. Elsevier, London.
McGreevy, P.D., Harman, A., McLean, A., Hawson, L., 2010. Over-flexing the horse's neck: A modern equestrian obsession? J. Vet. Behav. 5 (4), 180–186.
Mullard, J., Berger, J.M., Ellis, A.D., Dyson, S., 2017. Development of an ethogram to describe facial expressions in ridden horses (FEReq). J. Vet. Behav. 18, 7–12.
Niggli, W.M., 2003. Dressage: A Guideline for Riders and Judges. JA Allen, London.
Ödberg, F.O., Bouissou, M.F., 1999. The development of equestrianism from the baroque period to the present day and its consequences for the welfare of horses. Equine Vet. J. 31 (S28), 26–30.
Smiet, E., Van Dierendonck, M.C., Sleutjens, J., Menheere, P.P.C.A., Van Breda, E., De Boer, D., Back, W., Wijnberg, I.D., Van der Kolk, J.H., 2014. Effect of different head and neck positions on behavior, heart rate variability and cortisol levels in lunged Royal Dutch Sport horses. Vet. J. 202 (1), 26–32.

Stachurska, A., Bartyzel, K., 2011. Judging dressage competitions in the view of improving horse performance assessment. Acta Agriculturae Scandinavica, Section A-Anim. Sci. 61 (2), 92–102.
Uldahl, M., Clayton, H.M., 2019. Lesions associated with the use of bits, nosebands, spurs and whips in Danish competition horses. Equine Vet. J. 51 (2), 154–162.
Waran, N., McGreevy, P., Casey, R.A., 2007. Training methods and horse welfare. In: Waran, N. (Ed.), The welfare of horses. Springer, Dordrecht, pp. 151–180.
Warren-Smith, A.K., Curtis, R.A., Greetham, L., McGreevy, P.D., 2007. Rein contact between horse and handler during specific equitation movements. Appl. Anim. Behav. Sci. 108 (1-2), 157–169.
Williams, L.R., Warren-Smith, A.K., 2010. Conflict responses exhibited by dressage horses during competition. J. Vet. Behav.: Clin. Appl. Res. 4 (5), 215.
Young, T., Creighton, E., Smith, T., Hosie, C., 2012. A novel scale of behavioral indicators of stress for use with domestic horses. Appl. Anim. Behav. Sci. 140 (1-2), 33–43.