

The Humane Society Institute for Science and Policy
Animal Studies Repository

2-15-2013

Application of Qualitative Behavioural Assessment to Horses during an Endurance Ride


Patricia A. Fleming
Murdoch University

Cheree L. Paisley
Murdoch University

Anne L. Barnes
Murdoch University

Françoise Wemelsfelder
Scotland's Rural College

Follow this and additional works at: https://animalstudiesrepository.org/acwp_asie

 Part of the [Animal Studies Commons](#), [Behavior and Ethology Commons](#), and the [Comparative Psychology Commons](#)

Recommended Citation

Fleming, P. A., Paisley, C. L., Barnes, A. L., & Wemelsfelder, F. (2013). Application of Qualitative Behavioural Assessment to horses during an endurance ride. *Applied Animal Behaviour Science*, 144(1), 80-88.

This Article is brought to you for free and open access by the Humane Society Institute for Science and Policy. It has been accepted for inclusion by an authorized administrator of the Animal Studies Repository. For more information, please contact eyahner@humanesociety.org.



Application of Qualitative Behavioural Assessment to Horses during an Endurance Ride

Patricia A. Fleming¹, Cheree L. Paisley¹, Anne L. Barnes¹, Françoise Wemelsfelder²

¹ Murdoch University

² Scotland's Rural College

KEYWORDS

behavioural expression, equine, competition, endurance, QBA, ride place

ABSTRACT

Endurance horses are considered subject to a unique set of training and competing pressures due to the long distances travelled. The health and welfare of these horses could be compromised if they have not been adequately trained or are pushed beyond their limits, and there are increasing concerns regarding the capacity of horses to cope with the exercise demands placed on them, with high elimination rates for lameness and metabolic reasons. Veterinary inspections during these rides are important for assessing physiological measures, but the inclusion of behavioural assessments is also warranted. We investigated the application of Qualitative Behavioural Assessment (QBA) as a method for assessing demeanour in horses engaged in a 160-km endurance ride. We used a Free Choice Profiling (FCP) methodology to analyse footage of 10 horses collected during veterinary inspection (1) pre-ride, (2) midway through and (3) at the end of the ride. FCP allows each individual observer to develop his/her own unique set of descriptive terms for scoring the behavioural expression of animals. Observers ($n = 22$) reached consensus in their assessment of the behavioural expression of the 10 horses ($P < 0.001$). The first dimension of behavioural expression was characterised by terms such as 'calm', 'content' and 'relaxed' contrasting with 'agitated', 'angry' and 'annoyed'. Scores on this dimension did not differ between the three time points ($P = 0.372$), but did demonstrate significant individual differences ($P = 0.004$). This dimension may capture individual responses to the veterinary inspection procedures or the general endurance environment. On the second dimension, observers scored animals as more 'alert', 'curious' and 'excited' pre-ride and more 'tired', 'lazy' and 'sleepy' mid-ride and at the end of the ride ($P = 0.001$), which seems to indicate a more general effect of the race on the horses' state. There were also significant differences between individual horses on this behavioural dimension ($P = 0.028$). Interestingly, the behavioural assessment scores attributed to horses pre-ride on dimension 2 were correlated with the horse's final ride place ($P < 0.01$), suggesting that observers detected behavioural expressions reflecting the horses' interest in the event. These results support the potential of QBA as a useful tool for evaluating horses' emotional and attitudinal response to endurance events.

1. Introduction

Athletic competition can be stressful for horses (Cayado et al., 2006), and subsequently there is a need to evaluate the effect of various types of events and common equestrian practices on horses. Equestrian competition is governed by strict rules based on the tenet that the welfare of the horse is paramount (Federation Equestre Internationale (FEI), 2011). However, welfare encompasses both physical and mental health (Broom, 1991; Dawkins, 2004, 2006), and while the physical health of horses may be monitored during events, there are few measures that assess the mental state of horses during competition. An animal's mental state can be expressed through their behaviour (Broom, 1991). Thus, in conjunction with the many measures taken during physical examinations (e.g. Cayado et al., 2006; Frazier, 2000; König von Borstel et al., 2011; Munsters et al., 2011), behavioural assessments should be recognised as important for evaluating the mental state of competition horses.

There are a number of methods of evaluating behavior and welfare of horses at rest or in stables, including the use of ethograms (e.g. Hausberger et al., 2009; Lloyd et al., 2007) and the assessment of stereotyped behaviours as a psychological welfare indicator (Henderson, 2007; Mason and Latham, 2004). There are also measures of the response of horses to various tests, such as novel object tests (e.g. Visser et al., 2002). However, the focus on athletic performance may limit the repertoire of behaviours that can be assessed during competition, while human presence is also likely to influence how the horse conducts itself (König von Borstel et al., 2011; Visser et al., 2008). Qualitative Behavioural Assessment (QBA) is a method that uses human observers to assess and score the behavioural expression of animals (Wemelsfelder et al., 2000, 2001; Wemelsfelder and Lawrence, 2001). QBA takes an integrative approach by addressing the whole animal, where observers summarise all perceived details of an animal's posture and movement into descriptions of expressive demeanour (e.g. 'relaxed', 'anxious', 'playful', 'content'). Such terms do not describe what an animal is doing but how it goes about doing what it is doing – in other words, the animal's behavioural 'style' or 'body language' (Feaver et al., 1986; Stevenson-Hinde, 1983; Wemelsfelder, 1997, 2007). For example, immobility which is coupled with a rigid body stance, fixed, staring eyes and a tense mouth is very different from immobility associated with a lowered head and relaxed ears, eyes and tail (Minero et al., 2009). Integrating such subtle details can allow experienced animal workers to detect subtle shifts in demeanour, important for welfare assessment, that may be overlooked by isolating and quantifying individual physical behaviours (Meagher, 2009; Wemelsfelder, 1997, 2007; Whitham and Wielebnowski, 2009).

QBA builds on scientific approaches developed for assessing individual differences, temperament and personality, and extends these to the assessment of experience and welfare (Wemelsfelder, 2007; Wiseman-Orr et al., 2006). Body language descriptors such as 'content' and 'anxious' have an emotional connotation that refers to the animal's affective experience of its surroundings, and could potentially provide information that is highly relevant for the evaluation of an animal's well-being. The use of such terms may formalise skills that animal care professionals acquire through their daily work with animals, learning to filter large amounts of information to judge the wellbeing of individuals (Whitham and Wielebnowski, 2009). QBA has been shown to have merit as part of welfare assessments for a range of species, including horses (Minero et al., 2009; Napolitano et al., 2008), cattle (Brscic et al., 2009; Rousing and Wemelsfelder, 2006; Stockman et al., 2011, 2012a, 2012b), pigs (Rutherford et al., 2012; Temple et al., 2011; Wemelsfelder et al., 2000, 2001, 2009c), sheep (Wickham et al., 2012), poultry (Wemelsfelder, 2007) and dogs (Walker et al., 2010). QBA scores are also correlated with quantitative measures of behaviour (Minero et al., 2009; Napolitano et al., 2008, 2012; Rousing and Wemelsfelder, 2006) and with physiological indicators relevant to welfare (e.g. heart rate, core body temperature, plasma glucose and the neutrophil: lymphocyte ratio; Stockman et al., 2011, 2012b; Wickham et al., 2012).

Importantly, QBA is a method that can either be applied retrospectively, e.g. to assess animals on video footage and determine differences between treatments, or has the potential for immediate use, for example in on-farm welfare assessments (e.g. Brscic et al., 2009; Rousing and Wemelsfelder, 2006; Temple et al., 2011). With horses in endurance events, QBA could be used as part of the assessment of animals at regular observation points during competition, as well as a tool to evaluate differences in horse responses to training, competition, riding style, speeds, gear, environments, etc.

This study reports the application of QBA to horses involved in a 160-km competitive endurance ride. Endurance rides impose considerable physical and mental challenges to some horses, due to the distance travelled and the long duration of the event (Bergero et al., 2005; Frazier, 2000). Endurance competitions around the world are increasing in speed, and high elimination rates for lameness and metabolic reasons (Marlin et al., 2008; Nagy et al., 2010) have sparked concerns regarding the capacity of horses to cope with the exercise demands placed on them. In this study, our aim was to investigate the effect of endurance rides on the expressive demeanour of horses to see if, and how, application of QBA might reveal differences between individuals and across time that could potentially reflect the performance and welfare state of these animals.

2. Methods

2.1. Animals and experimental design

This study was based on video-footage of privately-owned endurance horses (of 20 senior competitors), collected during a 160-km endurance ride event at Collie, Western Australia (WA). The endurance ride was conducted under the rules set by the FEI governing endurance competitions (FEI, 2011), whereby the same horse and rider compete over a set course and distance against the clock. The total ride distance is separated into multiple phases or loops. Before being allowed to start the ride, between the loops, and at the end of the ride, the physical health and recovery of the horses is stringently evaluated at veterinary inspections to ensure the horse is fit to continue. The ride was conducted as vet-gate-into-hold (VGIH), where the arrival time is recorded for each leg of the ride and the rider has up to 30 min to present the horse at the 'vet gate' for veterinary inspections; the actual time taken to present is added to the ride time. Riders brought their horse for veterinary inspection once they judged that the horse had recovered sufficiently (heart rate and recovery criteria). Under VGIH, there is a subsequent compulsory rest period ('hold'), ranging from 30 to 60 min, before returning to the track for the next loop. Specific details for the Collie ride: pre-ride veterinary inspection was carried out in the afternoon/evening; the ride commenced at midnight, with horses finishing 10–17 h later. There were six ~30-km loops in this event, with veterinary inspections occurring after each.

Horses eliminated from the ride did not present for the end of ride veterinary inspections and therefore we only examined animals that successfully completed the event. Fourteen of the 20 horses successfully completed the event, and clear footage was obtained for pre-ride, midway through and end of ride veterinary inspections for 10 of these. All 10 horses included in this study had no adverse findings at any of the veterinary inspections (i.e. all horses passed veterinary inspection), and so there was insufficient variability in the data to enable statistical analyses relating their behavioural expression to the findings of those inspections. In our experience, presentation times varied according to the rider's level of competitiveness, with some riders presenting their horses to the vet gate immediately they were eligible on the basis of heart rate, while other riders did not do so. Presentation times would therefore not reflect the horses' abilities.

All horses were under the control of their owner and/or rider at all times, and so were fed, watered and cared for appropriately as judged by their owner and/or rider. Veterinarians were present at all times to

perform the inspections, and in case of emergency. This study had the approval of the Animal and Human Ethics Research Committees at Murdoch University, Perth, WA (R2153/08 and 2008/047 respectively). Owner consent was obtained to film the horses during the event.

Horses were filmed (15 fps; Panasonic SDR-H250 camcorders) during the veterinary inspection at a distance of 5–10 m (the cameraman stood outside the designated veterinary inspection area). Footage was collected during veterinary examination (horses were examined for heart rate; respiratory rate; and hydration: mucous membranes, capillary refill time, jugular recovery time, skin tenting) as well as the trot-out: the horses were then led out at a trot for approximately 60 m to check for any signs of lameness, as well as noting impulsion.

The footage included all horses that finished the event, for which we had succeeded in obtaining adequately clear footage for all three time points. Horses were a mix of seven geldings, two mares and one stallion, all Arabian or Arabian cross, and ranging from 7 to 16 years old. Eight animals were from WA and two were from Queensland, having been trucked over 4000 km to WA for the event several days earlier.

Footage was edited (Adobe Premier Pro CS3 and Adobe After Effects CS3, Chatswood, NSW, Australia) into clips 45–70 s in duration which included veterinary examination of the horse as well as the trot out. The clips did not deliberately include any identifying information, but the number on the horse or rider's identification bib was visible in some clips. The footage included thirty video clips, one for each of the 10 horses at each of three time points: pre-ride, midway through, and at the end of the ride.

2.2. Qualitative Behavioural Assessment (QBA)

QBA was applied to the acquired video footage using a Free-Choice Profiling (FCP) methodology, which relies on human observers generating descriptors and quantified scores for observed animals (Wemelsfelder et al., 2001). Twenty-two observers were recruited for this study as part of a laboratory session teaching Animal Science and Veterinary Science students about assessment methods for equine behaviour. To ascertain whether these observers were experienced with horses or not, they were given a survey asking whether they had a pet (any species), how often they currently came into contact with horses, the total time over their entire life in contact with horses, and whether they currently owned or leased a horse. Observers were classified as 'experienced' if they currently came into contact with horses daily or weekly, had cumulatively more than a year over their life in contact with horses and currently owned/leased a horse.

Observers were told that they would be shown 30 clips of horses filmed during veterinary inspection as part of an endurance ride, but they were not told the length of the ride or that they would see three clips of each horse at different stages of the ride (i.e. they were not told the number of horses they would be shown). FCP consists of two phases, term generation and quantification, and both these phases took place in one, 3-h session. For term generation, observers were shown 16 video clips of individual horses showing a range of behaviour thought to be representative of the overall repertoire of endurance horses, including horses filmed at the ride and also horses filmed in other situations (e.g. at their home pen, feeding, being led around to graze). At the end of each clip, observers were given 2 min to write down terms they felt suitably described the expressive qualities of the observed animals. Negative terms were converted to their positive form for ease of scoring (e.g. 'unhappy' was listed as 'happy', and an 'unhappy' horse would therefore be scored low for this term). These descriptive terms were then sorted alphabetically for each observer, and presented to observers with a 100-mm visual analogue scale adjacent to each term. In the subsequent quantification phase, observers scored the 30 randomly-arranged video clips on these personal rating scales; they were told to think of the distance between the

zero-point and their mark on the scale as reflecting the intensity of the animal's expression on each descriptive term.

The distance from the start of the visual analogue scale to where the observer had made a mark was measured in millimetres, and these measurements were entered into individual observer Excel files (Microsoft Excel 2003, North Ryde, NSW, Australia). Observer scoring patterns were analysed using a specialised GenStat software edition written for Franc, oise Wemelsfelder, including both Generalised Procrustes Analysis (GPA) and Principal Component Analysis (PCA) as described below. Briefly summarised, GPA calculates a consensus or 'best fit' profile between observer assessments through complex pattern matching. Because each observer scores the same footage, GPA captures the similarity in scoring patterns between observers. The level of consensus (i.e. the percentage of variation between observers explained) that was achieved was expressed as the Procrustes Statistic. Whether this consensus is a significant feature of the data set, or, alternatively, an artefact of the Procrustean calculation procedures, is determined through a randomisation test (Dijksterhuis and Heiser, 1995). This procedure rearranges at random each observer's scores and produces new permuted data matrices. By applying GPA to these permuted matrices, a 'randomised' profile is calculated. This procedure is repeated 100 times, providing a distribution of the Procrustes Statistic indicating how likely it is to find an observer consensus based on chance alone. Subsequently a one-way t-test is used to determine whether the actual observer consensus profile falls significantly outside the distribution of randomised profiles.

Through Principle Components Analysis (PCA), the number of dimensions of the consensus profile is reduced to several main dimensions (usually two or three). Interpretation of these consensus dimensions is made possible by selecting terms for each observer that correlated strongly with the consensus dimensions (Pearson's $R > 0.6$ for GPA 1 and $R > 0.5$ for GPA 2). It should be stressed that all these processes are the result of mathematical calculation procedures that in no way depend on interpretation by the researchers.

Each animal receives a quantitative score on each of the consensus dimensions. To statistically analyse behavioural expression scores, we undertook mixed-model ANOVA, with horse ID included as a random factor, ride stage as the fixed factor, and scores for each horse for each GPA dimension as the dependent variables. The data did not violate the assumptions of parametric analyses (Levene's test and Shapiro-Wilk's W test). To investigate the correlation between GPA scores for individual horses and their final ride place or their age (which were not continuous data), Spearman Rank Order Correlation analysis (r_s) was carried out for each stage of the ride.

Statistical analyses were carried out using GenStat 10.2 (VSN International, Hemel Hempstead, Hertfordshire, UK) and Statistica 8.0 (StatSoft-Inc, Tulsa, OK, USA). Data are presented as means ± 1 SD and a statistical level of $\alpha \leq 0.05$ is used throughout.

3. Results

The 22 observers generated, among them, a total of 108 unique descriptive terms (average 18 ± 5 terms per observer; range 11–27). All students had some type of companion animal, but there was a range of experience with horses, from daily contact with their own horse, through to little contact (only a few times a year). Around half of observers were identified as 'experienced' with horses ($n = 14$).

The GPA consensus profile explained 47.9% of the variation in observer scoring patterns, which differed significantly from the mean randomised profile ($35.0 \pm 0.1\%$, $t_{99} = 73.52$, $P < 0.001$). Two main GPA dimensions explained 57.5% of the variation in scores attributed to individual animals; 43.0% of variation in the data was captured by GPA dimension 1 and 14.5% by GPA dimension 2 (see Table 1 for a list of

terms associated with each GPA dimension). Examples of GPA dimension 1 and 2 terminologies are shown for two individual observers (Fig. 1). Although observers used different sets of terms, there is semantic consistency in how they applied those terms to scoring the 10 endurance horses observed. All observers scored terms that were strongly correlated with GPA dimension 1 ($R > 0.6$; e.g. Fig. 1a), but not all observers generated descriptive terms correlated with GPA dimension 2 above a correlation level of $R > 0.5$ (e.g. Fig. 1b). Both horse-experienced and non-experienced observers scored terms that were strongly correlated with GPA dimension 2.

Table 1. Terms used by observers to describe behavioural expression of horses filmed during vetting at the beginning, midway through, and at the end of a 160-km endurance ride (ride stage).

GPA dimension	Low values	High values	Effect of ride stage; correlation with relative placement
1 (43.0%) (terms correlated $r > 0.6$ correlation with consensus axis)	Calm (8), Content (4), Relaxed (4) , Patient (2), Trusting, Submissive Quiet, Willing, Comfortable, Happy, Cooperative, Obedient	Agitated (11), Angry (10), Annoyed (9) , Defensive (6), Uneasy (5), Impatient (4), Restless (3), Flighty (2), Anxious (2), Hesitant, Moody, Pissed Off, Hostile, Irritated, Bossy, Distressed, Highly Strung, Reluctant, Frustrated, Discontent, Aggressive, Stressed, Crazy, Grumpy, Excited, Pushy, Startled, Worried, Uncomfortable	Ride stage: NS: $F_{2,18} = 1.04$, $P = 0.372$ Relative placement: NS
2 (14.5%) (terms correlated $r > 0.5$ correlation with consensus axis)	Tired (6), Lazy (2), Sleepy (2) , Relaxed, Exhausted, Sad, Weary, Shattered	Alert (8), Curious (4), Excited (4) , Active (2), Inquisitive (2), Playful (2), Energetic (2), Distracted (2), Observant, Interested, Aware, Anticipation, Restless, Confident, Focussed, Hyper, Proud	Ride stage: $F_{2,18} = 10.01$, $P = 0.001$ Relative placement: Pre-ride correlated ($P = 0.001$)
The terms shown are those that had the highest correlation with each end of each GPA dimension axis (% of variation in behavioural expression accounted for by each dimension). Term order is determined firstly by the number of observers to use each term (in brackets if greater than one), and secondly by weighting of each term (i.e. correlation with the GPA dimension). Terms in bold were used to describe the GPA dimensions in the text and graphs. The last column shows a summary of statistical analyses for each dimension.			

GPA dimension 1 (43.0% of the variation) was characterized by terms such as ‘calm’, ‘content’ and ‘relaxed’ on the low end of the axis, and ‘agitated’, ‘angry’ and ‘annoyed’ at the high end. GPA dimension 1 scores did not differ between the three time points (MM-ANOVA ride stage: $F_{2,18} = 1.04$, $P = 0.372$; Fig. 2), but demonstrated individual differences (horse ID: $F_{9,18} = 4.43$, $P = 0.004$; Fig. 3). The strong effect of horse ID suggests that GPA dimension 1 captured behavioural responses that were relatively consistent over successive time points. GPA dimension 1 scores attributed to individual horses for the pre-ride clips were correlated with their scores for mid-ride clips ($P = 0.03$) but not the end of the ride ($P = 0.16$); mid-ride scores only showed a tendency to correlate with end of ride scores ($P = 0.07$). The effect of horse ID

can be seen through examination of the scores for individual horses (Fig. 3); for example, two geldings were consistently scored higher on this dimension at all stages of the ride (Fig. 3, grey symbols), i.e. these individuals were recorded as more ‘agitated/angry/annoyed’. These animals held their ears pinned back and had puckered nostrils during the veterinary procedures. The one stallion in the group (Fig. 3, white symbols) was initially scored as more ‘agitated/angry/annoyed’ but less so midway through and at the end of the ride.

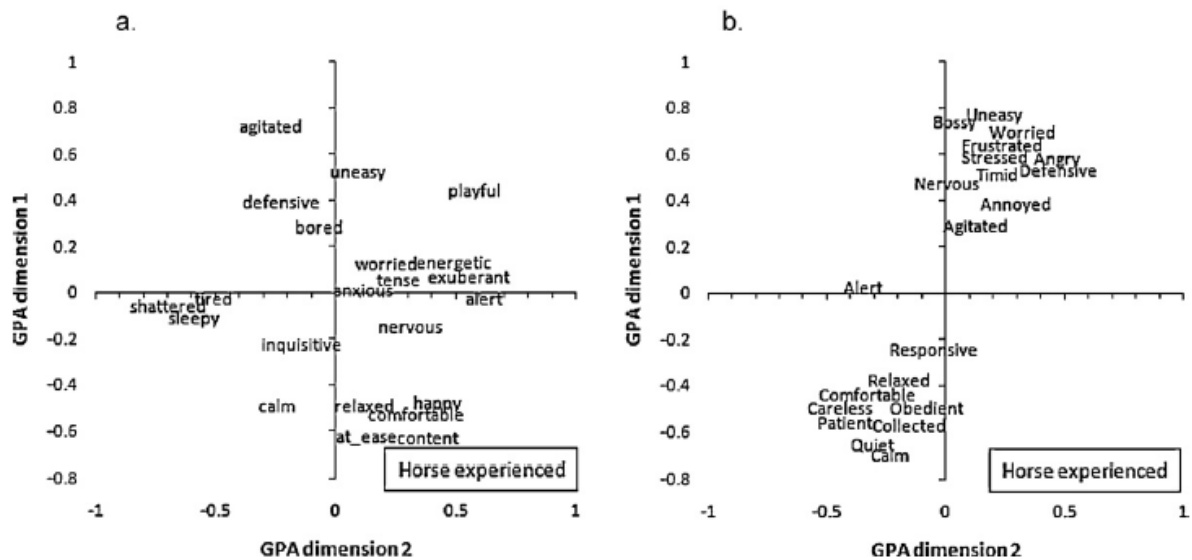


Fig. 1. Comparison of the terms used by two observers, where their descriptive terms generated under Free Choice Profiling are plotted according to their correlation with GPA consensus dimensions 1 (y-axes) and 2 (x-axes). Both observers were experienced with horses but observer (b) did not have terms that were strongly correlated with GPA dimension 2.

GPA dimension 2 (14.5% of the variation) was characterized by terms such as ‘tired’, ‘lazy’ and ‘sleepy’ on the low end of the axis and ‘alert’, ‘curious’ and ‘excited’ at the high end. Dimension 2 scores varied significantly between time points (GPA scores pre-ride: 0.04 ± 0.05 , mid-ride: -0.01 ± 0.04 , end of ride: -0.03 ± 0.04 ; ride stage: $F_{2,18} = 10.01$, $P = 0.001$; Fig. 2), with horses scored as significantly more ‘alert’, ‘curious’ and ‘excited’ during pre-ride veterinary inspection, and more ‘tired’, ‘lazy’ and ‘sleepy’ during mid-ride and end of ride veterinary inspection. There were also significant differences between individuals (horse ID: $F_{9,18} = 2.85$, $P = 0.028$; although these were not as marked as for GPA dimension 1). The GPA dimension 2 scores attributed to individual horses for the pre-ride clips were not correlated with their scores for midride ($P = 0.57$) or end of the ride ($P = 0.90$) and there was only a trend for correlation between mid-ride scores and end of ride scores ($P = 0.07$). There was a correlation between pre-ride dimension 2 scores and the individual horse’s final ride place ($r_s = -0.91$; $P < 0.01$; Fig. 3); however neither mid-ride nor end of ride GPA dimension 2 scores were significantly correlated with ride place.

There was no significant correlation between GPA dimension scores and age (tested separately for each ride stage).

4. Discussion

This study applied QBA to quantify the expressive demeanour of 10 horses before, midway through and at the end of a 160-km endurance ride. Twenty-two observers agreed significantly on their assessments of the horses, and distinguished two main dimensions of horse expression: ‘calm/content/relaxed’ – ‘agitated/angry/annoyed’ (GPA dimension 1 – 43%), and ‘tired/lazy/sleepy’ – ‘alert/curious/excited’ (GPA dimension 2 – 14.5%). There were significant differences between mean horse scores for the three ride stages on dimension 2, with horses being perceived as more ‘tired/lazy/sleepy’ as time progressed. Individual horse scores on dimension 1 were correlated across the first two stages of the ride, and showed a tendency to do so for the second and third stage, indicating a certain stability in the horses’ level of agitation throughout the ride. There was also a correlation between the horses’ pre-ride scores and final pace in the ride on dimension 2, with horses perceived to be more ‘alert/curious/excited’ completing the ride in shorter times than animals scored as more ‘tired/lazy/sleepy’.

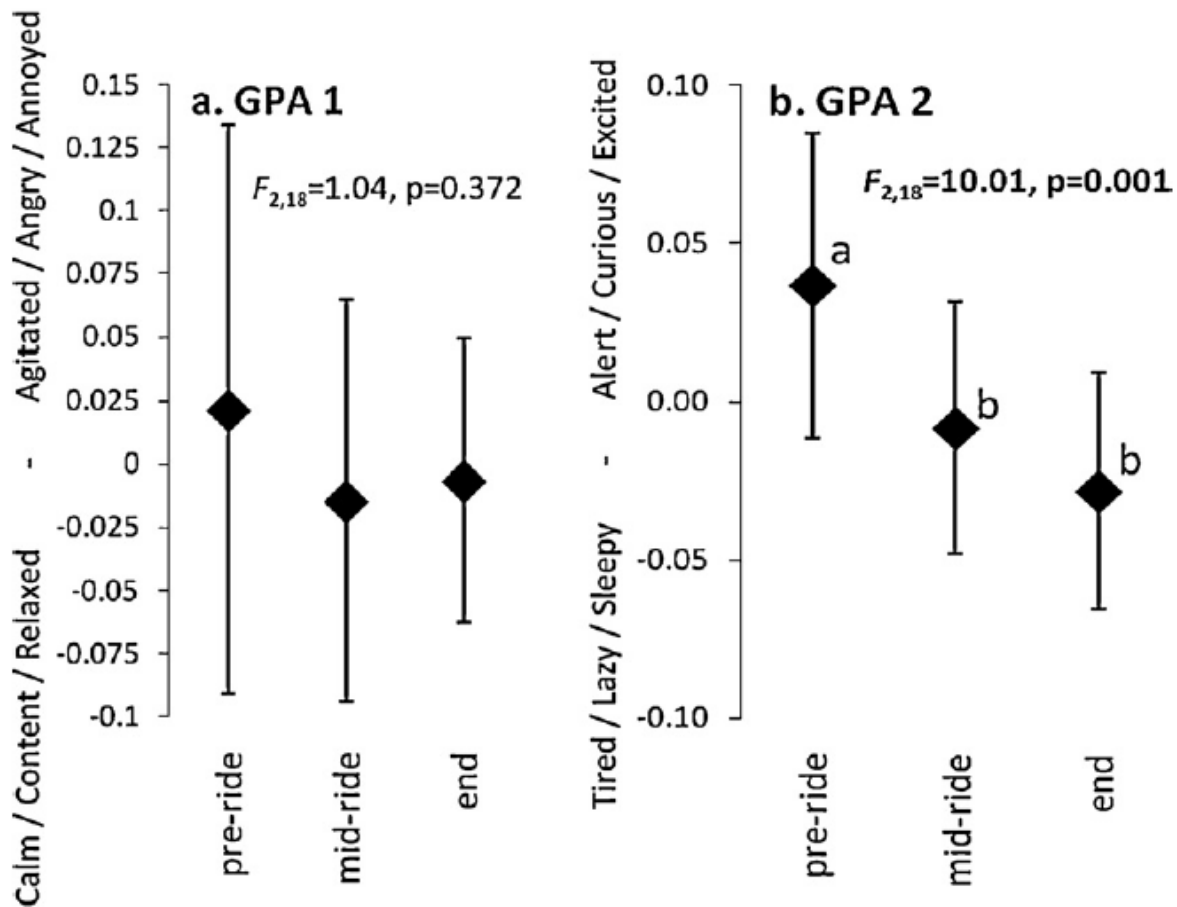


Fig. 2. Average (± 1 SD) GPA dimension scores for 10 horses filmed at the beginning, midway through, and end of a 160-km endurance ride. The effects of ride stage on each GPA dimension are indicated (results of mixed-model ANOVA), with letters in (b) linking ride stages that were not significantly different ($P > 0.05$) by Tukey’s HSD post hoc analysis.

Mean horse scores on GPA dimension1 did not differ significantly between different stages of the ride, while rankings of individual horses on this dimension were correlated across different stages, which seems to indicate that horses' responsiveness showed stability across the ride, and may have been affected by individual temperaments (AERC, 2008). Temperament has been defined as relative stability in responsiveness within individuals over time and across situations (Visser et al., 2001), and a criticism of behavioural studies is that assessments are often carried out with little background knowledge of the effect of temperament on an animal's response (Seaman et al., 2002). The results of the present study suggest that QBA is capable of repeatedly capturing individual response differences over time, and as such may be a useful additional tool in temperament research, particularly in applied situations where use of separate temperament tests is not feasible. The application of QBA to temperament research is novel and warrants further research.

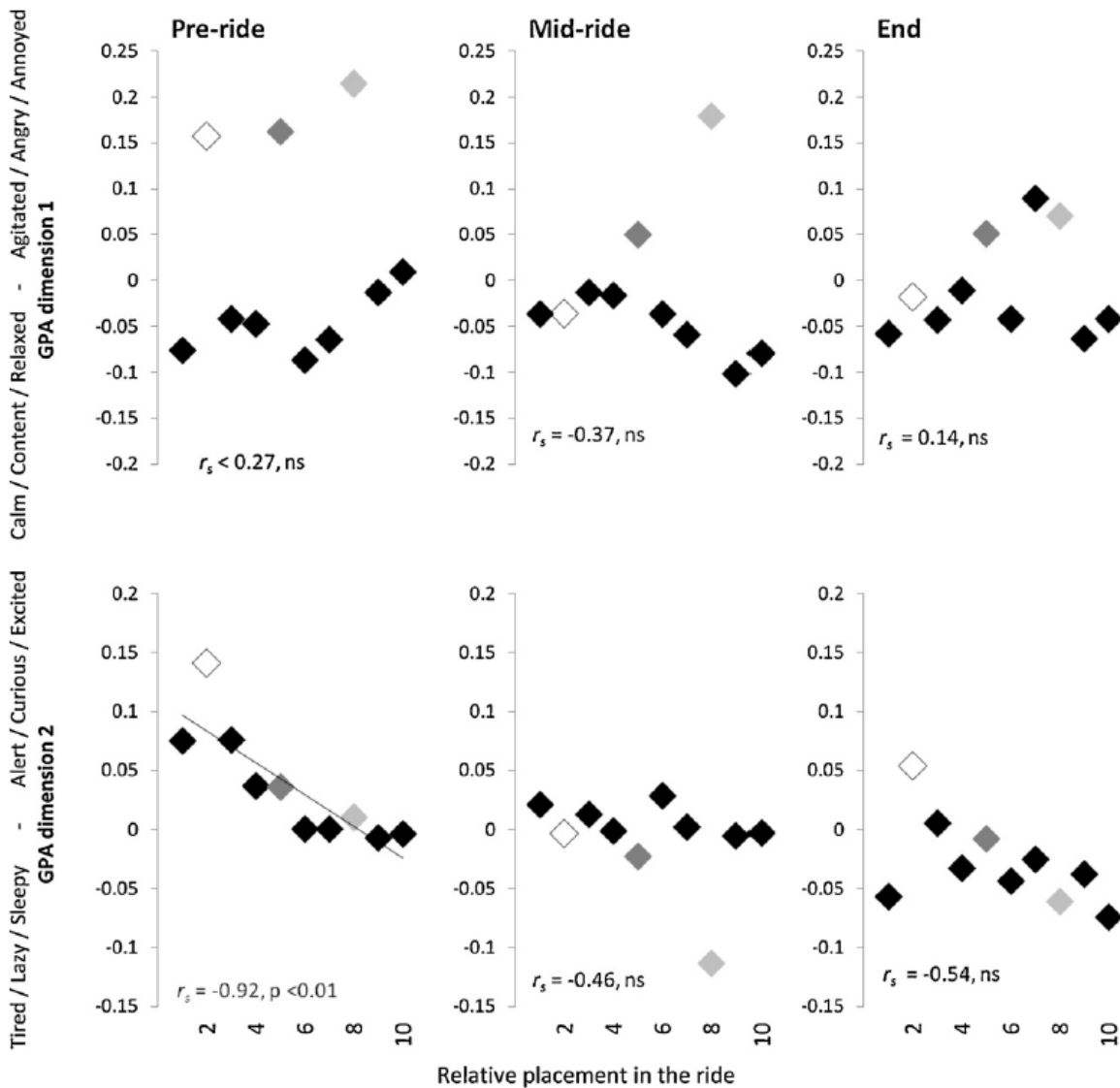


Fig. 3. The GPA dimension scores for each individual horse at the beginning, midway through, and end of the ride plotted against its eventual relative placement in the ride (i.e. time to completion). The one stallion is shown with a white symbol and two geldings whose behaviour stood out consistently are shown with grey symbols.

GPA dimension 2, capturing level of alertness, curiosity and excitement, addresses an important aspect of a horse's response to extensive exercise. There was a significant effect of stage of the ride on the horses' mean scores, with animals scoring as more 'tired/lazy/sleepy' midway through and at the end of the ride than at the start. Furthermore, one of the study's most interesting findings was that the horses' pre-ride scores on this dimension, before they had done any exercise, was correlated with their ultimate placing in the 160-km ride: horses scored as more 'alert/curious/excited' pre-ride finished earlier than those scored pre-ride as more 'tired/lazy/sleepy'. This finding suggests that horses may have different levels of interest (e.g. being aware of their environment – head up, looking around, neighing, watching other horses/people, responsive to being handled, reacting to commands), competitiveness and/or fitness (which our results indicated were unrelated to age, while all animals were of similar breeding) that could affect the outcome of a competitive event. The influence of their riders (i.e. differing levels of competitiveness) is also likely to be substantial, although we note that there was often a team of handlers present at the veterinary arena and it was not always the riders themselves that presented their horses. Results of the present study contrast with a previous study on race horse demeanour and performance. Hutson and Haskell (1997) found that 'pre-race' measures of arousal (tail elevation, neck angle with the jockey up, as well as a horse's resistance to the bit) were correlated to finishing order (although less so than starting price or weight carried). However, the authors found that winners (top 20%) were fitter but also more relaxed (relaxed tail, lowered neck, less resistance to the bit) than losers (last 20%), which were more aroused and required greater control. The differences between our study and that of Hutson and Haskell (1997) may relate to the myriad differences between endurance rides and sprinting performance.

As GPA dimension 2 scores were not correlated across different ride-stages, it is difficult to gather to what extent these differences were affected by temperament, and/or by circumstances before the start of the ride. Another finding was that not all observers generated terms that were strongly correlated to this dimension. Recognition of GPA dimension 2 behaviour was irrespective of an observer's experience with horses, since observers that were both horse experienced or not-experienced scored terms that correlated strongly with this dimension.

The present study demonstrates that QBA may provide a useful characterisation of behavioural expression in horses, capable of capturing details and differences that are potentially relevant to welfare, but might elude other methods. Key expressive terms such as 'curious', 'excited', 'tired', 'lazy', and 'sleepy' {note that we have not included the term 'alert' due to ambiguity in how different observers use this term; noted in the present study as well as QBA studies in sheep (Wickham et al., 2012) and cattle (Stockman et al., 2011)} could be useful in distinguishing the likely performance of horses under a range of events, but also reflect affective states (e.g. fatigue) that are difficult to assess using other methods. Napolitano et al. (2008) applied QBA to individual horses and ponies in an open field test, and found these animals' positions on different GPA dimensions correlated with both quantitative behaviours as well as owners' and instructors' perceptions of these horses' responsivity. The authors argue that QBA's multifaceted expressive characterisations of horses may better capture how individual animals differ than more rigid differentiations such as 'coping style' (e.g. active copers try to escape or remove aversive conditions vs. passive copers that do not respond to aversive stimuli and wait for change) or 'temperament' (e.g. exploration, aggression, sociality). Minero et al. (2009) found that yearling foals responded to a month-long handling treatment and were seen to be more 'suspicious/nervous' and 'impatient/reactive' before handling, and 'explorative/sociable' and 'calm/apathetic' after handling. These dimensions correlated meaningfully to quantitative measures of the horses, and the authors argue that QBA may add an interpretative layer to welfare studies encapsulating subtle differences that are difficult to quantify otherwise.

Together, previous and current findings indicate the potential relevance of qualitative measures of behavior to competitive horse sports for assessing various aspects of the horses' performance, such as individual variation, consistency over time, and shifts in affective experience. QBA can be quickly applied under any situation where working equines can be observed, regardless of breed or work type, and requires little disturbance to the animals. Similarly to the present study, Pritchard et al. (2005) qualitatively assessed demeanour in working horses, and reported that ridden horses were more likely to be scored as 'alert' and more likely to show a response (either friendly or avoidance/aggression) than draught or pack horses which were more likely scored as 'apathetic/severely depressed' or showing no response. These types of qualitative assessment meet the requirements listed by Sørensen et al. (2001) for satisfactory welfare indicators, since they capture responses that: (1) are relevant to an individual animal's welfare, (2) are sensitive to (changes in) an animal's immediate environment; (3) are not outside the control of stakeholders (as disease outbreaks would be), and (4) are relatively cheap and easy to quantify. To this list we would like to add that, as discussed above, qualitative measures (5) allow evaluation of brief and subtle responses that are not easily captured by other methods, but are nevertheless important to welfare (Whitham and Wielebnowski, 2009). Qualitative measures of behavior may encapsulate concepts not otherwise easily scored, for example a horse's level of 'interest' and 'exhaustion' in the present study, or of 'friendliness' in the study by Pritchard et al. (2005). Qualitative measures of behavior may therefore provide important information for welfare assessments. Supporting this contention, various authors have suggested QBA may be a promising and immediately applicable methodology for assessing positive welfare in animals (Boissy et al., 2007; Napolitano et al., 2009).

In terms of the present study, the more specific question is whether QBA could contribute usefully to endurance ride veterinary inspection. The American Endurance Ride Conference (AERC, 2008) include evaluation of the behaviour of horses during veterinary inspection, particularly in relation to determining fatigue. The AERC guidelines (AERC, 2008) includes a grading for 'attitude' and 'overall impression' on the veterinary card (there is no definition of these terms under the guidelines and the authors appear to recommend subjective assessment of these qualities). In a retrospective study of almost 3500 horses taking part in endurance events, Langdon Fielding et al. (2011) reported that 'overall impression' of horses recorded pre-ride was associated with elimination from the ride due to metabolic issues (e.g. heart rate, cardiac recovery index, abnormal gastrointestinal sounds), while that recorded in the first half of the ride was associated with overall elimination. They also reported that when 'attitude' was not recorded pre-ride, this was associated with an increased risk of lameness elimination and overall elimination, and they suggest veterinarians may have been reluctant to score ambiguous or negative results. These findings indicate that there is value in veterinarians recording details about horses' attitude and overall impression, and the present study suggests that QBA may be a valuable methodology for pursuing such goals.

In conclusion, we propose the use of QBA as a measure of horse behavioural expression during competition events, to address various aspects of horse performance and welfare. Validation and application of QBA has evolved from Free-Choice Profiling assessment of video footage to the development of standard fixed-term lists that assessors can use on site (e.g. at farms or races) to provide immediate information and feedback about the animals' state (Wemelsfelder et al., 2009a,b; Wemelsfelder and Millard, 2009). Such fixed-list evaluation could be useful as a relatively quick, non-invasive method, which, when linked to physical findings, could contribute to decisions regarding the physical and mental capacity of animals to continue an event. While many equestrians and officials may subconsciously assess their horses, the inclusion of agreed-upon qualitative terms in assessment scoring methods could provide an overt, quantifiable measure of behavioural expression.

Acknowledgements

Our thanks to all the endurance riders who permitted us to film their horses and to the observers who participated in this study. Thanks also to Sarah Wickham for the use of her Excel macro worksheets and Catherine Stockman for assistance with data analysis.

References

- AERC, 2008. American Endurance Ride Conference. Rider Handbook 3.0.
- Bergero, D., Assenza, A., Caola, G., 2005. Contribution to our knowledge of the physiology and metabolism of endurance horses. *Livestock Prod. Sci.* 92, 167–176.
- Boissy, A., Manteuffel, G., Jensen, M.B., Moe, R.O., Spruijt, B., Keeling, L.J., Winckler, C., Forkman, B., Dimitrov, I., Langbein, J., Bakken, M., Veissier, I., Aubert, A., 2007. Assessment of positive emotions in animals to improve their welfare. *Physiol. Behav.* 92, 375–397.
- Broom, D.M., 1991. Animal welfare: concepts and measurement. *J. Anim. Sci.* 69, 4167–4175.
- Brscic, M., Wemelsfelder, F., Tessitore, E., Gottardo, F., Cozzi, G., Van Reenen, C.G., 2009. Welfare assessment: correlations and integration between a Qualitative Behavioural Assessment and a clinical/health protocol applied in veal calves farms. *Proceedings of the 18th National Congress ASPA, Palermo, Italy. Ital. J. Anim. Sci.* 8, 601–603.
- Cayado, P., Muñoz-Escassi, B., Domínguez, C., Manley, W., Olabarri, B., Sánchez De La Muela, M., Castejon, F., Marañón, G., Vara, E., 2006. Hormone response to training and competition in athletic horses. *Equine Vet. J.* 38, 274–278.
- Dawkins, M.S., 2004. Using behaviour to assess animal welfare. *Anim. Welf.* 13, S3–S7.
- Dawkins, M.S., 2006. A user's guide to animal welfare science. *Trends Ecol. Evol.* 21, 77–82.
- Dijksterhuis, G.B., Heiser, W.J., 1995. The role of permutation tests in exploratory multivariate data analysis. *Food Qual. Pref.* 6, 263–270.
- Feaver, J., Mendl, M., Bateson, P., 1986. A method for rating the individual distinctiveness of domestic cats. *Anim. Behav.* 34, 1016–1025.
- Federation Equestre Internationale (FEI), 2011. Rules. <http://www.fei.org/disciplines/endurance>
- Frazier, D.L., 2000. Who speaks for the horse—the sport of endurance riding and equine welfare; AVMA Animal Welfare Forum: Equine Welfare. *J. Am. Vet. Med. Assoc.* 216, 1258–1261.
- Hausberger, M., Gautier, E., Biquand, V., Lunel, C., Jégo, P., 2009. Could work be a source of behavioural disorders? A study in horses. *PLoS One* 4, e7625.
- Henderson, A.J.Z., 2007. Don't fence me in: manging psychological well being for elite performance horses. *J. Appl. Anim. Welf. Sci.* 10, 309–329.
- Hutson, G.D., Haskell, M.J., 1997. Pre-race behaviour of horses as a predictor of race finishing order. *Appl. Anim. Behav. Sci.* 53, 231–248.
- König von Borstel, U., Euent, S., Graf, P., König, S., Gauly, M., 2011. Equine behaviour and heart rate in temperament tests with or without rider or handler. *Physiol. Behav.* 104, 454–463.
- Langdon Fielding, C., Meier, C.A., Balch, O.K., Kass, P.H., 2011. Risk factors for the elimination of endurance horses from competition. *J. Am. Vet. Med. Assoc.* 239, 493–498.
- Lloyd, A.S., Martin, J.E., Bornett-Gauci, H.L.I., Wilkinson, R.G., 2007. Evaluation of a novel method of horse personality assessment: rater-agreement and links to behaviour. *Appl. Anim. Behav. Sci.* 105, 205–222.
- Marlin, D.J., McEwen, J., Sluyter, F., 2008. Completion and treatment rates in modern endurance racing. In: *Proceedings of 4th International Equitation Science Conference, Dublin.*
- Mason, G.J., Latham, N.R., 2004. Can't stop, won't stop: is stereotypy a reliable animal welfare indicator? *Anim. Welf.* 13, 57–69.

- Meagher, R.K., 2009. Observer ratings: validity and value as a tool for animal welfare research. *Appl. Anim. Behav. Sci.* 119, 1–14.
- Minero, M., Tosi, M.V., Canali, E., Wemelsfelder, F., 2009. Quantitative and qualitative assessment of the response of foals to the presence of an unfamiliar human. *Appl. Anim. Behav. Sci.* 116, 74–81.
- Munsters, C.C., Visser, K.E., van den Broek, J., Sloet van Oldruitenborgh-Oosterbaan, M.M., 2011. The influence of challenging objects and horse-rider matching on heart rate, heart rate variability and behavioural score in riding horses. *Vet. J.* 192, 75–80.
- Nagy, A., Murray, J.K., Dyson, S., 2010. Elimination from elite endurance rides in nine countries: a preliminary study. *Equine Vet. J.* 42, 637–643.
- Napolitano, F., De Rosa, G., Braghieri, A., Grasso, F., Bordi, A., Wemelsfelder, F., 2008. The qualitative assessment of responsiveness to environmental challenge in horses and ponies. *Appl. Anim. Behav. Sci.* 109, 342–354.
- Napolitano, F., De Rosa, G., Grasso, F., Wemelsfelder, F., 2012. Qualitative behaviour assessment of dairy buffaloes (*Bubalus bubalis*). *Appl. Anim. Behav. Sci.* 141, 91–100.
- Napolitano, F., Knierim, U., Grasso, F., De Rosa, G., 2009. Positive indicators of cattle welfare and their applicability to on-farm protocols. *Ital. J. Anim. Sci.* 8, 355–365.
- Pritchard, J.C., Lindberg, A.C., Main, D.C., Whay, H.R., 2005. Assessment of the welfare of working horses, mules and donkeys, using health and behaviour parameters. *Prev. Vet. Med.* 12, 265–283.
- Rousing, T., Wemelsfelder, F., 2006. Qualitative assessment of social behaviour of dairy cows housed in loose housing systems. *Appl. Anim. Behav. Sci.* 101, 40–53.
- Rutherford, K.M.D., Donald, R.D., Lawrence, A.B., Wemelsfelder, F., 2012. Qualitative Behavioural Assessment of emotionality in pigs. *Appl. Anim. Behav. Sci.* 139, 218–224.
- Seaman, S.C., Davidson, H.P.B., Waran, N.K., 2002. How reliable is temperament assessment in the domestic horse (*Equus caballus*)? *Appl. Anim. Behav. Sci.* 78, 175–191.
- Sørensen, J.T., Sandøe, P., Halberg, N., 2001. Animal welfare as one among several values to be considered at farm level: the idea of an ethical account for livestock farming. *Acta Agric. Scand. A: An Suppl.* 30, 11–16.
- Stevenson-Hinde, J., 1983. Individual characteristics: a statement of the problem. In: Hinde, R.A. (Ed.), *Primate Social Relationships: An Integrated Approach*. Blackwell Scientific Publications, Oxford, UK, pp. 28–34.
- Stockman, C.A., Collins, T., Barnes, A.L., Miller, D., Wickham, S.L., Beatty, D.T., Blache, D., Wemelsfelder, F., Fleming, P.A., 2011. Qualitative behavioural assessment of cattle naïve and habituated to road transport. *Anim. Prod. Sci.* 51, 240–249.
- Stockman, C.A., McGilchrist, P., Collins, T., Barnes, A.L., Miller, D., Wickham, S.L., Greenwood, P.L., Cafe, L.M., Blache, D., Wemelsfelder, F., Fleming, P.A., 2012a. Qualitative Behavioural Assessment of Angus steers during pre-slaughter handling and relationship with temperament and physiological responses. *Appl. Anim. Behav. Sci.* 142, 125–133.
- Stockman, C.A., Collins, T., Barnes, A.L., Miller, D., Wickham, S.L., Beatty, D.T., Blache, D., Wemelsfelder, F., Fleming, P.A., 2012b. Flooring and driving conditions during road transport influence the behavioural expression of cattle. *Appl. Anim. Behav. Sci.* <http://dx.doi.org/10.1016/j.applanim.2012.11.003>
- Temple, D., Manteca, X., Velarde, A., Dalmau, A., 2011. Assessment of animal welfare through behavioural parameters in Iberian pigs in intensive and extensive conditions. *Appl. Anim. Behav. Sci.* 131, 29–39.
- Visser, E.K., van Reenen, C.G., Hopster, H., Schilder, M.B.H., Knaap, J.H., Barneveld, A., Blokhuis, H.J., 2001. Quantifying aspects of young horses temperament: consistency of behavioural variables. *Appl. Anim. Behav. Sci.* 74, 241–258.

- Visser, E.K., van Reenen, C.G., van der Werf, J.T.N., Schilder, M.B.H., Knaap, J.H., Barneveld, A., Blokhuis, H.J., 2002. Heart rate and heart rate variability during a novel object test and a handling test in young horses. *Physiol. Behav.* 76, 289–296.
- Visser, K.E., Reenan, C.G., Zetterqvist Blokhuis, M., Morgan, E.K., Hassmen, P., Rundgren, T.M.M., Blokhuis, H.J., 2008. Does horse temperament influence horse-rider cooperation? *J. Appl. Welf. Sci.* 11, 267–284.
- Walker, J., Dale, A., Waran, N., Clarke, N., Farnworth, M., Wemelsfelder, F., 2010. The assessment of emotional expression in dogs using a Free Choice Profiling methodology. *Anim. Welf.* 19, 75–84.
- Wemelsfelder, F., 1997. The scientific validity of subjective concepts in models of animal welfare. *Appl. Anim. Behav. Sci.* 53, 75–88.
- Wemelsfelder, F., 2007. How animals communicate quality of life: the qualitative assessment of behaviour. *Anim. Welf.* 16, 25–31.
- Wemelsfelder, F., Hunter, E.A., Mendl, M.T., Lawrence, A.B., 2000. The spontaneous qualitative assessment of behavioural expressions in pigs: first explorations of a novel methodology for integrative animal welfare measurement. *Appl. Anim. Behav. Sci.* 67, 193–215.
- Wemelsfelder, F., Hunter, T.E.A., Mendl, M.T., Lawrence, A.B., 2001. Assessing the ‘whole animal’: a free choice profiling approach. *Anim. Behav.* 62, 209–220.
- Wemelsfelder, F., Knierim, U., Schulze Westerath, H., Lentfer, T., Staack, M., Sandilands, V., 2009a. Qualitative Behaviour Assessment. In: Forkman, B., Keeling, L. (Eds.), *Assessment of Animal Welfare Measures for Layers and Broilers. Welfare Quality Reports No. 9, Sixth Framework Programme. University of Cardiff, Cardiff*, pp. 113–119.
- Wemelsfelder, F., Lawrence, A.B., 2001. Qualitative assessment of animal behaviour as an on-farm welfare-monitoring tool. *Acta Agric. Scand. A: Anim. Sci.* 51, 21–25.
- Wemelsfelder, F., Millard, F., 2009. Qualitative Behaviour Assessment. In: Forkman, B., Keeling, L. (Eds.), *Assessment of Animal Welfare Measures for Sows, Piglets and Fattening Pigs. Welfare Quality Reports No. 10, Sixth Framework Programme. University of Cardiff, Cardiff*, pp. 213–219.
- Wemelsfelder, F., Millard, F., De Rosa, G., Napolitano, F., 2009b. Qualitative Behaviour Assessment. In: Forkman, B., Keeling, L. (Eds.), *Assessment of Animal Welfare Measures for Dairy Cattle, Beef Bulls and Veal Calves. Welfare Quality Reports No. 11, Sixth Framework Programme. University of Cardiff, Cardiff*, pp. 215–224.
- Wemelsfelder, F., Nevison, I., Lawrence, A.B., 2009c. The effect of perceived environmental background on qualitative assessments of pig behaviour. *Anim. Behav.* 78, 477–484.
- Whitham, J.C., Wielebnowski, N., 2009. Animal-based welfare monitoring: using keeper ratings as an assessment tool. *Zoo Biol.* 28, 545–560.
- Wickham, S.L., Collins, T., Barnes, A.L., Miller, D., Beatty, D.T., Stockman, C., Blache, D., Wemelsfelder, F., Fleming, P.A., 2012. Behavioral assessment of transported sheep. Qualitative behavioral assessment of transport-naive and transport-habituated sheep. *J. Anim. Sci.* (Epub ahead of print).
- Wiseman-Orr, M.L., Scott, E.M., Reid, J., Nolan, A.M., 2006. Validation of a structured questionnaire as an instrument to measure chronic pain in dogs on the basis of effects on health-related quality of life. *Am. J. Vet. Res.* 67, 1826–1836.