QUANTITATIVE AND QUALITATIVE ASSESSMENT OF THE RESPONSE OF FOALS TO THE PRESENCE OF AN UNFAMILIAR HUMAN

Minero M.[a]*, Tosi M.V.[a], Canali E.[a], Wemelsfelder F.[b]

[a] *Dipartimento di Scienze Animali, Sezione di Zootecnica Veterinaria, Facoltà di Medicina Veterinaria, Università degli studi, Via Celoria 10, 20133 Milano, Italy.*
[b] *Scottish Agricultural College, Bush Estate, Penicuik, Midlothian EH26 0PH, Scotland, United Kingdom.*

Keywords: Horse; Behavioural expression; Qualitative measurements; Free Choice profiling; Generalised Procrustes analysis

*****Corresponding author: Michela Minero, Istituto di Zootecnica, Facoltà di Medicina Veterinaria, Università degli studi, Via Celoria 10, 20133 Milano, tel. 02.50318042, fax 02.50318030, e-mail michela.minero@unimi.it

Abstract

This work aimed to apply a combined qualitative and quantitative approach to the interpretation of an on-farm behaviour test for horses, and to examine whether one month of handling would affect the response of yearlings to an unfamiliar stationary human in their home environment. Throughout a one month period, 14 Thoroughbred Yearlings (16±0.22 months old) that had formerly experienced minimal contact with humans, were handled daily for about 45 minutes. The yearlings were tested twice, just before and just after the handling period. The behaviour of the horses during the tests was both video-recorded and directly recorded by the experimenter using an instantaneous time sampling recording method. Quantitative analysis of these data was achieved using Principal Component Analysis (PCA). Qualitative analysis took place from video clips using a Free Choice Profiling (FCP) methodology that requires observers to generate their own qualitative descriptors of behaviour, and in a second phase instructs these observers to quantify their personal descriptors on a Visual Analogue Scale. Observers were 21 veterinarians who were unaware that the horses had been handled in half of the clips and not in the other half. The data generated through FCP assessment were analysed using Generalised Procrustes Analysis (GPA). Any differences in behaviour that may have occurred before and after the handling period were evaluated by comparing horse scores on the main PCA and GPA factors using a Wilcoxon Matched-Pairs test. To compare qualitative and quantitative assessments, both the quantitative behaviour measures and the qualitative behaviour scores were correlated to the main PCA factors obtained from the quantitative analysis using Spearman's Rank correlation. PCA analysis revealed 3 main factors (explaining 30%, 23% and 21% of the total variation between horses respectively). The first factor showed high negative loadings for immobile behaviour and high positive loadings for contact and nibbling behaviour, and indicated that the horses tended to be more inclined to approach and contact the experimenter after handling (p = 0.08). GPA analysis revealed two main factors of expression (explaining 51.4% and 10.2% respectively). Both factors indicated significant qualitative differences in the behavioural style of yearlings before and after handling ($p < 0.05$ and $p < 0.01$ respectively), characterising yearlings as 'suspicious/nervous' and 'impatient/reactive' before handling, and as 'explorative/sociable' and 'calm/apathetic' after handling. The correlation between

GPA factor 1 scores with PCA factor 1 scores was highly significant (Spearman's r = 0.75; $p < 0.001$), while those between GPA factor 2 scores with PCA factor 2 and 3 scores were not significant (r = -0.255; ns and r = 0,251; ns, respectively). On the whole a meaningful relationship was found to exist between the quantitative and qualitative behavioural assessments of the horses' behaviour, indicating that these methods may be usefully combined in interpreting a behavioural test involving the presence of an unfamiliar human person.

1. Introduction

The quality of the human-horse relationship greatly affects the welfare of farmed horses, and in turn the horses' level of confidence influences their disposability to work with man. This hypothesis, well known in practice, has been confirmed by many scientific studies: rough or uneducated riders can inadvertently cause pain to the horses, soon causing a conditioned fear response of avoidance (Casey, 2002). Intensively handled foals are calmer and more tractable than untreated ones (Simpson, 2002), however their learning efficiency can deteriorate if they are pushed to work too hard (Rubin *et al.*, 1980). In light of these findings, any method for evaluating a horse's relationship with humans when assessing horse welfare on stud farms could be of considerable practical importance.

Behavioural tests have been frequently used in various animal species to assess the level of fear and reactivity towards humans (Hemsworth and Barnett, 2000). Hemsworth and Coleman (1998), for example, showed that general approach behaviour is influenced by the level of fear of man due to previous handling experiences, and Waiblinger and colleagues (2003) found that in cows, the avoidance distance towards an unfamiliar person reflects both the stockman's and the cow's role in the human-animal relationship. Various studies involving behavioural fear tests in horses have been published (Le Scolan *et al.*, 1997; Mackenzie and Thiboutot, 1997; Wolff *et al.*, 1997; Jezierski *et al.*, 1999; Anderson *et al.*, 1999; Hausberger and Muller, 2002, SØndergaard and Halekoh, 2003; Visser *et al.*, 2003), and it has been shown that handling foals affected their reactions to humans in a novel environment but not in the home environment (SØndergaard and Halekoh, 2003). Unfortunately it is not simple to estimate the validity of tests aimed to measure variables such as fear of humans, because the quality of a horse-human relationship is complex and we tend to lack thorough knowledge of whether and how the observed behaviours in a horse are affected by the animal's previous experience with man. Thus it is often difficult to establish the underlying motivation of an animal's behavioural response (Seaman *et al.*, 2002). For example, approach behaviours towards an unfamiliar human may not only be elicited by different levels of fear, but also by other motivational states such as the presence or absence of curiosity. In addition to validity and reliability, a crucial criterion for developing tests suitable to be carried out on farm is that they have to be feasible and adaptable to stud farms with different structural characteristics. This requirement often brings a cost of simplification, leaving the experimenter to wonder whether or not he/she inadvertently missed important bits of information; for example, it is difficult to interpret unambiguously postural signals such as "ears back" when they are isolated from the larger context in which they developed.

The qualitative assessment of behaviour integrates and summarises the different aspects of an animal's dynamic style of interaction with the environment, using expressive terms such as 'calm', 'friendly', 'anxious' or 'hostile' (Stevenson-Hinde et al., 1980). This type of assessment consists of a process of integrating measurement and interpretation and is highly sensitive to context, and it could therefore be a useful addition to classical ethological measures of animal behaviour in human approach tests (Wemelsfelder et al., 2000, 2001). Animal professionals (breeders, riders, veterinarians) frequently use qualitative terms to describe the temperament of horses and interpret their relationship with them, but can be in danger of creating an anthropomorphic picture that relies on popular unvalidated beliefs or has strong moral overtones. A qualitative research approach that facilitates the quantification of qualitative descriptors for the benefit of scientific computation could potentially bridge the gap that traditionally exists between these subjective judgements and scientific measurement approaches. Wemelsfelder et al. (2001) have developed an experimental Free-Choice-Profiling (FCP) methodology that combines procedures of qualitative interpretation with procedures of quantitative scoring (see methods for further details)[1]. Using this

[1] Even though this research approach includes a quantitative phase, we will continue to refer to it throughout the paper as 'qualitative assessment' of behaviour, in contrast with conventional quantitative measurements that tend not to include a phase of integrative, qualitative judgement

Formattato: Inglese (Regno Unito)

method, a previous study by Napolitano et al. (2007) found that qualitative assessments of behaviour in horses and ponies showed meaningful correlations with both subjective assessments and quantitative ethogram-based measures of the same animals.

In light of these considerations, the aim of this work was twofold: generally to apply a combined qualitative and quantitative assessment approach to the interpretation of an on-farm behavioural test for horses, and more specifically to examine whether and if so, how, one month of intensive handling would affect the response of yearlings to an unfamiliar stationary human in their home environment.

2. Material and methods

2.1 Animals, housing and management

Experimental subjects were 14 Thoroughbred Yearlings, 7 females and 7 males, ranging in age from 16 to 18 months at the beginning of the experiment. They belonged to different farms of Northern Italy and they were taken to the Yearling Training Centre one week before the start of the experiment. The horses, which formerly lived in groups in grassy paddocks, receiving minimal contact with humans (other than for de-worming and vaccination), were individually stabled in loose-boxes with straw bedding after reaching the training centre. Box stalls were 3.5 x 3.5 m wide, with concrete walls and a frontal sliding door. Each box had a window at 2.5 m from the floor, opposite to the frontal door. Horses were submitted to the same daily management routine, water was available ad libitum and they were fed hay and concentrate twice a day approximately at 6:30 a.m. and 3:30 p.m. Boxes were cleaned at approximately 9:00 a.m.

2.2 Behavioural testing and experimental handling procedures

The yearlings were prepared for the auctioneer sales over a period of one month. Throughout this period, they were handled daily for about 45 minutes to become accustomed to humans and receptive to subsequent training. The handling procedure consisted of haltering, leading outdoor to the paddock, brushing, picking up their feet and receiving veterinary examinations.

The yearlings were tested twice, once just before and once after the handling period. The day before the start of the handling period, all horses were individually tested in the presence of an unfamiliar person in their home box. The effect of familiarization was minimized through careful design of the order in which the animals were tested. The experimenter never tested horses in adjoining boxes, but followed test with another in a distant part of the stable, so that it can be reasonably assumed that the horses could not see or hear the experimenter before being tested. The test was adapted from that used by Jago *et al.* (1999) with young cattle. The experimenter, a female wearing blue overalls, approached the box 1 step/s, stood outside the front door for 2 seconds, then opened the door, entered the box and stood still with the hands by her sides in the corner of the box for 90 s. The yearlings were re-tested in the same way one month later, before leaving the centre for the auctions.

The behaviour of the horses during the tests was video-recorded with a semi-professional digital camcorder (Panasonic AG-E210). We placed the camcorder through the window of the home box, so that its placement allowed observation of the entire bodies of the yearling and the experimenter without interfering with them.

2.2.1 *Quantitative behaviour assessment*

The test described above was designed for on-farm use and implied that the experimenter (unfamiliar human) would observe the response of the horses directly during the test sessions. Thus the experimenter recorded the behaviour of the horses while approaching the box, opening the box door, entering the box and then every 30-90 seconds using an instantaneous time sampling recording method. The following behaviours were recorded: standing immobile, approaching person, in contact with person, sniffing person, nibbling person’s clothes, nibbling hay, vocalising, sniffing the environment and moving away from person.

To ensure that the results of quantitative behaviour analysis were comparable with those of the qualitative analysis, which was done from video clips of the same tests, the behaviour of horses observed in these video clips was also analysed quantitatively. This was done by three final year veterinary students (who were unfamiliar with the horses' treatments), using the same forms and behavioural recording method adopted in the direct observations.

2.2.2 Qualitative behaviour assessment

Observers

The observers were 21 Italian veterinarians from the School of Specialisation in Applied Ethology of the Veterinary Faculty in Milan. They were all experienced animal observers but not all of them had experience with horses. None of the observers were aware that the horses had been handled in half of the video clips and not in the other half. Before starting the observations, observers were told that the aim of the study was to evaluate the response of yearlings to a behavioural test using a qualitative behaviour assessment approach, and were subsequently instructed in the application of Free Choice Profiling procedures (see below). The need to generate independent individual assessments was emphasized and so they were asked to refrain from discussing any of their generated outcomes over the course of the experiment.

Video materials

To facilitate qualitative assessment of the horses' behaviour, 28 video clips were made (14 horses in 2 tests, before and after handling), each of 2 minute duration. These clips were put on a video tape in random order, with a space of 30 seconds between each video clip.

Free Choice Profiling Procedures

To facilitate the qualitative assessment of horse behaviour, a Free Choice Profiling (FCP) methodology was used. This methodology was originally developed in food science (Oreskovich *et al.*, 1991), and has been adopted for use in animal science by Wemelsfelder *et al.* (2001). It is characteristic of this method that it asks observers to generate their own qualitative descriptors of behaviour, rather than providing them with a pre-fixed list of terms, and then to quantify these personal descriptors using a Visual Analogue Scale. This ensures that assessors actively interpret their own observations rather than rely on provided interpretations.

FCP consists of two phases (for a detailed description, see Wemelsfelder *et al.*, 2001). In phase 1, observers generate their own descriptive terms. They were instructed to watch each clip without interruption, and then at the end of the clip to write down on a form all terms which in their opinion best described the way in which the horses had behaved. Thus, after having watched 28 clips in total, observers will have collected a list of qualitative descriptors assessing the behavioural style of the horses during the test. In phase 2, observers use this personal set of terms to quantitatively score the behavioural style of the observed horses. To this end, the experimenter provides each observer with a form on which a Visual Analogue Scale of 12.5 cm length (ranging from ‘minimum’ to ‘maximum’, with an uncategorised continuous line between these points) is added to each of this person’s terms. Observers then watch the same videos again, and at the end of each clip they score the observed horse on all of their terms, by ticking the scale at the appropriate point. This score is recorded as the measure of the distance in millimeters between the left ‘minimum’ point of the scale and the point where the observer’s thick crosses the line. Thus, for each of the 21 observers a data spreadsheet was created containing the scores of 28 horses on each of their personal terms.

2.3 Statistical analysis

2.3.1 Quantitative behaviour assessment

The inter-observer reliability and the concordance between direct and indirect observations were evaluated by Kendall’s coefficient of concordance. To analyse the relationships between the quantitative behavioural variables of the test, a principal component analysis (PCA) was used. A correlation matrix was used, and factor scores were calculated for horses when the factor’s Eigen value was greater than 1. Any differences in behaviour which may have occurred before and after the handling period were evaluated by comparing horse scores on the main PCA factors using a Wilcoxon Matched-Pairs test.

2.3.2 Qualitative behaviour assessment

The inter-observer agreement between the 21 observer score sheets was investigated using Generalised Procrustes Analysis (GPA), a multivariate statistical technique that does not rely on fixed variables (Gower, 1975; Gower and Dijksterhuis, 1994; Wemelsfelder *et al.*, 2000, 2001). GPA transforms individual observer scoring patterns into multidimensional configurations and, through a complex process of rotation and transformation, determines the "best fit" of these patterns, named the "consensus profile". This calculation is essentially a process of complex pattern recognition and takes place independently of the meaning of the terminologies used by observers. How well individual observer scores fit the consensus profile (i.e. the degree of agreement) is quantified by the Procrustes Statistic, and expressed in an 'observer plot'. 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 'randomised' data matrices. By applying GPA to these matrices, a 'randomised' profile is calculated. This procedure is repeated 100 times, providing a distribution of Procrustes Statistics 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.

The consensus profile has as many dimensions as the largest number of terms generated by any of the 21 observers. To facilitate interpretation, this number is reduced through Principal Component Analysis (PCA) to one or more main consensus factors explaining the majority of variation between the observed animals. Semantic interpretation of these main consensus factors subsequently takes place by correlating them to the original individual observer data matrices. This step of the analysis produces two-dimensional interpretative word-charts, one for each individual observer (21 in this case). In each chart, all terms of a particular observer are correlated with the principal axes of the consensus profile; the higher the correlation of a term, the more weight it has as a descriptor for that axis. If there is sufficient semantic convergence between the different observer word charts, then it becomes possible to select representative labels which interpret the main consensus factors, and which can be used to evaluate differences between individual horses. These differences are expressed graphically by the position of animals in a “horse plot”, which shows the distribution of the 14 yearlings before and after handling along the principal axes of the consensus profile. To investigate whether the behaviour of the 14 horses was perceived differently before and after the handling period, their scores on the main consensus factors were compared using a Wilcoxon Matched-Pairs test.

2.3.3 The relationship between quantitative and qualitative behaviour assessments

To compare the qualitative and quantitative assessments of horse behaviour, some form of ‘mapping’ these assessments on to each other is required (Rousing and Wemelsfelder, 2006; Napolitano *et al.*, 2007). This was achieved by using Principal Component Analysis in the first instance and Spearman’s Rank correlation in the second. Firstly, PCA was performed purely on the quantitative behavioural data. This resulted in the attribution of scores to individual horses on the main factors of this PCA. These PCA factors were subsequently used as the frame on to which both quantitative and qualitative assessments of individual horses were to be mapped. This was achieved by correlating the original quantitative values for each behavioural category, as well as the horse scores on the main consensus factors of the GPA analysis, to these PCA factor scores. The r-values of these correlations served as the coordinates with which each behavioural variable could be mapped on to the PCA factors in a 2-dimensional plot. The proximity of the various quantitative and qualitative variables in this plot can now be used to evaluate the degree of association between these variables. Rather than performing a PCA that included both quantitative and qualitative variables, we chose to use this approach in order to avoid interaction between these variables during PCA calculation procedures, and achieve a more objective way of investigating their association.

3. Results

3.1 Quantitative behaviour assessment

Table 1 shows the Kendall’s coefficient of concordance (W) for the degree of agreement between the experimenter who directly collected quantitative behavioural observations of the horses, and the observers who analysed this behaviour from video, for each recorded behaviour. The significance of observer agreement was calculated approximating a Chi-Squared distribution. These results show that for each behaviour the observers showed significant agreement on the ranking of the horses, indicating that direct and video-based observations were not substantially different. Given this result, it was decided to use the directly recorded behavioural measurements as a basis for subsequent quantitative analysis.

Table 2 shows the outcomes of the PCA of the recorded quantitative behavioural variables. The analysis identified three main factors with Eigenvectors greater than 1, which together explain 74.6% of the variation between horses. The first factor, accounting for 30.1% of the total variance, shows high negative loadings for immobile behaviour and high positive loadings for contact and nibbling behaviour, suggesting that horses scoring high on this factor can be described as more active and human oriented than horses with low scores. The meaning of the other two factors, accounting for 23.3% and 21.2% of the total variance respectively, seems more elusive. On the second factor, the tendency of approaching and moving away from the experimenter may indicate some form of heightened reactivity as opposed to animals who sniff the experimenter. The third factor may indicate a tendency to either approach or withdraw from the experimenter without engaging in closer contact as indicated by the first factor.

Fig. 1 shows the distribution of individual horses along the first two PCA factors. The Wilcoxon Matched-Pairs test did not provide evidence of a significant difference in the behaviour of the horses before and after the handling treatment, however it did indicate a tendency on the first factor for horses to be more inclined to approach and contact the experimenter after handling ($p = 0.08$). Fig. 2 shows the distribution of horses along the first and third PCA factors. As with the second factor, there was a tendency for horses to show more approaching and sniffing behaviour after treatment ($p = 0.09$).

3.2 Qualitative behaviour assessment

The Procrustes Statistic of the GPA consensus profile explained a significantly higher percentage of variation between observer matrices (58.35%) than the mean of 100 randomised profiles (39.77%; $t_{99} = 58.8$; $p < 0.001$), indicating the consensus to be a significant feature of the data set rather than an artefact of the Procrustean calculation procedures. The observer plot (Fig. 3) reflects the consensus among the 21 observers, as the majority of them fall within the 95% confidence region.

Two main factors of the consensus profile were identified, explaining 51.4% and 10.2% of the total variation between animals respectively. Of the observer word charts interpreting these factors, those of observers 12 and 21 are shown as examples in Fig. 4 and 5. These charts characterised the first factor of the consensus profile with terms ranging from tranquil/curious and resolute/sociable to suspicious/scared and shy/uncertain (for observers 12 and 21 respectively), while the second factor was described as ranging from impulsive/reactive and explorative/interested to static/indifferent and bored/lymphatic. To provide an overview of highly correlated terms for all observers, Table 3 lists for each observer the terms with the highest positive and negative correlation to factors 1 and 2 of the consensus profile. On the basis of this table, factor 1 was labelled as ranging from 'explorative/social' to 'suspicious/nervous', and factor 2 as ranging from 'calm/apathetic' to 'impatient/reactive'.

Fig. 6 shows the 'horse plot' of the qualitative behaviour assessment, in which individual horses before and after handling are positioned on the two main factors of the GPA consensus profile. These positions and the variation between them can be semantically interpreted with the qualitative labels derived from the word charts as discussed above. Fig. 6 suggests that observers perceived clear qualitative differences in the behavioural style of yearlings before and after handling. Before handling, most yearlings were characterised as 'suspicious/nervous' and/or 'impatient/reactive' in their response to the experimenter, while after handling they became significantly more 'explorative/sociable' and 'calm/apathetic' (factor 1: $p < 0.05$; factor 2: $p < 0.01$). It also appears that after handling the yearlings are slightly less variable in their response towards the experimenter than before handling.

3.3 The relationship between quantitative and qualitative behaviour assessments

Fig. 7 gives a visual representation of the association between quantitative behaviour measures and qualitative horse scores, both positioned in reference to the axes generated by PCA analysis of the quantitative behavioural variables (see *2.3.3*). The correlation between GPA factor 1 scores and PCA factor 1 scores was highly significant (Spearman r = 0.75; p < 0.001), indicating that horses engaging in close contact with the experimenter were assessed as 'explorative/sociable', while horses showing immobility behaviour were characterised by observers as 'suspicious/nervous'. The correlations of GPA factor 2 scores with PCA factor 2 and 3 scores were not significant (r = -0.255; ns and r = 0,251; ns, respectively). However the coordinates of GPA factor 2 (calm/apathetic – impatient/reactive) with PCA2 provide some support for the suggestion given in *3.1* that frequent movement away from and approach of the experimenter could be interpreted as a reactive style of interaction with humans, as opposed to calm animals who stop and sniff the experimenter.

On the whole, Fig. 7 indicates that a meaningful relationship exists between the quantitative and qualitative behavioural assessments of the horses' behaviour during the test, which suggests that qualitative assessments may assist in the interpretation of behaviours that potentially have different underlying motivations in different contexts.

4. Discussion and Conclusion

This study has demonstrated that quantitative and qualitative assessment methods could be meaningfully combined in interpreting a behavioural test involving the presence of an unfamiliar human person and demonstrating the effect of a month of intensive handling on the behaviour of horse yearlings. The results indicate that after the handling period, yearlings engaged more frequently in physical contact behaviours with the experimenter, and were qualitatively characterised by observers as significantly more 'explorativa/social' and 'calm/apathetic', than before handling. The significant association between quantitative and qualitative assessments found in this study on the one hand validates the qualitative assessments in showing these assessments not to be dissociated from standard quantitative measurements. On the other hand, the qualitative assessments add an interpretative element to the quantitative analysis, and in that sense also validate the meaning of this analysis. Given its integrative, 'whole animal' nature, qualitative behaviour assessment includes many aspects of behaviour that are difficult to record quantitatively, but are nevertheless important in understanding the animals' state. This may explain why the qualitative analysis distinguished significantly between handling treatments on both main factors, whereas the quantitative analysis only indicated distinction between treatments. Fig. 7 suggests that observers differentiated between 'calm' and 'reactive' horses on the basis of a complex pattern of approach/withdrawal and contact through sniffing; this pattern was difficult to capture quantitatively, but led to clear distinctions between horses when integrated qualitatively.

Such issues also arise when interpreting the state of animals that neither approach or withdraw in human interaction tests. Typical disadvantages of the behaviour tests where the reactions to a human stimulus are considered, include the difficult interpretation of the state of animals that neither approach nor withdraw (Waiblinger *et al.*, 2006). This is particularly evident when horses receive frequent human contact and may be ignoring the stimulus. The behaviour of 'standing immobile' for example, could signify different underlying motivational states such as fear, vigilance, curiosity, rest, defeat; it is only when observers interpret subtle details of the immobile posture such as muscular tension, orientation of body, head, eyes, and ears, breathing rhythm, and other details, that the nature of the animal's state can be gauged more precisely. In this particular study immobility was interpreted by a collective of 21 observers as a sign of 'nervousness'. This interpretation makes sense given the novelty of the test situation and the unfamiliarity of the human experimenter; fear responses are well known to be characteristic of various types of human-animal behaviour tests (Waiblinger *et al.*, 2006). The increased contact with the experimenter after handling on the other hand could have signified aggression, explorative curiosity, or a more engaged form of friendliness and sociability; in this study observers judged the horses' demeanour to show elements of both curiosity and friendliness.

The responses of horse yearlings to the entrance of an unfamiliar human into their home box described in this study are in accordance with the findings on responsiveness styles reported in cattle, where the avoidance distances in the stable were correlated with the amount of previous contact with humans (Waiblinger *et al.*, 2003). To gain a clearer view on whether a human interaction test can illuminate the quality of previous human contact, further research should be carried out on foals subjected to different handling practices. A common goal in the various human interaction tests is to develop them for use in animal welfare assessment on farm. In this case it is preferable to use observers used to working with the species under investigation, because it minimises the risk of misjudging the animals' expressions due to a lack of experience. Where possible, further validation of qualitative behaviour assessment can be sought by investigating its association with physiological parameters such as heart rate and heart rate variability measured by telemetry devices.

5. Acknowledgements

The authors are grateful to Roberta Polito for her expert practical work with horses, to Valentina Sanguineti and the 21 observers for their participation in the qualitative behaviour assessment.

References

Anderson, M.K., Friend, T.H., Evans, J.V., Bushong, D.M., 1999. Behavioral assessment of horses in therapeutic riding programs. Appl. Anim. Behav. Sci, 63, 11-24.

Casey, R.A., 2002. Clinical problems associated with the intensive management of performance horses. In: Waran, N. (Ed), The Welfare of Horses. Kluwer Academic Publishers, Cambridge, pp. 19-44.

Dijksterhuis, G.B., Heiser, W.J., 1995. The role of permutation tests in exploratory multivariate data analysis. Food Qual. Prefer. 6, 263-270.

Gower, J.C., 1975. Generalyzed procrustes analysis. Psychometrica 40, 35-51.

Gower, J.C., Dijksterhuis, G.B., 1994. Multivariate analysis of coffee images: a study in the simultaneous display of multivariate quantitative and qualitative variables for several assessors. Quality and Quantity 28, 165-184.

Hausberger, M., Muller, C., 2002. A brief note on some possible factors involved in the reactions of horses to humans. Appl. Anim. Behav. Sci 76, 339-344.

Hemsworth, P.H., Barnett, J.L., 2000. Human –Animal Interactions and Animal Stress. In: Moberg, G.P., Mench, J.A. (Ed), The Biology of Animal Stress. Cabi Publishing, Wallingford, pp. 309-336.

Hemsworth, P.H., Coleman, G.J., 1998. Human-livestock interactions: the stockperson and the productivity and welfare of intensively farmed animals. CAB International, Wallingford.

Jago, J.G., Krohn, C.C., Matthews, L.R., 1999. The influence of feeding and handling on the development of the human-animal interactions in young cattle. Appl. Anim. Behav. Sci 62, 137-151.

Jezierski, T., Jaworski, Z., Gorecka, A, 1999. Effects of handling on behaviour and heart rate in Konik horses: comparison of stable and forest reared youngstock. Appl. Anim. Behav. Sci 62, 1-11.

Le Scolan, N., Hausberger, M., Wolff, A., 1997. Stability over situations in temperamental traits of horses as revealed by experimental and scoring approaches. Behav. Process. 41, 257-266.

Mackenzie, S.A., Thiboutot, E., 1997. Stimulus reactivity tests for the domestic horse (*Equus caballus*). Equine Pract. 19, 21-22.

Napolitano, F., De Rosa, G., Braghieri, A., Grasso, F., Bordi, A., Wemelsfelder, F., 2007. The qualitative assessment of responsiveness to environmental challenge in horses and ponies. Appl. Anim. Behav. Sci, doi:101016/j.applanim.2007.03.009.

Oreskovich, D.C., Klein, B.P., Sutherland, J.V., 1991. Procrustes analysis and its applications to free-choice and other sensory profiling. In: Lawless, H.T., Klein, B.P. (Eds.), Sensory Science: Theory and Applications in Foods. Marcel Dekker, New York, pp. 353-393.

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.

Rubin, L., Oppegard, C.O., Hintz, H.F., 1980. The effect of varying the temporal distributionof conditioning trials on equine learning behavior. J. Anim. Sci. 50, 1184-1187.

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.

Simpson, B.S., 2002. Neonatal foal handling. Appl. Anim. Behav. Sci. 78, 303-317.

SØndergaard, E., Halekoh, U., 2003. Young horses' reactions to humans in relation to handling and social environment. App. Anim. Behav. Sci. 84, 265-280.

Stevenson-Hinde, J., Stillwell-Barnes, R., Kunk, M., 1980. Subjective assessment of rhesus monkeys over four successive years. Primates 21, 66-82.

Visser, E.K., Reenen, C.G., van. Schilder, M.B.H. Barneveld, A., Blokhuis, H.J, 2003. Learning performances in young horses using two different learning tests. Appl. Anim. Behav. Sci. 80, 311-326.

Waiblinger, S.; Boivin, X.; Pedersen, V. Tosi, M., Janczak, A.M., Visser, E.K. Jones, R.B., 2006. Assessing the human-animal relationship in farmed species: A critical review. . Anim. Behav. Sci. 101, 185-242.

Waiblinger, S., Menke, C., Folsch, D.W, 2003. Influences on the avoidance and approach behaviour of dairy cows towards humans on 35 farms. Appl. Anim. Behav. Sci. 84, 23-39.

Wemelsfelder, F., Hunter, T.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.

Wolff, A., Hausberger, M., Le Scolan, N., 1997. Experimental tests to assess emotionality in horses. Behav. Process. 40, 209-221.