Using qualitative behaviour assessment (QBA) to explore the emotional state of 1 horses and its association with human-animal relationship

Michela Minero[a*], Emanuela Dalla Costa[a], Francesca Dai[a], Elisabetta Canali[a], Sara Barbieri[a], Adroaldo Zanella[b], Riccardo Pascuzzo[c], Francoise Wemelsfelder[d]

[1] Università degli Studi di Milano, Department of Veterinary Medicine, Via Celoria 10, 20133 Milan, Italy 7

[2] Av. Prof. Dr. Orlando Marques de Paiva, 87 - Cidade Universitária, Butantã - São Paulo/SP, Brasil - CEP: 05508-270 9

[3] Politecnico di Milano, Department of Mathematics, MOX Laboratory for Modelling and Scientific Computing, Via Edoardo Bonardi 9, 20133 Milan, Italy

[4] Animal and Veterinary Sciences Group, SRUC, Roslin Institute Building, Midlothian EH25 9RG, UK 13

*Corresponding author: Michela Minero, Università degli Studi di Milano, Dipartimento di Medicina Veterinaria, Via Celoria 10, 20133 Milano, Italy – Tel. +39.02.50318037 - Fax: +39.02.50318030 – Email: michela.minero@unimi.it

1. Abstract

This study aimed to apply qualitative behaviour assessment (QBA) to horses farmed in single boxes, in order to investigate their emotional state and explore its association with indicators of human-animal relationship. A fixed list of 13 QBA descriptive terms was determined. Three assessors experienced with horses and skilled in measuring animal behaviour underwent a common training period, consisting of a theoretical phase and a practical phase on farm. Their inter-observer reliability was tested on a live scoring of 95 single stabled horses. Principal Component Analysis (PCA) was conducted to analyse QBA scores and identify perceived patterns of horse expression, both for data obtained in the training phase and from the on-farm study. Given the good level of agreement reached in the training phase (Kendall W=0.76 and 0.74 for PC1 and PC2 scores respectively), it was considered acceptable in the subsequent on-farm study to let these three observers each carry out QBA assessments on a sub-selection of a total of 355 sport and leisure horses, owned by 40 horse farms. Assessment took place immediately after entering the farms: assessors had never entered the farms before and were unaware of the different backgrounds of the farms. After concluding QBA scoring, the assessors further evaluated each horse with an avoidance distance test (AD) and a forced human approach test (FHA). A MANOVA test was used to assess the association of the AD and FHA tests with the on-farm QBA PC scores. The QBA approach described in this paper was feasible on farm and showed good acceptability by owners. In the analysis of on-farm QBA scores, the first Principal Component ranged from relaxed/at ease to uneasy/alarmed, the second Component ranged from curious/pushy to apathetic. Horses perceived as more relaxed/at ease with QBA showed less avoidance during the AD test ($P=0.0376$), and responded less aggressively and fearfully to human presence in the FHA test ($P < 0.0001$). Our results support the hypothesis that QBA is sensitive to the quality of human contact in horses.

Keywords: Horses; Qualitative Behaviour Assessment; Human-animal relationship; Welfare

1. Introduction

When asked, most horse people would claim that it is reasonably easy to recognise the affective state of a horse, but they would probably be unable to substantiate it or to explain how to do it. Scientists are challenged by the same question, as it is difficult to reliably establish the emotional state of horses, which, differently from humans, cannot report verbally if scientific suppositions match with their actual state. Tackling a similar challenge, researchers worldwide have been working at the development of a variety of methods to assess emotions in different animal species (Boissy et al., 2007; Fraser, 2009; Fureix and Meagher, 2015; Mendl et al., 2010; Millot et al., 2014; Murphy et al., 2014; Panksepp, 2011; Paul et al., 2005; Wemelsfelder, 2007). Qualitative behaviour assessment (QBA) is one of those scientific methods, originally developed by Wemelsfelder et al. (2000, 2001), that has been proven to contribute to the identification of the main dimensions of animal emotional states (Carreras et al., 2016; Mendl et al., 2010; Mullan et al., 2011; Rutherford et al., 2012; Temple et al., 2013). By its very nature, QBA is an intrinsically holistic and dynamic tool used for capturing the expressive quality of animal behaviour. When using QBA, an observer addresses the whole animal, focussing on details of how an animal is behaving; then he or she scores the animal on visual analogue scales corresponding to different behavioural descriptors (e.g. curious, aggressive). This method enables an experienced observer to capture (subtle) changes in the animal body language in relation to the environment, and to express them as quantitative measures that can be analysed statistically. Thus QBA facilitates the dialogue between horse professionals expressing subjective judgments and scientists needing to respect assumptions of scientific methods (Minero et al., 2009; Wemelsfelder, 2007).

Research only recently has begun to explore the value of applying QBA in the context of human-animal relationships. For example, QBA was used to explore the link between a stockperson' handling style and dairy calves' behavioural expressions (Ebinghaus et al., 2016; Ellingsen et al., 2014). Calves with more positive QBA 'mood' scores (e.g. enjoying, friendly) were typically handled by persons treating them patiently and calmly. Furthermore QBA, alongside other human-animal relationship measures, proved to be a suitable measure of animal reactivity to humans (Ebinghaus et al., 2016; Minero et al., 2016). In the case of donkeys, animals characterised by QBA as ‘relaxed’ and ‘at ease’, did not show any avoidance, tail tuck, or other negative reactions when approached by a human (Minero et al., 2016).

QBA descriptors can be individually generated by observers, as in the case of Free-Choice-Profiling methodology (Wemelsfelder et al., 2001), or they can be chosen by researchers first from literature and then discussed in focus groups of experts and tested on-farm (Andreasen et al., 2013). FCP is unsuitable for on-farm welfare assessment, as it requires a minimum of 10 observers and extensive data analysis; hence, the second approach using a fixed list of terms was adopted for on-farm QBA assessment in different animal species (Fleming et al., 2016). In horses, the Free-Choice-Profiling methodology of QBA has previously been applied to answer various research questions, for instance it was used to investigate ponies’ response to an open field test (Napolitano et al., 2008), to investigate the response of foals to the presence of an unfamiliar human (Minero et al., 2009), and to assess demeanour in horses engaged in a 160-km endurance ride (Fleming et al., 2013). The present study was done in the framework of the Animal Welfare Indicators (AWIN) research project funded by EU FP7 (AWIN, 2015). For the first time a fixed QBA term list was included in the AWIN welfare assessment protocol for horses as an on-farm measure for positive emotional state (Dalla Costa et al., 2016). The goal of the present study was to apply the fixed QBA term list for horses developed for the AWIN protocol (2015) to the study of human relationships with horses and their relation with the animals emotional state, allowing us to further investigate the feasibility and reliability of the AWIN QBA horse term list in on-farm conditions. Our null hypothesis was that patterns of emotional expression in stabled horses identified through QBA would not show any meaningful association with independent measures of the human-horse relationship.

2. Material and methods

2.1 Development of the QBA rating scale

An initial list of qualitative descriptors was created deriving terms from the scientific literature where qualitative expressions were used to describe horse behaviour. This list contained 36 English terms, which were then discussed during a face-to-face focus group with 18 horse professionals (veterinarians, breeders, horse welfare organisations members). The focus group took place at the premises of the Veterinary Faculty. After a general introduction to the Qualitative Behaviour Assessment method, the participants discussed and refined the original list of descriptors. They removed some terms, which they felt were difficult to interpret unambiguously or which they did not consider relevant to the assessment of horses on farm, and refined some of the terms' characterisations. Using this modified list of terms they then scored 10 videos of horses filmed individually for 1 min that showed a wide range of behavioural expressions. After this practical exercise and extensive discussion, the group agreed on a final list of 13 terms (Table 1) to be used for scoring individual horses on farm.

2.2 Training of assessors and inter-observer reliability

The assessors were three veterinarians experienced with horses and skilled in assessing animal behaviour. These assessors together attended two training sessions. In the first session, assessors were encouraged to discuss the concept of QBA and the meaning of each of the 13 QBA descriptors. In the second session, the assessors observed 20 horses in their home boxes, and through comparison and discussion of their individual scores for these horses on the 13 terms, calibrated their scoring to become more closely aligned (see Grosso et al., 2016). Final interobserver reliability of the QBA descriptors was tested by asking assessors to simultaneously and independently score 95 single stabled horses at eight horse facilities.

2.3 Farm visits

Each of the three trained assessors independently carried out QBA assessments on a sub-selection of a total of 40 horse facilities, so that each horse was assessed only once by one of the three assessors. In each facility, all the horses over 5 years were assessed individually, adding up to a total of 355 sport and leisure horses of different gender, breed and riding discipline (riding school=37%; training centre=24%; breeding farm=15%; hippodrome=3%; other (e.g. animal-assisted activity)=21%). QBA assessment took place immediately after entering the farms and letting the animals adapt to the observers' presence. Assessors had never entered the farms before and were unaware of the different backgrounds of the farms, so as not to be biased by any pre-existing prejudices regarding these backgrounds. They wore blue overalls and had not made any clinical examination nor treatment to horses during the month prior the assessments.

The assessor initially observed a horse from outside the box, without disturbing it, for 30 s. Then they entered the box, approaching the horse slowly and scratched the horse at the withers for 30 s, all the while observing the horse's responses. At the end of each horse observation period, they scored the list of QBA descriptors on visual analogue scales (VAS), where the ends of the scale represented the 'minimum' (this expressive quality is absent) and 'maximum' (this quality could not be present more strongly) of the expressive quality. The score was represented by the measure of the distance in millimetres between the left 'minimum' point of the scale and the point where the observer's thick crossed the line. Automated data recording and download of scores to excel files was made possible by use of a dedicated electronic application specifically developed at SRUC (Scotland's Rural College) in the UK.

In order to evaluate the quality of the human-horse relationship, after concluding QBA scoring the assessors performed and scored an avoidance distance test (AD) and a forced human approach test (FHA) (Dalla Costa et al., 2015). The AD test was performed from outside the box. When the horse was attentive to their presence, the assessor approached the animal walking at measured pace of one step per second. If the horse showed an avoidance response, this was recorded as 0, no avoidance was recorded as 1. In the FHA test, the assessor opened the box door, entered the box, and approached the horse slowly. If the horse stood still calmly, the assessor raised their hand, touched the withers and moved their hand along the back of the subject. The horse's reaction was scored from 0 to 2 (0=the horse showed aggressive behaviour; 1=the horse moved away as soon as he/she touched the withers; 2=the horse stood still calmly or showed positive signs of interest). Horses that were reported by their owners as having or having suffered back pain were not tested. Automated data recording and download to excel files was made possible by use of a dedicated electronic application specifically developed for the AWIN project (AWINHorse app).

2.4 Statistical analysis

IBM SPSS Statistics 24 software (IBM Corp., 2016) and R software (RCore Team, 2016) were used for statistical analysis.

The QBA scores generated by the three assessors scoring the 95 horses on all 13 descriptors during the training phase, were analysed together as part of one Principal Component Analysis (PCA, correlation matrix, no rotation). The PC scores attributed to the 95 horses on the first three main Principal Components were then tested for inter-observer reliability using Kendall Correlation Coefficient W. Chi-square test (94 df) was used for statistical significance of association between the observers, allowing rejection of the null hypothesis (non-association between the observers), when $P < 0.05$. To further analyse inter-observer reliability for each separate QBA descriptor, Kendall Correlation Coefficient W was calculated on the raw descriptor scores. Kendall W values can vary from 0 (no agreement at all) to 1 (complete agreement), with values higher than 0,6 showing substantial agreement (Eliasson et al., 2017).

The QBA scores generated by the three assessors for a total of 355 individual horses over 40 farm visits (93 horses by assessor 1, 147 horses by assessor 2, and 115 horses by assessor 3), were also analysed together using Principal Component Analysis (PCA, correlation matrix, no rotation). In order to estimate the association between indicators of the horses' human-animal relationship and their emotional state, the PC scores attributed to the animals on the first two main components of the PCA (explaining 55.549% of total variance) were analysed through a two-way MANOVA test. To explain in more detail, we considered the subdivision of the horses in six groups, according to their scores obtained in the avoidance distance (AD) test (0 or 1) and in the forced human approach (FHA) test (0, 1 or 2), obtaining unequal sizes of the observed classes. A Mardia's test (Mardia, 1970) was used to assess the multivariate normality of the distribution of PC scores within each group: in three cases of six, the assumption of normality was not met. In addition, a Box's M test (Johnson and Wichern, 2007) confirmed that the groups had homogeneous covariance matrices. Since MANOVA is quite robust to violations of normality (Johnson and Wichern, 2007), we performed a type III MANOVA on the PC scores, which is the most recommended type of analysis when dealing with unbalanced data (Milliken and Johnson, 2009). In this framework, we computed the Pillai statistic, as suggested by Tabachnick and Fidell (2013), to perform the hypothesis tests that aimed at assessing the effects of the AD and FHA on QBA PC scores, as well as their interaction. We found that the interaction was not statistically significant ($P > 0.05$), thus, we removed it from the model and performed the test again. Then, one-way ANOVAs (with p-values corrected by the Bonferroni method) were used as a post-hoc test to verify specific relationships between the humananimal tests and the two sets of PC scores separately.

3. Results

No safety issues were encountered during the QBA assessment or the performance of human-animal behaviour tests. No assessments had to be interrupted because of horse reactions and all owners showed good acceptance of the procedures adopted.

3.1 Inter-observer reliability in the training phase

Table 2 shows the percentage of variation explained by the first three Principal Components explaining the majority of the variance between horses, and the level of agreement between the scores generated by the three assessors on each of these components. The distribution of loadings of

QBA terms on each PC was highly similar to that for the on-farm assessments reported below, and will therefore not be repeated here. In summary, the highest and lowest loading terms describing each of the components were for PC1: at ease/relaxed to uneasy/ alarmed; for PC2: look for contact/curious to apathetic and for PC3: curious to apathetic.

Table 3 shows the Kendall W values for each of the separate QBA descriptors. The three assessors reached satisfactory agreement (values larger than 0.60) in scoring all descriptors, with the exception of apathetic, which had a value of 0.56.

3.2 QBA assessment of horses on farm

Given the high levels of agreement between assessors both for PC scores and scores on separate descriptors, it was considered to be acceptable for the 3 assessors to independently visit and assess horses at different farms, and subsequently analyse all collected scores together in one PCA.

This PCA identified three main Principal Components with Eigen value greater than 1, together explaining 65% of the variation between horses. Table 4 shows the outcomes for these PCs, as well as the loading of QBA terms on each PC. From these loadings it can be seen that PC1 ranges from relaxed/at ease to uneasy/alarmed, PC2 from curious/pushy to apathetic, and PC3 from happy to 'looking for contact'. Fig. 1 shows the distribution of QBA term loadings along PC 1 and 2.

3.3 Influence of Human-horse relationship on horse emotional state

The results of the two-way MANOVA suggested that the horses' responses to the Avoidance Distance (AD) test were very close to being significantly linked to their scores on both QBA Principal Components ($P=0.0565$). In particular, looking at the post-hoc tests, we found a significant difference with respect to the first Principal Component (adjusted $P=0.0376$) and no difference with respect to the second Principal Component (adjusted $P=1$). Regarding the Forced Human Approach (FHA) test, we found a significant difference to their scores on both QBA Principal Component ($P < 0.0001$), which was confirmed also by the post-hoc test performed on the two Principal Components separately (both adjusted $P < 0.0001$). The results of the post-hoc analysis are summarised in Fig. 2.

The upper part of Fig. 2 shows significant associations between the horses' PC1 scores and their scores for the AD and FHA tests, indicating that horses perceived as more relaxed/at ease were more frequently scored 1 (no avoidance) during the avoidance test (AD) and responded less aggressively and fearfully to human presence (higher scores in the FHA test). The lower part of Fig. 2 shows a significant association only between the horses' PC2 scores and their FHA scores, indicating that horses perceived as curious/pushy responded more aggressively to human presence.

4. Discussion

The present study was based on an interest in the association between the emotional state of horses and their human-animal relationship. To achieve this aim we developed a qualitative behaviour assessment procedure for horses farmed in single boxes, and investigated the association of the horses' QBA scores with their scores on Avoidance Distance and Forced Human Approach tests. Our findings were that firstly, the approach described in this paper was feasible on farm and showed good acceptability by owners; secondly, trained assessors showed good inter-observer reliability scoring horses with QBA, and thirdly, we found a significant association between the first two QBA components and the horses' reactions to two human-animal interaction tests.

Fixed lists of QBA descriptors are currently used in several farm animal species to assess their welfare (Andreasen et al., 2013; Brscic et al., 2010; Fleming et al., 2013; Grosso et al., 2016; Minero et al., 2016; Munsterhjelm et al., 2015; Napolitano et al., 2012; Phythian et al., 2016; Rousing and Wemelsfelder, 2006); their inclusion in a protocol to assess horse welfare, together with other relevant measures, was reported for the first time in the AWIN welfare assessment protocol for horses (AWIN, 2015). The barren environment of single boxes might limit the expression of affective states of horses, and prevents the evaluation of their behaviour, in relation with other animals. The two phase assessment procedure proposed here allowed to overcome some of these issues. Animals were observed in the home environment both when they were on their own and when experiencing a pleasant stimulus (grooming at the withers). The rationale behind the choice of using positive stimulation was based on suggestions by Keeling et al. (2008) that repeated disruption of reward cycles cause long term negative effects on welfare and could result in less positive behaviour during a pleasant situation (Costa et al., 2012; Keeling et al., 2008). For example, in a complete cycle (e.g. feeding, drinking, play, etc.) an organism passes through appetitive, consummatory and post- consummatory phases and is characterised by positive affective states, whereas repeated experience of disrupted cycles alters long term affective state and mood. One can thus expect that only horses enjoying good welfare and no disruption of reward cycles would be characterised by positive QBA descriptors and behaviour when experiencing a positive situation such as grooming. In horses, grooming is associated with pleasure and it was shown to have positive affective and physiological effects (Feh and de Mazières, 1993; Lynch et al., 1974; Normando et al., 2002; Thorbergson et al., 2016). Albeit correct and useful, the construct underlying this approach can be denied under specific circumstances: horses experiencing or having experienced back pain would likely find unpleasant being touched at the withers, making it difficult to infer about their original affective state. To control for this possible bias, we did not assess horses that were reported by their owners as having or having suffered back pain. No assessments had to be interrupted because of horse reactions and owners always showed good acceptability of the procedures adopted. It should be considered that in the case of horses kept in groups, an adaptation of the assessment procedure would be needed. It should also be noted that stallions might show different posture and facial expressions when groomed at withers compared to female and geldings (McDonnell, 2003).

Since QBA relies on observer's assessment, improving and assessing the reliability of all assessors is paramount in the process of validating new QBA procedures. Our results indicate that during the training phase, observers ranked the different horses in similar ways when using the QBA descriptors. The good inter-observer reliability in assessing single horses using QBA, both on overall PC scores and single descriptors, suggest that the training of assessors described here and grounded on previous experiences with other animal species (Grosso et al., 2016; Minero et al., 2016) was effective in reaching a satisfactory reliability of observers. The agreement on the use of single terms can be considered important as part of an effort to increase overall agreement between observers, however QBA outcomes should primarily consider the dynamic patterns of demeanour captured by multi-variate analysis tools such as PCA. Assessors reached excellent agreement on the first two Principal Components and a good agreement on the third Component.

Consistent with previous findings in other species (Ellingsen et al., 2014; Rousing and Wemelsfelder, 2006), the Principal Component Analysis of horse scores in the on-farm study revealed two main dimensions of the affective state of horses. The first Principal Component ranged from at ease/relaxed to uneasy/alarmed: horses with high positive scores on this Component could be described as in a positive affective state. The second component, ranging from curious/pushy to apathetic, could be interpreted as more indicative of the horses' arousal level. These findings map well in the overall picture where different methods to assess emotions in animals repeatedly highlighted dimensions of valence and arousal of affective states (Mendl et al., 2009; Paul et al., 2005). Differently from other methods, QBA can be applied during on-farm assessments and can be used to facilitate the dialogue between owners and assessors (Minero et al., 2009; Wemelsfelder, 2007), possibly increasing the engagement of owners in the process of improving animal welfare.

Horses reactions to human-animal interaction tests were significantly linked to Qualitative Behaviour Assessment. In particular, a high score on QBA descriptors like relaxed, friendly, at ease, loading high on the first component, was found to be pronouncedly associated with an absence of signs of avoidance, and positive signs of interest towards an interacting human. Horses achieving higher scores in the tests had a better relationship with humans and a more positive affective state. Conversely, horses showing an aggressive reaction to a forced human approach were described as more pushy when assessed beforehand with QBA. Horses achieving low scores in the FHA test (more aggressive behaviour during the test) had a poorer relationship with humans and were described as being more aroused. These results add to those reported by other authors, that animals having a positive bond with humans are safer and easier to handle, whilst negative handling leads to poorer mood and an aroused state (Breuer et al., 2000; Ellingsen et al., 2014; Waiblinger et al., 2006). It can also be suggested that poor handling increases fear of humans in horses, influencing their mood and level of arousal, and drive them into a negative feedback cycle that progressively leads them to become more aggressive and unsafe to handle.

5. Conclusions

The QBA assessment procedure proposed here allowed to capture expressions of affective states of horses in their home box and proved to be feasible on-farm. The good inter-observer reliability achieved, both on overall PC scores and single descriptors, suggest that a phased procedure for the training of assessors is effective in reaching a satisfactory agreement between observers. QBA was useful to identify horses in a more positive affective state and, in line with previous findings in dairy cows (Brscic et al., 2010; Ellingsen et al., 2014) and lambs (Serrapica et al., 2017), we can support the hypothesis that QBA is sensitive to the quality of human contact. Our results suggest that high quality relations with humans are a potential tool to provide good animal welfare, also in terms of positive emotions.

Conflict of interest

All authors of the manuscript "Using qualitative behaviour assessment (QBA) to explore the emotional state of horses and its association with human-animal relationship" declare no actual or potential conflict of interest including financial, personal or other relationships with other people or organizations within three years of beginning the submitted work that could inappropriately influence, or be perceived to influence, their work.

Author contributions statement

The first two authors contributed equally to the manuscript. E.DC. F.D. F.W. and M.M. conceived the experiment(s), E.D.C. F.D. and M.M. conducted the experiment(s), E.DC. F.W. M.M. and R.P. analysed the results. F.W. and M.M supervised the research. All authors contributed to preparation and revision of the manuscript.

Acknowledgements

The authors wish to thank the EU VII Framework program (FP7-KBBE-2010-4) for financing the Animal Welfare Indicators (AWIN) project. They also thank Philipp Scholz for his help in data collection and a group of master students of Mathematical Engineering of Politecnico di Milano for their help with data analysis. The authors are grateful to Tomasz Krzyzelewski at SRUC EGENES and Alberto Carrascal from DAIA Intelligent Solutions for their programming expertise and collaboration in developing the AwinHorse app.

References

AWIN, 2015. AWIN welfare assessment protocol for horses. http://dx.doi.org/10.13130/AWIN_HORSES_2015.

Andreasen, S., Wemelsfelder, F., Sandøe, P., Forkman, B., 2013. The correlation of Qualitative Behavior Assessments with Welfare Quality ® protocol outcomes in on onfarm welfare assessment of dairy cattle. Appl. Anim. Behav. Sci. 143, 9–17. http://dx.doi.org/10.1016/j.applanim.2012.11.013.

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. http://dx.doi.org/10.1016/j.physbeh.2007.02.003.

Breuer, K., Hemsworth, P., Barnett, J., Matthews, L., Coleman, G., 2000. Behavioural response to humans and the productivity of commercial dairy cows. Appl. Anim. Behav. Sci. 66, 273–288. http://dx.doi.org/10.1016/S0168-1591(99)00097-0.

Brscic, M., Wemelsfelder, F., Tessitore, E., Gottardo, F., Cozzi, G., Van Reenen, C.G., 2010. Welfare assessment: correlations and integration between a Qualitative Behavioural Assessment and a clinical/health protocol applied in veal calves farms. Ital. J. Anim. Sci. 8, 601–603. http://dx.doi.org/10.4081/ijas.2009.s2.601.

Carreras, R., Mainau, E., Arroyo, L., Moles, X., González, J., Bassols, A., Dalmau, A., Faucitano, L., Manteca, X., Velarde, A., 2016. Housing conditions do not alter cognitive bias but affect serum cortisol, qualitative behaviour assessment and wounds on the carcass in pigs. Appl. Anim. Behav. Sci. 185, 39–44. http://dx.doi.org/10.1016/j.applanim.2016.09.006.

Costa, E.D., Bonaita, C., Pedretti, S., Govoni, E., Guzzeloni, A., Canali, E., Minero, M., 2012. Inter-observer reliability of three human-horse relationship tests.

Dalla Costa, E., Dai, F., Murray, L.A.M., Guazzetti, S., Canali, E., Minero, M., 2015. A study on validity and reliability of on-farm tests to measure human-animal relationship in horses and donkeys. Appl. Anim. Behav. Sci. 163, 110–121. http://dx.doi.org/10.1016/j.applanim.2014.12.007.

Dalla Costa, E., Dai, F., Lebelt, D., Scholz, P., Barbieri, S., Canali, E., Zanella, A.J., Minero, M., 2016. Welfare assessment of horses: the AWIN approach. Anim. Welf. 25, 481–488. http://dx.doi.org/10.7120/09627286.25.4.481.

Ebinghaus, A., Ivemeyer, S., Rupp, J., Knierim, U., 2016. Identification and development of measures suitable as potential breeding traits regarding dairy cows' reactivity towards humans. Appl. Anim. Behav. Sci. 185, 30–38. http://dx.doi.org/10.1016/j.applanim.2016.09.010.

Eliasson, K., Palm, P., Nyman, T., Forsman, M., 2017. Inter – and intra – observer reliability of risk assessment of repetitive work without an explicit method. Appl. Ergon. 62, 1–8. http://dx.doi.org/10.1016/j.apergo.2017.02.004.

Ellingsen, K., Coleman, G.J., Lund, V., Mejdell, C.M., 2014. Using qualitative behaviour assessment to explore the link between stockperson behaviour and dairy calf behaviour. Appl. Anim. Behav. Sci. 153, 10–17. http://dx.doi.org/10.1016/j.applanim.2014.01.011.

Feh, C., de Mazières, J., 1993. Grroming at a preferred site reduces heart rate in horses. Anim. Behav. 46, 1191–1194.

Fleming, P.A., Paisley, C.C.L., Barnes, A.L., 2013. Application of Qualitative Behavioural Assessment to horses during an endurance ride. Appl. Anim. Behav. Sci. 144, 80–88.

Fleming, P.A., Clarke, T., Wickham, S.L., Stockman, C.A., Barnes, A.L., Collins, T., Miller, D.W., 2016. The contribution of qualitative behavioural assessment to appraisal of livestock welfare. Anim. Prod. Sci. http://dx.doi.org/10.1071/AN15101.

Fraser, D., 2009. Animal behaviour, animal welfare and the scientific study of affect. Appl. Anim. Behav. Sci. 118, 108–117. http://dx.doi.org/10.1016/j.applanim.2009.02.020.

Fureix, C., Meagher, R.K., 2015. What can inactivity (in its various forms) reveal about affective states in non-human animals? A review. Appl. Anim. Behav. Sci. 171, 8–24. http://dx.doi.org/10.1016/j.applanim.2015.08.036.

Grosso, L., Battini, M., Wemelsfelder, F., Barbieri, S., Minero, M., Dalla Costa, E., Mattiello, S., 2016. On-farm Qualitative Behaviour Assessment of dairy goats in different housing conditions. Appl. Anim. Behav. Sci. 180, 51–57. http://dx.doi.org/10.1016/j.applanim.2016.04.013.

Johnson, R., Wichern, D., 2007. Applied Multivariate Statistical Analysis, 6th edition. Prentice Hall, New Jersey.

Keeling, L.J., Algers, B., Blokhuis, H.J., Boissy, A., Lidfors, L., Mendl, M., Oppermann-Moe, R., Paul, E., Uvnas-Moberg, K., Zanella, A.J., 2008. Looking on the bright side of life: reward, positive emotions and animal welfare. In: Boyle, L., Hanlon, N., O'Connell, A. (Eds.), Proceedings of the 42nd International Congress of the ISAE, Dublin, p. 3.

Lynch, J., Fregin, G., Mackie, J., Monroe, R., 1974. Heart rate changes in the horse to human contact. Psychopharmacology (Berl.) 11, 472–478.

Mardia, K., 1970. Measures of multivariate skewness and kurtosis with applications. Biometrika 57, 519–530.

McDonnell, S., 2003. Practical Field Guide to Horse Behavior: the Equid Ethogram. The Blood-Horse, Inc., Lexington, USA.

Mendl, M., Burman, O.H., Parker, R.M., Paul, E.S., 2009. Cognitive bias as an indicator of animal emotion and welfare: emerging evidence and underlying mechanisms. Appl. Anim. Behav. Sci. 118, 161–181. http://dx.doi.org/10.1016/j.applanim.2009.02.023.

Mendl, M., Burman, O.H.P., Paul, E.S., 2010. An integrative and functional framework for the study of animal emotion and mood. Proc. Biol. Sci. 277, 2895–2904. http://dx.doi.org/10.1098/rspb.2010.0303.

Milliken, G., Johnson, D., 2009. Analysis of Messy Data: Volume 1, Designed Experiments, 2nd edition. Chapman & Hall/CRC Press, Boca Raton, FL.

Millot, S., Cerqueira, M., Castanheira, M.F., Øverli, Ø., Martins, C.I., Oliveira, R.F., 2014. Use of conditioned place preference/avoidance tests to assess affective states in fish. Appl. Anim. Behav. Sci. 154, 104–111. http://dx.doi.org/10.1016/j.applanim.2014.02.004.

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. http://dx.doi.org/10.1016/j.applanim.2008.07.001.

Minero, M., Dalla Costa, E., Dai, F., Murray, L.A.M., Canali, E., Wemelsfelder, F., 2016. Use of Qualitative Behaviour Assessment as an indicator of welfare in donkeys. Appl. Anim. Behav. Sci. 174. http://dx.doi.org/10.1016/j.applanim.2015.10.010.

Mullan, S., Edwards, S., Butterworth, A., Whay, H., Main, D.C., 2011. A pilot investigation of possible positive system descriptors in finishing pigs. Anim. Welf. 20, 439–449.

Munsterhjelm, C., Heinonen, M., Valros, A., 2015. Application of the Welfare Quality® animal welfare assessment system in Finnish pig production, part II: associations between animal-based

and environmental measures of welfare. Anim. Welf. 24, 161–172.
http://dx.doi.org/10.7120/09627286.24.2.161.

Murphy, E., Nordquist, R.E., van der Staay, F.J., 2014. A review of behavioural methods to study emotion and mood in pigs, Sus scrofa. Appl. Anim. Behav. Sci. 159, 9–28.
http://dx.doi.org/10.1016/j.applanim.2014.08.002.

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. http://dx.doi.org/10.1016/j.applanim.2007.03.009.

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.
http://dx.doi.org/10.1016/j.applanim.2012.08.002.

Normando, S., Haverbeke, A., Meers, L., Ödbreg, F., Ibañez Talegón, M., 2002. Heart rate reduction by grooming in horses (Equus caballus). In: Havemeyer Workshop on Horse Behavior and Welfare. Holar, Iceland. pp. 23–26.

Panksepp, J., 2011. Toward a cross-species neuroscientific understanding of the affective mind: do animals have emotional feelings? Am. J. Primatol. 73, 545–561.
http://dx.doi.org/10.1002/ajp.20929.

Paul, E.S., Harding, E.J., Mendl, M., 2005. Measuring emotional processes in animals: the utility of a cognitive approach. Neurosci. Biobehav. Rev. 29, 469–491.
http://dx.doi.org/10.1016/j.neubiorev.2005.01.002.

Phythian, C.J., Michalopoulou, E., Cripps, P.J., Duncan, J.S., Wemelsfelder, F., 2016. Onfarm qualitative behaviour assessment in sheep: repeated measurements across time, and association with physical indicators of flock health and welfare. Appl. Anim. Behav. Sci. 175, 23–31.
http://dx.doi.org/10.1016/j.applanim.2015.11.013.

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. http://dx.doi.org/10.1016/j.applanim.2005.12.009.

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. http://dx.doi.org/10.1016/j.applanim.2012.04.004.

Serrapica, M., Boivin, X., Coulon, M., Braghieri, A., Napolitano, F., 2017. Positive perception of human stroking by lambs: qualitative behaviour assessment confirms previous interpretation of quantitative data. Appl. Anim. Behav. Sci. 187, 31–37. http://dx.doi.org/10.1016/j.applanim.2016.11.007.

Tabachnick, B., Fidell, L., 2013. Using Multivariate Statistics, 6th edition. Pearson, Boston.

Temple, D., Manteca, X., Dalmau, A., Velarde, A., 2013. Assessment of test-retest reliability of animal-based measures on growing pig farms. Livest. Sci. 151, 35–45. http://dx.doi.org/10.1016/j.livsci.2012.10.012.

Thorbergson, Z., Nielsen, S., Beaulieu, R., Doyle, R., 2016. Physiological and behavioural responses of horses to wither scratching and patting the neck when under saddle. J. Appl. Anim. Welf. Sci. 19, 245–259. http://dx.doi.org/10.1080/10888705.2015.1130630.

Waiblinger, S., Boivin, X., Pedersen, V., Tosi, M.V., 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. http://dx.doi.org/10.1016/j.applanim. 2006.02.001.

Wemelsfelder, F., Hunter, E., Mendl, M., Lawrence, A., 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., Mendl, M.T., Lawrence, A.B., 2001. Assessing the whole animal: a free choice profiling approach. Anim. Behav. 62, 209–220. http://dx.doi. org/10.1006/anbe.2001.1741.

Wemelsfelder, F., 2007. How animals communicate quality of life: the qualitative assessment of behaviour. Anim. Welf. 16, 25–31.

Table 1 List of AWIN QBA descriptors for horses and their characterisations.

Descriptors	
Aggressive	Hostile, attacking, wants to fight/attack, dominance, defensive aggression, (i.e. may display the following: bite/kick, position of ears flat-back against head, dilated nostrils, turns the hindquarters towards object of aggression, intention to harm, tail-swishing)
Alarmed	Worried/tense, apprehensive, jumpy, nervous, watchful, on guard against a possible threat/danger (i.e. rigid stance, startled reaction to loud noise, looking around/vigilant, moving ears)
Annoyed	Irritated, displeased, bothered by something, disturbed, upset, troubled, exasperated (i.e. may display rapid tail-swishing, stomping)
Apathetic	Having or showing little or no emotion; disinterested, indifferent, isolated, depressed, unresponsive, motionless
At ease	Calm, carefree, peaceful
Curious	Inquisitive, desire to investigate (i.e. approach person/object of curiosity, engaged in exploratory behaviour; possibly displaying head and neck extended toward object of curiosity, with ears pricked forward)
Friendly	Affectionate, kind, not hostile, receptive, positive feelings toward people, confident (i.e. the horse approaches the person, may sniff or interact in some way)
Fearful	Afraid, hesitant, timid, not confident, not necessarily linked with something going on in the environment (i.e. you may see the body tremble, flared nostrils, tail clamped)
Happy	Feeling, showing or expressing joy, pleased, lively, playful, satisfied
Look for	Actively looking for interaction, interested, close proximity, eager to approach

contact	
Relaxed	Not tense or rigid, easy-going, tranquil
Pushy	Assertive or forceful (i.e. not leaving space, head butting out of the way, exhibits dominant behaviour, may be mouthy or nippy)
Uneasy	Afflicted, uncomfortable, unsettled, restless

Table 2 Training phase. Percentage of variation explained, and level of inter-observer agreement achieved, for the first three PCA Components. For each Component, the terms reported in bold between brackets are the ones with highest and lowest loadings.

	PC1 (at ease–uneasy)	PC2 (look for contact–apathetic)	PC3 (curious–apathetic)
% of variation explained	48%	17%	9%
Kendall's W (n=3, df=94)	0.76	0.74	0.50
Chi-sq P	P < 0.001	P < 0.001	P < 0.001

Table 3 Kendall's W correlation coefficients for individual QBA descriptors in the training phase.

Descriptors	Kendall's W
Pushy	0.77
At ease	0.72
Aggressive	0.71
Alarmed	0.71
Uneasy	0.71
Look for contact	0.70
Relaxed	0.68
Annoyed	0.67
Happy	0.64
Friendly	0.61
Fearful	0.61
Curious	0.60
Apathetic	0.56

Table 4 On-farm assessment. Outcomes for the first 3 Principal Components, and loadings of QBA terms on these components (highest and lowest loadings in bold).

Principal Component	Eigen value	% Of variance explained	Cumulative variance explained
PC1	5.081	39.083	39.083
PC2	2.140	16.465	55.549
PC3	1.322	10.171	65.720
QBA descriptors	PC1	PC2	PC3
Aggressive	**−0.488**	0.456	0.194
Alarmed	**−0.706**	0.354	0.396
Annoyed	**−0.606**	0.411	0.267
Apathetic	−0.155	**−0.592**	−0.234
At_ease	**0.811**	0.054	0.393
Curious	0.576	0.622	−0.275
Fearful	**−0.546**	0.178	−0.014
Friendly	**0.805**	0.208	−0.169
Happy	**0.613**	0.325	**0.568**
Look_for_contact	0.459	**0.589**	**−0.512**
Pushy	−0.326	**0.601**	−0.251
Relaxed	**0.826**	−0.064	0.306
Uneasy	**−0.801**	0.79	−0.025

Fig. 1. On-farm assessment. Loadings of QBA descriptors along the first two PCA Components.

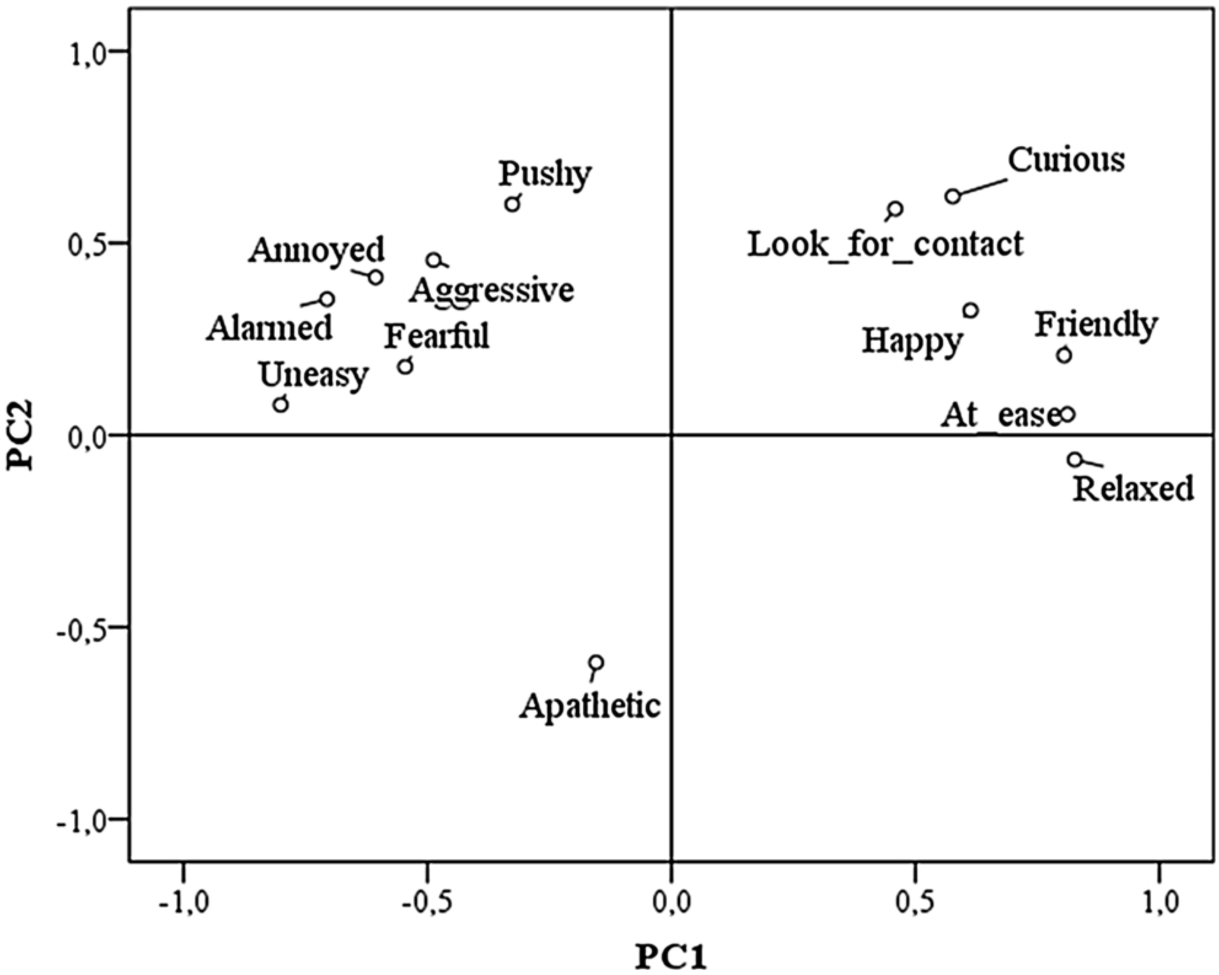

Fig. 2. On-farm assessment. Interactions between the first two QBA components (PC1 and PC2) and different scores of human-animal tests.