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

Michela Minero[1]*, Emanuela Dalla Costa[1], Francesca Dai[1], Elisabetta Canali[1], Sara Barbieri[1], Adroaldo Zanella[2], Riccardo Pascuzzo[3], Francoise Wemelsfelder[4]

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

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

[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

*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 to 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 (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). For data from the training phase, Kendall Correlation Coefficient W was used to assess inter-observer reliability both at Principal Component (PC) and individual descriptor level. A MANOVA test was used to assess the association of the AD and FHA tests with the QBA PC scores. The QBA approach described in this paper was feasible on farm and showed good acceptability by farmers. 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 more frequently showed no 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 (Paul et al., 2005; Boissy et al., 2007; Wemelsfelder, 2007; Fraser, 2009; Mendl et al., 2010; Panksepp, 2011; Millot et al., 2014; Murphy et al., 2014; Fureix and Meagher, 2015). Qualitative behaviour assessment (QBA) is one of those scientific methods, originally developed by Wemelsfelder and colleauges (2000, 2001), that has been proven to contribute to the identification of the main dimensions of animal emotional states (Carreras et al., 2016; Mullan et al., 2011; Rutherford et al., 2012; Temple et al., 2013; Mendl et al., 2010). 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 (Minero et al., 2016; Ebinghaus 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).

The QBA descriptors can be individually generated by observers, as in the case of the Free-Choice-Profiling methodology (FCP), or they are 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 assessment in different animal species (Grosso et al., 2016; Minero 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). Recently, QBA was included in the AWIN welfare assessment protocol for horses as an on-farm measure for positive emotional state (AWIN, 2015). This protocol was developed by the Animal Welfare Indicators (AWIN) research project funded by EU FP7, and described by Dalla Costa and colleagues (2016b).

Here for the first time we applied a fixed list of terms to the qualitative behaviour assessment on horses stabled in single boxes to investigate their emotional state and to explore its association with the human-animal 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 inter-observer 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 (riding school = 37%; training centre = 24%; breeding farm = 15%; hippodrome = 3%; other (e.g. animal-assisted activity) = 21%), as reported by (Dalla Costa et al., 2016a). 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. QBA assessment took

122 place immediately after entering the farms and letting the animals adapt to the observers' presence.
123 Assessors had never entered the farms before and were unaware of the different backgrounds of the
124 farms, so as not to be biased by any pre-existing prejudices regarding these backgrounds. They wore
125 blue overalls and had not made any clinical examination nor treatment to horses during the month
126 prior the assessments.
127 The assessor initially observed a horse from outside the box, without disturbing it, for 30 s. Then they
128 entered the box, approaching the horse slowly and scratched the horse at the withers for 30 s, all the
129 while observing the horse's responses. At the end of each horse observation period, they scored the
130 list of QBA descriptors on visual analogue scales (VAS), where the ends of the scale represented the
131 'minimum' (this expressive quality is absent) and 'maximum' (this quality could not be present more
132 strongly) of the expressive quality. The score was represented by the measure of the distance in
133 millimetres between the left 'minimum' point of the scale and the point where the observer's thick
134 crossed the line. Automated data recording and download of scores to excel files was made possible
135 by use of a dedicated electronic application specifically developed at SRUC (Scotland's Rural
136 College) in the UK.
137 In order to evaluate the quality of the human-horse relationship, after concluding QBA scoring the
138 assessors performed and scored an avoidance distance test (AD) and a forced human approach test
139 (FHA) (Dalla Costa et al., 2015). The AD test was performed from outside the box. When the horse
140 was attentive to their presence, the assessor approached the animal walking at measured pace of one
141 step per second. If the horse showed an avoidance response, this was recorded as 0, no avoidance was
142 recorded as 1. In the FHA test, the assessor opened the box door, entered the box, and approached the
143 horse slowly. If the horse stood still calmly, the assessor raised their hand, touched the withers and
144 moved their hand along the back of the subject. The horse's reaction was scored from 0 to 2 (0 = the
145 horse showed aggressive behaviour; 1 = the horse moved away as soon as he/she touched the withers;
146 2 = the horse stood still calmly or showed positive signs of interest). Horses that were reported by
147 their owners as having or having suffered back pain were not tested. Automated data recording and

download to Excel file 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 (R Core Team, 2016) were used for statistical analysis.

The QBA scores generated by the three assessors scoring all the 13 descriptors on 95 horses during the training phase, were analysed together using a 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 (55.549% of variance explained) 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 human-animal 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, and the level of agreement between the scores generated by the three assessors on each of these components. 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

199 Given the high levels of agreement between assessors both for PC scores and scores on separate
200 descriptors, it was considered to be acceptable for the 3 assessors to independently visit and assess
201 horses at different farms, and subsequently analyse all collected scores together in one PCA.
202 This PCA identified three main Principal Components with Eigen value greater than 1, together
203 explaining 65% of the variation between horses. Table 4 shows the outcomes for these PCs, as well
204 as the loading of QBA terms on each PC. From these loadings it can be seen that PC1 ranges from
205 relaxed/at ease to uneasy/alarmed, PC2 from curious/pushy to apathetic, and PC3 from happy to
206 'looking for contact'. Figure 1 shows the distribution of loadings along PC 1 and 2.
207
208 *3.3 Influence of Human-horse relationship on horse emotional state*
209 The results of the two-way MANOVA suggested that the horses' responses to the Avoidance Distance
210 (AD) test were very close to being significantly linked to their scores on both QBA Principal
211 Components (P=0.0565). In particular, looking at the post-hoc tests, we found a significant difference
212 with respect to the first Principal Component (adjusted P=0.0376) and no difference with respect to
213 the second Principal Component (adjusted P=1). Regarding the Forced Human Approach (FHA) test,
214 we found a significant difference to their scores on both QBA Principal Component (P<0.0001),
215 which was confirmed also by the post-hoc test performed on the two Principal Components separately
216 (both adjusted P<0.0001). The results of the post-hoc analysis are summarised in Figure 2.
217 The upper part of Figure 2 shows significant associations between the horses' PC1 scores and their
218 scores for the AD and FHA tests, indicating that horses perceived as more relaxed/at ease were more
219 frequently scored 1 (no avoidance) during the avoidance test (AD) and responded less aggressively
220 and fearfully to human presence (higher scores in the FHA test). The lower part of Figure 2 shows a
221 significant association only between the horses' PC2 scores and their FHA scores, indicating that
222 horses perceived as curious/pushy responded more aggressively to human presence.
223
224 **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 (Rousing and Wemelsfelder, 2006; Brscic et al., 2009; Napolitano et al., 2012; Andreasen et al., 2013; Munsterhjelm et al., 2015; Phythian et al., 2016; Fleming et al., 2013; Grosso et al., 2016; Minero et al., 2016); 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 and colleagues (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 (Dalla 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 (Lynch et al., 1974; Feh and de Mazières, 1993; 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 (Rousing and Wemelsfelder, 2006; Ellingsen et al., 2014), 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

277 positive affective state. The second component, ranging from curious/pushy to apathetic, could be
278 interpreted as more indicative of the horses' arousal level. These findings map well in the overall
279 picture where different methods to assess emotions in animals repeatedly highlighted dimensions of
280 valence and arousal of affective states (Mendl et al., 2009; Paul et al., 2005). Differently from other
281 methods, QBA can be applied during on-farm assessments and can be used to facilitate the dialogue
282 between owners and assessors (Wemelsfelder, 2007; Minero et al., 2009), possibly increasing the
283 engagement of owners in the process of improving animal welfare.
284 Horses reactions to human-animal interaction tests were significantly linked to Qualitative Behaviour
285 Assessment. In particular, a high score on QBA descriptors like relaxed, friendly, at ease, loading
286 high on the first component, was found to be pronouncedly associated with an absence of signs of
287 avoidance, and positive signs of interest towards an interacting human. Horses achieving higher
288 scores in the tests had a better relationship with humans and a more positive affective state.
289 Conversely, horses showing an aggressive reaction to a forced human approach were described as
290 more pushy when assessed beforehand with QBA. Horses achieving low scores in the FHA test (more
291 aggressive behaviour during the test) had a poorer relationship with humans and were described as
292 being more aroused. These results add on those reported by other authors, that animals having a
293 positive bond with humans are safer and easier to handle, whilst negative handling leads to poorer
294 mood and an aroused state (Breuer et al., 2000; Waiblinger et al., 2006; Ellingsen et al., 2014). It can
295 also be suggested that poor handling increases fear of humans in horses, influencing their mood and
296 level of arousal, and drive them into a negative feedback cycle that progressively leads them to
297 become more aggressive and unsafe to handle.
298
299 **5. Conclusions**
300 The QBA assessment procedure proposed here allowed to capture expressions of affective states of
301 horses in their home box and proved to be feasible on-farm. The good inter-observer reliability
302 achieved, both on overall PC scores and single descriptors, suggest that a phased procedure for the

303 training of assessors is effective in reaching a satisfactory reliability of observers. QBA was useful to
304 identify horses in a more positive affective state and, in line with previous findings in dairy cows
305 (Brscic et al., 2009; Ellingsen et al., 2014) and lambs (Serrapica et al., 2017), we can support the
306 hypothesis that QBA is sensitive to the quality of human contact. Our results suggest that high quality
307 relations with humans are a potential tool to provide good welfare, also in terms of positive emotions.
308

309 **Conflict of interest**

310 All authors of the manuscript "Using qualitative behaviour assessment (QBA) to explore the
311 emotional state of horses and its association with human-animal relationship" declare no actual or
312 potential conflict of interest including financial, personal or other relationships with other people or
313 organizations within three years of beginning the submitted work that could inappropriately influence,
314 or be perceived to influence, their work.
315

316 **Acknowledgements**

317 The authors wish to thank the EU VII Framework program (FP7-KBBE-2010-4) for financing the
318 Animal Welfare Indicators (AWIN) project. They also thank Philipp Scholz for his help in data
319 collection and a group of master students of Mathematical Engineering of Politecnico di Milano for
320 their help with data analysis. The authors are grateful to Tomasz Krzyzelewski at SRUC EGENES
321 and Alberto Carrascal from DAIA Intelligent Solutions for their programming expertise and
322 collaboration in developing the AwinHorse app.
323

324 **Author contributions statement**

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

328 and revision of the manuscript.

329

330 **References**

331 Andreasen, S.N., F. Wemelsfelder, P. Sandøe, and B. Forkman. 2013. The correlation of Qualitative
332 Behavior Assessments with Welfare Quality® protocol outcomes in on-farm welfare assessment
333 of dairy cattle. *Applied Animal Behaviour Science*. 143:9–17.
334 doi:10.1016/j.applanim.2012.11.013.

335 AWIN. 2015. AWIN welfare assessment protocol for donkeys.

336 Boissy, A., G. Manteuffel, M.B. Jensen, R.O. Moe, B. Spruijt, L.J. Keeling, C. Winckler, B. Forkman,
337 I. Dimitrov, J. Langbein, M. Bakken, I. Veissier, and A. Aubert. 2007. Assessment of positive
338 emotions in animals to improve their welfare. *Physiology & behavior*. 92:375–97.
339 doi:10.1016/j.physbeh.2007.02.003.

340 Breuer, K., P. Hemsworth, J. Barnett, L. Matthews, and G. Coleman. 2000. Behavioural response to
341 humans and the productivity of commercial dairy cows. *Applied Animal Behaviour Science*.
342 66:273–288. doi:10.1016/S0168-1591(99)00097-0.

343 Brscic, M., F. Wemelsfelder, E. Tessitore, F. Gottardo, G. Cozzi, and C.G. Van Reenen. 2009.
344 Welfare assessment: correlations and integration between a Qualitative Behavioural Assessment
345 and a clinical/ health protocol applied in veal calves farms. *Italian Journal of Animal Science*.
346 8:601–603. doi:10.4081/ijas.2009.s2.601.

347 Carreras, R., E. Mainau, L. Arroyo, X. Moles, J. González, A. Bassols, A. Dalmau, L. Faucitano, X.
348 Manteca, and A. Velarde. 2016. Housing conditions do not alter cognitive bias but affect serum
349 cortisol, qualitative behaviour assessment and wounds on the carcass in pigs. *Applied Animal*
350 *Behaviour Science*. 185:39–44. doi:10.1016/j.applanim.2016.09.006.

351 Dalla Costa, E., C. Bonaita, S. Pedretti, E. Govoni, A. Guzzeloni, E. Canali, and M. Minero. 2012.
352 Inter-observer reliability of three human-horse relationship tests. *In* 8th International Equitation
353 Science Conference. 170.

354 Dalla Costa, E., F. Dai, D. Lebelt, P. Scholz, S. Barbieri, E. Canali, and M. Minero. 2016a. Initial
355 outcomes of a harmonized approach to collect welfare data in sport and leisure horses. *Animal*.
356 1–7. doi:10.1017/S1751731116001452.
357 Dalla Costa, E., F. Dai, D. Lebelt, P. Scholz, S. Barbieri, E. Canali, A.J. Zanella, and M. Minero.
358 2016b. Welfare assessment of horses: the AWIN approach. *Animal Welfare*.
359 Dalla Costa, E., F. Dai, L.A.M.L.A.M. Murray, S. Guazzetti, E. Canali, and M. Minero. 2015. A
360 study on validity and reliability of on-farm tests to measure human–animal relationship in horses
361 and donkeys. *Applied Animal Behaviour Science*. 163:110–121.
362 doi:10.1016/j.applanim.2014.12.007.
363 Ebinghaus, A., S. Ivemeyer, J. Rupp, and U. Knierim. 2016. Identification and development of
364 measures suitable as potential breeding traits regarding dairy cows' reactivity towards humans.
365 *Applied Animal Behaviour Science*. 185:30–38. doi:10.1016/j.applanim.2016.09.010.
366 Eliasson, K., P. Palm, T. Nyman, and M. Forsman. 2017. Inter - and intra - observer reliability of risk
367 assessment of repetitive work without an explicit method. *Applied Ergonomics*. 62:1–8.
368 doi:10.1016/j.apergo.2017.02.004.
369 Ellingsen, K., G.J. Coleman, V. Lund, and C.M. Mejdell. 2014. Using qualitative behaviour
370 assessment to explore the link between stockperson behaviour and dairy calf behaviour. *Applied*
371 *Animal Behaviour Science*. 153:10–17. doi:10.1016/j.applanim.2014.01.011.
372 Feh, C., and J. de Mazières. 1993. Grooming at a preferred site reduces heart rate in horses. *Animal*
373 *Behaviour*. 46:1191–1194.
374 Fleming, P.A., C.C.L. Paisley, and A.L. Barnes. 2013. Application of Qualitative Behavioural
375 Assessment to horses during an endurance ride. *Animal Behaviour*. 144:80–88.
376 Fraser, D. 2009. Animal behaviour, animal welfare and the scientific study of affect. *Applied Animal*
377 *Behaviour Science*. 118:108–117. doi:10.1016/j.applanim.2009.02.020.
378 Fureix, C., and R.K. Meagher. 2015. What can inactivity (in its various forms) reveal about affective
379 states in non-human animals? A review. *Applied Animal Behaviour Science*. 171:8–24.

doi:10.1016/j.applanim.2015.08.036.

Grosso, L., M. Battini, F. Wemelsfelder, S. Barbieri, M. Minero, E. Dalla Costa, and S. Mattiello. 2016. On-farm Qualitative Behaviour Assessment of dairy goats in different housing conditions. *Applied Animal Behaviour Science*. 180:51–57. doi:10.1016/j.applanim.2016.04.013.

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

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

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

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

Mcdonnell, S.M. 2003. A practical field guide to horse behavior. The Equid Ethogram. Eclipse Press. 60-76 pp.

Mendl, M., O.H.P. Burman, R.M.A. Parker, and E.S. Paul. 2009. Cognitive bias as an indicator of animal emotion and welfare: Emerging evidence and underlying mechanisms. *Applied Animal Behaviour Science*. 118:161–181. doi:10.1016/j.applanim.2009.02.023.

Mendl, M., O.H.P. Burman, and E.S. Paul. 2010. An integrative and functional framework for the study of animal emotion and mood. *Proceedings. Biological sciences / The Royal Society*. 277:2895–904. doi:10.1098/rspb.2010.0303.

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

Millot, S., M. Cerqueira, M.F. Castanheira, Ø. Øverli, C.I. Martins, and R.F. Oliveira. 2014. Use of conditioned place preference/avoidance tests to assess affective states in fish. *Applied Animal*

406 *Behaviour Science*. 154:104–111. doi:10.1016/j.applanim.2014.02.004.
407 Minero, M., E. Dalla Costa, F. Dai, L.A.M. Murray, E. Canali, and F. Wemelsfelder. 2016. Use of
408 Qualitative Behaviour Assessment as an indicator of welfare in donkeys. *Applied Animal*
409 *Behaviour Science*. 174. doi:10.1016/j.applanim.2015.10.010.
410 Minero, M., M.V. Tosi, E. Canali, and F. Wemelsfelder. 2009. Quantitative and qualitative
411 assessment of the response of foals to the presence of an unfamiliar human. *Applied Animal*
412 *Behaviour Science*. 116:74–81. doi:10.1016/j.applanim.2008.07.001.
413 Mullan, S., S. Edwards, A. Butterworth, H. Whay, and D.C. Main. 2011. A pilot investigation of
414 possible positive system descriptors in finishing pigs. *Animal Welfare*. 20:439–449.
415 Munsterhjelm, C., M. Heinonen, and A. Valros. 2015. Application of the Welfare Quality® animal
416 welfare assessment system in Finnish pig production, part II: Associations between animal-
417 based and environmental measures of welfare. *Animal Welfare*. 24:161–172.
418 doi:10.7120/09627286.24.2.161.
419 Murphy, E., R.E. Nordquist, and F.J. van der Staay. 2014. A review of behavioural methods to study
420 emotion and mood in pigs, Sus scrofa. *Applied Animal Behaviour Science*. 159:9–28.
421 doi:10.1016/j.applanim.2014.08.002.
422 Napolitano, F., G. De Rosa, A. Braghieri, F. Grasso, A. Bordi, and F. Wemelsfelder. 2008. The
423 qualitative assessment of responsiveness to environmental challenge in horses and ponies.
424 *Applied Animal Behaviour Science*. 109:342–354. doi:10.1016/j.applanim.2007.03.009.
425 Napolitano, F., G. De Rosa, F. Grasso, and F. Wemelsfelder. 2012. Qualitative behaviour assessment
426 of dairy buffaloes (Bubalus bubalis). *Applied Animal Behaviour Science*. 141:91–100.
427 doi:10.1016/j.applanim.2012.08.002.
428 Normando, S., A. Haverbeke, L. Meers, F. Ödbreg, and M. Ibañez Talegón. 2002. Heart rate reduction
429 by grooming in horses (Equus caballus). *In* Havemeyer Workshop on Horse Behavior and
430 Welfare. Holar, Iceland. 23–26.
431 Panksepp, J. 2011. Toward a cross-species neuroscientific understanding of the affective mind: Do

animals have emotional feelings? *American Journal of Primatology*. 73:545–561. doi:10.1002/ajp.20929.

Paul, E.S., E.J. Harding, and M. Mendl. 2005. Measuring emotional processes in animals: The utility of a cognitive approach. *Neuroscience and Biobehavioral Reviews*. 29:469–491. doi:10.1016/j.neubiorev.2005.01.002.

Phythian, C.J., E. Michalopoulou, P.J. Cripps, J.S. Duncan, and F. Wemelsfelder. 2016. On-farm qualitative behaviour assessment in sheep: Repeated measurements across time, and association with physical indicators of flock health and welfare. *Applied Animal Behaviour Science*. 175:23–31. doi:10.1016/j.applanim.2015.11.013.

Rousing, T., and F. Wemelsfelder. 2006. Qualitative assessment of social behaviour of dairy cows housed in loose housing systems. *Applied Animal Behaviour Science*. 101:40–53. doi:10.1016/j.applanim.2005.12.009.

Rutherford, K.M.D., R.D. Donald, A.B. Lawrence, and F. Wemelsfelder. 2012. Qualitative Behavioural Assessment of emotionality in pigs. *Applied Animal Behaviour Science*. 139:218–224. doi:10.1016/j.applanim.2012.04.004.

Serrapica, M., X. Boivin, M. Coulon, A. Braghieri, and F. Napolitano. 2017. Positive perception of human stroking by lambs: Qualitative behaviour assessment confirms previous interpretation of quantitative data. *Applied Animal Behaviour Science*. 187:31–37. doi:10.1016/j.applanim.2016.11.007.

Tabachnick, B., and L. Fidell. 2013. Using multivariate statistics. 6th editio. Pearson, Boston.

Temple, D., X. Manteca, A. Dalmau, and A. Velarde. 2013. Assessment of test-retest reliability of animal-based measures on growing pig farms. *Livestock Science*. 151:35–45. doi:10.1016/j.livsci.2012.10.012.

Thorbergson, Z., S. Nielsen, R. Beaulieu, and R. Doyle. 2016. Physiological and Behavioral Responses of Horses to Wither Scratching and Patting the Neck When Under Saddle. *Journal of Applied Animal Welfare Science*. 19:245–259. doi:10.1080/10888705.2015.1130630.

458 Waiblinger, S., X. Boivin, V. Pedersen, M.-V. Tosi, A.M. Janczak, E.K. Visser, and R.B. Jones. 2006.
459 Assessing the human–animal relationship in farmed species: a critical review. *Applied Animal*
460 *Behaviour Science*. 101:185–242. doi:10.1016/j.applanim.2006.02.001.
461 Wemelsfelder, F. 2007. How animals communicate quality of life: the qualitative assessment of
462 behaviour. *Animal Welfare*. 16:25–31.
463 Wemelsfelder, F., E. Hunter, M. Mendl, and A. Lawrence. 2000. The spontaneous qualitative
464 assessment of behavioural expressions in pigs: first explorations of a novel methodology for
465 integrative animal welfare measurement. *Applied animal behaviour science*. 67:193–215.
466 Wemelsfelder, F., T.E. Hunter, M.T. Mendl, and A.B. Lawrence. 2001. Assessing the "whole
467 animal": A free choice profiling approach. *Animal Behaviour*. 62:209–220.
468 doi:10.1006/anbe.2001.1741.
469