



HAL
open science

Validation of Qualitative Behaviour Assessment for Dairy Cows at Pasture

Lydiane Aubé, Marie-Madeleine Mialon, Isabelle Veissier, Alice de Boyer Des Roches

► **To cite this version:**

Lydiane Aubé, Marie-Madeleine Mialon, Isabelle Veissier, Alice de Boyer Des Roches. Validation of Qualitative Behaviour Assessment for Dairy Cows at Pasture. 2024. hal-04775130

HAL Id: hal-04775130

<https://hal.inrae.fr/hal-04775130v1>

Preprint submitted on 9 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Validation of qualitative behaviour assessment for dairy cows at pasture**

2 L. Aubé^{ab,1}, M.M. Mialon^a, I. Veissier^a, A. de Boyer des Roches^{ab}

3 ^aUniversité Clermont Auvergne, INRAE, VetAgro Sup, UMR Herbivores, 63122 Saint-Genès-

4 Champanelle, France

5 ^bChaire Bien-Être Animal, VetAgro Sup, 1 avenue Bourgelat, 69280 Marcy-l'Etoile, France

6 Corresponding author: isabelle.veissier@inrae.fr

7

8 **Abstract**

9 Qualitative Behaviour Assessment (**QBA**) has been validated as a measure of emotional state in
10 dairy cows kept indoors. We investigated the validity of QBA as a measure of emotional state of
11 cows at pasture. We focused on intra- and inter-observer reliability and the selectivity of QBA.
12 First, 5 observers watched 20 videos of dairy cows at pasture 4 times (resulting in 400
13 observations). The observers performed QBA using the fixed list of 20 terms proposed in Welfare
14 Quality. The first axis of the principal component analysis performed on these observations
15 represented emotional valence (**PC1**, from irritable to content) and a second axis represented
16 arousal (**PC2**, from calm to active). Kendall's concordance coefficients (**KW**) and interclass
17 correlation coefficients (**ICC**) within and between observers were calculated for PC1 and PC2.
18 Intra-observer reliability was mainly high ($KW > 0.75$ and $ICC > 0.75$), except for 3 observers during
19 the first session for PC1, for whom reliability was moderate. Inter-observer reliability was low to
20 moderate for PC1 and moderate to high for PC2. Second, two observers performed live QBA on a
21 herd of cows at pasture during 3 contexts designed to induce different emotional valences and
22 arousal levels: **AM**, in the morning after milking (when cows are active); **PM**, in the afternoon
23 before milking (when cows are mainly resting); and during handling to collect cows at pasture for
24 afternoon milking (when cows may be stressed). The effect of context on PC1 and PC2 was
25 investigated using linear mixed effects models. AM and PM contexts had higher scores on PC1
26 than handling context, indicating a more positive emotional state during AM and PM than during
27 handling. PM context had lower scores on PC2 than AM and handling contexts. Thus, QBA at
28 pasture was able to discriminate between contexts that should differ in emotional valence and
29 arousal. Thirdly, an observer assessed the behavioural responses of cows to handling by direct
30 observations followed by QBA. Pearson correlation coefficients were calculated between the

31 number of behavioural responses and the coordinates on PC1 and PC2. The more cows were
32 observed trotting, the more the cow's mood was perceived as negative ($r=-0.71$) and the more
33 cows were observed galloping and turning, the more the cow was perceived as excited ($r=0.77$
34 and 0.60). In conclusion, the QBA appears to be a valid measure of the emotional and arousal
35 state of dairy cows at pasture, but inter-observer reliability could be improved.

36 **Keywords:** emotional state; welfare assessment, reliability, grazing cattle, handling

37

38 1 Introduction

39 Qualitative Behaviour Assessment (**QBA**) has been developed as a measure of emotional state
40 (Forkman and Keeling, 2009; Wemelsfelder et al., 2001). QBA is an integrative assessment of the
41 'whole animal', where the focus is not on what behaviour or how many behaviours the animal
42 performs, but on how the animal behaves and interacts with its environment (Wemelsfelder,
43 1997; Wemelsfelder et al., 2001). QBA consists of observing an individual animal or a group of
44 animals over a period of time and assessing the animal's behaviour based on the animal's dynamic
45 body language, using either adjectives spontaneously generated by the observers (free choice
46 profiling) or a fixed list of adjectives (or descriptors) (Fleming et al., 2016). For example, the
47 Welfare Quality protocol includes a QBA with a fixed list of adjectives such as 'excited', 'content',
48 'anxious' (Forkman and Keeling, 2009). The QBA, which is carried out by observing the animals in
49 their daily environment without disturbance, is intended to measure the general emotional state
50 of the animals. QBA can also be conducted during specific situations/contexts to assess how the
51 situation is perceived by the animals. For example, QBA has been carried out on veal calves during
52 handling to assess the impact of the handling style of the keeper (Ellingsen et al., 2014).

53 The validation of QBA as a welfare indicator has mainly focused on observer reliability and
54 selectivity. Inter-observer reliability has been confirmed in many studies (Ceballos et al., 2021;
55 Phythian et al., 2013; Rousing and Wemelsfelder, 2006). However, some authors found low inter-
56 observer reliability (Bokkers et al., 2012; Czycholl et al., 2016; Gutmann et al., 2015). Therefore,
57 inter-observer reliability must be assessed before using QBA. The selectivity of a measure
58 corresponds to the ability of the measure to quantify what it is supposed to quantify (Knierim et
59 al., 2021). The selectivity of QBA has been validated through convergent validity by determining
60 the correlation of QBA results with other welfare indicators (e.g. proportion of lame animals in

61 sheep (Phythian et al., 2016), clinical indicators of disease (e.g. in dairy cows, de Boyer des Roches
62 et al, 2018), physiological measures (e.g. in cattle, Stockman et al., 2011) or behavioural measures
63 (e.g. in polar bears, Skovlund et al., 2023; in salmon, Wiese et al., 2023) known to reflect animal
64 welfare and emotional state. The selectivity of QBA has also been validated through construct
65 validity, by investigating whether QBA is able to discriminate between different
66 contexts/situations that are thought to induce different emotional and arousal states in animals
67 (e.g. in salmon, Wiese et al., 2023). QBA has been validated in a wide range of species, including
68 farm animals (e.g. in Welfare Quality protocols for cattle, pigs and poultry), zoo animals (e.g. polar
69 bears, Skovlund et al., 2023) and companion animals (e.g. dogs, Arena et al., 2019), and is now
70 widely used as a welfare indicator (in Welfare Quality and AWIN protocols for cattle, pigs, poultry,
71 sheep, goats and donkeys). In the Welfare Quality protocols (cattle, pigs and poultry), QBA is the
72 only measure used to assess positive emotional states.

73 QBA has been developed and validated as a measure of emotional state for indoor cows in either
74 tie stalls, cubicles or straw-bedded farms (Forkman and Keeling, 2009). More recently, the
75 Welfare Quality protocol has been applied to cattle at pasture, including QBA (e.g. in dairy cows,
76 Armbrecht et al., 2019; Wagner et al., 2018; in dual-purpose cattle, Hernandez et al., 2017).
77 However, to our knowledge, it has not yet been determined whether QBA developed for indoor
78 cows is relevant and valid for grazing cows. The validation of QBA at pasture seems necessary to
79 have a valid measure of emotional state, as even at pasture cows can be exposed to risks that
80 may affect their welfare (Aubé et al., 2022).

81 In order to determine whether QBA could be included in a protocol to assess cow welfare at
82 pasture, we propose to investigate the validity (both reliability and selectivity) of this measure
83 applied to dairy cows at pasture. The first objective of this study was to determine if QBA of cows

84 at pasture could be interpreted in a similar way to that of cows indoors. The second objective was
85 to assess observer reliability for QBA at pasture. The third objective was to assess the selectivity
86 of the measure through construct validity by comparing three contexts that are thought to differ
87 in the emotional and arousal state of cows, and through convergent validity by investigating the
88 relationship between QBA and quantitative behavioural reactivity during handling.

89 **2 Materials and methods**

90 The study took place at the INRAE farm 'Le Pin-au-Haras' (France, 48.448N, 0.098E, DOI:
91 10.15454/1.5483257052131956E12). This study was part of an experiment on the feeding
92 management of dairy cows' in a rotational grazing system. The experimental protocol was
93 assessed by the Rennes Institutional Animal Care and Use Committee of Rennes and approved by
94 the French Ministry of Higher Education, Research and Innovation (APAFIS agreement #20846-
95 2019052810487566 v2).

96 *2.1 Animals, housing, and management*

97 The housing and management of the cows are described in Aubé et al. (2023). Briefly, a herd of
98 144 cows (48 Holstein, 48 Normande, 48 Jersey) was housed indoors during winter (from
99 December to March) and turned out to pasture in spring after calving. The present study took
100 place when the cows were at pasture (from April to November 2021). The cows were at pasture
101 all day long, except when they were brought indoors for milking, between 07h00 and 08h30 and
102 between 16h00 and 17h30. The cows grazed on plots of 10 ha. They changed plot after 7 to 12
103 days of grazing.

104 *2.2 Qualitative behaviour assessment (QBA)*

105 *2.2.1 Observation conditions*

106 2.2.1.1 *Qualitative behaviour assessments from videos.*

107 To assess intra- and inter-observer reliability, twenty videos (1 min each) were taken of cows
108 grazing in different contexts: cows mainly resting or ruminating (n = 5 videos), cows mainly grazing
109 (n = 10), part of the herd grazing and part resting or ruminating (n = 2), cows released to pasture
110 for the first time after wintering indoors (n = 1), cows being bothered by flies in the sun (n = 1),
111 and cows reacting to a truck passing near the pasture (n = 1). Videos included from 4 to
112 approximately 100 cows. To assess intra- and inter-observer reliability, each video was watched
113 twice (at least 24 h apart) during two sessions (4 months apart, one in March, one in July) by 5
114 observers (A, B, C, D and E), resulting in a total of 400 observations (20 videos x 4 times x 5
115 observers). Prior to the experiment, observers A and C were trained by two scientists who
116 followed the training programme organised by Welfare Quality and regularly train people in the
117 use of the dairy cow protocol. Subsequently, Observer A trained Observers B, D and E. The QBA
118 training of Observers B, D and E was part of a larger training on 15 animal-based indicators and 3
119 resource-based indicators (not described in this paper), which took about 15 h spread over 5 half
120 days. Of these 15 h of training, about 30 min were spent in the classroom explaining the
121 theoretical background of QBA and about 1 hour was spent in the field (observing a herd of cows
122 at pasture in the morning, in the afternoon and during handling). The observers were female (A,
123 C) or male (B, D and E). Observer A was an ethologist, Observer C was a veterinarian and Observers
124 B, D and E were technicians from the experimental farm who had worked daily with cattle for
125 several years.

126 2.2.1.2 *Live QBA*

127 Live QBA was performed by two observers (Observer A was always present) on the whole herd at
128 pasture on 33 occasions and in different contexts: after returning to pasture from morning milking

129 (AM context, n = 8), before afternoon milking (PM context, n = 9) and during the collection of
130 cows at pasture for afternoon milking (handling context, n = 16) (see Table 1 for dates, contexts
131 and observers involved).

132 For the AM context, the two observers remained in the pasture plot as the cows arrived from
133 morning milking until the cows dispersed to graze (about 5 min), after which they performed the
134 observation. For the PM context, as the cows were already present in the pasture plot when the
135 observers arrived, the observer entered the pasture and stood still for about 5 min to allow the
136 cows to get used to the presence of the observers. Prior to the AM and PM observations, the two
137 observers discussed and virtually divided the pasture plot into 4 observation areas, which they
138 defined together in order to observe all animals. The boundaries of the observation areas (using
139 environmental cues such as trees or water troughs) were chosen by the two observers so that
140 each area contained approximately one quarter of the total herd (i.e. between 30 and 40 cows).
141 The time allocated to each observation area was determined according to the Welfare Quality
142 protocol (i.e. 5 min/area for 4 areas for a total of 20 min of observation). The two observers spent
143 5 min in each observation area, visually scanning the designated area to observe all cows present
144 in the area.

145 For the handling context, the two observers stood behind the herd in the pasture plot (in relation
146 to the exit of the pasture plot) in order to see the herders leading the cows to the exit. Observation
147 began when the first stockperson entered the pasture and ended when the last cow of the herd
148 left the pasture. As the herd moved forward, the two observers walked quietly behind to follow
149 and observe the herd. Throughout the handling process, the two observers visually scan the entire
150 herd. The duration of the handling context observation depended only on the duration of the
151 handling process itself (i.e. from 3 to 17 min).

152 2.2.2 *Scoring procedure*

153 QBA was scored according to the Welfare Quality (2009) protocol. Briefly, each observer observed
154 the group of animals (either 1 min on video or 3 to 20 min live), then assigned a value to 20
155 adjectives on a 125 mm visual analogue scale, based on how well the adjective describes the
156 group. The adjectives were: Active, Agitated, Apathetic, Bored, Calm, Content, Distressed, Fearful,
157 Friendly, Frustrated, Happy, Indifferent, Inquisitive, Irritable, Lively, Playful, Positively occupied,
158 Relaxed, Sociable, and Uneasy.

159 2.3 *Quantitative behaviour measures during cow collection*

160 Cow reactivity to handling at herd level was assessed on 16 occasions (by observer A concurrently
161 with the QBA in handling context) during cow collection in the pasture for afternoon milking (see
162 Table 1 for detailed dates). During cow collection, observer A measured the duration of the entire
163 handling process using a chronometer and recorded the occurrence (all occurrence sampling,
164 Altmann, 1974) of the following responses to the stockperson's actions: cow freezing, startled,
165 kicking, trotting, galloping, slipping, falling, moving backwards, turning around, vocalising
166 (ethogram in Table 2).

167 The intra-observer reliability of observer A for cow behavioural responses to handling (galloping
168 and trotting behaviours only because they were the most frequent responses) was tested on 30
169 videos (of 15 s each) recorded during cow collection and viewed twice by the observer. This
170 observer was reliable for these two behaviours with interclass correlation coefficient > 0.95 and
171 percentage of agreement (number of observations where observer A agreed with the total
172 number of observations and multiplied by 100) $> 95\%$ for both galloping and trotting behaviour.

173 2.4 *Statistical analyses*

174 Statistical analyses were performed using R software version 4.2.2 (R Core Team, 2022). For all
175 analyses, the statistical unit is a QBA observation made by an observer.

176 2.4.1 *Principal component Analysis on QBA and adjectives weights to produce a QBA Score*

177 In order to summarise the information from the QBA of each video observation (n = 400), as in
178 Welfare Quality, a Principal Component Analysis (**PCA**, based on the correlation matrix and
179 without rotation) was performed on the values assigned to the 20 adjectives (from 0 to 125). The
180 PCA was performed using the "ade4" package (Chessel et al., 2004). The coordinates of the live
181 observations on the PCA first axis (**PC1**) and second axis (**PC2**) were further used as the output of
182 the QBA to test intra- and inter-observer reliability. A linear model was applied with the
183 coordinate of observations on PCA first axis (PC1) as explanatory variable and each adjective value
184 as predictors (Welfare Quality, 2009; Brscic et al. (2019)). The effect (i.e. estimate) of an adjective
185 in the linear model is used as the weight for that adjective to calculate the QBA score (see formula
186 below). We checked whether the weights of the adjectives found in our study were similar to the
187 weights of the adjectives used in Welfare Quality for Dairy Cattle. First, we checked that the
188 weights had the same sign. Second, we calculated the Pearson correlation between the weights.
189 Thirdly, the absolute value of the weight of each adjective was used to rank the 20 adjectives to
190 assess the relative importance of the adjectives (from 1, adjective with the highest absolute
191 weight, to 20, adjective with the lowest absolute weight). We then compared the ranking of the
192 20 adjective weights between our study and Welfare Quality (2009). We considered the rankings
193 of the absolute value of the weight to be close if they were within a range of 2 (e.g. rank 5 vs. rank
194 3).

195 The QBA score is a weighted sum where the adjective scores are multiplied by their weights and
196 then summed (Brscic et al., 2019).

197
$$QBA\ Score = constant + \sum_{k=1}^{20} \omega_k N_k$$

198 where N_k is the value (in mm) on the scale of a given adjective k for one observation, ω_k is the
199 weight given to the adjective and 'constant' is a fixed value corresponding to the intercept of the
200 linear model described above.

201 We calculated the QBA score of the video observations, using the adjectives' weights and the
202 constant from the present study. To compare our results with those of Welfare Quality (2009),
203 we calculated a QBA score using the weights and the constant defined in Welfare Quality (which
204 differ from ours because the PCA and the linear model were applied to a different set of
205 observations). We will refer to the pasture score (calculated with the weights and constant
206 obtained in the present study) vs. Welfare Quality score (calculated according to the weights and
207 constant defined in Welfare Quality). Pearson correlation coefficient was calculated to assess the
208 relation between the two scores.

209 2.4.2 *Intra- and inter-observer reliability of QBA*

210 Intra- and inter-observer reliability was assessed using the coordinates on PC1 or PC2 of each
211 observation made by each observer from the PCA previously performed on the video observations
212 (see 0).

213 Kendall's concordance coefficients (**W**) and interclass correlation coefficients (**ICC**) were
214 calculated, using the 'irr' package (Gamer et al., 2019). W and ICC comparing PC1 or PC2 obtained
215 by the same observer between two days of video observations indicate the intra-observer
216 reliability (calculated for each observer for each session). W and ICC calculated on PC1 or PC2
217 obtained by the different observers on the same day of video observation indicate inter-observer
218 reliability. For W , values less than 0.4 are considered to indicate poor reliability, values between

219 0.41 and 0.70 - moderate reliability, values between 0.71 and 0.90 – high reliability, and values
220 greater than 0.91 - excellent reliability (Martin and Bateson, 1993). For ICC, values less than 0.5
221 are considered indicative of poor reliability, values between 0.5 and 0.75 - moderate reliability,
222 values between 0.75 and 0.9 - good reliability, and values greater than 0.90 - excellent reliability
223 (Koo and Li, 2016).

224 2.4.3 *Construct validity*

225 Live QBA observations (AM, PM, and handling contexts) from both observers were added as
226 supplementary observations to the PCA previously conducted with the video observations. The
227 coordinates of the live observations on both PC1 and PC2 were compared between the three
228 contexts (AM, PM, and handling context) with a linear mixed effects model (using the R package
229 lme4) with context as a fixed effect and observer as a random effect. We hypothesised that
230 emotional valence during handling would be less positive and arousal would be lower in PM cows
231 than in the other two contexts. We based our hypothesis on the fact that motivation for cows to
232 be milked is low (Prescott et al., 1998), whereas cows are highly motivated for access to pasture
233 (Von Keyserlingk et al., 2017). We also assumed that emotional valence would be quite similar
234 between the two pasture contexts (AM and PM), whereas arousal levels would likely differ, with
235 cows being more active during AM observations (cows mostly grazing) and less active during PM
236 observations (cows mostly resting and ruminating, personal observation). Variations in PC1 or PC2
237 with context would therefore indicate that the QBA is valid for discriminating emotional and
238 arousal states of animals (construct validity).

239 2.4.4 *Convergent validity*

240 The Pearson correlation coefficients calculated between the frequency of each behaviour and the
241 coordinates of PC1 and PC2 during cow collection are used to assess convergent validity.

242 2.4.5 *Stability over the grazing season*

243 The stability of the QBA results (PC1 and PC2), obtained from the live observations (AM, PM and
244 handling contexts) of observer A only, was examined visually by examining the different
245 observations made on the herd throughout the season. In addition, the variability (amplitude:
246 maximum value minus minimum value obtained) of QBA per context over the season was
247 compared with the total variability of QBA per visit, all contexts together.

248 **3 Results**

249 3.1 *Principal component analysis of QBA from video and adjectives' weights*

250 The PCA was performed on 399 video observations (instead of 400 initially due to one missing
251 observation). The first axis explained 31% of the variation (eigenvalue, 6.1) and a second one
252 explained 27% of the variation (eigenvalue, 5.4). Adjectives with absolute loadings > 0.6 on PC1
253 were: content, happy, positively occupied (positive loadings) vs. frustrated irritable, uneasy, and
254 distressed (negative loadings) (Fig. 1 and Table 3). PC1 seems thus to be related to emotional
255 valence: the higher the value on PC1 the more positive the emotional valence. Adjectives with
256 absolute loadings > 0.6 on PC2 were: active, agitated, friendly, playful, lively, inquisitive and
257 sociable (positive loadings) vs. calm (negative loading). PC2 thus seems to relate to arousal: the
258 higher the value on PC2 the more active the animals.

259 For 16 out of the 20 adjectives, the sign of the weight in the present study was the same as in
260 Welfare Quality (2009). For the adjectives 'indifferent', 'sociable', 'playful' and 'friendly', the sign
261 of the weight differed from that in Welfare Quality (Table 3). In addition, the adjective weights
262 from our video observations and the Welfare Quality weights were highly correlated ($r = 0.86$).
263 The rankings of the absolute value of the weight were close (within a range of 2 for 12 adjectives),
264 except for 'restless', 'bored', 'calm' and 'lively' (for which the weights at pasture had a higher rank

265 than in Welfare Quality) and 'indifferent', 'friendly', 'apathetic' and 'excited' (for which the
266 weights from Welfare Quality had a higher rank than at pasture) (Table 3).

267 Pasture scores ranged from 0 to 81.9 and Welfare Quality scores ranged from 0 to 86.4. Mean
268 Pasture score was 53.1 (SD, 18.1) while mean Welfare Quality score was 55.6 (SD, 19.0). Pearson
269 correlation coefficient between the Pasture score and Welfare Quality score was 0.95.

270 *3.2 Intra and Inter-observer reliability of QBA*

271 Intra-observer reliability for PC1 and PC2 was high to excellent for observers A and B. Intra-
272 observer reliability of observers C, D and E for PC1 was moderate in Session 1 and high to excellent
273 in Session 2 except for Observer D (moderate reliability regarding ICC on Session 2). Intra-observer
274 reliability of observers C, D and E for PC2 was high to excellent but moderate for Observer D on
275 Session 1 regarding ICC (Table 4)

276 Inter-observer reliability as assessed from W (Kendall's concordance coefficients) was moderate
277 for PC1 in both sessions and high for PC2 in both sessions. Inter-observer reliability as assessed
278 from ICC was moderate in both sessions for PC1 and PC2 except for PC2 on Session 2 Day 1 for
279 which it was considered high (ICC = 0.76) and for PC1 on Session 1 Day 1 for which it was
280 considered low (ICC = 0.47) (Table 4).

281 *3.3 Construct validity*

282 The observations at pasture during AM and PM had higher coordinates on PC1 (i.e. reflecting a
283 better emotional state) than the observations during handling (Table 5).

284 The observations at pasture during the PM context had lower coordinates on PC2, i.e. reflecting
285 lower level of arousal, than observations during the AM and handling contexts (Table 5).

286 *3.4 Convergent validity*

287 During the handling observations (n = 16), no kicking, no falling, no moving backwards and no
288 freezing were observed. There was one case of startle and two cases of slipping. These responses
289 were therefore not included in the statistical analysis. Only, turning around, galloping, trotting
290 and vocalizing were retained for analysis. Turning frequency per observation ranged from a
291 minimum of 0 to a maximum of 14 (median = 2.5). Trotting frequency ranged from 0 to 37 (median
292 = 21). Galloping frequency ranged from 0 to 36 (median = 10). Vocalizing frequency ranged from
293 0 to 8 (median = 0). The duration of the handling process ranged from 3.2 to 17.4 min (median =
294 11.3 min).

295 The frequency of galloping and turning around observed during handling were positively
296 correlated with PC2 (arousal) ($r = 0.77$ and 0.60 , $P = 0.001$ and 0.015). The duration of handling
297 and the number of trotting behaviours were negatively correlated with PC1 (valence) ($r = -0.70$
298 and -0.71 , $P = 0.012$ and 0.004). Vocalization frequency was not significantly correlated to either
299 PC1 or PC2 ($r < 0.4$, $P > 0.05$).

300 3.5 *Stability over the season*

301 The handling context had always negative coordinates on PC1 (except for one observation at visit
302 3), whereas the AM and PM contexts always had positive coordinates on PC1 (Fig. 1). Throughout
303 the season, PC1 values (i.e. coordinates of the observations on PC1) ranged from 2.2 to 3.2 for
304 AM, from 0.5 to 2.9 for PM and from -4.0 to 0.7 for handling contexts. The amplitudes of PC1
305 values were thus, 1.0 for AM, 2.4 for PM and 4.7 for handling. The amplitude of PC1 values all
306 contexts considered, were 6.2, 4.8, 4.3, 4.1, 5.9, 6.4 and 5.0 for Visits 1 to 7 respectively.

307 The AM and PM contexts always had negative coordinates on PC2, AM and the handling context
308 had either on positive or negative coordinates on PC2 depending on the visit (Fig. 1). Over the
309 whole season, PC2 values (i.e. coordinates of the observations on PC2) ranged from -1.0 to 1.0 for

310 AM, from -3.0 to 2.9 for PM and from -0.8 to 2.5 for handling context. The amplitudes of PC2
311 values were thus, 2.0 for AM, 2.9 for PM and 3.2 for handling. The amplitude of PC2 values all
312 contexts considered, were 2.5, 3.8, 1.3, 1.0, 2.5, 3.2 and 5.4 for Visits 1 to 7 respectively.

313 **4 Discussion**

314 Our study shows that QBA has the potential to capture the arousal and emotional state of cows
315 at pasture. However, inter-observer reliability should be improved.

316 ***The use of the 20 QBA adjectives as defined in Welfare Quality for indoor cows makes sense for***
317 ***cows at pasture.***

318 Our results showed that PCA axes from QBA of cows at pasture (from video observations) are
319 similar to PCA axes previously obtained from QBA performed on cows indoors (Forkman and
320 Keeling, 2009; Gutmann et al., 2015; Russell et al., 2023; Winckler, 2014) with PC1 being the
321 'mood' axis and PC2 being the 'arousal' axis. In addition, our PCA axes from QBA of cows at pasture
322 are similar to those found from observations of beef cattle either at pasture in summer or indoors
323 throughout the year (Cooke et al., 2023; except that the axes were reversed in their study (PC1 :
324 'arousal' and PC2 : 'mood')). Similar results were also found for goats including goats at pasture
325 (observations in eight farms where goats were housed indoors and eight farms where goats were
326 at pasture) (Grosso et al., 2016). Furthermore, the adjectives' weights that we obtained for cows
327 at pasture were similar to those proposed in the Welfare Quality protocol for cows indoor. As a
328 result, the pasture scores calculated from adjectives' weights from our observations were highly
329 correlated with the scores calculated as in the Welfare Quality protocol, and the distributions of
330 the two scores were similar. QBA can therefore be used for dairy cows at pasture as it is described
331 in the Welfare Quality protocol (same adjectives, same formula to calculate the score).

332 ***Our construct and convergent validity results support the idea that QBA has the potential to***
333 ***capture emotional state and arousal of cows at pasture.***

334 QBA at pasture is selective in that it is able to discriminate between different contexts that we
335 would expect to be different in terms of both emotional valence and arousal. Indeed, in the
336 present study, the handling situation, which we expected to be less positive in terms of emotional
337 state for the animals compared to situations in which the animals were not disturbed (morning
338 or afternoon observations), obtained lower scores on PC1. In addition, QBA was able to
339 discriminate between contexts in terms of arousal, with afternoon observations (cows are mainly
340 resting and ruminating at this time) scoring lower on PC2 than handling observations and morning
341 observations (cows are all grazing at this time). As QBA at pasture can discriminate between
342 emotional and arousal contexts, we assume that it can also discriminate between farms with
343 different emotional states of the animals.

344 Gutmann et al. (2015) also found differences in PC1 and PC2 for QBA performed at three different
345 times of the day (early morning, late morning, early afternoon) for cows indoors. This raises the
346 question of the moment of day for QBA, as different moments may give different results. QBA
347 should either be done at different times of the day to get a global view or always at the same time
348 to compare situations (same herd at different times or different herds). If QBA is to be carried out
349 only once during the day, it is advisable to carry out the QBA of grazing cows after morning milking
350 (as in the Welfare Quality protocol for indoor cows), when the cows are coming back to pasture
351 from milking. At this time, all cows are generally active and grazing, making it more likely to
352 observe the cow's interactions with her environment and conspecifics (personal observations). In
353 addition, according to Gutman et al. (2015), performing QBA at the beginning of a welfare
354 assessment day could minimise potential observer bias arising from the assessment of other

355 welfare measures. This is already done in the Welfare Quality protocol for dairy cows; indeed,
356 QBA is carried out at the beginning of the assessment day, immediately after avoidance distance
357 tests and before individual measures carried out in the home pen.

358 In our study, duration of handling and trotting frequency correlated negatively with PC1. Thus,
359 the longer the handling and the more cows were observed to trot in response to the handlers,
360 the more negative the mood of the cow was perceived by the observer. Galloping and trotting
361 frequencies were positively correlated with PC2, i.e. the more cows were observed galloping away
362 from the handler and turning in response to the handlers, the more the cows were perceived as
363 excited by the observer. These results confirm that QBA is selective in assessing the arousal and
364 emotional state of cows, at least during handling. This is in line with a previous study aimed at
365 identifying appropriate measures of reactivity of dairy cows towards humans (Ebinghaus et al.,
366 2016). Values obtained on PC1 (mood axis) from QBA during two handling tests on individual cows
367 (tolerance to standardised tactile interaction (**TTI**) and release behaviour after restraint (**RB**))
368 were correlated with reactivity scores during these tests (Ebinghaus et al., 2016). The more
369 reactive cows were to TTI and RB, the more negative emotional state was obtained from QBA.
370 Ebinghaus et al. (2016) concluded that QBA can be used to assess cow reactivity during handling,
371 which is consistent with our results.

372 As dairy cows at pasture are usually handled at least twice a day for milking, the quality of handling
373 may be an issue for their welfare. We therefore suggest that QBA during handling can be included
374 as a new measure in a future protocol to assess handling quality. As the duration of the handling
375 process in our study ranged from 3 to 17 min, it seems very feasible to include such observations
376 in a protocol. However, the duration and conditions of handling may vary between farms, so QBA
377 during handling may not always be feasible.

378 ***Intra-observer reliability was acceptable but inter-observer reliability should be improved.***

379 Intra-observer reliability for PC1 and PC2 was mainly good (except moderate for three observers
380 at the first session), which is in line with previous studies investigating intra-observer reliability of
381 QBA. For example, intra-observer reliability for both PC1 and PC2, as assessed by Spearman rank
382 correlations, was found to be high for QBA used to assess social behaviour in cows (Rousing and
383 Wemelsfelder, 2006). Similarly, the intra-observer reliability of eight observers was considered
384 high for PC1 ($W = 0.71$ to 0.89) and PC2 ($W = 0.74$ to 0.90) for QBA used to assess maternal
385 protective behaviour in dairy cows (Ceballos et al., 2021).

386 Inter-observer reliability was low to moderate for PC1 and moderate to high for PC2, so we
387 consider that inter-observer reliability was close to acceptable but not fully validated, especially
388 for PC1. This result is not surprising as other studies in cattle have found W coefficients indicating
389 low ($W \leq 0.40$) or moderate ($W \leq 0.70$) inter-observer reliability (Andreasen et al., 2013; Forkman
390 and Keeling, 2009; Gutmann et al., 2015). For example, in indoor dairy cows, inter-observer
391 reliability was low for PC1 ($W = 0.38$) and moderate for PC2 ($W = 0.46$) (Forkman and Keeling,
392 2009).

393 In our study and in previous studies (Battini et al., 2021; Forkman and Keeling, 2009; Gutmann et
394 al., 2015), inter-observer reliability seemed to be better for the axis representing arousal than for
395 the axis representing mood. This may be because it is easier to assess the activity, which is
396 external, than an animal's mood, which is an internal state and seems more open to
397 interpretation.

398 However, acceptable inter-observer reliability of QBA has been found in others studies, probably
399 due to extensive training of the observers. For example, in sheep a high inter-observer reliability
400 was obtained after 3 h of training dedicated to QBA alone (Phythian et al., 2013). In contrast, in

401 our study, the training on QBA was part of a larger training on 15 animal-based indicators,
402 together with 3 resource-based indicators (not described in this paper), which was conducted in
403 about 15 h spread over 5 half days, with about 1.5 h dedicated to QBA. Increasing the training
404 time with the observer on QBA could probably improve inter-observer reliability.

405 In our study, inter-observer reliability seemed to be better for the second session than for the
406 first. In fact, the W coefficients were all slightly better (or at least equal) during the second session
407 of video observations for both axes than during the first. This suggests that the observers became
408 more reliable with practice. With more time for training before session 1, we would probably have
409 achieved good inter-observer reliability. Providing brief definitions of each adjective, as has been
410 done for several species (donkeys, Minero et al., 2016; goats, Grosso et al., 2016; horses, Minero
411 et al., 2018; sows, Ibach et al., 2024; and shelter dogs, Stubbsjøen et al., 2020), to avoid
412 misunderstanding of adjectives by observers, could also improve inter-observer agreement.
413 Definition of adjectives and extensive training of observers should be considered to improve the
414 quality of QBA.

415 **PC1 (“mood”) values appeared to be stable across the grazing season for the AM and PM**
416 **contexts but not for the handling context, while PC2 (“arousal”) scores did not appear to be**
417 **stable for any context.**

418 The variation (amplitude) in PC1 scores for the AM and PM contexts was lower than the variation
419 between visits, but the variation in PC1 scores for the handling context seemed to be quite similar
420 to the variation between visits. This suggests that PC1 scores were quite stable over the season,
421 except for the handling context. The variation for the handling context may be due to the pair of
422 handlers not being the same between visits. The variation in PC2 scores (amplitude) within the
423 AM, PM and handling contexts seemed to be close to the variations in PC2 scores between visits

424 (all contexts considered). This suggests that PC2 scores were not stable over the season whatever
425 the context. As the QBA score calculated in the Welfare Quality protocol is based on PC1 only, this
426 measure has the potential to be a stable measure for cows at pasture under undisturbed contexts.
427 We only observed one herd of cows live, the results should be confirmed by observing different
428 herds and contexts.

429 **5 Conclusion**

430 In conclusion, the QBA appears to be a valid measure of the emotional and arousal state of dairy
431 cows at pasture. The time and context of observation during the day must be chosen according
432 to the aim of the study, because QBA results vary during the day (afternoon vs. morning) and in
433 different contexts (undisturbed vs. handling situation). Before using QBA, inter-observer reliability
434 must be checked and training must be intensified if observers' reliability is not good.

435 **Declaration of interest**

436 None.

437 **Acknowledgements**

438 The authors thank Estelle Mollaret (Vetagro Sup), Yves Carbonnier (INRAE), Loïc Leloup (INRAE)
439 and Laurent Chanteloup (INRAE) for observing the cows, Luc Delaby and Frédéric Launay for
440 affording us access to the experimental unit, and the staff from INRAE Le Pin
441 (<https://doi.org/10.15454/1.5483257052131956E12>) for taking care of the animals. We also
442 thank Marianne Berthelot (Anses) for the discussions about statistical analysis. The experiment
443 was co-financed by the Banque des Territoires and Conseil Régional de Normandie, as part of the
444 France 2030 Territoires d'Innovation LIT OUESTEREL project.

445 **References**

- 446 Altmann, J., 1974. Observational study of behavior: sampling methods. *Behaviour* 49, 227–266.
- 447 Andreasen, S.N., Wemelsfelder, F., Sandøe, P., Forkman, B., 2013. The correlation of Qualitative
448 Behavior Assessments with Welfare Quality® protocol outcomes in on-farm welfare
449 assessment of dairy cattle. *Applied animal behaviour science* 143, 9–17.
- 450 Arena, L., Wemelsfelder, F., Messori, S., Ferri, N., Barnard, S., 2019. Development of a fixed list of
451 terms for the Qualitative Behavioural Assessment of shelter dogs. *PLOS ONE* 14,
452 e0212652. <https://doi.org/10.1371/journal.pone.0212652>
- 453 Armbrecht, L., Lambertz, C., Albers, D., Gauly, M., 2019. Assessment of welfare indicators in dairy
454 farms offering pasture at differing levels. *Animal* 13, 2336–2347.
455 <https://doi.org/10.1017/S1751731119000570>
- 456 Aubé, L., Mialon, M.M., Mollaret, E., Mounier, L., Veissier, I., de Boyer des Roches, A., 2022.
457 Review: Assessment of dairy cow welfare at pasture: measures available, gaps to address,
458 and pathways to development of ad-hoc protocols. *animal* 16, 100597.
459 <https://doi.org/10.1016/j.animal.2022.100597>
- 460 Aubé, L., Mollaret, E., Mialon, M.M., Mounier, L., Veissier, I., de Boyer des Roches, A., 2023.
461 Measuring the human–animal relationship in cows by avoidance distance at pasture.
462 *Applied Animal Behaviour Science* 265, 105999.
463 <https://doi.org/10.1016/j.applanim.2023.105999>
- 464 Battini, M., Renna, M., Giammarino, M., Battaglini, L., Mattiello, S., 2021. Feasibility and Reliability
465 of the AWIN Welfare Assessment Protocol for Dairy Goats in Semi-extensive Farming
466 Conditions. *Frontiers in Veterinary Science* 8, 1210.
467 <https://doi.org/10.3389/fvets.2021.731927>

468 Bokkers, E.A.M., Vries, M., Antonissen, I., Boer, I.J.M., 2012. Inter- and intra-observer reliability of
469 experienced and inexperienced observers for the Qualitative Behaviour Assessment in
470 dairy cattle. *Animal Welfare* 21, 307–318. <https://doi.org/10.7120/09627286.21.3.307>

471 Brscic, M., Dam Otten, N., Contiero, B., Kirchner, M.K., 2019. Investigation of a standardized
472 qualitative behaviour assessment and exploration of potential influencing factors on the
473 emotional state of dairy calves. *Animals* 9, 757.

474 Ceballos, M.C., Góis, K.C.R., Sant’Anna, A.C., Wemelsfelder, F., da Costa, M.P., 2021. Reliability of
475 qualitative behavior assessment (QBA) versus methods with predefined behavioral
476 categories to evaluate maternal protective behavior in dairy cows. *Applied Animal
477 Behaviour Science* 236, 105263.

478 Chessel, D., Dufour, A.B., Thioulouse, J., 2004. The ade4 package-I-One-table methods. *R news* 4,
479 5–10.

480 Consortium, W.Q., 2009. Welfare Quality® assessment protocol for cattle. Welfare Quality
481 Consortium, Lelystad, the Netherlands.

482 Cooke, A.S., Mullan, S., Morten, C., Hockenull, J., Le-Grice, P., Cocq, K.L., Lee, M.R.F., Cardenas,
483 L.M., Rivero, M.J., 2023. Comparison of the welfare of beef cattle in housed and grazing
484 systems: hormones, health and behaviour. *The Journal of Agricultural Science* 161, 450–
485 463. <https://doi.org/10.1017/S0021859623000357>

486 Czycholl, I., Kniese, C., Büttner, K., Beilage, E. grosse, Schrader, L., Krieter, J., 2016. Interobserver
487 reliability of the ‘Welfare Quality® Animal Welfare Assessment Protocol for Growing Pigs.’
488 *SpringerPlus* 5, 1114. <https://doi.org/10.1186/s40064-016-2785-1>

489 de Boyer des Roches, A., Lussert, A., Faure, Marion., Herry, Vincent., Rainard, Pascal., Durand,
490 Denys., Wemelsfelder, F., Foucras, G., 2018. Dairy cows under experimentally-induced

491 *Escherichia coli* mastitis show negative emotional states assessed through Qualitative
492 Behaviour Assessment. *Applied Animal Behaviour Science* 206, 1–11.
493 <https://doi.org/10.1016/j.applanim.2018.06.004>

494 Ebinghaus, A., Ivemeyer, S., Rupp, J., Knierim, U., 2016. Identification and development of
495 measures suitable as potential breeding traits regarding dairy cows' reactivity towards
496 humans. *Applied Animal Behaviour Science* 185, 30–38.
497 <https://doi.org/10.1016/j.applanim.2016.09.010>

498 Ellingsen, K., Coleman, G.J., Lund, V., Mejdell, C.M., 2014. Using qualitative behaviour assessment
499 to explore the link between stockperson behaviour and dairy calf behaviour. *Applied*
500 *Animal Behaviour Science* 153, 10–17. <https://doi.org/10.1016/j.applanim.2014.01.011>

501 Fleming, P.A., Clarke, T., Wickham, S.L., Stockman, C.A., Barnes, A.L., Collins, T., Miller, D.W., 2016.
502 The contribution of qualitative behavioural assessment to appraisal of livestock welfare.
503 *Anim. Prod. Sci.* 56, 1569–1578. <https://doi.org/10.1071/AN15101>

504 Forkman, B., Keeling, L.J., 2009. Assessment of animal welfare measures for dairy cattle, beef bulls
505 and veal calves (Welfare Quality Report No. 11). School of City and Regional Planning,
506 Cardiff University, Cardiff, UK.

507 Gamer, M., Lemon, J., Singh, I.F.P., 2019. irr: Various Coefficients of Interrater Reliability and
508 Agreement. R package version 0.84.1.

509 Grosso, L., Battini, M., Wemelsfelder, F., Barbieri, S., Minero, M., Dalla Costa, E., Mattiello, S.,
510 2016. On-farm Qualitative Behaviour Assessment of dairy goats in different housing
511 conditions. *Applied Animal Behaviour Science* 180, 51–57.

512 Gutmann, A.K., Schwed, B., Tremetsberger, L., Winckler, C., 2015. Intra-day variation of
513 Qualitative Behaviour Assessment outcomes in dairy cattle. *Animal Welfare* 24, 319–326.
514 <https://doi.org/10.7120/09627286.24.3.319>

515 Hernandez, A., Berg, C., Eriksson, S., Edstam, L., Orihuela, A., Leon, H., Galina, C., 2017. The
516 Welfare Quality® assessment protocol: how can it be adapted to family farming dual
517 purpose cattle raised under extensive systems in tropical conditions? *Animal Welfare* 26,
518 177–184. <https://doi.org/10.7120/09627286.26.2.177>

519 Ibach, S., Chou, J.-Y., Battini, M., Parsons, T.D., 2024. A systematic approach to defining and
520 verifying descriptors used in the Qualitative Behavioural Assessment of sows. *Animal*
521 *Welfare* 33, e8. <https://doi.org/10.1017/awf.2024.6>

522 Knierim, U., Winckler, C., Mounier, L., Veissier, I., 2021. Developing effective welfare measures
523 for cattle. pp. 81–102. <https://doi.org/10.19103/AS.2020.0084.05>

524 Minero, M., Dalla Costa, E., Dai, F., Canali, E., Barbieri, S., Zanella, A., Pascuzzo, R., Wemelsfelder,
525 F., 2018. Using qualitative behaviour assessment (QBA) to explore the emotional state of
526 horses and its association with human-animal relationship. *Applied Animal Behaviour*
527 *Science* 204, 53–59. <https://doi.org/10.1016/j.applanim.2018.04.008>

528 Minero, M., Dalla Costa, E., Dai, F., Murray, L.A.M., Canali, E., Wemelsfelder, F., 2016. Use of
529 Qualitative Behaviour Assessment as an indicator of welfare in donkeys. *Applied Animal*
530 *Behaviour Science* 174, 147–153. <https://doi.org/10.1016/j.applanim.2015.10.010>

531 Phythian, C., Michalopoulou, E., Duncan, J., Wemelsfelder, F., 2013. Inter-observer reliability of
532 Qualitative Behavioural Assessments of sheep. *Applied Animal Behaviour Science* 144,
533 73–79. <https://doi.org/10.1016/j.applanim.2012.11.011>

534 Phythian, C.J., Michalopoulou, Eleni., Cripps, P.J., Duncan, J.S., Wemelsfelder, Françoise., 2016.
535 On-farm qualitative behaviour assessment in sheep: Repeated measurements across
536 time, and association with physical indicators of flock health and welfare. *Applied Animal
537 Behaviour Science*, Behavioural indicators of health 175, 23–31.
538 <https://doi.org/10.1016/j.applanim.2015.11.013>

539 Prescott, N.B., Mottram, T.T., Webster, A.J.F., 1998. Relative motivations of dairy cows to be
540 milked or fed in a Y-maze and an automatic milking system. *Applied Animal Behaviour
541 Science* 57, 23–33. [https://doi.org/10.1016/S0168-1591\(97\)00112-3](https://doi.org/10.1016/S0168-1591(97)00112-3)

542 R Core Team, 2022. R: A language and environment for statistical computing.

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

546 Russell, A., Randall, L., Kaler, J., Eyre, N., Green, M., 2023. Use of qualitative behavioural
547 assessment to investigate affective states of housed dairy cows under different
548 environmental conditions. *Frontiers in veterinary science* 10, 1099170.
549 <https://doi.org/10.3389/fvets.2023.1099170>

550 Skovlund, C.R., Kirchner, M.K., Contiero, B., Ellegaard, S., Manteca, X., Stelvig, M., Tallo-Parra, O.,
551 Forkman, B., 2023. Qualitative Behaviour Assessment for zoo-housed polar bears (*Ursus
552 maritimus*): Intra- and inter-day consistency and association to other indicators of
553 welfare. *Applied Animal Behaviour Science* 263, 105942.
554 <https://doi.org/10.1016/j.applanim.2023.105942>

555 Stockman, C.A., Collins, T., Barnes, A.L., Miller, D., Wickham, S.L., Beatty, D.T., Blache, D.,
556 Wemelsfelder, F., Fleming, P.A., 2011. Qualitative behavioural assessment and

557 quantitative physiological measurement of cattle naïve and habituated to road transport.
558 Animal Production Science 51, 240–249.

559 Stubbsjøn, S.M., Moe, R.O., Bruland, K., Lien, T., Muri, K., 2020. Reliability of observer ratings:
560 Qualitative behaviour assessments of shelter dogs using a fixed list of descriptors.
561 Veterinary and Animal Science 10, 100145.

562 Von Keyserlingk, M.A., Cestari, A.A., Franks, B., Fregonesi, J.A., Weary, D.M., 2017. Dairy cows
563 value access to pasture as highly as fresh feed. Scientific reports 7, 44953.

564 Wagner, K., Brinkmann, J., March, S., Hinterstoißer, P., Warnecke, S., Schüler, M., Paulsen, H.M.,
565 2018. Impact of daily grazing time on dairy cow welfare—Results of the Welfare Quality®
566 protocol. Animals 8, 1.

567 Wemelsfelder, F., 1997. The scientific validity of subjective concepts in models of animal welfare.
568 Applied Animal Behaviour Science, Basic and Applied Aspects of Motivation and Cognition
569 53, 75–88. [https://doi.org/10.1016/S0168-1591\(96\)01152-5](https://doi.org/10.1016/S0168-1591(96)01152-5)

570 Wemelsfelder, F., Hunter, T.E., Mendl, M.T., Lawrence, A.B., 2001. Assessing the ‘whole animal’:
571 a free choice profiling approach. Animal Behaviour 62, 209–220.

572 Wiese, T.R., Rey Planellas, S., Betancor, M., Haskell, M., Jarvis, S., Davie, A., Wemelsfelder, F.,
573 Turnbull, J.F., 2023. Qualitative Behavioural Assessment as a welfare indicator for farmed
574 Atlantic salmon (*Salmo salar*) in response to a stressful challenge. Front. Vet. Sci. 10.
575 <https://doi.org/10.3389/fvets.2023.1260090>

576 Winckler, C., 2014. Inter-observer agreement for qualitative behaviour assessment in dairy cattle
577 in three different countries, in: Proceedings of the 6th International Conference on the
578 Assessment of Animal Welfare at the Farm and Group Level. Wageningen Academic, pp.
579 181–181. https://doi.org/10.3920/9789086867981_149

Preprint not peer reviewed

582 **Table 1:** Weather conditions, pasture conditions, observers, hour of observation of qualitative behaviour assessment (QBA) performed on a herd grazing herd (n
 583 = 144 dairy cows) and handlers who collect the cows on the pasture plot during the handling context for each observation.

visit	Date	Weather conditions on the test day						Pasture conditions		Context of observation										
		Average temperature (°C)	Minimum temperature (°C)	Maximum temperature (°C)	Rainfall (mm)	Solar radiation (Joule/cm ²)	Average wind (m/s)	Average humidity (%)	Days in the pasture plot	Pasture access	AM ⁴	PM ⁵	Handling ⁶	Observers	Hour	Observers	Hour	Observers	Hour	Handlers
											Observers	Hour	Observers	Hour	Observers	Hour	Observers	Hour	Handlers	
1	15-Apr.	4.3	-1.5	10	0	1937	2.2	70	4	Day only ¹	A and B	10:35	A and B	15:17	A and B	15:54				
2	18-May.	11.7	7.7	16.5	5.5	2200	1.9	75	2	Day and night			A and B	15:30	A and B	16:10	F and G			
2	20-May.	10.8	2.5	17.1	0	1859	2.4	77	4	Day and night	A and B	11:08								
3	21-Jun.	16.5	13.3	22.2	29	1910	1.1	86	4	Day and night			A and E	15:33	A and E	15:45	F and H			
3	24-Jun.	14.3	8.9	20.1	0.5	2937	1.3	77	1	After morning milking ²			A and E	15:34	A and E	15:42	F and I			
3	25-Jun.	14.6	9.6	17.6	10.5	994	1.3	91	2	Day and night	A and E	10:16								
4	28-Jul.	16.5	12.5	21.5	0.5	1823	2	77	5	Day and night	A and D	10:57	A and D	15:29	A and D	15:47	F and J			
5	30-Aug.	15.8	11.3	19.5	0	858	1.9	87	3	Day and night					A and D	15:48	G and J			
5	31-Aug.	16.6	13.7	21.3	0	1859	2.2	76	4	Day and night	A and E	11:11	A and E	15:30	A and E	15:55	F and K			
5	2-Sep.	16.8	10.6	23.6	0	2275	2.1	73	6	Day and night	A and E	10:59	A and E	15:24	A and E	15:48	F and L			
5	3-Sep.	18.1	10.6	25.4	0	1929	0.8	77	7	Day and night					A and D	15:50	H, G and J			
6	4-Oct.	11.9	7	17.6	8.5	1248	1.8	89	5	After morning milking ²					A and D	15:50	M and J			
6	5-Oct.	12.1	10.8	15.6	3	1349	2.7	85	6	Day and night					A and D	15:58	H and N			
6	6-Oct.	11.1	8.4	16.7	0	1100	1.5	84	7	Day and night	A and B	10:50	A and B	15:32	A and B	15:54	M and J			
6	7-Oct.	11.4	5.4	16.6	1	773	0.4	90	8	Day and night					A and B	16:19	H and N			
7	2-Nov.	7.2	4.7	10.1	0	720	1	93	2	Day and night					A and D	16:00	M and L			
7	3-Nov.	6.7	2.6	11.6	5.5	875	1	92	3	Day and night					A and D	15:54	F and O			
7	4-Nov.	6.8	3.8	9.9	9	590	1.4	96	4	Day and night	A and E	10:40	A and E	15:33	A and E	15:48	I and M			

¹ Pasture access between morning and evening milking only

² Cows spent night indoor and were released on pasture after morning milking

³ At pasture, right after coming back to pasture from morning milking (AM context, n = 8)

⁴ At pasture, before the afternoon milking (PM context, n = 9)

⁵ During the gathering of cows at pasture for afternoon milking (handling context, n = 16)

584

585 **Table 2:** Ethogram used to describe cows' responses to handling (quantitative behaviour assessment)

Response	Description	Reference
Slipping	Lowering of the animal's body by slipping or folding of the leg/legs, resulting in loss of balance, with no part of the body other than the hooves and/or legs in contact with the floor.	Welfare Quality (2009) - Fattening cattle at slaughterhouse
Falling	Toss of balance in which parts of the body other than feet and legs are in contact with floor surface.	Welfare Quality ((2009) - Fattening cattle at slaughterhouse
Moving backwards	The animal moves backwards as a reaction to handling. Moves at least two steps backwards. When an animal takes a few steps backwards to achieve balance or changes position in relation to other animals when crowding it is not considered as moving backwards	Welfare Quality (2009) - Fattening cattle at slaughterhouse; Hultgren et al. (2014)
Turning around	The animal turns around, by itself or as a reaction to the handling regime. When/if the animal turns back again to the former direction, the behaviour should not be recorded again.	Welfare Quality ((2009) - Fattening cattle at slaughterhouse
Vocalizing	The cow makes vocal sound	Grandin (1998) ; Grandin (2001) ; Hultgren et al. (2014)
Galloping	The cow is running with both fore and hindlegs placed forward alternately	Bokkers et al 2008
Trotting	The cow is running with the left foreleg and right hindleg are placed forward simultaneously, followed by the right foreleg and left hindleg etc.	Bokkers et al 2009
Kicking	Kicks with hind leg towards stockperson	Hultgren et al. (2014)
Freezing	The route is free in front or behind the animal but the animal refuses to move forwards or backwards within 4 seconds from being touched/coerced by the handler. If the animal takes more than one step and stops again, or moves backwards, a 'freeze' is recorded again when a new driving attempt is made. An animal that stops but continues to walk when the handler drives it forwards is not frozen.	Welfare Quality (2009) - Fattening cattle at slaughterhouse; Hultgren et al. (2014)
Startle	The animal flinches, jumps or bucks in response to stimulus.	Gibbons et al (2009)

586

587 **Table 3:** Measure on the scale in mm for each adjective for QBA performed on video of dairy cows at
 588 pasture, loading on PC1 and PC2, weights from the Welfare Quality protocol for dairy cows and weight from
 589 the present study, rank of each adjectives in Welfare Quality protocol and from the present study at pasture
 590 (according to their absolute weights) (n = 399).

variable	Pasture context				Welfare Quality					
	Measure on the scale (mm)				Loadings on PC1	Loadings on PC2	Weights (Estimate)	Rank*	Weights (Estimate)	Rank
	Mean	SD	min	max						
active	71	29.8	1	123	0.1	0.84	0.0014	18	0.0077	16
relaxed	66	34.2	1	125	<i>0.52</i>	<i>-0.52</i>	0.0061	12	0.01	14
fearful	7	8.9	0	95	<i>-0.35</i>	<i>0.42</i>	-0.0161	5	-0.0129	7
agitated	35	33	0	123	-0.49	0.74	-0.0061	13	-0.0162	3
calm	78	34.1	0	122	0.59	-0.63	0.007	11	0.0088	15
content	86	26.2	1	124	0.82	0.2	0.0127	6	0.0121	8
indifferent	36	25.4	0	116	<i>0.24</i>	<i>-0.46</i>	0.0038	14	-0.0112	11
frustrated	12	16.7	0	115	-0.83	-0.11	-0.0201	2	-0.0161	4
friendly	31	23.4	1	101	-0.14	0.63	-0.0024	16	0.0117	10
bored	13	18.2	0	120	<i>-0.48</i>	<i>-0.4</i>	-0.0108	8	-0.0109	12
playful	19	24.3	0	116	-0.05	0.75	-0.0008	20	0.0011	18
positively occupied	82	30.6	1	125	0.74	0.28	0.0099	10	0.0118	9
lively	61	35.7	1	123	0.28	0.81	0.0032	15	0.0003	20
inquisitive	41	34.4	0	124	0.08	0.78	0.0009	19	0.0005	19
irritable	14	18.4	0	120	-0.87	-0.04	-0.0191	3	-0.0218	1
uneasy	12	18.6	0	120	-0.83	-0.03	-0.0181	4	-0.0103	13
sociable	30	24.2	0	106	-0.13	0.67	-0.0022	17	0.0053	17
apathetic	9	10.2	0	98	<i>-0.26</i>	<i>-0.3</i>	-0.0103	9	-0.0156	5
happy	87	26.5	0	122	0.8	0.21	0.0122	7	0.0147	6
distressed	6	12.7	0	117	-0.79	-0.17	-0.0253	1	-0.0203	2

*rank determined from absolute value of weight (1 being the adjective with the maximal absolute weight)

In bold, absolute loadings > 0.6

In italics, absolute loadings never > 0.6 on both axe 1 and axe 2

In grey the ranks that differ of more than 2 ranks between WELFARE QUALITY and pasture ranking

591

592

593 **Table 4:** Intra- and inter-observer reliability of coordinates on PC1 and PC2 obtained from video
 594 observations (W, Kendall's concordance coefficient; ICC, intraclass correlation coefficient).

	PCA axis 1 (PC1)				PCA axis 2 (PC2)				
	Session 1		Session 2		Session 1		Session 2		
	ICC	Kendall	ICC	Kendall	ICC	Kendall	ICC	Kendall	
intra-observer reliability									
Observer									
A	0.81	0.84	0.82	0.86	0.96	0.96	0.90	0.96	
B	0.88	0.90	0.91	0.90	0.89	0.93	0.92	0.92	
C	0.73	0.68	0.93	0.91	0.86	0.90	0.88	0.94	
D	0.61	0.66	0.67	0.76	0.74	0.86	0.86	0.90	
E	0.75	0.69	0.95	0.89	0.90	0.94	0.89	0.94	
Inter-observer reliability									
Day									
1	0.47	0.56	0.62	0.62	0.72	0.89	0.76	0.89	
2	0.52	0.52	0.59	0.67	0.68	0.88	0.72	0.88	

595

596

597 **Table 5:** Results of the linear mixed models (with estimated coefficients and associated statistics) testing
 598 the effect of the context of direct observations on the coordinates on axis 1 (PC1) and axis 2 (PC2) of the
 599 Principal Component Analysis. Observers were included as a random effect.

Model for PC1*			Confidence interval		
Fixed effects: coefficients of the model	Estimate	SE	2.50%	97.50%	t value
Handling context (Intercept)	-2.1	0.43	-3.0	-1.1	-4.8
AM pasture context	4.1	0.30	3.6	4.8	13.6
PM pasture context	3.5	0.29	3.0	4.1	12.0
Random effects: standard deviations	Variance	SD			
Observer	0.61	0.78	0.24	1.75	
Residuals	0.95	0.97	0.81	1.16	

Model for PC2**			Confidence interval		
Fixed effects: coefficients of the model	Estimate	SE	2.50%	97.50%	t value
PM pasture context (Intercept)	-1.7	0.35	-2.4	-1.0	-4.8
Handling context	2.0	0.35	1.3	2.7	5.8
AM pasture context	2.1	0.40	1.3	2.8	5.2
Random effects: standard deviations	Variance	SD			
Observer	0.16	0.41	0.00	1.05	
Residuals	1.35	1.16	0.97	1.39	

*R Formula: lmer(PC1 ~ context+ (1|observer))

**R Formula: lmer(PC2 ~ context+ (1|observer))

SE, standard error; SD, standard deviation

600

601 **FIGURE CAPTIONS**

602 **Fig. 1.** Two-dimensional plot of the Principal Component Analysis (PCA) showing the loading of all 20
603 adjectives on PC1 (Valence) and PC2 (Arousal) from Qualitative Behaviour Assessment of dairy cows at
604 pasture observed on video (n = 399 observations in total, 20 videos watched by n = 5 observers in 4
605 sessions).

606 **Fig. 1.** Coordinates on PC1 and PC2 of the QBA from live observations performed by observer A throughout
607 the grazing season. Dots shape represents the context of observation (□: AM observation at pasture; ×: PM
608 observations at pasture; ▽: observation during herd handling).

609