

Scotland's Rural College

## **On-farm qualitative behaviour assessment in sheep: repeated measurements across time, and association with physical indicators of flock health and welfare**

Phythian, CJ; Michalopoulou, E; Cripps, PJ; Duncan, JS; Wemelsfelder, F

*Published in:*  
Applied Animal Behaviour Science

*DOI:*  
[10.1016/j.applanim.2015.11.013](https://doi.org/10.1016/j.applanim.2015.11.013)

First published: 28/11/2015

*Document Version*  
Peer reviewed version

[Link to publication](#)

### *Citation for published version (APA):*

Phythian, CJ., Michalopoulou, E., Cripps, PJ., Duncan, JS., & Wemelsfelder, F. (2015). 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.  
<https://doi.org/10.1016/j.applanim.2015.11.013>

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

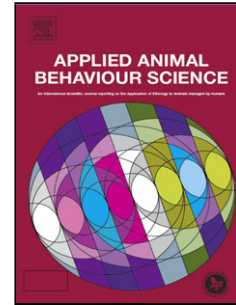
### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

## Accepted Manuscript

Title: On-farm qualitative behaviour assessment in sheep: repeated measurements across time, and association with physical indicators of flock health and welfare

Author: Clare J. Phythian Eleni. Michalopoulou Peter J. Cripps Jennifer S. Duncan Françoise. Wemelsfelder



PII: S0168-1591(15)00317-2  
DOI: <http://dx.doi.org/doi:10.1016/j.applanim.2015.11.013>  
Reference: APPLAN 4162

To appear in: *APPLAN*

Received date: 17-5-2014  
Revised date: 7-10-2015  
Accepted date: 14-11-2015

Please cite this article as: Phythian, C.J., Michalopoulou, Ei., Cripps, P.J., Duncan, J.S., Wemelsfelder, Fe., 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* (2015), <http://dx.doi.org/10.1016/j.applanim.2015.11.013>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

## Highlights

- QBA assessments of sheep flocks over a year were compared to flock health measures
- Two dimensions of sheep expression, summarised as ‘mood’ and ‘responsiveness’
- Flock scores on both dimensions showed high consistency across the year
- Flock ‘mood’ scores correlated to flock lameness and ‘dull physical demeanour’
- Results support QBA as a meaningful, complementary sheep flock welfare indicator

Accepted Manuscript

1 **On-farm qualitative behaviour assessment in sheep: repeated measurements across time, and**  
2 **association with physical indicators of flock health and welfare**

3

4 Phythian, Clare J.<sup>1a</sup>, Michalopoulou, Eleni.<sup>2</sup>, Cripps, Peter J.<sup>2</sup>, Duncan, Jennifer S.<sup>2</sup>, and  
5 Wemelsfelder, Françoise.<sup>3</sup>

6 <sup>1</sup> Department of Epidemiology and Population Health, University of Liverpool, Leahurst, CH64  
7 7TE, United Kingdom

8 <sup>2</sup> School of Veterinary Science, University of Liverpool, Leahurst CH64 7TE, United Kingdom

9 <sup>3</sup> Animal and Veterinary Sciences Group, SRUC, Roslin Institute Building, Bush Estate, EH25  
10 9RG, United Kingdom

11

12 Corresponding author: [clare.phythian@nmbu.no](mailto:clare.phythian@nmbu.no)

13 <sup>a</sup> Current address: Section for Small Ruminant Research, Department of Production Clinical  
14 Science, NMBU, 4325 Sandnes, Norway

15 Telephone: +47 51603510

16 **Abstract**

17 Qualitative Behavioural Assessment (QBA) is a ‘whole-animal’ methodology that assesses the  
18 expressive qualities of animal behaviour using terms such as ‘tense’, ‘relaxed’, ‘anxious’, and  
19 ‘content’. The reliability and validity of QBA as an indicator for on-farm welfare assessment in  
20 pigs, cattle, poultry and sheep has been examined in a number of ways. However, the use of QBA  
21 on farms over longer periods of time has not yet been examined. The aim of this study was to  
22 investigate whether and how on-farm QBA of sheep varies over the different seasons of the year,  
23 and whether it is associated with physical measures of sheep health and welfare such as lameness.  
24 A trained assessor visited each of 12 farms six times within a one year period at two month  
25 intervals, and made group level assessments of approximately 100 sheep selected *ad hoc*  
26 (assuming homogeneity within the flock). The sheep flocks were assessed with a list of twelve  
27 QBA descriptive terms previously developed for sheep. Following QBA, the same sheep were also  
28 assessed with seven physical indicators of health and welfare (‘dull physical demeanour’,  
29 lameness, breech and abdominal soiling, pruritis, wool loss, and coughing). QBA scores from all  
30 visits were analysed together, and also in combination with the physical measures, with Principal  
31 Component Analysis (PCA - correlation matrix, no rotation). The effect of visit on PCA flock  
32 scores was analysed with random-effects multiple linear regression models. The association  
33 between PCA flock scores and physical measures was investigated using Spearman rank  
34 correlation ( $r_s$ ), and the correlation of flock rankings across visits was examined with Kendall  
35 Coefficient of Concordance. PCA distinguished two main dimensions of sheep expression: PC1  
36 (47% variation) ranging from content/relaxed/thriving to distressed/dull/dejected (summarised as  
37 ‘mood’) and PC2 (21%), which ranged from anxious/agitated/responsive to relaxed/dejected/dull  
38 (summarised as ‘responsiveness’). No significant effect of visit on PC1 scores was found

39 (p=0.155), and PC1 flock scores correlated at  $W=0.84$  ( $p<0.001$ ) across the 6 visits, indicating high  
40 consistency of characterisations of individual flock mood over the year. However there was an  
41 effect of visit on PC2 scores ( $p<0.001$ ), and PC2 flock scores were correlated at  $W=0.60$  ( $p<0.001$ )  
42 across visits, indicating that the presence of young lambs may have had a consistently relaxing  
43 effect on flocks. There was also an effect of visit period on lameness ( $p=0.025$ ), and on breech  
44 ( $p<0.001$ ) and abdominal ( $p=0.0048$ ) soiling. With the exception of lameness and breech and  
45 abdominal soiling, the physical indicators were observed at a low prevalence ( $<2\%$ ) across the  
46 study farms. The highest lameness levels were observed during the winter period (mean 17.86%,  
47 95% CI 7.83 – 27.90) whilst breech soiling was highest in spring (mean 23.83%, 95% CI 11.86 –  
48 35.81). An effect of farm type was found on lameness scores ( $p=0.0176$ ) and an effect of flock size  
49 on abdominal soiling scores ( $p=0.025$ ). PC1 ‘mood’ scores were negatively correlated to the  
50 proportion of lame sheep ( $n=72$ ;  $rS=-0.72$ ,  $p<0.001$ ), and to the proportion of animals with dull  
51 physical demeanour ( $rS=-0.70$ ,  $p<0.001$ ), while PC2 ‘responsiveness’ scores showed a weak  
52 correlation with breech soiling ( $rS=0.42$ ,  $p<0.001$ ). In summary, these results suggest that QBA  
53 has the potential to serve as a sensitive, meaningful indicator for on-farm welfare assessment in  
54 sheep.

55

56 **Keywords:** Qualitative Behaviour Assessment; animal welfare; consistency; animal-based  
57 outcomes; sheep.

## 58 1. Introduction

59 Qualitative Behavioural Assessment (QBA) is a ‘whole-animal’ methodology that assesses the  
60 expressive qualities of animal behaviour using terms such as ‘tense’, ‘relaxed’, ‘anxious’, and  
61 ‘content’. Thus it addresses an animal’s ‘body language’, including both negative and positive  
62 aspects of well-being, and has the potential to integrate and help interpret specific clinical measures  
63 of physical and psychological health (Napolitano *et al.*, 2009; Wemelsfelder & Lawrence, 2001;  
64 Wiseman-Orr *et al.*, 2006). This methodology has been applied to assess animals on-farm and  
65 during transport both individually and at group-level, with different livestock species such as pigs,  
66 cattle, poultry and sheep (e.g. Bassler *et al.*, 2013; Rousing and Wemelsfelder, 2006; Stockman *et*  
67 *al.*, 2011; Temple *et al.*, 2011; Wickham *et al.*, 2012). Generally good levels of inter-observer  
68 reliability (but not always, see Bokkers *et al.*, 2012), meaningful associations with other measures  
69 (but not always, see Andreasen *et al.*, 2013), as well as short assessment times, suggest this method  
70 has the potential to be an effective welfare indicator that can be readily applied in the field.

71  
72 In common with other global pasture-based production systems, sheep managed under British  
73 farming systems spend a considerable part of the production cycle outdoors at pasture being kept in  
74 specific management groups. Therefore, groups of sheep often require gathering and handling to  
75 facilitate close inspection and assessment of the health and welfare of both the individual sheep  
76 and the flock. Since disturbance by humans, dogs and handling can alter ovine behavioural  
77 expression (Boivin *et al.*, 2000, Le Neindre *et al.*, 1996) and mask painful conditions (Fitzpatrick  
78 *et al.*, 2006), it is possible that some sheep with welfare issues may be missed when gathered for  
79 closer examination. Furthermore, the practicalities of assessment need to be considered for  
80 different management systems. The gathering and handling of extensively-managed sheep and

81 those managed over multiple locations can be time and labour consuming and also may not be  
82 appropriate at certain periods of the production calendar, for example, when ewes have young  
83 lambs at foot or during the mating period. Therefore, a welfare indicator that does not involve  
84 major disturbance, requires few resources, and offers valid information on the health and wellbeing  
85 of groups of animals, could offer clear benefits for sheep, producers and assessors.

86

87 One major concern in the development of on-farm welfare assessment protocols is the challenge of  
88 interpreting fluctuations shown by welfare indicators across time. Such fluctuations may be part of  
89 normal day-to-day or seasonal variations in welfare, may reflect more serious deviations of basic  
90 welfare, or could reflect the effects of varying times and contexts on repeat assessments. Thus, if  
91 repeated assessments of the same farm do not show similar levels of animal demeanour, it is  
92 difficult to know whether this difference reflects normal baseline variation, a welfare problem, or a  
93 problem of intra-observer reliability (Temple *et al.*, 2013). The aim of this study was to apply QBA  
94 to the repeated assessment of sheep at flock level in a one-year longitudinal study, to investigate  
95 whether and how the sheep's expressive demeanour would be perceived by an experienced  
96 assessor to vary across 6 visits at two-monthly intervals. To evaluate these assessments against  
97 other welfare indicators, seven physical measures of sheep health and welfare were also examined.

98

## 99 **2. Materials and methods**

100

### 101 **2.1 Design of longitudinal study**

102 A longitudinal on-farm study performed during the period of May 2009 – April 2010 was  
103 conducted on twelve farms, located in North-West England and North Wales, which had



104 previously participated in a sheep welfare research project. Farms were selected according to their  
105 location, farm type and owner's informed consent to participate. Selection provided a sample of  
106 eleven commercial flocks and one small-holding, including farms from hill, upland and lowland  
107 areas (for details see Table 1). At each visit, each farm was asked to provide a sample of 70-100  
108 sheep that were selected *ad hoc* by the farmer and left undisturbed for assessment. This sample size  
109 was not related to a farm's flock size, but was based on previous experience of the assessor  
110 regarding the feasibility of completing the protocol of qualitative and quantitative assessments  
111 within the time limits of a day visit. The exact numbers of sheep selected at each farm for each  
112 visit were recorded. The study was approved by the University of Liverpool Ethics Committee  
113 (ethical review reference number RETH000287).

114  
115 During the one year study, flocks were repeatedly assessed by one sheep veterinary surgeon who  
116 performed all QBA and physical indicator assessments on all farms throughout the study. Repeated  
117 sampling of twelve sheep flocks over 6 visits spread out over one year produced 72 on-farm  
118 assessments. Flocks were visited at an interval of approximately 60 days, to coincide with key  
119 periods in the sheep production cycle (Table 2). At each visit, the selected group of sheep was  
120 firstly assessed using twelve QBA descriptors (relaxed, dejected, thriving, agitated, responsive,  
121 dull, content, anxious, bright, tense, vigorous and distressed), which had previously been  
122 developed and tested for inter-observer reliability by Phythian *et al.* (2013a). Due to their  
123 integrative, qualitative nature, it is impossible to define QBA terms in precise physical terms such  
124 as is done for conventional ethograms (however very recently QBA studies have begun to provide  
125 brief qualitative characterisations of individual terms to enhance observer agreement). Detailed  
126 instructions for how to score QBA terms were developed for the Welfare Quality® protocols for

127 cattle, pigs and poultry (Welfare Quality®, 2009), including careful reflection on, and, where more  
128 than one assessor are involved, discussion of, the meaning of individual terms. These instructions  
129 were followed in the present study.

130

131 The assessor quietly approached the sample group and performed assessments from a distance by  
132 standing at the boundary of a field, or several metres from groups of housed animals. The exact  
133 sizes of fields and assessment areas were not measured, but a number of observation points was  
134 selected according to the relative size of the field and sample group, after which a 5 minute period  
135 was allowed to let sheep get accustomed to the presence of the assessor. The mean number of  
136 sheep assessed in any one group was 77, and ranged from a minimum group size of 24, which  
137 represented all the flock of a small-holding farm, to a maximum group size of 137 animals on a  
138 commercial farm. Minimal disturbances of the sheep by assessor movements, particularly in  
139 situations where scrutiny of individual animals was difficult, were found to be helpful and  
140 considered acceptable. The observer then spent 5 minutes at each of the observation points,  
141 visually scanning the designated observation area to assess the entire sample group of sheep. When  
142 observations were completed, the groups' predominant behavioural expressions were scored on  
143 each of the QBA terms along a visual analogue scale (VAS) of 125 mm length, labelled from  
144 'zero' to 'maximum' expression. This entire process of QBA assessment, of up to 120 sheep, took  
145 on average about 30 minutes per farm.

146

147 Following completion of QBA, seven additional physical indicators of sheep health and welfare  
148 were assessed at group-level by the same assessor. Whilst these physical measures were taken in  
149 the same observation area as QBA, the exact observation points from which they were made

150 differed. The group of sheep was briefly observed at a distance for five minutes, and then the  
151 assessor entered the assessment area to count the number of animals observed to be affected by the  
152 following physical indicators (as described in Phythian et al., 2012): coughing (defined as  
153 observation of one, or a combination, of the following signs: paroxysmal coughing, and respiratory  
154 distress including abdominal effort associated with breathing or wheezing), lameness (any, or a  
155 combination, of the following signs: ‘nodding’ of head in unison with short stride, grazing on  
156 knees, uneven gait, arching of back during locomotion, non-weight bearing on affected limb when  
157 standing, extreme difficulty rising, and reluctance to move once standing, as described in Kaler *et*  
158 *al.*, 2009), breech soiling (discrete/solid plaques or more diffusely soiled areas of contamination by  
159 faecal matter, mud or soil of the perineum and/or tailhead, and/or superficial gluteal region, and/or  
160 caudal aspect of the hindlimb(s) as far as the hock), abdominal soiling (discrete/solid plaques or  
161 more diffusely soiled areas of contamination by faecal matter, and/or mud, and/or soil over the  
162 ventral abdomen), pruritis (one, or a combination, of the following signs: rubbing or scratching  
163 against walls/posts/fences/other objects, restlessness, stamping of feet, biting and nibbling of own  
164 body), wool loss (observation of small discrete areas extending to diffuse areas of fleece loss), and  
165 ‘dull physical demeanour’ (defined as “an animal with lowered head carriage, showing behavioural  
166 separation from the rest of the group, and unresponsive to the presence of other sheep or the  
167 observer”).

168

## 169 **2.2 Statistical analysis**

170 QBA data for each farm were recorded by measuring the distance in millimetres between the zero  
171 point of the VAS scale and the mark on the line made on the scale for each term, to provide a value  
172 between 0 and 125. For physical health and welfare indicators, the percentage (%) of sheep in a

173 group showing signs of coughing, wool loss, pruritis, lameness, breech and abdominal soiling, and  
174 ‘dull physical demeanour’, was calculated for each farm assessment.

175

176 QBA data recorded over the 6 visits were analysed together using Principal Component Analysis  
177 (PCA – correlation matrix, no rotation) in Minitab version 16 (Minitab, Inc, State College, PA).  
178 PCA identifies the least number of components that explain most of the variance in the data  
179 (Jolliffe, 2002). QBA PC1 and PC2 accounted for a cumulative variance  $\geq 68\%$ , hence two  
180 components were retained in subsequent analyses. The correlation between each QBA term and  
181 PC1 and PC2 is contained in the loading values, which reflect the weighting of each term within  
182 each component. In addition, a combined PCA analysis (correlation matrix, no rotation) was  
183 performed by analysing data on all 19 variables (12 QBA terms and 7 physical indicators) gathered  
184 over the 6 visits.

185

186 To investigate whether PC1 and PC2 scores for the 12 flocks differed in ranking over the 6 visits,  
187 Kendall coefficient of concordance (W) was calculated (n=12). Correlations of PC1 and PC2 flock  
188 scores with the outcomes of physical health and welfare indicators were examined using  
189 Spearman’s rank correlation coefficient (rS, n=72), and the distributions of QBA PC1, PC2 and  
190 physical indicator scores for each farm over the 6 visits were examined graphically.

191

192 To investigate whether there was a significant effect of visit period on QBA and physical indicator  
193 scores, mixed effects linear regression models were fitted in Stata version 13.1 (StataCorp LP,  
194 College Station, TX). Visit period (n=6; Table 1) was included as a fixed effect, and farm identity  
195 as a random effect. Farm type (categorised for the purposes of analysis as 1. lowland or 2. hill and

196 upland) and flock size (categorised as  $\leq 100$ , 101 – 350, 351 – 650, 651 – 950 and 951 – 1250  
197 sheep) were also included as covariates in the mixed effects models. Visit period (1 to 6) was  
198 included as a repeated measure within-farm. To ensure the robustness of regression models, an  
199 auto-regression correlation was also fitted in the order of 1. Models examined the effect of visit  
200 period on both the QBA PC1 and PC2 scores, and on the combined QBA/physical indicator PC1  
201 and PC2 scores. However, due to paucity of data for several physical indicators only those  
202 indicators observed at a prevalence  $>2\%$  (lameness, and breech and abdominal soiling) were  
203 included. The models' outcomes were described using coefficient  $\beta$  (indicating the magnitude of  
204 the effect), a 95% confidence interval (CI), and Wald p-values (Long and Freese, 2006). To assess  
205 the effect of visit period, the baseline ( $\beta=0$ ) for comparison of coefficient values for each visit  
206 period was set as visit 1 (May-June 2009). Lowland farms and flocks with less than 100 sheep  
207 were set as the baseline values ( $\beta=0$ ) for comparing the effects of farm type and flock size  
208 respectively.

209

### 210 3. Results

211 A total of 5740 sheep (aged  $> 1$  year) and lambs (aged  $> 12$  weeks) were assessed, using QBA and  
212 7 physical indicators of sheep health and welfare. Over the six visits the total number of sheep  
213 presented for assessment on each farm were: farm 1 = 481, farm 2 = 552, farm 3 = 216, farm 4 =  
214 447, farm 5 = 428, farm 6 = 567, farm 7 = 439, farm 8 = 525, farm 9 = 553, farm 10 = 529, farm  
215 11 = 471 and farm 12 = 532. The total number of sheep assessed per visit varied: visit 1: n = 1182,  
216 visit 2; n = 1133, visit 3, n = 990; visit 4, n = 780; visit 5, n = 709; and visit 6, n = 946.

217

#### 218 3.1 QBA and physical health indicator outcomes

219 PCA identified two principal components (PC) which together explained 68% of the variation  
220 between farms (Fig. 1). PC1 (47% variation) ranged from ‘content/relaxed/thriving’ to  
221 ‘distressed/dull/dejected’ (summarised as ‘mood’), while PC2 (21%) ranged from  
222 ‘anxious/agitated/responsive’ to ‘relaxed/dejected/dull’ (summarised as ‘responsiveness’). The  
223 proportion of sheep observed with signs of each physical indicator varied between individual  
224 farms. Across the entire study period, at the level of the individual farm, ‘dull physical  
225 demeanour’ ranged from 0 (minimum) to 15% (maximum), coughing from 0% to 38.55%, wool  
226 loss from 0% to 11.54%, pruritis 0 to 2.88%, lameness 1.23% to 61.86%, whilst breech and  
227 abdominal soiling ranged from 0 to 59.68% and 0 to 100% respectively. In addition, as can be  
228 seen from Table 3, there was seasonal variation between different visits in the mean proportion  
229 of affected sheep for all study farms.

230

### 231 **3.2 Effects of visit period**

232 Overall, regression modelling identified no significant effect of visit periods 1-6 on PC1 flock  
233 scores when QBA scores were evaluated independently ( $p=0.155$ ), nor when analysed together  
234 with physical indicator outcomes ( $p=0.1982$ ). This result indicates that the perceived mood of  
235 sheep flocks was relatively stable across the year. By contrast, there was a significant effect of visit  
236 period on PC2 scores for both the independent QBA PC2 scores ( $p<0.001$ ), and the combined  
237 QBA/physical indicator PC2 scores ( $p<0.001$ ). More detailed results for the random-effects models  
238 are shown in Table 4, which show that the lowest  $\beta$  coefficient values for PC2 scores were  
239 associated with visit 1 (May/June 2009) and visit 6 (March/April 2010), indicating that flocks  
240 appeared relatively more relaxed and less agitated over the lambing and post-lambing period than  
241 at other times of year. The significant correlations between the rankings of flocks on PC1 and PC2

242 across the 6 visits indicate that the relative characterisations of flock expression on the two QBA  
243 dimensions did not significantly change over the year (PC1:  $W=0.84$ ,  $p<0.001$ ; PC2:  $W=0.60$ ,  
244  $p<0.001$ ). As regards physical indicators, there was an effect of visit period on lameness and  
245 breech and abdominal soiling scores. The highest level of lameness (17.86%) was recorded in the  
246 winter months (visit 5), breech soiling (23.83%) was highest at visit 1 in the spring period (Table  
247 4), and the highest levels of abdominal soiling (20.17%) occurred during the autumn/winter period  
248 (visit 4).

249

### 250 **3.3 Effects of farm type and flock size**

251 An effect of farm type was found on QBA PC1 scores in which hill/upland flocks received higher  
252 PC1 scores ( $\beta$  2.47, 95% CI 0.54 – 4.38,  $p=0.017$ ), and were thus perceived as more  
253 ‘content/relaxed/thriving’, compared to the lowland flocks in this sample ( $\beta$  -1.21, 95% CI -2.57–  
254 0.14). There was also an effect of farm type on lameness scores, indicating that hill/upland flocks  
255 showed lower levels of lameness ( $\beta$  -9.70, 95% CI -17.72 – -1.68,  $p=0.0176$ ) than lowland flocks  
256 ( $\beta$  18.32, 95% CI 12.66 – 23.99). There was one effect of flock size: larger flocks with 951–1250  
257 sheep showed higher levels of abdominal soiling ( $\beta$  28.28, 95% CI 1.94 – 29.17  $p=0.025$ ) than  
258 flocks with less than 100 sheep ( $\beta$  0, 95% CI -11.12 – 11.11). No significant interaction effects  
259 were found.

260

### 261 **3.4 Associations between QBA and physical health indicators**

262 Fig. 2 shows PC1 and PC2 of the combined PCA of QBA and physical indicator scores. This graph  
263 illustrates a close alignment of the negative end of PC1 (distressed/dull/dejected) with the  
264 prevalences of lameness ( $rS=-0.72$ ,  $p<0.001$ ), and ‘dull physical demeanour’ ( $rS=-0.70$ ,  $p<0.001$ ).

265 This association is supported by significant correlations of lameness with individual QBA terms  
266 ‘distressed’ ( $rS=0.50$ ,  $p<0.001$ ), ‘dull’ ( $rS=0.57$ ,  $p<0.001$ ), and ‘dejected’ ( $rS=0.57$ ,  $p<0.001$ ), and  
267 by a correlation of ‘dull physical demeanour’ with individual QBA terms ‘distressed’ ( $rS=0.70$ ,  
268  $p<0.001$ ), ‘dull’ ( $rS=0.74$ ,  $p<0.001$ ), and ‘dejected’ ( $rS=0.66$ ,  $p<0.001$ ). In addition there was a  
269 weak but significant correlation between the negative end of PC2 (relaxed/dejected/dull) and  
270 breech soiling ( $rS=0.42$ ,  $p<0.001$ ).

271  
272 Fig. 3 presents a visual image of the distributions of QBA PC1 and PC2 scores, and lameness  
273 and breech soiling percentages, for each of the 12 farms across all 6 visits. This overview allows  
274 closer investigation of the extent to which the PC scores of different farms remained stable, or  
275 shifted up or down the expressive dimensions. Thus some farms (e.g. farm 3 in Fig.3) did not  
276 vary much in their PC1 and PC2 positions over the year, whereas other farms showed more  
277 variation over time (e.g. farm 4). The majority of flocks stayed located on either the positive or  
278 negative side of PC1, and thus appeared to be quite consistent in general mood (as supported by  
279 a high Kendall W for PC1 of 0.84). On PC2 there was more variation for some farms (e.g. farms  
280 4 and 5), reflected in a somewhat lower, but still significant, Kendall W value (0.60). The effect  
281 of farm type on QBA PC1 scores and lameness scores reported above can also be seen in this  
282 graph: hill farm flocks (farms 3, 9, 10, 11, and 12) were assessed as relatively content, relaxed  
283 and thriving compared to lowland flocks (farms 1, 2, 4, 5, 6, and 7), and also showed  
284 consistently lower levels of lameness across visits than lowland flocks. Particularly in flocks 1  
285 and 4 the association between low PC1 scores/negative mood and high levels of lameness is  
286 evident.

287



#### 288 4. Discussion

289 This study applied qualitative behaviour assessment (QBA) to the on-farm assessment of sheep  
290 welfare, using a list of QBA terms developed for sheep by Phythian *et al.*, (2013a). The emphasis  
291 in this study was on repeated QBA assessment of sheep flocks across time, over 6 visits at two-  
292 monthly intervals, on 12 hill and lowland farms in the UK. The study's aims were to assess  
293 whether and how QBA, applied by an experienced assessor, was capable of detecting differences  
294 in sheep behavioural expression over time, and was associated with physical health measures  
295 taken at the same time points. Multivariate analysis identified two main dimensions of sheep  
296 expression: PC1, ranging from content/relaxed/thriving to distressed/dull/dejected (summarised  
297 as 'mood'), and PC2, ranging from anxious/agitated/responsive to relaxed/dejected/dull  
298 (summarised as 'responsiveness'). These dimensions correspond well to those found (dim1:  
299 content/relaxed/bright to distressed/dejected/tense; dim2: agitated/responsive/anxious to  
300 dull/dejected/relaxed) in a study by Phythian *et al.* (2013a), in which 13 veterinary and farm  
301 assurance assessors provided QBA, using the same terms as the current study, of 12 video clips  
302 showing sheep in varying indoor and outdoor situations and housing conditions. This  
303 convergence supports the relevance of these dimensions for characterising sheep expressions in  
304 varying on-farm conditions.

305  
306 Previous QBA studies (e.g. Rutherford *et al.*, 2012) have frequently found two main dimensions  
307 of behavioural expression, where the first dimension corresponds to a distinction between  
308 positive and negative experience, and the second dimension appears to distinguish between low  
309 and high levels of arousal/activation in these experiences. The dimensions identified in the  
310 present study concur with this pattern, with positive and negative descriptors aligning on

311 opposite sides of PC1, and high-arousal/activation terms (e.g.: agitated, responsive) placed on the  
312 opposite side to low-arousal terms (e.g.: relaxed, dull) on PC2. The four quadrants thus formed  
313 appear to fit in well with the integrative functional framework for emotion and mood proposed  
314 by Mendl *et al.* (2010), supporting our summarising labels of PC1 and PC2 term-loadings as  
315 ‘mood’ and ‘responsiveness’. This is not the place to discuss the relationship of QBA with  
316 cognitive or motivational theoretical frameworks of emotion, and we do not wish to suggest that  
317 QBA studies actually measure any cognitive or motivational states inferred by such frameworks;  
318 however QBA assessments of animals’ dynamic expressive demeanour do generally appear to be  
319 compatible with these frameworks (Boissy *et al.*, 2007; Mellor, 2012).

320

#### 321 **4.1 Detecting differences over time**

322 This study’s outcomes indicate that on average, PC1 flock scores (whether QBA scores were  
323 evaluated independently or combined with physical indicator outcomes) did not differ across the  
324 6 visits made in the course of a year, while the rankings of flocks on PC1 were highly correlated  
325 across visits. Together these findings indicate a high consistency of QBA assessments of the  
326 flocks’ mood over the course of the year. To our knowledge, only one previous on-farm study  
327 with pigs has investigated repeated application of QBA over a longer time period, finding  
328 moderate correlation ( $r_S=0.50$ ) between the PC1 scores generated by two visits (1 year apart) to  
329 the same 15 intensive pig farms (Temple *et al.*, 2013). The authors of the pig study rightly note  
330 that when using one assessor, it is not possible to tell whether lack of good correlation is due to  
331 poor intra-observer reliability or to a genuine change in the animals’ state. Weaker correlations  
332 may also reflect a lack of sufficient between-farm variation in the same production system. The  
333 present sheep study assessed the same farms 6 rather than 2 times a year, and included farms

334 from different production systems, allowing closer examination of variation between and within  
335 farms. The high consistency found in PC1 farm scores across visits, both within and between  
336 farms, suggests that the flocks' mood remained relatively stable on most farms across the study  
337 period.

338

339 PC2 flock scores (responsiveness) did show significant variation across visits, with the lowest  
340 scores (i.e. most relaxed/dejected/dull) recorded post-lambing in May/June (visit 1), and at the  
341 2010 lambing season (visit 6). An explanation for this might be that the presence of lambs less  
342 than 12 weeks old had a relaxing effect on sheep, due to physiological changes associated with  
343 maternal bonding behaviours such as licking and grooming of lambs (Dwyer *et al.*, 2008). It is  
344 also conceivable that variations in sheep responsivity reflected arbitrary differences in how sheep  
345 were selected for assessment by farmers. However, there were no signs of deliberate bias in how  
346 farmers selected sample animals. Moreover, as management practices tend to affect the whole  
347 flock rather than specific animals, repeated assessment of different groups of sheep from the  
348 same farm was considered to provide a representative sample of the flock. This view is supported  
349 by our finding that the rankings of individual sheep flocks on PC2, like those on PC1, were  
350 significantly correlated across the 6 visits.

351

352 Levels of lameness and breech and abdominal soiling varied over the year. The mean lameness  
353 level of 7.23% observed during the first visit period (May/June 2009) is close to the mean  
354 lameness estimate of 7.10% previously identified during a cross-sectional study on 40 English  
355 and Welsh sheep farms (Phythian *et al.*, 2013b). However, at 13.39% the mean level of lameness  
356 for all flocks observed across the whole study period is considerably higher than previously

357 reported, but does concur with King and Green's (2011) conclusions regarding the true  
358 prevalence of flock lameness in England. The high lameness levels observed in the current study  
359 may reflect the convenience-based approach to farm sampling, or be due to specific disease  
360 outbreaks in these flocks. High prevalence of lameness appears to coincide with periods of  
361 gathering, handling, and housing of animals, which favour the transmission of infectious causes  
362 of ovine lameness such as footrot (Raadsma and Egerton, 2013). Indeed in the current study the  
363 highest levels of lameness were recorded during autumn, coinciding with the mating period, and  
364 in winter, when most sheep flocks were housed. Follow-up veterinary examination after  
365 assessments were completed indicated that farms with high lameness prevalence showed high  
366 levels of infectious footrot, and problems with controlling an outbreak of contagious ovine  
367 digital dermatitis were thought to explain the very high (up to 61.86%) lameness levels recorded  
368 on farm 1. The increased level of breech soiling at the spring visit may be due to changes in  
369 nutrition, and to the greater parasite challenges of the spring grazing season. On the other hand,  
370 the higher level of abdominal soiling observed in the winter period may relate to the wet weather  
371 conditions experienced by some sheep flocks that had not yet been winter housed.

372

#### 373 **4.2 Effects of farm type and flock size**

374 There were effects of farm type on both QBA and physical indicator scores. Flocks on  
375 hill/upland farms were assessed as more content, relaxed and thriving (PC1) than lowland flocks.  
376 This may reflect a difference between extensive and intensive management practices, but could  
377 also reflect the effects on welfare of differences in physical health. Indeed, we found that hill  
378 farm flocks showed consistently lower levels of lameness than lowland farms, and there was a  
379 significant negative correlation between PC1 mood scores and lameness (for further discussion

380 of this association see below). Thus the hill farms selected for this study appeared to have a  
381 higher health and welfare status than the lowland farms, however, the study's non-random  
382 sampling approach, and the small sample of 12 farms, prevent a more general interpretation of  
383 these findings in terms of sheep welfare in different production systems. The difficulty of  
384 establishing larger samples with sufficient farm-level variation is a common feature of applied  
385 on-farm research (Andreasen *et al.*, 2013).

386

387 There was no significant effect of farm type or flock size on the sheep's responsiveness (PC2);  
388 however visual inspection of PC2 scores (Fig. 3) suggests some meaningful variation at  
389 individual farm level: sheep on a small-holding in a hill area (farm 3), which were regularly  
390 handled and petted, appeared consistently more relaxed than sheep on hill farms under more  
391 extensive management (farms 9-12). This difference may reflect the positive effect of handling  
392 on welfare (Boivin *et al.*, 2003), and suggests that QBA can generate expressive dimensions on  
393 which management practices can be meaningfully evaluated. However further study with larger  
394 sample sizes is required to support these findings. There was one effect of flock size on physical  
395 health, with large flocks (950-1250 sheep) showing a higher prevalence of abdominal soiling  
396 than sheep in small flocks (less than 100 sheep), which may reflect the greater exposure of large  
397 extensively managed commercial flocks to environmental and climatic conditions. However here  
398 too, further study with larger sample sizes is required.

399

#### 400 **4.3 Associations between QBA and physical health indicators**

401 Of the seven physical indicators of sheep health and welfare recorded, only the proportion of lame  
402 sheep on a farm, and the proportion of sheep recorded with 'dull physical demeanour', correlated

403 significantly with PC1 mood scores, indicating that flocks with high levels of lameness and dull  
404 physical postures were perceived as more distressed/dull/dejected than other flocks. Indeed visual  
405 inspection of individual farm scores (Fig. 3) indicates that farms with consistently low mood scores  
406 (e.g. farms 1 and 4) showed high levels of lameness. Lameness is a key welfare issue for sheep,  
407 and dull physical demeanour is an attribute commonly assessed by stock-people and veterinarians  
408 to detect sickness and disease. The significant association of QBA expressive dimensions with  
409 these health measures supports that QBA addressed important aspects of sheep welfare, and  
410 provided complementary information to help interpret the wider welfare impact of these health  
411 problems. Lameness is associated with pain, however the present study described lame sheep as  
412 more distressed, dull, and dejected than non-lame sheep, suggesting that lameness and the pain  
413 underlying it also had a more generally deleterious effect on the sheep's emotional state. The same  
414 can be said for dull physical demeanour; this is a specific physical measure usually associated with  
415 sickness and pain, however the present study suggests these clinical signs also had a more  
416 generally deleterious emotional effect.

417  
418 One could argue that the QBA terms 'dull', 'dejected', and 'responsive' and the physical indicator  
419 'dull physical demeanour' can hardly be considered independent measures, and may have been  
420 subject to what Greenwald *et al.*, (1986) call 'theory confirmation bias'. However the two  
421 measures were embedded in very different assessment and scoring procedures, and were not taken  
422 closely together. QBA was always performed right at the start of the assessment, ensuring  
423 independence of any physical measurement afterwards. Moreover, PC1 mood scores were not the  
424 result of direct measurement, but the outcome of a multivariate analysis, creating more analytical  
425 distance between the two types of measure. Their significant correlation was thus not pre-given,

426 but can still be considered a meaningful outcome. A good degree of association between different  
427 types of measure confers an aspect of internal validity known as convergent validity (Abramson  
428 and Abramson, 2008). Yet to confer validity this association does not have to be maximal, in that  
429 the different types of measure address the animal welfare construct from different angles: that of  
430 psychological well-being (QBA), and physical health (lameness, dull physical demeanour).

431  
432 Coughing, wool-loss, abdominal soiling, and skin irritation were observed too infrequently to  
433 enable meaningful correlation with QBA. Breech soiling did occur frequently, and there was a  
434 weak but significant correlation between breech soiling and PC2, indicating that flocks that were  
435 more relaxed were also more soiled. The reasons for this association are likely to be complex, but  
436 may be related to a co-variance in time: both measures increased significantly in spring. As  
437 discussed above, sheep may have become more relaxed after lambing, and this may have affected  
438 their behaviour in a way (e.g. increased lying) that led to more breech soiling in wet weather  
439 conditions. Breech soiling can also be influenced by seasonal dietary changes, or indicate the  
440 presence of endo-parasitism or blowfly myiasis (French *et al.*, 1994), which may have caused the  
441 sheep to slow down, but on these farms it did not appear to affect their welfare such that they  
442 became dull and dejected, as was the case with lameness.

443  
444 Thus, combining QBA with other indicators is likely to provide a fuller, more complex picture of  
445 animal health and welfare (Wemelsfelder & Farish, 2004; Wemelsfelder & Mullan, 2014).  
446 Through patterns of correlation, QBA can help to interpret health and behaviour measures in terms  
447 of an animal's well-being, as has for example been reported for sheep (Wickham *et al.*, 2012),  
448 horses (Minero *et al.*, 2009), and pigs (Rutherford *et al.*, 2012). An on-farm assessment study of 43  
449 Danish dairy cattle farms found no meaningful correlations between QBA and measures of the

450 Welfare Quality® assessment protocol (Andreasen *et al.*, 2013). However, as discussed above for  
451 the Temple *et al.* (2013) on-farm pig study, this may have at least partly been due to a lack of  
452 sufficient between-farm variation in the study's sample. A risk factor analysis of on-farm  
453 assessments of 89 commercial broiler farms in various EU countries reported some meaningful  
454 associations between QBA and other measures (Bassler *et al.*, 2013). The present study included  
455 farms from different sheep production systems; that despite its small sample size it found  
456 significant associations between QBA and physical health measures is encouraging, and should  
457 stimulate more research.

458

#### 459 **4.4 Methodological considerations**

460 For reasons of feasibility and cost, all assessments in this study were performed by the same  
461 assessor. It could be argued therefore that the repeated farm assessments were not independent, and  
462 may have artificially inflated the stability of PC1 mood scores across the year. Indeed, as noted by  
463 Temple *et al.* (2013), when using a single assessor it is not possible to distinguish observer from  
464 farm effects, other than by consideration of a study's larger context and totality of findings.  
465 However, PC2 responsiveness scores did fluctuate across visits, which suggests that the assessor  
466 was not simply repeating previous scoring patterns, but was sensitive to the possibility of change  
467 across visits. That the QBA dimensions found here were extremely similar to those identified from  
468 video by experienced veterinary and farm assurance inspectors, also supports their relevance  
469 (Phythian *et al.*, 2013a).

470

471 Another concern is that qualitative judgments are sensitive to environmental context, which can  
472 both be a strength and a potential source of bias (Wemelsfelder *et al.*, 2009; Tuytens *et al.*,



473 2014; Fleming *et al.*, 2015). Random variation in on-farm conditions may have affected QBA  
474 scoring levels, and it is difficult to distinguish this from meaningful variation. To counter such  
475 effects, it is important to use more than one assessor, and investigate the inter-observer reliability  
476 of on-farm QBA assessments (e.g. Andreasen *et al.*, 2013). However this is costly and may not  
477 be feasible, and it is thus advisable to always apply QBA in combination with other health and  
478 welfare measures. The potential for inadvertent observer bias is not exclusive to qualitative  
479 methods, but applies equally to quantitative measures (Tuytens *et al.*, 2014). To increase the  
480 robustness of on-farm welfare assessments it is crucial to use trained pools of assessors, who are  
481 experienced in assessing sheep over the seasons of the year, across a range of production  
482 systems, and are able to distinguish meaningful from random fluctuations in sheep expression. If  
483 further research were to uphold the positive results found in this study, QBA could potentially be  
484 applied as a day-to-day management tool on sheep farms, and be used to communicate welfare  
485 values to farmers, shepherds, and consumers.

486

#### 487 **4.5 Conclusion**

488 The results of this study generally indicate that QBA was capable of identifying expressive  
489 dimensions that distinguished meaningfully between sheep demeanour within- and between farms  
490 and across the seasons of the year, and correlated significantly with important physical indicators  
491 of sheep health. A strong negative correlation was found between PC1 ‘mood’ scores and levels of  
492 lameness and ‘dull physical demeanour’, indicating that the latter clinical signs of compromised  
493 health also had a wider deleterious effect on the sheep’s emotional state.

494

#### 495 **Acknowledgments**

496 This study was conducted as part of the Defra funded AW1025 study (Development of indicators  
497 for the on-farm assessment of sheep welfare). The authors gratefully acknowledge all the farmers  
498 involved for their time and effort in participating with this study.

499

## 500 **References**

501 Abramson, Z.H., Abramson, J.H. 2008. Validity. In: Research methods in community medicine:  
502 surveys, epidemiological research, programme evaluation, clinical trials. Sixth edition. John Wiley  
503 and Sons Ltd, Chichester, England, pp 161-178.

504 Andreassen, S.N., Wemelsfelder, F., Sandøe, P., Forkman, B. 2013. The correlation of Qualitative  
505 Behavior Assessments with Welfare Quality<sup>®</sup> protocol outcomes in on-farm welfare assessment of  
506 dairy cattle. *Applied Animal Behaviour Science* 143, 9-17.

507 Bassler, A.W., Arnould, C. Butterworth, A., Colin, L., De Jong, I.C., Ferrante, V., Ferrari, P.,  
508 Haslam, S., Wemelsfelder, F., Blokhuis, H.J. 2013. Potential risk factors associated with contact  
509 dermatitis, lameness, negative emotional state, and fear of humans in broiler chicken flocks.  
510 *Poultry Science* 92, 2811-2826.

511 Bokkers, E.A.M., de Vries, M., Antonissen, I.C.M.A., de Boer, I.J.M. 2012. Inter- and intra-  
512 observer reliability of experienced and inexperienced observers for the Qualitative Behaviour  
513 Assessment in dairy cattle. *Animal Welfare* 21, 307-318.

514 Boissy, A., Manteuffel, G., Jensen, M.B., Moe, R.O., Spruijt, B., Keeling, L.J., Winckler, C.,  
515 Forkman, B., Dimitrov, I., Langbein, J., Bakken, M., Veissier I., Aubert, A. 2007. Assessment of  
516 positive emotions in animals to improve their welfare. *Physiology and Behaviour* 92, 375-397.

- 517 Boivin, X., Tournadre, H., Le Neindre, P. 2000. Hand-feeding and gentling influence early-weaned  
518 lambs' attachment responses to their stockperson. *Journal of Animal Science* 78, 879-884.
- 519 Boivin, X., Lensink, J., Tallet, C., Veissier, I. 2003. Stockmanship and farm animal welfare.  
520 *Animal Welfare* 12, 479-492.
- 521 Dwyer, C.M. 2008. Genetic and physiological determinants of maternal behavior and lamb  
522 survival: Implications for low-input sheep management. *Journal of Animal Science* 86, 246-258.
- 523 Fitzpatrick, J., Scott, M., Nolan, A. 2006. Assessment of pain and welfare in sheep. *Small*  
524 *Ruminant Research* 62, 55-61.
- 525 Fleming, P.A., Wickham, S.L., Stockman, C.A., Verbeek, E., Matthews, L., Wemelsfelder, F.  
526 2015. The sensitivity of QBA assessments of sheep behavioural expression to variations in visual  
527 or verbal information provided to observers. *Animal* 9, 878–887.
- 528 French, N.P., Wall, R., Morgan, K.L. 1994. Lamb tail docking: a controlled field study of the  
529 effects of tail amputation on health and productivity. *Veterinary Record* 134, 463-467.
- 530 Greenwald, A.G., Pratkanis, A.R., Leippe, M.R., Baumgardner, M.H. 1986. Under what conditions  
531 does theory obstruct research progress. *Psychological Bulletin* 93, 216-299.
- 532 Joliffe, I.T. 2002. *Principal component analysis*. Second edition. Springer series in statistics.  
533 Springer-Verlag, New York.
- 534 King, E.M., Green, L.E. 2011. Assessment of farmer recognition and reporting of lameness in  
535 adults on 35 lowland sheep flocks in England. *Animal Welfare* 20, 321-328.

- 536 Le Neindre, P., Bovin, X., Boissy, A. 1996. Handling of extensively kept animals. *Applied Animal*  
537 *Behaviour Science* 49, 73-81.
- 538 Long, J.S., Freese, L. 2006. Regression models for categorical dependent variables. Using Stata.  
539 Stata Press Publications, Texas, USA.
- 540 Mellor, D.J. 2012. Animal emotions, behaviour and the promotion of positive welfare states. *New*  
541 *Zealand Veterinary Journal* 60, 1-8.
- 542 Mendl, M., Burman, O.H.P., Paul, E.S. 2010. An integrative and functional framework for the  
543 study of animal emotion and mood. *Proceedings of the Royal Society B: Biological Sciences* 277,  
544 2895–2904.
- 545 Minero, M., Tosi, M.V., Canali, E., Wemelsfelder, F. 2009. Quantitative and qualitative  
546 assessment of the response of foals to the presence of an unfamiliar human. *Applied Animal*  
547 *Behaviour Science* 116, 74-81.
- 548 Napolitano, F., Knierim, U., Grasso, F., De Rosa, G. 2009. Positive indicators of cattle welfare and  
549 their applicability to on-farm protocols. *Italian Journal of Animal Science* 8, 355-365.
- 550 Phythian, C.J., Cripps, P.J., Michalopoulou, E., Jones, P.H., Grove-White, D., Clarkson, M.J.,  
551 Winter, A.C., Stubbings, L.A., Duncan, J.S. 2012. Reliability of indicators of sheep welfare  
552 assessed by a group observation method. *The Veterinary Journal* 193, 257–263.
- 553 Phythian, C., Michalopoulou, E., Duncan, J., Wemelsfelder, F. 2013a. Inter-observer reliability of  
554 Qualitative Behavioural Assessments of sheep. *Applied Animal Behaviour Science* 144, 73–79.

- 555 Phythian, C.J., Cripps, P.J., Grove-White, D., Jones, P.H., Michalopoulou, E., Duncan, J.S. 2013b.  
556 Observing lame sheep: evaluating test agreement between group-level and individual animal  
557 methods of assessment. *Animal Welfare* 22, 417-422.
- 558 Raadsman H.W., Egerton, J.R. 2013. A review of footrot in sheep: Aetiology, risk factors and  
559 control methods. *Livestock Science* 156, 106-114.
- 560 Rousing, T., Wemelsfelder, F. 2006. Qualitative assessment of social behaviour of dairy cows  
561 housed in loose housing systems. *Applied Animal Behaviour Science* 101, 40-53.
- 562 Rutherford K.M., Donald R.D., Lawrence A.B., Wemelsfelder F. 2012. Qualitative Behavioural  
563 Assessment of emotionality in pigs. *Applied Animal Behaviour Science* 139, 218-224.
- 564 Stockman, C.A., Collins, T., Barnes, A.L., Miller, D., Wickham, S.L., Beatty., D.T., Blache, D.,  
565 Wemelsfelder, F., Fleming, P.A., 2011. Qualitative behavioural assessment and quantitative  
566 physiological measurement of cattle naive and habituated to road transport. *Animal Production*  
567 *Science* 51, 240-249.
- 568 Temple, D., Manteca, A., Verlade, A., Dalmau, A. 2011. Assessment of animal welfare through  
569 behavioural parameters in Iberian pigs in intensive and extensive conditions. *Applied Animal*  
570 *Behaviour Science* 131, 29-39.
- 571 Temple, D., Manteca, X., Dalmau, A., Velarde, A. 2013. Assessment of test-retest reliability of  
572 animal-based measures on growing pig farms. *Livestock Production Science* 151, 35-45.

- 573 Tuytens, F.A.M., de Graaf S., Heerkens, J.L.T., Jacobs, L., Nalon, E., Ott, S., Stadig, L., Van  
574 Laer, E., Ampea, B. 2014. Observer bias in animal behaviour research: can we believe what we  
575 score, if we score what we believe? *Animal Behaviour* 90, 273-280.
- 576 Wemelsfelder, F., Lawrence, A.B. 2001. Qualitative assessment of animal behaviour as an on-farm  
577 welfare-monitoring tool. *Acta Agriculturae Scandinavica Section A-Animal Science* 30, 21-25.
- 578 Wemelsfelder, F., Farish, M. 2004. Qualitative categories for the interpretation of sheep welfare: a  
579 review. *Animal Welfare* 13, 261-268.
- 580 Wemelsfelder, F., Nevison, I., Lawrence, A.B. 2009. The effect of perceived environmental  
581 background on qualitative assessments of pig behaviour. *Animal Behaviour* 78, 477-484.
- 582 Wemelsfelder, F., Mullan, S. 2014. Applying ethological and health indicators to practical animal  
583 welfare assessment. In: *Animal Welfare: focusing on the future* (Ed. D.J. Mellor), OIE Scientific  
584 and Technical Review, 33, 111-120.
- 585 Wickham, S., Collins, T., Barnes, A.L., Miller, D.W., Beatty, D.T., Stockman, C., Blache, D.,  
586 Wemelsfelder, F. 2012. Qualitative behavioural assessment of transport-naïve and transport  
587 habituated sheep. *Journal of Animal Science* 90, 4523-4535.
- 588 Wiseman-Orr, M.L., Scott, E.M., Reid, J., Nolan, A.M. 2006. Validation of a structured  
589 questionnaire as an instrument to measure chronic pain in dogs on the basis of effects on health-  
590 related quality of life. *American Journal of Veterinary Research* 67, 1826-1836.

591

592 **Figure Captions**

593 **Figure 1:** Loadings of the 12 QBA terms on PC1 and PC2 of the PCA analysis. Axes reflect  
594 arbitrary scaling values.

595

596 **Figure 2:** Loadings of the combined PCA analysis of 12 QBA terms and 7 physical health  
597 indicators on PC1 and PC2. Axes reflect arbitrary scaling values.

598

599 **Figure 3:** Distributions of QBA PC1 and PC2 scores, and of lameness and breech soiling  
600 percentages, for the 12 sheep farms across the 6 visits. Visit periods were: 1. May-June 2009, 2.  
601 July-August 2009, 3. September-October 2009, 4. November-December 2009, 5. January-  
602 February 2010, and 6. March-April 2010.

603

**Table 1:** Overview of assessment visit periods

Visit	Study period	Season	Production stage
1	May – June 2009	Spring/Summer	Post-lambing
2	July – August 2009	Summer	Weaning
3	September – October 2009	Autumn	Mating
4	November – December 2009	Autumn/Winter	Early pregnancy
5	January – February 2010	Winter	Mid-pregnancy
6	March – April 2010	Spring	Lambing



**Table 2:** Study farm details

Farm ID	Farm type	Flock size	Farm purpose	Farm assurance scheme member	Farming system
1	Lowland	850	Commercial	No	Conventional
2	Lowland	260	Commercial	No	Conventional
3	Hill	24	Small-holding	No	Conventional
4	Lowland	250	Commercial	No	Conventional
5	Lowland	210	Commercial	Yes	Conventional
6	Lowland	280	Commercial	No	Conventional
7	Lowland	600	Commercial	No	Conventional
8	Upland	450	Commercial	Yes	Organic
9	Hill	320	Commercial	Yes	Conventional
10	Hill	800	Commercial	Yes	Conventional
11	Hill	1100	Commercial	Yes	Conventional
12	Hill	1260	Commercial	Yes	Conventional

**Table 3:** Prevalence of observed physical health and welfare indicators in 12 sheep flocks over 6 visits.

Indicator	Mean proportion of sheep affected (%) at each of the six assessment visits (95% CI)						Mean proportion (%) affected across study
	1	2	3	4	5	6	
	May-June 2009	July-August 2009	Sept-Oct 2009	Nov-Dec 2009	Jan-Feb2010	March-April 2010	
Dull demeanour	0.17 (-0.07 – 0.41)	0.71 (0.02 – 1.40)	1.62 (0.01 – 3.21)	0.64 (-0.78 – 2.06)	3.10 (0.04 – 6.17)	3.70 (-0.65 – 8.05)	1.52 (0.77 – 2.77)
Coughing	0	0.18 (-0.08 – 0.43)	0	0.13 (-0.16 – 0.41)	0.28 (-0.11 – 0.67)	0.11 (-0.12 – 0.33)	0.87 (0 – 2.03)
Pruritis	0	0.15 (-0.10 – 0.39)	0.05 (-0.07 – 0.17)	0	0	0.29 (-0.18 – 0.77)	0.12 (0.01 – 0.24)
Wool loss	1.86 (-0.65 – 4.37)	0.18 (-0.09 – 0.44)	0.40 (-0.07 – 0.88)	0	0.85 (0.13 – 1.56)	2.64 (-0.12 – 5.41)	0.74 (0.29 – 1.18)
Lameness	7.23 (3.68 – 10.78)	14.59 (8.79– 20.41)	16.77 (5.62– 27.92)	12.51 (7.14 – 17.89)	17.86 (7.83 – 27.90)	11.29 (4.45 – 18.11)	13.39 (10.53 – 16.22)
Breech soiling	23.83 (11.86 – 35.81)	12.42 (8.69– 16.15)	7.91 (3.25– 12.56)	13.92 (3.08– 24.76)	7.76 (1.28– 14.24)	14.63 (7.97 – 21.27)	13.41 (10.29 – 16.53)
Abdominal soiling	9.24 (-4.37– 22.86)	0	0.11 (-0.14 – 0.37)	20.17 (-0.74 – 41.09)	0.90 (-0.60 – 2.40)	0.37 (-0.50 – 1.20)	5.13 (1.09 – 9.18)

**Table 4:** Regression parameters describing the effect of visit period (1 to 6) on PC1 and PC2 flock scores, for both the QBA and combined QBA/physical indicator analyses, and for physical indicators of lameness and breech and abdominal soiling.

Principal Component	Study visit	$\beta$	95% CI	p-value
<b>QBA PC1<sup>a</sup></b>	<b>For all 6 visits</b>	-	-	<b>p=0.155</b>
<b>QBA PC1</b>	1	0.21	-1.11 – 1.54	-
	2	-0.55	-1.40 – 0.29	0.197
	3	-0.54	-1.47 – 0.39	0.255
	4	0.11	-0.84 – 1.06	0.817
	5	-0.56	-1.51 – 0.40	0.251
	6	0.27	-0.68 – 1.23	0.576
<b>QBA PC2</b>	<b>For all 6 visits</b>	-	-	<b>p&lt;0.001</b>
<b>QBA PC2</b>	1	-1.25	-2.02 – -0.49	-
	2	1.82	1.14 – 2.49	<0.001
	3	2.04	1.19 – 2.89	<0.001
	4	1.93	0.99 – 2.87	<0.001
	5	1.01	0.02 – 1.99	0.046
	6	0.73	-0.29 – 1.74	0.160
<b>Combined PC1<sup>b</sup></b>	<b>For all 6 visits</b>	-	-	<b>p=0.1982</b>
<b>Combined QBA/ physical indicators PC1</b>	1	-0.33	-1.82 – 1.16	-
	2	0.52	-0.45 – 1.51	0.292
	3	0.60	-0.49 – 1.69	0.277
	4	-0.23	-1.34 – 0.86	0.679
	5	0.87	-0.24 – 1.99	0.125
	6	0.20	-0.91 – 1.32	0.723
<b>Combined PC2</b>	<b>For all 6 visits</b>	-	-	<b>p&lt;0.001</b>
<b>Combined QBA/physical indicators PC2</b>	1	-1.46	1.36 – 2.79	-
	2	2.07	1.39 – 3.13	<0.001
	3	2.26	1.39 – 3.27	<0.001
	4	2.33	0.46 – 2.39	<0.001

	5	1.42	-0.31– 1.65	0.004
	6	0.67	-0.31– 1.65	0.178
	<b>For all 6 visits</b>	-	-	<b>p=0.0249</b>
<b>Lameness</b>	1	7.23	0.72– 13.74	-
	2	7.37	0.49– 14.23	0.036
	3	9.53	2.90– 16.17	0.005
	4	5.28	-1.37 – 11.93	0.120
	5	10.63	3.98– 17.28	0.002
	6	4.05	-2.59– 10.70	0.232
	<b>For all 6 visits</b>	-	-	<b>p=0.0008</b>
<b>Breech soiling</b>	1	23.83	16.92– -4.16	-
	2	-11.40	-18.65– -4.17	0.002
	3	-15.92	-24.21– -7.64	0.001
	4	-9.90	-18.49– -1.33	0.024
	5	-16.06	-24.74– -7.40	0.001
	6	-9.21	-17.90– -.51	0.038
	<b>For all 6 visits</b>	-	-	<b>p=0.0048</b>
<b>Abdominal soiling</b>	1	9.24	0.53 – 17.95	-
	2	-9.24	-21.64– 3.15	0.144
	3	-9.12	-21.44– 3.17	0.146
	4	10.93	-1.39– 23.25	0.082
	5	-8.34	-20.66 – 3.97	0.184
	6	-8.87	-21.19– 3.45	0.158

<sup>a</sup> Principal Component obtained through PCA of QBA scores only

<sup>b</sup> Principal Component obtained through PCA of QBA and physical indicator scores

Figure 1

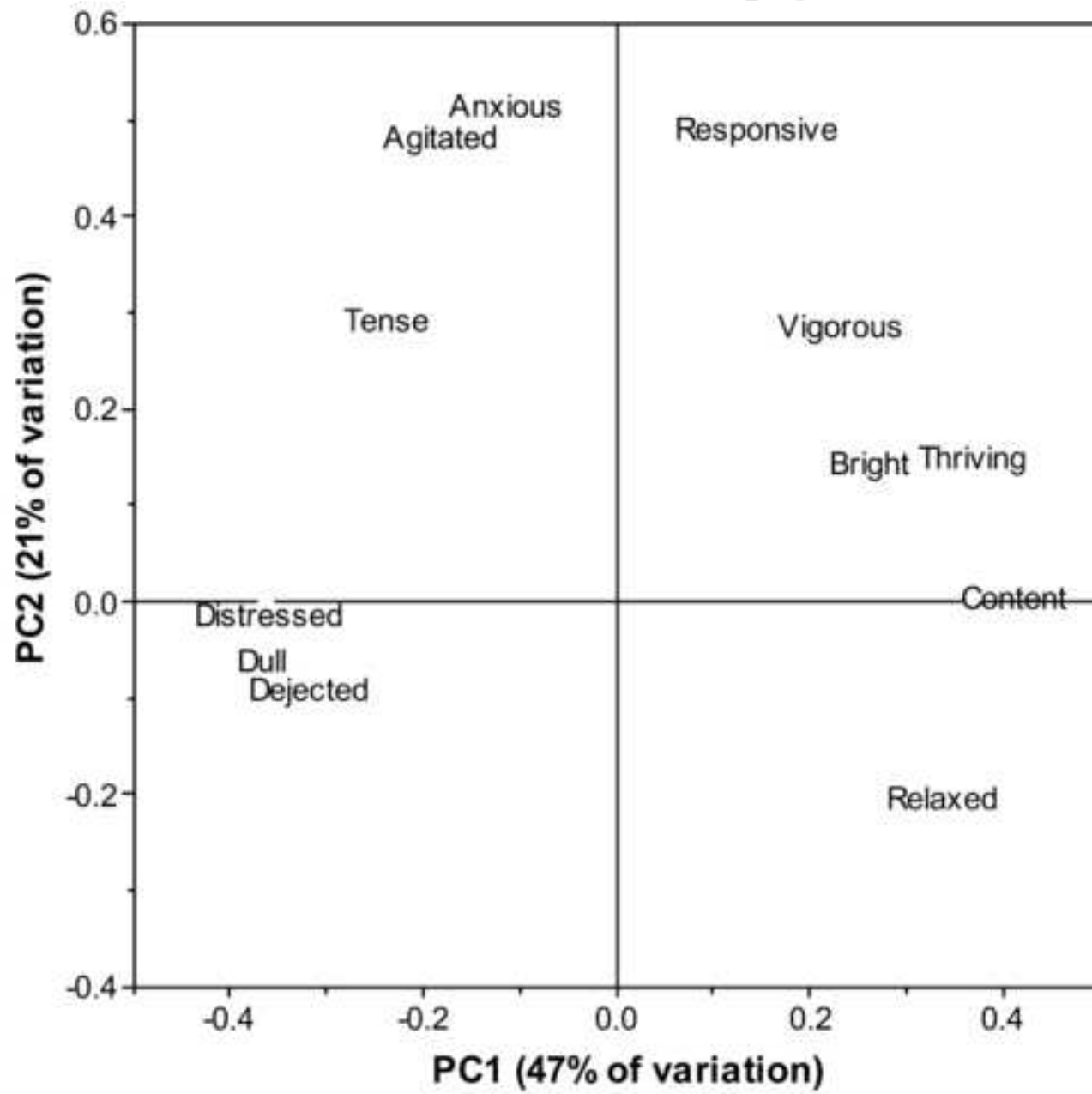


Figure 2

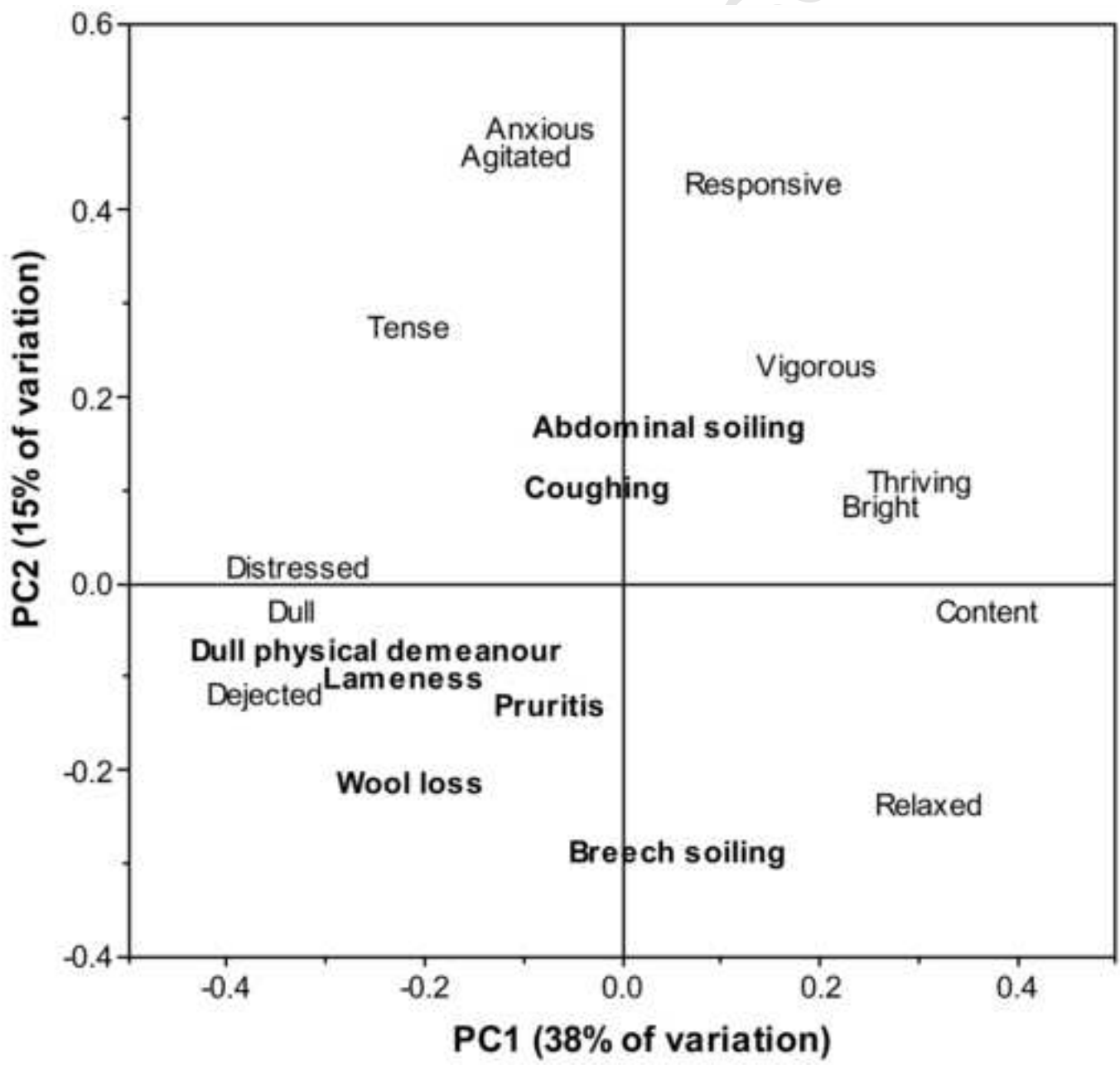


Figure 3

