

**The Humane Society Institute for Science and Policy**  
**Animal Studies Repository**

2-15-2013

# Inter-Observer Reliability of Qualitative Behavioural Assessments of Sheep


Clare Phythian  
*University of Liverpool*

Eleni Michalopoulou  
*University of Liverpool*

Jennifer Duncan  
*University of Liverpool*

Françoise Wemelsfelder  
*Scotland's Rural College, francoise.wemelsfelder@sruc.ac.uk*

Follow this and additional works at: [http://animalstudiesrepository.org/acwp\\_asie](http://animalstudiesrepository.org/acwp_asie)

 Part of the [Animal Studies Commons](#), [Behavior and Ethology Commons](#), and the [Comparative Psychology Commons](#)

---

## Recommended Citation

Phythian, C., Michalopoulou, E., Duncan, J., & Wemelsfelder, F. (2013). Inter-observer reliability of qualitative behavioural assessments of sheep. *Applied Animal Behaviour Science*, 144(1), 73-79.

This Article is brought to you for free and open access by the Humane Society Institute for Science and Policy. It has been accepted for inclusion by an authorized administrator of the Animal Studies Repository. For more information, please contact [eyahner@humanesociety.org](mailto:eyahner@humanesociety.org).

# Inter-Observer Reliability of Qualitative Behavioural Assessments of Sheep

Clare Phythian<sup>1</sup>, Eleni Michalopoulou<sup>1</sup>, Jennifer Duncan<sup>1</sup>, Françoise Wemelsfelder<sup>2</sup>

<sup>1</sup> *University of Liverpool*

<sup>2</sup> *Scotland's Rural College*

## KEYWORDS

Sheep, Qualitative Behaviour Assessment, inter-observer reliability

## ABSTRACT

Qualitative Behaviour Assessment (QBA) is whole-animal methodology that assesses the expressive qualities of animal demeanour using descriptors such as 'relaxed', 'anxious' or 'content'. This study aimed to examine the inter-observer reliability of 12 fixed-list QBA descriptors for sheep that had been generated in a previous Free-Choice Profiling study by experienced animal welfare inspectors, based on the same video footage used in the current study. The 12 QBA terms were scored by two different assessor groups consisting of two veterinary science students and four veterinary surgeons (Group 1), and seven farm assurance inspectors (Group 2). The two assessor groups met and received training on different dates, and viewed the same 12 video clips of sheep showing a wide range of behavioural expressions in varying indoor and outdoor situations and housing systems. For each clip assessors scored the sheep expressions on each of the 12 QBA terms using a Visual Analogue Scale. Principal Component Analysis (covariance matrix, no rotation) was used to analyse assessor scores, both for Groups 1 and 2 separately, and together in one all-assessor group. All three analyses identified that over 70% of variance in sheep expression was described by two main components. For the all-assessor analysis the first Principal Component (PC1, 49% of variation) ranged from 'content/relaxed/bright' to 'distressed/dejected/tense', while PC2 (31%) ranged from 'agitated/responsive/anxious' to 'dull/dejected/relaxed'. The 2 Principal Components identified by Groups 1 and 2 when analysed separately, were highly similar. The level of agreement between assessors in each group was evaluated using Kendall's coefficient of concordance (*W*). This produced *W* values of 0.83 (PC 1) and 0.84 (PC 2) for the all-assessor group, values of 0.90 (PC 1) and 0.86 (PC 2) for Group 1, and of 0.78 (PC 1) and 0.91 (PC 2) for Group 2. All values were significant at  $P < 0.001$ . These results indicate that two trained assessor groups achieved high levels of inter-observer agreement using a list of 12 pre-fixed QBA terms developed for sheep. This study is the first to investigate QBA as a tool for assessing sheep behavioural expression, and its results support further exploration of the feasibility and validity of applying this method to the assessment of sheep welfare.

## 1. Introduction

Quantitative behavioural observations are used routinely in the assessment of sheep health and welfare. Clinical observations and assessments by veterinary surgeons and stockpeople often rely on broad

changes in behaviour to identify particular diseases and conditions of suffering (Gougoulis et al., 2010). Key behavioural signs or outcomes that are used for identifying individual sheep with a sub-optimal health and welfare condition include alterations in gait (Kaler et al., 2009), signs of pruritus (Berriatua et al., 2001), or separation of individual sheep from the rest of the flock (Lynch et al., 1992). However, relatively few research studies have examined the validity, reliability and feasibility of applying these measures for on-farm assessments of sheep welfare (Napolitano et al., 2009b, Phythian et al., 2011; Stubbsj en et al., 2011). As part of a Department of the Environment, Food and Rural Affairs (Defra) funded research project, animal-based indicators for the on-farm assessment of sheep welfare have been tested against the criteria of validity, reliability and feasibility (Phythian et al., 2011). The FAWC Five Freedoms (FAWC, 1994) were used as the framework for the development of welfare indicators for sheep through a scientific review and consultation of an expert panel (Phythian et al., 2011). The provisions of the Five Freedoms suggest the inputs required for achieving good welfare and so it may not be surprising that many of the welfare measures identified focus on the physical outcomes of the provision of on-farm resources and management actions, such as the presence of lameness and ecto-parasitism (Phythian et al., 2011). These animalbased outcomes have been tested as quantitative indicators of sheep welfare by assessing behavioural changes in gait and posture or signs of pruritus (Phythian et al., 2012). However, such quantitative measures may not capture an animal's long-term experience of well-being or 'quality of life' (Wiseman-Orr et al., 2006). Qualitative methods of assessment have recently come to be recognised as basically reliable measures which can, in combination with quantitative measures, make an important contribution to animal welfare science (Meagher, 2009). As a result, there is increasing interest in methods that assess animals beyond their physical appearance and are capable of addressing quality of life (FAWC, 2009).

Qualitative Behavioural Assessment (QBA) is an approach which examines the expressive body language of the whole animal as a means of addressing the animal's experience of particular situations and conditions, using qualitative descriptors such as 'relaxed', 'tense', 'anxious' or 'content' (Wemelsfelder, 2007). Such descriptors, given their emotional connotation, have the potential to make a valuable contribution to assessments of animal welfare, and may be particularly useful for addressing positive aspects of animal welfare that go beyond the mere absence of pain and distress (Boissy et al., 2007; Napolitano et al., 2009a). QBA has shown good levels of inter-observer reliability when applied to a range of species, including pigs (Wemelsfelder et al., 2000, 2001, 2012), poultry (Wemelsfelder et al., 2009a), dairy cattle and buffaloes (Rousing and Wemelsfelder, 2006; Napolitano et al., 2012), horses (Napolitano et al., 2008; Minero et al., 2009), and dogs (Walker et al., 2010). Scientific validation studies have demonstrated meaningful correlations of QBA with quantitative ethogram measures (e.g. Rousing and Wemelsfelder, 2006; Rutherford et al., 2012), and with physiological indicators and conditions relevant to welfare (e.g.: heart rate, core body temperature, plasma glucose and the neutrophil:lymphocyte ratio, Stockman et al., 2011; treatment with anxiolytic drug Azaperone, Rutherford et al., 2012.), and have shown QBA not to be unduly distorted by background environmental conditions (Wemelsfelder et al., 2009c). QBA is included within the WelfareQuality® on-farm welfare assessment protocols for cattle, poultry and pigs as a measure of positive emotional state (Wemelsfelder and Millard, 2009; Wemelsfelder et al., 2009a,b), and two on-farm studies support that QBA may contribute valuable information indicating how animals cope with conditions on farm (Brscic et al., 2009; Temple et al., 2011).

Many sheep in the United Kingdom and further afield are managed outdoors for a large part of the production cycle, so assessment of individual sheep can require gathering and handling of the flock to facilitate observations of certain behaviours. Handling and human disturbance are well-recognised to affect the behaviour and welfare of sheep (Boivin et al., 2003, Dwyer, 2009), and in extensive farming systems the requirement for gathering and handling of large groups of sheep is time consuming and labour intensive. Thus people involved with the daily on-farm care of sheep are in need of practical ways

to assess sheep welfare so as to promptly identify any signs of pain, distress and ill-health and to inform key management decisions that have impacts on flock health, welfare and production. The advantage of QBA is that it can be applied to both individual and groups of animals without major interference with these animals, and is performed in situ, i.e. in the environment where animals are regularly managed. Such a methodology would offer clear advantages for daily management, farm certification, and other welfare assessment purposes.

QBA involves scoring an animal's behavioural expressions using a range of positive and negative descriptive terms. The descriptive terms used in QBA may be developed either by a process known as Free-Choice Profiling (FCP), in which each assessor generates their own terminology for scoring animal demeanour (Wemelsfelder et al., 2001, 2012), or alternatively assessors can use a pre-determined list of terms for assessment. The use of a fixed-list of QBA terms provides a more standardised approach for farm assessment schemes compared to FCP, and in the long-term is more feasible in practice (Wemelsfelder and Millard, 2009; Wemelsfelder et al., 2009a,b). Therefore, the objective of the present study was to investigate the inter-observer reliability of 12 fixed-list QBA terms for sheep, applied by two assessor teams with experience and backgrounds relevant to on-farm welfare assessment, as a first step in examining the reliability and validity of QBA as a practical assessment tool for sheep.

## **2. Materials and methods**

### *2.1. Experimental design*

#### *2.1.1. Assessors*

Two groups of assessors took part in this study. Group 1 consisted of two veterinary science students, and four veterinary surgeons recruited from the University of Liverpool School of Veterinary Science. Group 2 consisted of seven experienced farm assurance assessors of the Soil Association. None of these assessors had prior experience with the QBA of sheep. Group 1 participated in the study in June 2009, and Group 2 in December 2009.

#### *2.1.2. Video footage*

Twelve video clips of approximately 1 min length were used to test the inter-observer reliability of QBA. Eleven of these video clips were shot by Françoise Wemelsfelder using a digital Panasonic NV-DX1E camcorder (Panasonic, Berkshire, UK). The twelfth clip was provided by Cathy Dwyer of the SAC. These 12 clips included 6 clips either showing or focusing on an individual animal, and 6 clips showing groups of sheep, with 3 clips of small groups of 3 or 4 animals, and 3 clips of larger flocks of more than 10 animals. To record animals from a distance without disturbing them, or to be able to focus on an individual animal in a group, a tripod and zoom function were used during recording. Footage included animals in lowland, upland and hill production systems, in intensive indoor housing systems and extensive outdoor grazing situations, and showed adult ewes and rams as well as young lambs. All clips were recorded with the aim of covering as many aspects of the expressive repertoire of sheep as could be found on local experimental and commercial sheep farms, so as to test inter-observable reliability on wide-ranging patterns of behaviour and expression.

Thus the 12 clips showed: 1. A small group of alert vigilant sheep grazing on moorland, 2. A severely lame individual sheep in a lowland field; 3. A hungry individual sheep closely searching a familiar human person in an indoor pen; 4. A large flock bleating loudly while being moved into a holding pen at upland pasture; 5. A mildly lame individual sheep amongst healthy grazing sheep in a lowland field; 6. A small group of relaxed sheep absorbed in grazing on moorland (different sheep from clip 1); 7. An individual sheep with mastitis in a lowland field; 8. A ewe, bleating loudly and stamping her feet, observing her lamb

being held and weighted by a research assistant in upland pasture; 9. A ewe lying down dozing in the sun in upland pasture with her two lambs playing around her; 10. An individual ram enclosed in a small indoor individual holding pen (but able to see, hear and touch neighbouring sheep similarly enclosed); 11. A large flock of sheep standing and panting in an outdoor pen after having been moved along country roads for several miles, and 12. A highly alert flock of sheep in an indoor housing pen with deep straw.

Assessors were asked to assess all sheep visible in a clip (see below for more details), with the exception of clips 5, 8, 9, and 10; in these clips assessors were told they should focus on the individual animal central to the clip even though other sheep were visible in certain parts of the clip.

### *2.1.3. QBA fixed list terms*

The QBA terms used for scoring sheep in this study were generated in a previous collaborative study involving the Scottish Agricultural College (SAC), Quality Meat Scotland (QMS), and the Scottish Society for the Prevention of Cruelty to Animals (SSPCA). A team of 10 experienced SSPCA inspectors participated in a Free Choice Profiling (FCP) study, using the same 12 video clips as described above to interpret and score sheep expressions. FCP asks assessors to generate their own descriptors, and to subsequently use these personal descriptors for scoring the expressions of observed animals (Wemelsfelder et al., 2001). The inspectors reached significant agreement on their assessments ( $P < 0.001$ ), and the (unpublished) outcomes of the analysis indicated which of their self-generated terms discriminated best between the variance in observed sheep expressions. Through discussion of these terms the inspectors selected 7 positive and 7 negative terms that they felt adequately covered the main dimensions of sheep expression identified by the analysis and were also practicable for use in the field. These terms were: relaxed, bright, thriving, responsive, satisfied, vigorous, and content; and tense, dejected, dull, low-demeanour, anxious, agitated, and distressed.

This list of 14 terms formed the basis for the QBA training of Group 1, and during training this group decided to remove one negative term (low-demeanour) and one positive term (satisfied) from the list, mainly because they felt these terms were synonymous to other terms on the list and added no extra information. As a result, 12 QBA terms for sheep expression remained and were used for reliability-testing in the present study. These were: relaxed, bright, thriving, responsive, vigorous, and content; and tense, dejected, dull, anxious, agitated, and distressed.

### *2.1.4. Assessment procedures*

Group 1 and Group 2 assessors met together for one day, each group at a different date, at the University of Liverpool School of Veterinary Science. The first half of the day was spent in a three-hour QBA training session led by Françoise Wemelsfelder. This training consisted of an explanation of QBA methodology, including explanations of how to apply QBA to both individual and groups of sheep. This was followed by an hour-long participative group-discussion of the meaning of the 12 terms on the QBA scoring list, and a brief practice of the use of Visual Analogue Scales (VAS) for scoring QBA terms (based on a video clip not used in the study). The actual scoring of clips took place in the second half of the day. The 12 video clips were shown to assessors in the order shown above, and at the end of each clip assessors took however long they needed to score the expressions of the sheep they had just observed on a VAS for each of the 12 QBA terms. Each VAS was 125 mm in length, labelled 'minimum' at one end and 'maximum' at the other end. 'Minimum' was explained as the level where an expression was not present at all, and maximum as the level where an animal could not show an expression more strongly. The distance in between these two points was explained as reflecting the overall intensity of an expression shown by the animal(s) over the length of a clip. Assessors were told not to discuss or disclose their scorings to others during the entire testing period.

## 2.2. Data analysis

The distance in millimetres between the minimum point on the left side of the scale and the mark on the line made by the assessor was measured using a ruler. Data was entered into a Microsoft Excel worksheet and analysed in Minitab version 15.1 (Minitab Inc, State

College, PA) using Principal component Analysis (PCA) with a covariance matrix (no rotation). Three different PCA's were executed: one including the scores of 6 assessors in Group 1, one including the scores of 7 assessors in Group 2, and one including the scores of all 13 assessors in an 'all-assessor' PCA. PCA generates two or more main dimensions describing the variance between sheep expressions, and each clip receives a score on each of these dimensions. Since more than one assessor's scores were included in each PCA in this study, each clip received as many scores on each dimension of a PCA as there were assessors included in that PCA. To determine inter-observer reliability, these clip scores were correlated using Kendall's coefficient of concordance (Martin and Bateson, 2007).

## 3. Results

PCA of scores produced by Group 1, Group 2 and the 'all-assessor' group found for each group two Principal Components with Eigenvalues greater than 1 (see Fig. 1). For the 'all-assessor' analysis the first Principal Component (PC1, explaining 49% of variation between sheep expressions) ranged from 'content/relaxed/bright' to 'distressed/dejected/tense', while PC2 (31%) ranged from 'agitated/responsive/anxious' to 'dull/dejected/relaxed'. For Group 1 (veterinarian students and surgeons) PC1 (52%) ranged from 'bright/content/vigorous' to 'dull/dejected/distressed', while PC2 (28%) ranged from 'anxious/agitated/responsive' to 'relaxed/content/dull'. For Group 2 (soil association inspectors) PC1 (48%) ranged from 'relaxed/content/thriving' to 'distressed/agitated/tense', while PC2 (33%) ranged from 'responsive/bright/vigorous' to 'dull/dejected/relaxed'.

Kendall's coefficient of concordance ( $W$ ) values for PC1 and PC2 scores in Group 1, Group 2, and 'all-assessor' analyses are shown in Table 1. This table shows that  $W$  values for PC1 range from 0.78 to 0.90, and for PC2 from 0.84 to 0.91, with all values significant at  $P < 0.001$ . Thus all three assessor groups reached excellent agreement ( $W > 0.80$ ) on their scores for PC1 and PC2, with the exception of a slightly weaker result for Group 2 on PC1 ( $W = 0.78$ ).

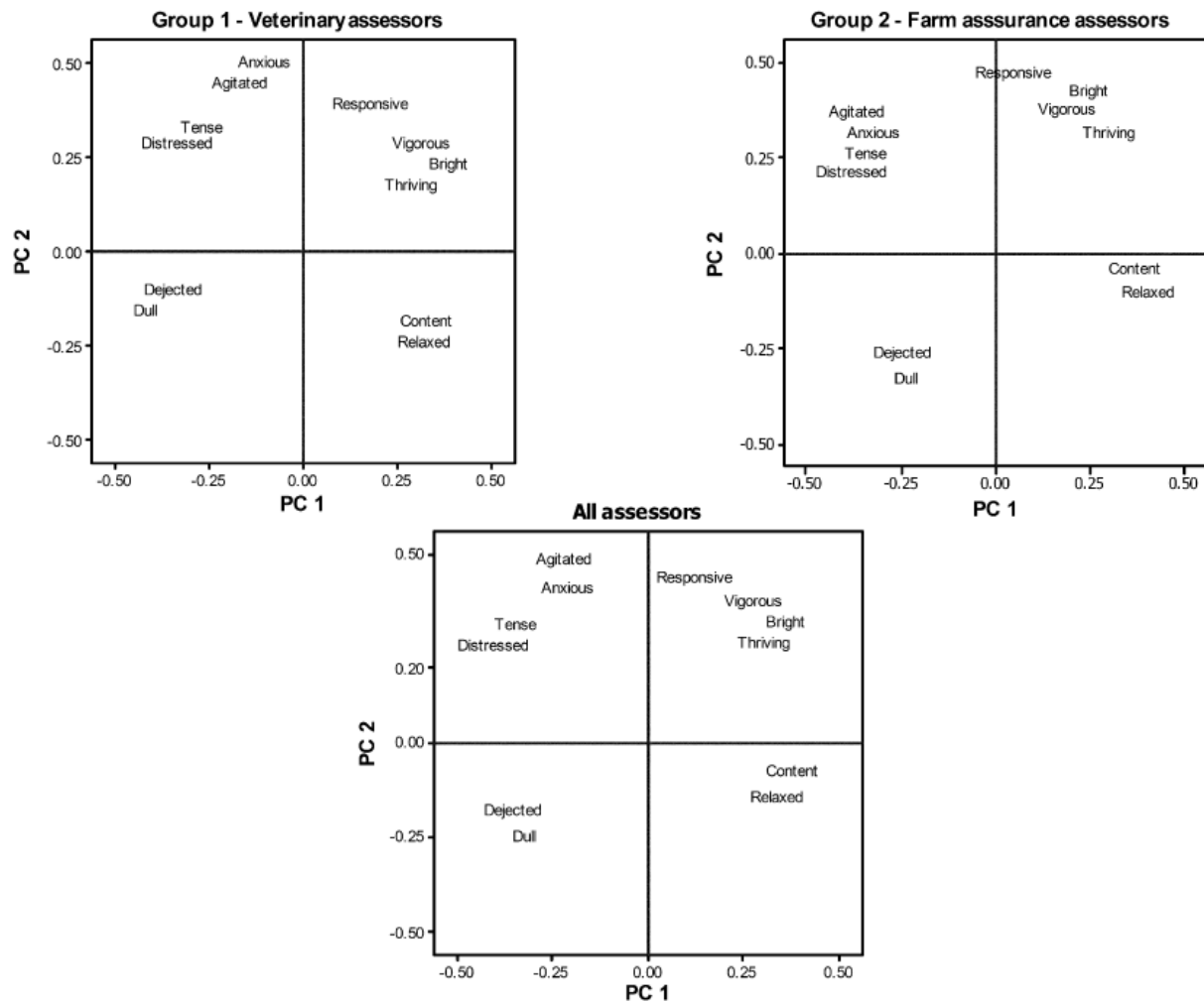
**Table 1. Kendall's coefficient of concordance ( $W$ ) values for scores attributed by assessors to observed sheep on Principal Components 1 and 2.**

Assessors	PC 1	PC 2
Veterinary (Group 1; $n = 6$ )	0.90	0.86
Farm assurance (Group 2; $n = 7$ )	0.78	0.91
All assessors (Groups 1 + 2; $n = 13$ )	0.83	0.84

## 4. Discussion

An important aspect for any measurement tool is that it is consistent when applied by different assessors (Kaler et al., 2009). Accordingly the study presented in this paper tested the level of inter-observer reliability of a fixed-list of 12 QBA terms for sheep that had been previously generated by experienced animal welfare inspectors. Six veterinary and seven farm assurance assessors used the 12 QBA descriptors to score the expressions of sheep observed in 12 video-clips that reflected as many aspects of the expressive repertoire of sheep as could be filmed in a range of intensive and extensive, indoor and outdoor management conditions. The two assessor groups, analysed both separately and together,

identified two highly similar dimensions of sheep expression together explaining more than 70% of the variation, and attributed observed sheep with scores on these dimensions that were highly correlated. Both teams thus showed good inter-observer reliability in their application of the fixed list of QBA terms.



**Fig. 1. QBA term charts for Group 1 (veterinary assessors), Group 2 (farm assurance assessors), and the all-assessor analysis. Axes reflect the level at which QBA terms load on to Principal Components 1 and 2 of each analysis.**

These results concur with the high levels of inter-observer agreement found for application of QBA fixed-list terms to cattle, pigs and poultry (Wemelsfelder and Millard, 2009; Wemelsfelder et al., 2009a,b). They also concur with key qualitative descriptors for sheep demeanour that were identified from the literature by Wemelsfelder and Farish (2004), such as agitated, distressed, responsive and relaxed. Previous QBA studies (e.g. Temple et al., 2011) frequently found two main dimensions of behavioural expression, where the first dimension corresponded to a distinction between positive and negative experience (e.g. enjoying/content vs. frustrated/bored), and the second dimension distinguished between low and high levels of arousal in these experiences (e.g.: calm vs. sociable, listless vs. agitated). The dimensions identified in the present study by the two assessor groups concur with this pattern, with positive and negative descriptors aligning on opposite sides of Principal Component 1, and high-arousal terms (e.g.:

vigorous, agitated) placed on the opposite side to low-arousal terms (e.g.: relaxed, dull) on Principal Component 2. The 12 QBA terms all appear to effectively discriminate between observed sheep, and to coherently characterize the four quadrants of the overall expressive pattern.

The high level of assessor agreement on these dimensions is likely to be the result of the relevance and good understanding of the descriptive terms that were selected. A participative discussion by assessor groups of each of the 12 descriptive fixed-list terms was part of their training sessions, and assessors reported they found this very useful for clarifying the meaning of terms, particularly those with potentially ambiguous meaning. The term 'responsiveness', for example, could be interpreted in terms of physical activity or in terms of an animal's attentiveness to perceived stimuli, and both these meanings could be interpreted in positive and negative ways. This term had been originally selected by the SSPCA inspector team who were familiar with and liked this term for judging certain aspects of sheep demeanour, however this ease of understanding does not necessarily transfer to other assessor groups. It is vital that assessors in a study take time to discuss the terms on a list and develop a common understanding of these terms, or, if necessary, add or delete terms that suit them better. Standardising lists of QBA terms makes this method more feasible to use in practice, but this is only effective if assessors apply these terms in a consistent way. The advantage of a Free Choice Profiling (FCP) approach is that assessors, having generated their own terms, are completely familiar with their meaning. It may be fruitful, if conditions permit, to use FCP to allow teams of assessors to develop and standardise their own fixed list of terms for practical application. Thus we do not propose that the list of sheep QBA terms tested in this study should be considered final, to be adopted without further change by other groups. Rather the approach taken here can be taken as a starting-point for further development and testing of QBA descriptor lists that suit particular goals and situations.

Another important goal of QBA training is to highlight the need for assessors to focus on the overall expressive qualities of behaviour, rather than on specific behavioural features or clinical signs they are familiar with, such as those characterising lameness. It emerged from discussions held by assessors during training that the farm assurance inspectors were more familiar with assessing cattle, pig and poultry than with assessing sheep, and it is possible that this may account to some extent for the difference in agreement levels between the two assessor groups on PC1 and PC2 scores. Farm assurance inspectors showed lower (but still good) agreement on PC1, with this dimension explaining a lower percentage of variation, than was the case for veterinary assessors. By contrast, their agreement on PC2 was higher, and explained a higher percentage of variation, than was the case for veterinarians. It could be that farm assurance inspectors were not quite as much at ease with distinguishing positive and negative expressions in sheep as veterinarians were, and instead focused more on differences in vigour and responsiveness. In addition they may have transposed some of their experience with other livestock species to judging the sheep. It would therefore be a useful part of QBA assessor training to pay explicit attention to an assessor's prior experience with, and interest in, different livestock species.

Testing the inter-observer reliability of QBA terms in the present study was based on video-clip recordings, and this had a number of advantages. Use of a zoom lens permitted close-up and undisturbed recording of the sheep's expressions, producing advantageous angles of observation that were equally available to all assessors and improved the chance of finding agreement. Using video footage reduced the time, effort and resources required to train assessors, and will make it easier in future to conduct similar trainings and reliability studies. However, advantageous as this may be, it is not the situation faced by animal welfare assessors when they go out on farms, where sheep are often away at a distance and walking towards them for a better view is highly likely to disturb them. When a team of assessors visit a farm, it is difficult to control exactly what assessors see of a sheep's demeanour, and it may be more difficult to reach agreement on how they score the sheep. An important next step would



therefore be to examine the inter-observer reliability of QBA fixed-list assessments when these assessments are made out on the farm by a number of experienced assessors. Previous testing of QBA fixed lists for beef cattle and poultry, done for the benefit of developing the WelfareQuality® protocol, was indeed done in on-farm conditions, and produced good results (Wemelsfelder et al., 2009a,b). Similar reliability studies now need to be undertaken for sheep.

## 5. Conclusion

Two groups of assessors with veterinarian and farm assurance backgrounds achieved high levels of inter-observer reliability applying QBA fixed-list terms to the assessment of a wide range of sheep expressions shown in a variety of conditions. The present study was video-based, and an important next step is to extend testing to less controlled on-farm conditions. This paper is the first to be published on applying QBA to sheep, and concurs with positive results found for other livestock species. QBA has the potential to contribute valuable information to assessments of farm animal quality of life, and to validate this potential for sheep it will be necessary to cross-validate QBA outcomes with other indicators of health and welfare measured on farm.

## Acknowledgements

This study was funded as part of the Defra AW1025 research project – development of indicators for the on-farm assessment of sheep welfare. We gratefully acknowledge Quality Meat Scotland (QMS) and the Scottish Society for Prevention of Cruelty to Animals (SSPCA) for sharing their fixed-list QBA terms for sheep, and also wish to thank the Soil Association farm assurance inspectors for giving their time to participate in this study.

## References

- Berriatua, E., French, N.P., Broster, C.E., Morgan, K.L., Wall, R., 2001. Effect of infestation with *Psoroptes ovis* on the nocturnal rubbing and lying behaviour of housed sheep. *Appl. Anim. Behav. Sci.* 71, 43–55.
- Boissy, A., Manteuffel, G., Jensen, M.B., Moe, R.O., Spruijt, B., Keeling, L.J., Winckler, C., Forkman, B., Dimitrov, I., Langbein, J., Bakken, M., Veissier, I., Aubert, A., 2007. Assessment of positive emotions in animals to improve their welfare. *Physiol. Behav.* 92, 375–397.
- Boivin, X., Lensink, J., Tallet, C., Veissier, I., 2003. Stockmanship and farm animal welfare. *Anim. Welfare* 12, 479–492.
- Brscic, M., Wemelsfelder, F., Tessitore, E., Gottardol, F., Cozzi, G., Van Reenen, C.G., 2009. Welfare assessment: correlations and integration between a Qualitative Behavioural Assessment and a clinical/health protocol applied in veal calves farms. *Ital. J. Anim. Sci.* 8, 601–603.
- Dwyer, C.M., 2009. Welfare of sheep: providing for welfare in an extensive environment. *Small Rumin. Res.* 86, 14–21.
- FAWC, 1994. Report on the welfare of sheep. PB 1755. Farm Animal Welfare Council (FAWC) Publication, London.
- FAWC, 2009. Report on Farm Animal Welfare in Great Britain: Past, Present and Future. Farm Animal Welfare Council (FAWC) Publication, London. <http://www.fawc.org.uk/pdf/ppf-report091012.pdf>
- Gougoulis, D.A., Kyriazakis, I., Fthenakis, G.C., 2010. Diagnostic significance of behaviour changes of sheep: a selected review. *Small Rumin. Res.* 92, 52–56.
- Kaler, J., Wassink, G.J., Green, L.E., 2009. The inter- and intra-observer reliability of a locomotion scoring scale for sheep. *Vet. J.* 180, 189–194.

- Lynch, J.J., Hinch, G.N., Adams, D.B., 1992. *The Behaviour of Sheep: Biological Principles and Implications for Production*. CAB International, Wallingford.
- Martin, P.R., Bateson, P., 2007. How good are your measures. In: *Measuring Behaviour: An Introductory Guide*, third ed. Cambridge University Press, Cambridge, pp. 72–73.
- Meagher, R.K., 2009. Observer ratings: validity and value as a tool for animal welfare research. *Appl. Anim. Behav. Sci.* 119, 1–14.
- Minero, M., Tosi, M.V., Canali, E., Wemelsfelder, F., 2009. Quantitative and qualitative assessment of the response of foals to the presence of an unfamiliar human. *Appl. Anim. Behav. Sci.* 116, 74–81.
- Napolitano, F., Knierim, U., Grasso, F., De Rosa, G., 2009a. Positive indicators of cattle welfare and their applicability to on-farm protocols. *Ital. J. Anim. Sci.* 8, 355–365.
- Napolitano, F., De Rosa, G., Braghieri, A., Grasso, F., Bordi, A., Wemelsfelder, F., 2008. The qualitative assessment of responsiveness to environmental challenge in horses and ponies. *Appl. Anim. Behav. Sci.* 109, 342–354.
- Napolitano, F., De Rosa, G., Ferrante, V., Grasso, F., Braghieri, A., 2009b. Monitoring the welfare of sheep in organic and conventional farms using an ANI 35 L derived method. *Small Rumin. Res.* 83, 49–57.
- Phythian, C.J., Michalopoulou, E., Jones, P.H., Winter, C.J., Clarkson, M.J., Stubbings, L.A., Grove-White, D., Cripps, P.J., Duncan, J.S., 2011. Validating indicators of sheep welfare through a consensus of expert opinion. *Animal* 5, 943–952.
- Phythian, C.J., Cripps, P.J., Michalopoulou, E., Jones, P.H., Grove-White, D., Clarkson, M.J., Winter, A.C., Stubbings, L.A., Duncan, J.S. 2012. Reliability of indicators of sheep welfare assessed by a group observation method. *Vet. J.* <http://dx.doi.org/10.1016/j.tvjl.2011.12.006>
- Rousing, T., Wemelsfelder, F., 2006. Qualitative assessment of social behaviour of dairy cows housed in loose housing systems. *Appl. Anim. Behav. Sci.* 101, 40–53.
- Rutherford, K.M.D., Donald, R.D., Lawrence, A.B., Wemelsfelder, F., 2012. Qualitative Behavioural Assessment of emotionality in pigs. *Appl. Anim. Behav. Sci.* 139, 218–224.
- Stockman, C.A., Collins, T., Barnes, A.L., Miller, D.S., Wickham, L., Beatty, D., Blache, C., Wemelsfelder, F., Fleming, P.A., 2011. Qualitative behavioural assessment and quantitative physiological measurement of cattle naïve and habituated to road transport. *Anim. Prod. Sci.* 51, 240–249.
- Stubsjøen, S.M., Hektoen, L., Valle, P.S., Janczak, A.M., Zanella, A.J., 2011. Assessment of sheep welfare using on-farm registrations and performance data. *Anim. Welfare* 20, 239–251.
- Temple, D., Manteca, X., Velarde, A., Dalmau, A., 2011. Assessment of animal welfare through behavioural parameters in Iberian pigs in intensive and extensive conditions. *Appl. Anim. Behav. Sci.* 131, 29–39.
- Walker, J., Dale, A., Waran, N., Clarke, N., Farnworth, M., Wemelsfelder, F., 2010. The assessment of emotional expression in dogs using a Free Choice Profiling methodology. *Anim. Welfare* 19, 75–84.
- Wemelsfelder, F., 2007. How animals communicate quality of life: the qualitative assessment of behaviour. *Anim. Welfare* 16, 25.
- Wemelsfelder, F., Farish, M., 2004. Qualitative categories for the interpretation of sheep welfare: a review. *Anim. Welfare* 13, 261–268.
- Wemelsfelder, F., Hunter, E.A., Mendl, M.T., Lawrence, A.B., 2000. The spontaneous qualitative assessment of behavioural expressions in pigs: first explorations of a novel methodology for integrative animal welfare measurement. *Appl. Anim. Behav. Sci.* 67, 193–215.
- Wemelsfelder, F., Hunter, T.E.A., Mendl, M.T., Lawrence, A.B., 2001. Assessing the ‘whole animal’: a Free-Choice Profiling approach. *Anim. Behav.* 62, 209.
- Wemelsfelder, F., Knierim, U., Schulze Westerath, H., Lentfer, T., Staack, M., Sandilands, V., 2009a. Qualitative behaviour assessment. In: *Assessment of animal welfare measures for layers and broilers*. Welfare Quality reports No. 9, Cardiff, pp. 113–119.

- Wemelsfelder, F. and Millard, F. 2009. Qualitative behaviour assessment. In: Forkman, B. and Keeling, L. (eds): Assessment of Animal Welfare Measures for Sows, Piglets and Fattening Pigs. Welfare Quality Reports No. 10, Sixth Framework Programme. University of Cardiff, Cardiff. pp. 213–219.
- Wemelsfelder, F., Millard, F., De Rosa, G., Napolitano, F., 2009b. Qualitative behaviour assessment. In: Assessment of animal welfare measures for dairy cattle, beef bulls and veal calves. WelfareQuality® Reports No. 11, Cardiff, pp. 215–224.
- Wemelsfelder, F., Nevison, I., Lawrence, A.B., 2009c. The effect of perceived environmental background on qualitative assessments of pig behaviour. *Anim. Behav.* 78, 477–484.
- Wemelsfelder, F., Hunter, A.E., Paul, E.S., Lawrence, A.B., 2012. Assessing pig body language: agreement and consistency among pig farmers, veterinarians, and animal activists. *J. Anim. Sci.* 90, 1–14.
- Wiseman-Orr, M.L., Scott, E.M., Reid, J., Nolan, A.M., 2006. Validation of a structured questionnaire as an instrument to measure chronic pain in dogs on the basis of effects on health-related quality of life. *Am. J. Vet. Res.* 67, 1826–1836.