

WellBeing International  
**WBI Studies Repository**

7-2016

## On-Farm Qualitative Behaviour Assessment of Dairy Goats in Different Housing Conditions

Lilia Grosso  
*University of Milan*

Monica Battini  
*University of Milan*

Françoise Wemelsfelder  
*Scotland's Rural College*

Sara Barbieri  
*University of Milan*

Michela Minero  
*University of Milan*

*See next page for additional authors*

Follow this and additional works at: [https://www.wellbeingintludiesrepository.org/acwp\\_asie](https://www.wellbeingintludiesrepository.org/acwp_asie)



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

---

### Recommended Citation

Grosso, L., Battini, M., Wemelsfelder, F., Barbieri, S., Minero, M., Dalla Costa, E., & Mattiello, S. (2016). On-farm Qualitative Behaviour Assessment of dairy goats in different housing conditions. *Applied Animal Behaviour Science*.

This material is brought to you for free and open access by WellBeing International. It has been accepted for inclusion by an authorized administrator of the WBI Studies Repository. For more information, please contact [wbisr-info@wellbeingintl.org](mailto:wbisr-info@wellbeingintl.org).



---

**Authors**

Lilia Grosso, Monica Battini, Françoise Wemelsfelder, Sara Barbieri, Michela Minero, Emanuela Dalla Costa, and Silvana Mattiello

# On-Farm Qualitative Behaviour Assessment of Dairy Goats in Different Housing Conditions

Lilia Grosso<sup>1</sup>, Monica Battini<sup>1</sup>, Françoise Wemelsfelder<sup>2</sup>, Sara Barbieri<sup>1</sup>, Michela Minero<sup>1</sup>, Emanuela Dalla Costa<sup>1</sup>, Silvana Mattiello<sup>1</sup>

<sup>1</sup> University of Milan

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

## KEYWORDS

Qualitative Behaviour Assessment, dairy goat, inter-observer reliability, indoor housing, pasture, animal welfare

## ABSTRACT

*This study reports the results of the first investigation on the use of Qualitative Behaviour Assessment (QBA) in dairy goats, using a fixed-list of descriptors specifically developed for this species. It aimed to verify whether QBA can be reliably used by observers with different backgrounds to differentiate between the emotional states of goats kept under different environmental conditions. Two trained observers simultaneously assessed 16 dairy goat farms (8 "Housed" (H) farms, where animals were observed in free stall pens, and 8 "Pasture" (P) farms, where animals were observed in open ranges), using a list of 16 QBA descriptors that were based on literature studies and discussed within a focus group of goat experts. One H farm was removed from analysis due to procedural error. The QBA scores were analysed together using Principal Component Analysis (PCA, correlation matrix, no rotation). Observer agreement for farm scores on PCA Components (PCs) and on separate QBA terms was investigated using Pearson and Spearman correlations respectively. The effects of housing system and observer on PC scores were analysed using analysis of variance (treatments = observer, housing system, and their interaction; block = farm). PCA identified three main components explaining 60.87% of the total variation between goat farms: PC1 (29.04%) ranged from "content/calm" to "frustrated/aggressive", suggesting a relationship to the animals' general mood; PC2 (19.70%) ranged from "curious/attentive" to "calm/bored", suggesting a relationship to the animals' level of arousal, and PC3 (12.13%) ranged from "sociable/playful" to "alert/agitated". The two observers showed a good level of agreement on the three PCA dimensions (PC1:  $r = 0.75$ ,  $P = 0.001$ ; PC2:  $r = 0.67$ ,  $P = 0.006$ ; PC3:  $r = 0.69$ ,  $P = 0.004$ ), and also on 7 out of 16 separate QBA descriptors ( $P < 0.05$ ). Two additional descriptors showed a moderate level of agreement ( $P = 0.10$ ). These results indicate an integrated PCA approach to QBA to be more robust. There were significant effects of housing system on both PC1 (ANOVA;  $P = 0.05$ ) and PC2 ( $P = 0.02$ ), indicating goats on P farms to be more "content/calm" and "curious/attentive" than goats on H farms. There was a significant observer effect on PC2 ( $P = 0.04$ ), and a significant observer by housing interaction on PC3 ( $P = 0.009$ ). In sum, these results suggest that QBA can be a reliable welfare indicator, used by observers with different backgrounds; however, further development of QBA training procedures is required to extend inter-observer reliability to all main expressive dimensions emerging from the analysis.*

## 1. Introduction

During the last decade, the assessment of animal welfare at farm level has received increasing attention, in response to consumer demand for assurance schemes of high quality animal products, including animal welfare. Most of the indicators developed for welfare assessment have been focused on negative aspects, and a need to develop more positive welfare indicators has been identified by several recent reviews (Boissy et al., 2007; Yeates and Main, 2008). Particularly the inclusion of positive qualitative indicators may play a key role in the communication of animal welfare to stakeholders, and therefore deserves further attention (FAWC, 2009). In a recent review on animal-based welfare indicators for dairy goats, Battiniet al. (2014) identified Qualitative Behaviour Assessment (QBA) as a promising approach to evaluate positive emotional state in this species.

QBA is a “whole-animal” method for evaluating the expressive quality of animal behaviour, using qualitative descriptors such as “tense”, “content”, or “relaxed” (Wemelsfelder et al., 2000, 2001; Wemelsfelder, 2007). Such descriptors have an emotional connotation, and can give information that is directly relevant to animal welfare, and complements the information provided by quantitative welfare indicators (Wemelsfelder et al., 2001). QBA offers advantages in terms of on-farm feasibility, in that it does not require any restraint or intervention in the lives of animals, can be applied at herd-level, and, once on farm, is not time-costly or labour-intensive. The inter-observer reliability and biological validity of QBA applied under controlled experimental conditions have been well-documented for a range of species (e.g. Stockman et al., 2011; Rutherford et al., 2012; Wemelsfelder and Mullan, 2014); however, the on-farm use of pre-fixed QBA term lists, such as during on-farm animal welfare inspections, requires further development and validation.

Good on-farm observer agreement has been reported for laying hens (Wemelsfelder et al., 2009), beef cattle (Wemelsfelder et al., 2009; Wemelsfelder and Millard, 2009), dairy cattle (Andreasenet al., 2013), dairy buffalo (De Rosa et al., 2015), and donkeys (Minero et al., 2016). Studies of on-farm video footage showed good observer agreement for sheep (Phythian et al., 2013), but not for dairy cattle (Bokkers et al., 2012; Gutmann et al., 2015). Few studies as yet have correlated on-farm QBA assessments to other measures taken on farm. Andreasen et al. (2013) did not find QBA assessments of Danish dairy farms to correlate to any outcomes of the Welfare Quality® protocol applied on the same farms a few days later, however Phythian et al. (2016) did find a good correlation between QBA assessments made on UK sheep farms and the proportion of lame sheep on these same farms as determined at a later point in time. On-farm QBA of sheep flocks on a range of farm types was found to be highly consistent across 6 visits in a year (Phythian et al., 2016); however, a video study by Gutmann et al. (2015) found the general mood of dairy cattle to vary significantly across different times of day, raising concerns about the need to standardise on-farm assessment times.

There is thus a need for further studies on the use of QBA in on-farm welfare assessment: the present study focuses on QBA assessment of dairy goats in different housing conditions, and tests the inter-observer agreement between two assessors with different professional background and experience. To date, only Muriet al. (2013) have reported a first attempt to apply QBA to goats: QBA was included in a comprehensive welfare assessment protocol for intensively farmed dairy goats. It was applied at group level and consisted of five descriptors (resting, aggressive, inquisitive/interested, fearful, calm/indifferent), which were modified from the terms used in the Welfare Quality® Protocol for dairy cows. This study was encouraging in that it found some significant correlations between QBA descriptors and health indicators and stockmanship. However, the Muri study used a limited number of descriptors, which did not adequately cover the expressive repertoire of goats. The goal of the current study was to apply QBA to goats in either pasture or housed management systems, and test its reliability for use by assessors from differing backgrounds.

## **2. Materials and methods**

### *2.1. Experimental design*

#### *2.1.1. Development of QBA fixed list descriptors for dairy goats*

A pre-fixed list of descriptors was used for this study, as this approach is considered more feasible for applying QBA to practical on-farm welfare assessment than the original Free-Choice Profiling approach, in which each observer generates his/her own descriptors (Wemelsfelder et al., 2009). The existing scientific literature on goat behaviour and welfare was reviewed in order to identify a list of potential QBA descriptors for dairy goats. A list of 32 descriptors was produced and then discussed in April 2013 by a panel of 10 Italian goat experts (farmers, veterinarians, technicians and researchers). This panel removed 20 descriptors from the list, either because they were considered too prone to anthropomorphism (e.g. angry), too generic (e.g. active), or too similar to other terms (e.g., explorative-curious, agitated-nervous, calm-relaxed). Furthermore, the term “interested” was replaced by “attentive”, and two new terms (bored, irritated) were added. Once the Italian goat experts had reached agreement on the use and definition of 15 descriptors, the discussion was extended to international level, involving nine goat experts engaged in the European Animal Welfare Indicators (AWIN) project, who added one new attribute (suffering).

The final list of descriptors thus included 16 fixed terms: aggressive, agitated, alert, bored, apathetic, attentive, content, curious, frustrated, playful, irritated, fearful, sociable, suffering, calm, and lively. For each descriptor a brief definition was provided in order to facilitate its interpretation by different observers (Table 1).

#### *2.1.2. Farms and animals*

QBA was applied on 16 Italian commercial dairy goat farms. In all farms, animals were housed in pens with straw litter, while in 8 farms goats had free access to pasture from spring to autumn. These 8 farms (“Pasture”, P) were assessed outdoors, at pasture, whereas the other 8 farms (“Housed”, H) were assessed indoors. All farms were visited in May 2013. Observations were performed on all lactating animals. Only farms with more than 30 female adult goats were selected (mean  $91.0 \pm 80.7$ , min 38, max 370 lactating goats). Before farm visits, the farmers were contacted and received basic information about the research.

#### *2.1.3. Observers*

Two independent observers conducted the QBA observations on-farm. Observer A (Obs-A) was a female veterinarian with work experience in extensive and organic goat practices. Observer B (Obs-B) was a female post-doctoral animal scientist, specialized in farm animal welfare, and familiar with intensive dairy goat farms. Neither of these observers had previous experience with QBA. Furthermore neither observer was familiar with any of the 16 selected farms, and so their judgment could not be biased by any previously conceived views regarding a specific farm.

Before starting the on-farm data collection, both observers received training in applying QBA to goats at group level. To practice, they both scored 15 video clips of two minutes each, showing groups of goats in different situations, using the 16 descriptors. After watching and scoring each video, the assessors compared their scores for the different terms, and discussed any discrepancies. They would then each watch and score the videos again, and repeat this process, until agreement on the interpretation and quantification of each descriptor was reached.

**Table 1. List and characterisation of the 16 QBA descriptors for dairy goats.**

Descriptor	Definition
AGGRESSIVE	An aggressive goat bites other goats (especially the ears), voluntarily attacks or threatens other goats with the intention of hurting or disturbing them, butts the belly or the head of other goats. She is intentionally noxious to other goats. The aggressive behaviour can be related to dominance, fear or resource protection
AGITATED	An agitated goat is restless, not at ease, highly reactive, she can move her ears, vocalize or nervously move around.
ALERT	An alert goat is ready to react to a potential danger or to something that frightens her. She can emit acoustic or visual alarm signals (e.g. vocalizations, snorts, stamping, ears in upright position, stiff body)
APATHETIC	An apathetic goat shows little or no movements or reactions to stimuli and often remains isolated from the group, depressed
ATTENTIVE	An attentive goat is concentrated on something that is happening or is going to happen, waiting for an event, she looks around but often concentrates her gaze towards a specific direction or signal
BORED	A bored goat is wearied, dull, she is uninterested in the surrounding environment, feeling tired of something that has continued for too long; lack in stimulation; she may be looking for something to do
CALM	A calm goat is quiet, relaxed and she feels at ease
CONTENT	A content goat is appeased, gratified, happy, comfortable, at ease, satisfied about the situation, positively engaged in something
CURIOUS	A curious goat is explorative, intrigued by something, attracted by the surrounding environment and by novelties (e.g. people, goats in oestrus, objects), engaged in exploratory behaviour
FEARFUL	A fearful goat is a scared and shy animal. She may look for shelter or for a way out and crouches down or may tend to hide in the middle of the group. A whole group may run around
FRUSTRATED	A frustrated goat is annoyed and impatient because prevented from achieving something (e.g. queuing at feeding rack or water place, passive behaviour)
IRRITATED	An irritated goat is bothered or annoyed by something (e.g. flies, pruritus, noise) that can disturb or upset or trouble or exasperate her
LIVELY	A lively goat is busy in different activities. She is active, animated, full of life and energy
PLAYFUL	A playful goat jumps, performs ritualized non-aggressive fights (sparring), plays and makes noise with objects, climbs or tries to climb. They stimulate each other and laterally run together
SOCIABLE	A sociable goat is friendly with other goats. She has affiliative (e.g. grooming, sniffing, resting in pairs) and playful contacts with other goats
SUFFERING	A suffering goat is feeling pain, often with contracted muscles, possibly in antalgic postures

#### 2.1.4. Data collection

On farm, the QBA assessment was performed by direct observation carried out during an activity period of the goats. In H farms, goats were observed in their home pen 60 min after feed distribution; whereas in P farms goats were observed in open pasture. All observations were performed 60 min before or after milking procedures. The two observers assessed the goats independently and simultaneously, without interfering with each other or the animals. The assessment was always performed on the whole herd and not on individual animals.

**Table 2. Time taken at each observation point depending on the number of points selected for a farm.**

Number of observation points	Duration of observation from each point (min)	Total observation time (min)
1	10	10
2	5	10
3	6.5	19.5
4	5	20
5	4	20
6	3	18
7	2.5	17.5
8	2.5	20

The QBA on-farm assessment procedures followed those developed for the Welfare Quality® protocol (Welfare Quality®, 2009). Observation of animals was carried out from one or more locations around the pen or grazing area (observation points) from which a good view of animals and environment was possible. Between one and eight observation points were selected for each farm by Obs-B, depending on the size of the herd and the complexity of the housing environment. Total observation time for each farm ranged from a minimum of 10 to a maximum of 20 min, depending on the number of observation points selected. If only one observation point was necessary for having a good view of the herd, observation time was 10 min; if two or more observation points were required, total observation time was never more than 20 min, with the time spent at each point proportionate to the number of points selected (Table 2). While selecting observation points, observers spent some time moving around pens or grazing area, to allow the animals to become familiar with them. During observation, the farmer was asked to keep out of the goats' sight, to avoid influencing their behaviour.

When observation of animals from all selected points was completed, the two observers scored the 16 descriptors using Visual Analogic Scales (VAS) embedded in a specific QBA application for Android devices, developed by Scotland's Rural College (SRUC, Edinburgh). Thus one assessment was made for each farm, integrating all observations made at that farm from the different observation points. To score each term, observers touched the tablet screen across the VAS at the appropriate point. Each VAS ranged from 0 mm (this expressive quality is absent) to 125 mm (this quality could not be present more strongly). The measure for each term was the distance in millimetres from 0 to the point where the VAS was touched. Thus, for each observer and each farm, an excel data spreadsheet was automatically created containing the scores of assessed farms on each of the 16 descriptors. To ensure the independence of each observer, silence was strictly maintained during observation and scoring procedures.

## *2.2. Statistical analysis*

Data were analysed using SPSS version 20 (IBM SPSS Inc., Chicago, USA). Due to procedural errors in executing the assessment, one H farm was removed from the data set. The QBA scores gathered by the two observers for the remaining 15 goat farms were analysed together, using Principal Component Analysis (PCA, correlation matrix, no rotation). Farm scores on the three main Principal Components (PCs) were normally distributed, and so to calculate inter-observer agreement, the scores generated by Obs-A and Obs-B on each of these PCs were correlated using Pearson correlations ( $r$ ). In addition, agreement in how farms were ranked on each of the 16 QBA descriptors was calculated using Spearman

correlations ( $\rho$ ). The effects of housing system, observer, and their interaction, on farm PC scores were analysed using two-way analyses of variance, blocked for farm.

### 3. Results

PCA of the QBA scores for 15 goat farms identified five main factors with eigenvalues greater than 1 (4.65, 3.15, 1.94, 1.52 and 1.10 for PC1, PC2, PC3, PC4 and PC5, respectively). The first three Components together explained 60.87% of the variation between farms (29.04, 19.70 and 12.13 for PC1, PC2 and PC3, respectively). Fig. 1 shows the distribution of the descriptors along the first two PCA factors. The first Principal Component accounted for 29.04% of the total variance and, ranging from “content/calm” to “frustrated/aggressive”, appears to distinguish generally between positive and negative mood. PC2, accounting for 19.70% of the total variance, seems related to the level of arousal/attention in the animals, ranging from “curious/attentive” to “calm/bored”. The third Component, explaining 12.13% of the total variance, ranged from “sociable/playful” to “alert/agitated”, and seems to characterise the goats’ level of sociable interaction (Fig. 2; Table 3).

Table 4 shows the correlations between farm scores generated by the two observers on the three PCs, indicating a good level of inter-observer reliability.

Table 5 shows the correlations between the two observers’ scores on separate QBA terms, indicating that seven out of 16 terms were significantly correlated ( $P < 0.05$ ), and an additional two terms approached significant correlation: “aggressive” ( $P = 0.102$ ) and “attentive” ( $P = 0.103$ ).

The housing system had a significant effect on farm scores on PC1 ( $F_1 = 4.45$ ;  $P = 0.05$ ) and PC2 ( $F_1 = 6.85$ ;  $P = 0.02$ ). Goats observed at pasture obtained significantly higher scores on both PCs, indicating that goats in P farms were assessed as having a more positive emotional state than goats in H farms. A significant observer effect was found on PC2 ( $F_1 = 5.45$ ;  $P = 0.04$ ): Obs-A attributed on average 29.04% of the total variance and, ranging from “content/calm” to “frustrated/aggressive”, appears to distinguish generally between positive and negative mood. PC2, accounting for 19.70% of the total variance, seems related to the level of arousal/attention in the animals, ranging from “curious/attentive” to “calm/bored”. The third Component, explaining 12.13% of the total variance, ranged from “sociable/playful” to “alert/agitated”, and seems to characterise the goats’ level of sociable interaction (Fig. 2; Table 3).

Table 4 shows the correlations between farm scores generated by the two observers on the three PCs, indicating a good level of inter-observer reliability.

Table 5 shows the correlations between the two observers’ scores on separate QBA terms, indicating that seven out of 16 terms were significantly correlated ( $P < 0.05$ ), and an additional two terms approached significant correlation: “aggressive” ( $P = 0.102$ ) and “attentive” ( $P = 0.103$ ).

The housing system had a significant effect on farm scores on PC1 ( $F_1 = 4.45$ ;  $P = 0.05$ ) and PC2 ( $F_1 = 6.85$ ;  $P = 0.02$ ). Goats observed at pasture obtained significantly higher scores on both PCs, indicating that goats in P farms were assessed as having a more positive emotional state than goats in H farms. A significant observer effect was found on PC2 ( $F_1 = 5.45$ ;  $P = 0.04$ ): Obs-A attributed on average significantly higher scores of “curious/attentive-ness” to goats than Obs-B. Furthermore, there was a significant observer by housing interaction on PC3 ( $F_1 = 9.26$ ;  $P = 0.009$ ). Obs-A attributed significantly higher scores of “sociable/playful-ness” to P farms than H farms, whereas for Obs-B it was the opposite (higher scores for H farms than P farms).



Fig. 1. Biplot showing term loadings and farm scores on PC1 and PC2. □ Pasture farms, Obs-A; ■ Housed farms, Obs-A; △ Pasture farms, Obs-B; ▲ Housed farms, Obs-B.

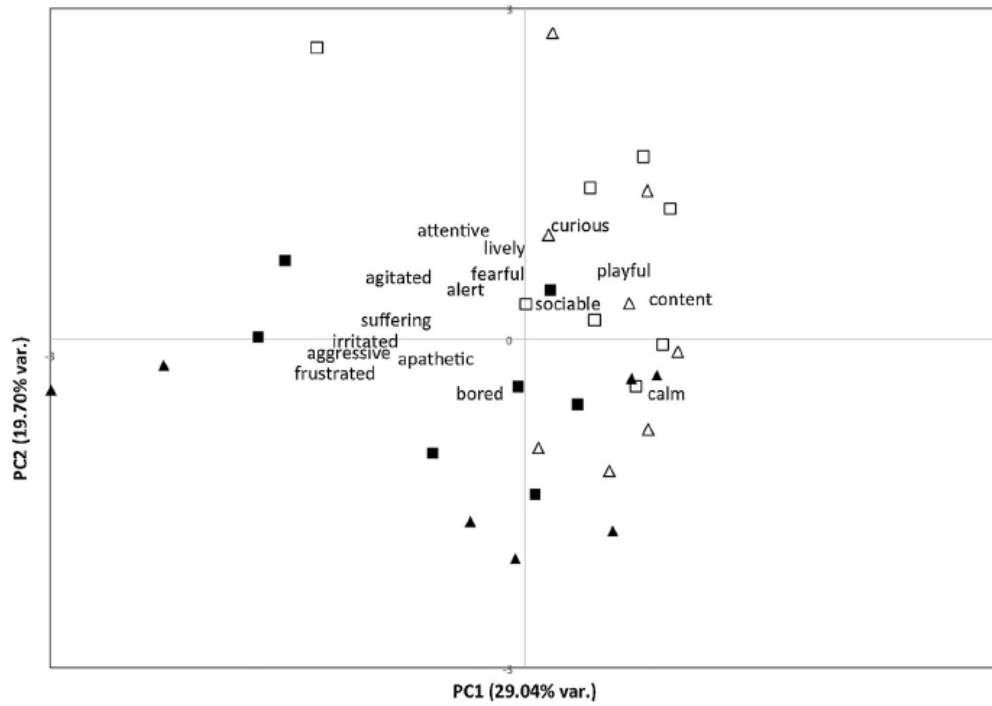
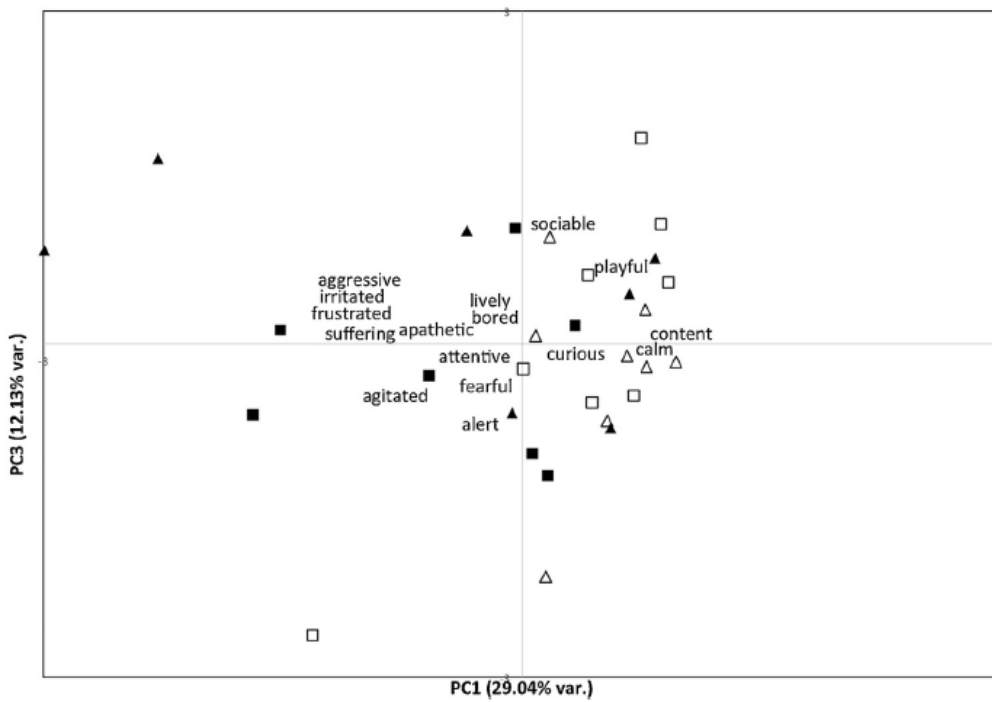


Fig. 2. Biplot showing term loadings and farm scores on PC1 and PC3. □ Pasture farms Obs-A; ■ Housed farms, Obs-A; △ Pasture farms, Obs-B; ▲ Housed farms, Obs-B.



**Table 3. PCA of the QBA descriptors. Loadings with positive or negative values higher than 0.6 are typed in bold.**

DESCRIPTOR	PC1	PC2	PC3
Aggressive	<b>-0.73</b>	-0.06	0.39
Agitated	<b>-0.64</b>	0.48	-0.31
Alert	-0.33	0.46	<b>-0.64</b>
Bored	-0.21	-0.37	0.20
Apathetic	<b>-0.61</b>	-0.09	0.15
Attentive	-0.27	<b>0.74</b>	-0.10
Content	<b>0.79</b>	0.34	0.12
Curious	0.23	<b>0.76</b>	-0.07
Frustrated	<b>-0.77</b>	-0.14	0.20
Playful	0.40	0.51	0.59
Irritated	<b>-0.73</b>	-0.03	0.36
Fearful	-0.23	0.52	-0.26
Sociable	0.08	0.37	<b>0.72</b>
Suffering	<b>-0.69</b>	0.09	0.04
Calm	<b>0.77</b>	-0.37	-0.08
Lively	0.20	<b>0.71</b>	0.29

**Table 4. Pearson correlations between farm scores generated by the two observers on each PC.**

PCs	Pearson Correlation r	P value
PC1 Obs-A and PC1 Obs-B	0.75	0.001
PC2 Obs-A and PC2 Obs-B	0.67	0.006
PC3 Obs-A and PC3 Obs-B	0.69	0.004

#### 4. Discussion

The aim of the present study was to determine whether two trained observers with different backgrounds could successfully apply QBA to assess the expressive demeanour of dairy goats kept in either “Housed” or “Pasture” conditions on 15 goat farms. PCA identified three main components: PC1: content/calm-frustrated/aggressive; PC2: curious/attentive-calm/bored, and PC3: sociable/playful-alert/agitated, which together explained 60.87% of the variation between farms. ANOVA of farm PC scores found that the goats’ demeanour on Housed and Pasture farms differed along both PC1 and PC2, indicating Pasture goats to display more “content/calm” and “curious/attentive” demeanour than Housed goats. This finding demonstrates the efficacy of QBA in differentiating goats’ demeanour in different housing systems, and supports the hypothesis that pasture allows animals to engage with their environment more positively (e.g., Casamassima et al., 2001).

These results are in agreement with previous studies which found QBA to discriminate between different housing systems, for example in pigs (Temple et al., 2011), and dairy cattle (Popescuet et al., 2014). Although animals in extensive systems face a range of welfare challenges (e.g., variability in climate conditions, parasitic diseases; Sevi et al., 2009; Dwyer, 2009; Goddard et al., 2006; Goddard, 2013), the present study suggests that such systems were generally experienced by goats more positively than

housed systems, with grazing goats appearing more content and calm than animals kept indoors. Research in ruminants has shown that access to pasture enhances complex natural behaviour patterns, such as exploratory and interactive behaviours (Casamassima et al., 2001), while housed conditions can prevent the expression of grazing behaviour (Dwyer, 2009; Braghieri et al., 2011). Housed farms received higher scores for negative moods such as “aggressive”, “irritated” and “suffering”, which may be due to the restricted availability of space for behavioural expression (Miranda-de la Lama and Mattiello, 2010), as stocking density in indoor housing systems tended to be higher than on pasture. It is however not clear a priori what might be the main factors affecting the animals’ emotional state in the different systems; in addition to whether or not animals had access to outdoor pasture, concomitant factors such as stocking density, feeding schedules, and quality of handling procedures may have contributed to the animal’s state. Some Housed farms appeared on the positive side of PC1, indicating that it is possible for such farms to achieve positive emotional states in goats similar to those observed in Pasture farms, and that therefore a simple indoor/outdoor dichotomy does not apply. It should also be noted that the findings of this study may have been affected by incidental factors such as time of day at which observations were carried out (Gutmann et al., 2015), and weather conditions at the time of observation. It is generally a concern in the application of animal-based indicators that they may vary across different time periods and environmental conditions, making it advisable to standardize on-farm assessment protocols as much as possible, and to repeat assessments across the seasons of the year (Phythian et al., 2016).

**Table 5. Spearman rank correlations between farm scores generated by the two observers for each of the 16 descriptors separately.**

Descriptors	Spearman’s correlation coefficients ( $\rho$ )	P value
Aggressive	0.44	0.102
Agitated	0.40	0.135
Alert	0.57	0.027
Bored	0.11	0.689
Apathetic	0.40	0.139
Attentive	0.44	0.103
Content	0.78	0.001
Curious	0.05	0.849
Frustrated	0.56	0.019
Playful	0.73	0.002
Irritated	0.30	0.277
Fearful	0.37	0.176
Sociable	0.69	0.004
Suffering	0.82	0.000
Calm	0.09	0.741
Lively	0.83	0.000

The farm scores generated by the two observers for the three main PCs were all correlated above 0.67, indicating good agreement in the relative ranking of farms on expressive dimensions (Martin and Bateson, 2007). Closer inspection of correlations between observers’ scores for separate terms indicated that they agreed well on the use of some terms, but not of others, a finding also reported by other studies (e.g. Wemelsfelder et al., 2009; Bokkers et al., 2012). However, as QBA is fundamentally a “whole animal” approach, addressing expressive patterns of demeanour rather than separate emotional states, the aim is not to reduce terms to the smallest possible number on the basis of univariate analysis, but to create a list

of positive and negative terms that together reliably produce meaningful dimensions of expression. To optimize the robustness of those dimensions, it is nevertheless important to note which terms achieve low reliability, and to consider how this may be improved. A primary way of improving a term's reliability is to revise how it is characterized. Such characterization however is not the same as defining categories of physical behavior in a conventional ethogram. Qualitative terms do not primarily refer to discrete physical behaviours, but to multi-layered whole-animal expressions which overlap and involve different physical behavioural elements in complex ways. Describing such meanings is not easy, and is likely to require a process of continuous revision and refinement based on growing experience. Thus the outcomes of the current study were used to adjust the QBA term list included in the final AWIN welfare assessment protocol for goats (AWIN, 2015).

It is important to underline that the two observers differed in their professional backgrounds and expertise. One was a veterinarian, used to work in pasture-based and organic farms, whereas the other was an animal scientist, experienced in working with intensive farms, making it likely that they would differ in their perspective of goat welfare on Housed vs Pasture farms. Research by Tuytens et al. (2014) indicates that such differences can lead to bias in both qualitative and quantitative assessments of animal behaviour, and need to be addressed through appropriate training. In the current study, observers were unavoidably aware of the type of farm they were assessing, yet there was no effect of observer on mean score values on the main principal component PC1 (positive vs negative mood), indicating that the applied training procedures had been effective in overcoming any differences between observers in how they perceived the animals' general mood, and the effect of housing system on this mood. That observers in the current study were experienced professionals rather than under-graduate veterinary science students as in the Tuytens study, may have contributed to their agreement in this respect.

There was however a significant observer effect on the mean value of PC2 farm scores, as well as a significant interaction of observer and housing system on PC3. Obs-A, the veterinarian used to working in pasture-based and organic farms, attributed on average higher "curious/attentive" scores to all visited farms (PC2), and higher "sociable/playful" scores to Pasture farms (PC3), than Obs-B, the animal scientist accustomed to intensive goat farms. These discrepancies are likely to be related to the differences in observers' background. However, it is not necessarily clear whether this is a matter of undesirable bias or of desirable experience. One explanation is that Obs-A was more sensitized to positive expressive features such as curiosity and play, and so scored them more strongly than Obs-B, particularly in the pasture-based farming environment she knew best. It seems quite understandable that time spent with animals in particular systems affects an assessor's sensitivity to these animals' demeanour, which could be an advantage and improve the assessor's perceptive ability. However, equally such experience could selectively skew the assessor's perception, which would be a disadvantage and cause for undesirable bias. It is perhaps most likely that people's background experience provides them with a mix of perceptive advantages and disadvantages, and the important task is to balance and overcome these through training. One could suggest it may be best to try avoiding background effects by using observers with little previous experience, such as students possessing only a basic knowledge of animal behaviour and husbandry, and provide them with appropriate QBA training. However, such restrictions on which assessors to use may not be practically feasible, and students may also have varying background experiences. If possible, it may be best to balance the backgrounds and levels of experience of different members of an on-farm assessor team.

In the majority of on-farm studies the first principal component (PC1) reflects a difference between positive and negative mood, or the valence of emotion, and provides salient information on animal welfare (e.g. Wemelsfelder et al., 2009; Andreasen et al., 2013; Phythian et al., 2016). Farm scores on this component tend to be most reliably scored, as this study also found, and are therefore used as QBA

measures in on-farm animal welfare monitoring schemes such as Welfare Quality® (2009) and AWIN (2015). However, we should also aim to achieve better robustness for scores on second and third principal components, which may well provide valuable additional information (Meagher, 2009). As the current study indicates, even when the relative ranking of scores on these components shows good agreement, there may be discrepancies in how observers quantify their assessments on the visual analogue scales (Fleming et al., 2015). That such discrepancies can and do occur is well known, and it is the subject of much research and discussion in human research (Kazdin, 1977; Torrance et al., 2001), and more recently also in animal research (Tuytens et al., 2014; Fleming et al., 2015). Efforts must be made to align the “scoring styles” of different observers, as was done in this study, by discussing the meaning of terms and practicing the use of visual analogue scales; however more can be done, for example by practicing scoring from videos that serve to illustrate particular individual terms (e.g. showing “bored” or “content” goats), or by actually going on-farm and practice scoring animal expressions “in situ”. As useful as videos are, they do not represent the real situation faced by observers when visiting a farm, where there are many disturbances and context factors potentially causing observers to score differently. Training directly on farm may be costly in terms of time and resources, but assessors are likely to achieve important skills for the use of QBA as an on-farm welfare assessment tool, which will pay off in the longer term.

## **5. Conclusion**

This is the first study on the use of QBA in dairy goats and it reports the inter-observer reliability of the method in different housing systems, using a fixed-list of QBA descriptors specifically developed for this species. Two observers with different professional backgrounds achieved good agreement in their relative ranking of farms on three main dimensions of goat expression, with assessed goats housed in pasture-based farms significantly more “calm/content” and “curious/attentive” than permanently housed goats. These results support that QBA can detect expressive information that is relevant to the evaluation of on-farm animal welfare. There was no observer effect on PC1 farm scores; however, the finding of such an effect on PC2 and PC3 farm scores indicates that further development of QBA training procedures is required to extend inter-observer reliability to all main expressive dimensions emerging from the analysis.

## **Conflict of interest**

None declared.

## **Acknowledgments**

The present document results from the Animal Welfare Indicators (AWIN) Project, which has been co-financed by the European Commission, within the VII Framework Program (FP7-KBBE-2010-4, Grant no. 266213). The authors are also grateful to all the experts involved in the Italian focus group, to all the AWIN goat experts and to all the farmers who collaborated with this research.

## **References**

- AWIN, 2015. AWIN Welfare Assessment Protocol for Goats Milano., [http://dx.doi.org/10.13130/AWIN\\_goats\\_2015](http://dx.doi.org/10.13130/AWIN_goats_2015).
- Andreasen, S.N., Wemelsfelder, F., Sandøe, P., Forkman, B., 2013. The correlation of Qualitative Behaviour Assessments with Welfare Quality® protocol outcomes in on-farm welfare assessment of dairy cattle. *Appl. Anim. Behav. Sci.* 143, 9–17.

- Battini, M., Vieira, A., Barbieri, S., Ajuda, I., Stilwell, G., Mattiello, S., 2014. Animal-based indicators for on-farm welfare assessment for dairy goats: a review. *J. Dairy Sci.* 97, 6625–6648.
- 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.
- Bokkers, E., de Vries, M., Antonissen, I., de Boer, I., 2012. Inter- and intra-observer reliability of experienced and inexperienced observers for the Qualitative Behaviour Assessment in dairy cattle. *Anim. Welfare* 21, 307–318.
- Braghieri, A., Pacelli, C., Girolami, A., Napolitano, F., 2011. Time budget, social and ingestive behaviours expressed by native beef cows in Mediterranean conditions. *Livestock Sci.* 141, 47–52.
- Casamassima, D., Sevi, A., Palazzo, M., Ramacciato, R., Colella, G.E., Bellitti, A., 2001. Effects of two different housing systems on behavior, physiology and milk yield of Comisana ewes. *Small Rumin. Res.* 41, 151–161.
- De Rosa, G., Grasso, F., Winckler, C., Bilancione, A., Pacelli, C., Masucci, F., Napolitano, F., 2015. Application of the Welfare Quality protocol to dairy buffalo farms: prevalence and reliability of selected measures. *J. Dairy Sci.* 98, 6886–6896.
- Dwyer, C.M., 2009. Welfare of sheep: providing for welfare in an extensive environment. *Small Rumin. Res.* 86, 14–21.
- FAWC, 2009. Farm Animal Welfare in Great Britain: Past, Present and Future (accessed on 17.12.15) <https://www.gov.uk/government/publications/fawc-report-on-farm-animal-welfare-in-great-britain-past-present-and-future>.
- Fleming, P.A., Wickham, S.L., Stockman, C.A., Verbeek, E., Matthews, L., Wemelsfelder, F., 2015. The sensitivity of QBA assessments of sheep behavioural expression to variations in visual or verbal information provided to observers. *Animal* 9, 878–887.
- Goddard, P., Waterhouse, T., Dwyer, C., Stott, A., 2006. The perception of the welfare of sheep in extensive systems. *Small Rumin. Res.* 62, 215–225.
- Goddard, P.J., 2013. Small ruminant welfare: levelling the playing field or raising the bar-A European perspective. *Small Rumin. Res.* 110, 108–111.
- Gutmann, A., Schwed, B., Tremetsberger, L., Winckler, C., 2015. Intra-day variation in Qualitative Behaviour Assessment outcomes in dairy cattle. *Anim. Welfare* 24, 319–326.
- Kazdin, A.E., 1977. Artifact, bias and complexity of assessment: ABCs of reliability. *J. Appl. Behav. Anal.* 10, 141–150.
- Martin, P.R., Bateson, P., 2007. *Measuring Behaviour: an Introductory Guide*. Cambridge University Press, Cambridge.
- 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., Dalla Costa, E., Dai, F., Murray, L.A.M., Canali, E., Wemelsfelder, F., 2016. Use of Qualitative Behaviour Assessment as an indicator of welfare in donkeys. *Appl. Anim. Behav. Sci.* 174, 147–153, <http://dx.doi.org/10.1016/j.applanim.2015.10.010>.
- Miranda-de la Lama, G.C., Mattiello, S., 2010. The importance of social behavior for goat welfare in livestock farming. *Small Rumin. Res.* 90, 1–10.
- Muri, K., Stubbsj en, S., Valle, P., 2013. Development and testing of an on-farm welfare assessment protocol for dairy goats. *Anim. Welfare* 22, 385–400.
- Phythian, C., Michalopoulou, E., Duncan, J., Wemelsfelder, F., 2013. Inter-observer reliability of Qualitative Behavioural Assessments of sheep. *Appl. Anim. Behav. Sci.* 144, 73–79.
- Phythian, C.J., Michalopoulou, E., Cripps, P.J., Duncan, J.S., Wemelsfelder, F., 2016. On-farm qualitative behaviour assessment in sheep: repeated measurements across time, and association with physical indicators of flock health and welfare. *Appl. Anim. Behav. Sci.* 175, 23–31.

- Popescu, S., Borda, C., Diugan, E.A., Niculae, M., Stefan, R., Sandru, C.D., 2014. The effect of the housing system on the Welfare Quality of dairy cows. *Ital. J. Anim. Sci.* 13, 15–22.
- 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.
- Sevi, A., Casamassima, D., Pulina, G., Pazzona, A., 2009. Factors of welfare reduction in dairy sheep and goats. *Ital. J. Anim. Sci.* 8 (Suppl. 1), 81–101.
- Stockman, C.A., Collins, T., Barnes, A.L., Miller, D., Wickham, S.L., Beatty, D.T., Blache, D., 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.
- 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.
- Torrance, G.W., Feeny, D., Furlong, W., 2001. Visual analog scales: do they have a role in the measurement of preferences for health states? *Med. Decis. Making* 21, 329–334.
- Tuytens, F.A.M., de Graaf, S., Heerkens, J.L.T., Jacobs, L., Nalon, E., Ott, S., Stadig, L., Van Laer, E., Ampe, B., 2014. Observer bias in animal behaviour research: can we believe what we score, if we score what we believe? *Anim. Behav.* 90, 273–280.
- Welfare Quality®, 2009. Welfare Quality® Assessment Protocol for cattle, Welfare Quality® Consortium, Lelystad.
- Wemelsfelder, F., Millard, F., 2009. Qualitative behaviour assessment. In: Forkman, B., Keeling, L. (Eds.), *Assessment of Animal Welfare Measures for Sows, Piglets and Fattening Pigs*. Welfare Quality® Reports 10. University of Cardiff, Cardiff, pp. 213–219.
- Wemelsfelder, F., Mullan, S., 2014. Applying ethological and health indicators to practical animal welfare assessment. In: Mellor, D.J. (Ed.), *Animal Welfare: Focusing on the Future*, 33. OIE Scientific and Technical Review, pp. 111–120.
- Wemelsfelder, F., Hunter, E., Mendl, M., Lawrence, A., 2000. The spontaneous qualitative assessment of behavioural expressions in pigs: first explorations of a novel methodology for integrative animal welfare measurement. *Appl. Anim. Behav. Sci.* 67, 193–215.
- Wemelsfelder, F., Hunter, T.E.A., Mendl, M.T., Lawrence, A.B., 2001. Assessing the whole animal: a free choice profiling approach. *Anim. Behav.* 62, 209–220.
- Wemelsfelder, F., Knierim, U., Schulze Westerath, H., Lentfer, T., Staack, M., Sandilands, V., 2009. Qualitative behaviour assessment. In: Forkman, B., Keeling, L. (Eds.), *Assessment of Animal Welfare Measures for Layers and Broilers*. Welfare Quality® Reports 9. University of Cardiff, Cardiff, pp. 113–119.
- Wemelsfelder, F., 2007. How animals communicate quality of life: the qualitative assessment of behaviour. *Anim. Welfare* 16, 25–31.
- Yeates, J.W., Main, D.C.J., 2008. Assessment of positive welfare: a review. *Vet. J.* 175, 293–300.