A protocol to assess the welfare of dairy ewes: From science to the field

Pierre-Alexandre Jolly, Elise Rouyer, Laetitia Bru, Lorie Marin, Elise Meillac, Gabrielle Foubert, Mathilde Roman, Andrew Ponter, Olivier Patout, Alline de Paula Reis

Small Ruminant Research 232 (2024) 107209

Contents lists available at ScienceDirect

Small Ruminant Research

journal homepage: www.elsevier.com/locate/smallrumres

A protocol to assess the welfare of dairy ewes: From science to the field

Pierre-Alexandre Jolly [a], Elise Rouyer [a], Laetitia Bru [d], Lorie Marin [c], Elise Meillac [c], Gabrielle Foubert [c], Mathilde Roman [a], Andrew Arthur Ponter [a,b], Olivier Patout [c], Alline de Paula Reis [a,b,*]

[a] Ecole Nationale Vétérinaire d'Alfort, BREED, 94700 Maisons-Alfort, France
[b] Université Paris-Saclay, UVSQ, INRAE, BREED, 78350 Jouy-en-Josas, France
[c] Association Vétérinaires Eleveurs du Millavois, CAP du Cres, 12100 Millau, France
[d] Fromageries Papillon, 12250 Roquefort-sur-Soulzon, France

ARTICLE INFO

Key-words:
Animal welfare
Assessment protocol
Dairy industry
Ewe
Sheep

ABSTRACT

There is growing societal demand for improved animal welfare. The stakeholders in the industry are also increasingly interested in information and expertise in this area. There is a scientific consensus that an objective assessment should be the foundation for discussions and actions in favour of better animal welfare. However, only a few protocols exist and most of them have been developed for meat-producing sheep and/or are more adapted for an extensive farming context and/or do not include all the five freedoms of animal welfare. Therefore, references are still scarce especially in the field of dairy ewe welfare. This makes the transition from science to practice difficult. The present study was carried out bearing two objectives in mind: (1) to develop a feasible protocol (EBBEL – Evaluation du Bien-être des Brebis en Elevage Laitier / Assessment of sheep welfare in dairy farms) to assess the welfare of dairy ewes and (2) to implement the protocol on a large scale (n = 81 dairy ewe farms under semi-intensive system) in the Roquefort region of France. The animals and the farms were examined in terms of satisfaction of 26 welfare indicators. The results were very informative about the points of excellence but also about those that need to be improved. The database that we developed can serve the scientific community when comparing different production contexts. Beyond the technical results, stakeholder participation in the design process was clearly positive for acceptance of the protocol in the field.

1. Introduction

Animal welfare is defined as the physical and psychological health of animals. Animals should be in harmony with their environment, and their "nature" must be respected (Hughes, 1976). According to the EFSA (2012), animal welfare in farms is satisfactory when animals are in good health, comfortable, in good nutritional status, safe, exhibit normal behaviour and are not in pain, fearful or stressed.

The increase in societal demand for improved animal welfare (Eurobarometer, 2016) and the increased interest of livestock farmers and the livestock industry in welfare means that it is necessary to develop tools to assess farm animal welfare to initiate discussion with farmers and set up action plans. There is a consensus of opinion that animal welfare assessment must rely on an objective approach, the measurements must be valid, repeatable, reproducible and feasible (EFSA, 2012) and applied on live animals (animal-based measurements). In addition, the concept of five freedoms proposed by Brambell and Barbour (1965) - the freedom from hunger/thirst, freedom from pain/injury/disease, freedom from fear/distress, freedom from discomfort and freedom to express normal behaviour – is a suitable framework to develop a holistic assessment of welfare. However, some difficulties remain: the subjectivity of some observations, especially behavioural measurements (Fleming et al., 2015) and the impossibility to assess some key parameters directly on the animals without the inclusion of environment-based measurements (EFSA, 2014). Several protocols exist to assess farm animal welfare: Welfare Quality for cows, pigs, poultry (Welfare Quality®, 2009), AWIN for sheep, goats and horses (Dwyer et al., 2015), BIENE for meat-producing sheep (IDELE, 2015) and recently others, such as those of Munoz et al. (2019) and Marcone et al. (2022).

* Correspondence to: Ecole nationale vétérinaire d'Alfort, Département de productions animales et santé publique, Unité pédagogique de zootechnie, 7, av. Général de Gaulle, 94700 Maisons-Alfort, France.
E-mail address: alline.reis@vet-alfort.fr (A. de Paula Reis).

https://doi.org/10.1016/j.smallrumres.2024.107209
Received 14 April 2023; Received in revised form 30 December 2023; Accepted 24 January 2024
Available online 29 January 2024

The sheep version of the AWIN protocol was developed using 39 meat sheep farms and 9 dairy sheep farms. One of the advantages of the AWIN protocol is the evaluation of the five freedoms. On the other hand, the visit is time-consuming and can last several hours. The AWIN individual animal assessment requires that the animals be moved to a handling corridor or race and that certain constraining procedures, such as assessing the colour of the ocular mucosa are performed.

On the other hand, the protocol proposed by Marcone et al. (2022) is specific for the dairy context but does not cover the five freedoms of welfare. The assessment is therefore not complete.

In the context of producing milk products using non pasteurised milk (i.e. more than 80 types of sheep cheese are produced using raw milk in the world), it is very important to be strict about the health of the udder / teats and their cleanliness to limit udder infections which adversely affects animal welfare.

Therefore the authors intended to develop a more specific method to assess dairy sheep welfare using severe thresholds for udder health and teat integrity which is beneficial for both raw milk quality and animal welfare. In view of this, the authors concluded that it was necessary to propose a protocol that was both more specific to the semi-intensive-intensive dairy context, covering the five freedoms of welfare but not too time-consuming, and with minimal disruption to the routine of the animals and the farm.

The present work was performed in the region of Roquefort (France) with two aims:

1) to develop a standardised rapid, simple and easy to perform field assessment protocol for dairy ewe welfare including indicators of the five freedoms of animal welfare,
2) to build a reference database (EBBEL database) which could serve as a basis for further studies or comparisons and to help dairy processors reinforce dairy sheep welfare programs.

A multidisciplinary steering committee was organised to discuss questions relating to the relevance for this work linking science and application. The experiment was performed in two phases: the first was a pre-experimental study (Study 1) performed on 21 farms to test the proposed protocol, identify weak points and propose technical improvements. The second phase, involved the use of the final protocol on a larger population to generate a database (Study 2.1) as well as to verify the reproducibility of individual measurements (Study 2.2).

2. Material and methods

2.1. Multidisciplinary steering committee

A multidisciplinary scientific committee was formed by animal scientists, veterinarians, veterinary and animal science students, farmers and members of the dairy industry. The role of the steering committee was to define the scope of the welfare assessment protocol, to further spread the implementation of the protocol and to encourage the use of the conclusions of the study. The veterinarians of the steering committee also monitored assessor training.

The criteria for farm inclusion in the pre-experimental study (Study 1; n = 21) were: farmer consent for external assessor evaluation of animals, the presence of lactating ewes and the availability of farm records including milk production and mortality. In Study 2, the final protocol was applied on 81 farms which is representative of the farm population supplying milk for Roquefort cheese production (total number = 1 465 farms) (level of confidence: 95% and margin of error: 10.6%).

2.1.1. Selection of indicators and measurements

A review of the literature was performed in order to identify existing animal welfare indicators, their validity and description quality (Jolly, 2020). The steering committee selected the most pertinent indicators and, when required, proposed adaptations to clarify the description of the measurement method or to improve the feasibility under the conditions of the present study.

2.1.2. Pre-experimental study (Study 1)

The objective of study 1 was to set up an animal welfare assessment protocol appropriate for dairy sheep farms.

2.1.3. Data collection

2.1.3.1. Implementation of the assessment protocol. The farms were assessed in January (winter), when the animals were housed. In order to limit disturbance and to provide a clear view of the animals, the individual measurements were carried out in the milking parlour. The behavioural indicators: Human-Animal Relationship (HAR), Qualitative Behaviour Assessment (QBA), quantitative social behaviour and abnormal behaviour were recorded on the group of animals (present in the main enclosure). In the Study one, QBA, social behaviour and HAR were performed, in this order, before the individual measurements. The video recordings were performed without the presence of a human in the barn. The farm records were obtained from farm registers.

2.1.3.2. Training program. Assessors were specifically trained for the protocol. The training program included theoretical concepts of animal welfare and assessment. Moreover, on farm practical training under the supervision of a qualified instructor was supplied.

2.1.4. Farms and animals

Study 1 was carried out on 21 farms with herd sizes ranging from 135 to 530 adult Lacaune lactating ewes (representing a total population of 7 377 milking ewes (1 328 primiparous (18%) and 6 049 multiparous (82%)) in the Roquefort production area (France).

2.1.5. Sampling method

2.1.5.1. Sampling for Qualitative Behaviour Assessment. Only the main enclosure (i.e.: the more crowded) was observed for the Qualitative Behaviour Assessment. The enclosure was divided into two halves and the assessor, standing discreetly outside the pen, observed the two halves of the enclosure consecutively.

2.1.5.2. Sampling for the quantitative behaviour analysis from videos. A portable camera (NBD, 4k, Ultra HD 48MP with a large angle view lens, digital 16x zoom, screen 3.0″) was discreetly placed outside the pen and installed on a tripod at a height of approximately 1.5 m thus providing an overview of the entire pen. The assessor left the building for 30 min during recording. The first 20 min of each 30 min video was not analysed since it was the time necessary for the ewes to become accustomed to the camera and only the last 10 min were used for the assessment. The animals present in the half of the pen closest to the camera were counted and their social behaviour was observed.

2.1.5.3. Sampling for the flight distance test (Human Animal Relationship). One randomly chosen animal, situated 3 m from the entrance of the main enclosure was considered as the target. As sheep present strong allelomimetism, the test should be performed only once.

2.1.5.4. Sampling for individual measurements. Individual observations were performed on the milking platform. Sampling was determined in two steps (calculations were preferably performed before the visit).

Step one: the sample size was defined based on the size of the lactating herd following the recommendations of the AWIN protocol (Dwyer et al. 2015) (Table 1).

Step two: the objective of this second step was to avoid bias due to a specific order of passage or the animals' health status and to equally distribute the sample over the entire morning or afternoon milking

Table 1
Suggested and minimum sample size for the individual assessment based on the number of lactating ewes on the farm.

Farm size - n lactating ewes	Suggested sample*	Minimum sample**	Farm size - n lactating ewes	Suggested sample*	Minimum sample**
<15	all animals	all animals	300-349	73	56
15-19	13	13	350-399	76	57
20-24	17	16	400-449	78	57
25-29	20	19	450-499	80	58
30-34	23	21	500-599	81	59
35-39	26	24	600-699	83	60
40-44	29	26	700-799	85	61
45-49	31	28	800-899	86	62
50-59	33	29	900-999	87	63
60-69	37	32	1000-1099	88	63
70-79	41	35	1100-1199	89	64
80-89	44	37	1200-1299	89	64
90-99	47	39	1300-1399	90	65
100-124	49	41	1400-1499	90	65
125-149	55	44	1500-1599	91	65
150-174	59	47	1600-1699	91	65
175-199	63	49	1700-1799	91	66
200-224	65	51	1800-1899	92	66
225-249	68	53	1900-1999	92	66
250-299	70	54	>2000	92	66

Source: AWIN protocol (Dwyer et al. 2015).
* Assuming a 50% prevalence, IC 95% and accuracy 10%
** Assuming a 50% prevalence, IC90% and accuracy 10%

period. The number of animals to be assessed at each milking session was calculated as follows:

Number of animals to be observed per milking platform side at each rotation = sample size / (number of ewes in production / number of milking places per platform side).

The method of assessment of the different measures are described in detail in the related sections.

2.1.6. *Behavioural measurements*

The behavioural assessment should be performed at the beginning of the visit. The Qualitative Behaviour Assessment (QBA) should be the first behavioural observation followed by the video recording and the HAR.

2.1.6.1. Qualitative Behaviour Assessment (QBA). The QBA was observed on site and adapted from the AWIN protocol (Dwyer et al., 2015). The assessor observed half of the main pen for 5 min and then a further 5 min for the other half of the pen. In the situation where a farm had multiple pens of lactating ewes, the assessor only observed the most crowded pen (considered as the main pen of the farm). The assessor noted 20 descriptors from 0 (absent) to 125 (strongly present). The higher the score of an adjective describing the flock (0 to 125), independently of a positive or negative connotation, the higher the importance of the adjective to describe the animals.

The descriptors were defined as follows (Dwyer et al., 2015):

Alert (AL): observant and vigilant.

Active (ACT): the animal is physically active. Engaged in task: eating, moving, walking or fighting.

Relaxed (REL): at ease, free from anxiety, agitation or tension. The animal appears to be unthreatened.

Fearful (FEAR): attention is focused on one specific object/being which is either real or perceived threat. Animal may also be fleeing.

Calm (CA): placid and sedate. If physically active the animal's movements are smooth and unhurried.

Content (CON): Satisfied and at peace. The animal's needs are met, or the animal is successfully working towards their completion.

Agitated (AGI): excessive cognitive and/or motor activity due to tension or anxiety. The animal is uneasy and if moving their actions are twitchy.

Aggressive (AGG): hostile and tense. Attacking/ready to attack, usually unprovocked or to compete for resource.

Vigorous (VIG): the animal is carrying out task in an energetic or forceful way. If stationary or moving slowly the animal expresses an inner strength and energy. May imply good physical health.

Frustrated (FRU): dissatisfied. Unable to fulfil satisfaction and achieve goal.

Subdued (SUB): submissive and docile. Often removed from social group and self-absorbed.

Physically uncomfortable (PHY): giving the impression of pain or other physical discomfort through posture or movement.

Listless (LIS): lack of vigour and energy. Animal appears lacklustre.

Inquisitive / Curious (CU): curious, interested and intrigued by the environment or other animals.

Bright (BRI): alert, lively and aware of environment.

Wary (WAR): the ewes are shy, cautious, apprehensive and possibly distrustful.

Tense (TEN): uneasy and/or on edge. Posture may show physical tension.

Sociable (SOC): seeking and interacting with other sheep. The sheep appears to be enjoying / taking comfort from their contact. The sheep is choosing to be part of a flock and not fully isolate themselves.

Apathetic (APA): unresponsive and dull.

Assertive (ASS): displaying confidence or determination.

Defensive (DEF): ready to potentially defend herself or lamb from harm / perceived threat.

The QBA results were summarised using a Principal Components Analysis (PCA).

2.1.6.2. Quantitative social behaviour and abnormal behaviour. The animals were filmed and the videos were prepared for analysis as previously described (2.3.1.2 Sampling method) to measure flock social and abnormal behaviour.

The activity of the ewes present in the selected part of the pen was characterised (eating, ruminating, laying or other activities: self-scratching, walking, isolating from the flock, vocalising). The proportion of ewes eating, ruminating or performing other activities was calculated for each farm.

Social behaviours were classified into 2 categories:

- Agonistic behaviours (AGO): kicking with front or hind legs, head-butts, pushing with or without displacement, chasing and mounting,

- Affiliative behaviours (AFF): head or jump games, licking, scratching or headrest on another animal.

Abnormal behaviour (ABN) was assessed by counting the number of animals with stereotypic activity (repetitive pacing or circling, star gazing, wool pulling, excessive scratching for more than 1.5 min).

The frequency of agonistic behaviour (FAGO), affiliative behaviour (FAFF) and abnormal behaviour (FABN) for 100 ewes was calculated using formulas (1) to (3) below:

Frequency of agonistic behaviour (FAGO) = ($\sum$ agonistic events / N animals present in the half pen observed) x 100 (1)

Frequency of affiliative behaviour (FAFF) = ($\sum$ affiliative events / N animals present in the half pen observed) x 100 (2)

Frequency of abnormal behaviour (FABN) = ($\sum$ abnormal events / N animals present in the half pen observed) x 100 (3)

2.1.6.3. Human Animal Relationship (HAR). The Human Animal Relationship (HAR) was assessed in the flock by evaluating the flock's flight behaviour in response to a moving person known to the sheep (the farmer). The objective in Study 1 was to identify key aspects of the test which needed to be standardised in order to improve the quality of the results. The standardisation needed to ensure the feasibility of the measurement whatever the shape of building or the layout of the building. Eleven farms were used to standardise the method and their results were not included in the results of the Study.

In the final method tested in Study 1, the assessor approached the main pen accompanied by the farmer but did not enter. The farmer was then provided with a tape measure and asked to formally identify a target ewe. The ewe should be easily accessible and be three meters from the fence. This ewe was used as the reference point. The farmer slowly entered the pen (smooth movement to jump over the fence) and moved towards the target ewe. The test was stopped when the ewe or the flock moved away (more than two steps). Two measurements were recorded at this moment:

1) the flight distance (FLID), defined as the distance in metres between the farmer's foot (closest to the ewe) and the target ewe when the flight movement was started. It was considered zero when the farmer could touch the ewe. The maximum distance was three meters (this meant that the ewe or the flock moved away as soon as the farmer entered the pen).
2) the type of flight (FLIT): considered as "absent" if the target ewe could be touched by the farmer; "walking" if the ewe or the flock walked away, or "running" if the ewe or the flock ran away.

The lower the FLID and the calmer the FLIT, the better the HAR.

2.1.6.4. Individual Animal-based measurements. Individual animal-based measurements representing the five freedoms of animal welfare were performed on the ewes and the individual results were later summarised into indicators at the herd level.

2.1.7. Body condition score

Body condition score was assessed by palpation on a scale of 0 to 5 points, in increments of 0.25 (Munoz et al., 2019). Ewes with a score <2 were considered lean, between 2 and 3.5 satisfactory and >3.5 fat. The "percentage of ewes with satisfactory body condition score" (%SBCS) was calculated for each farm.

2.1.8. Cleanliness

Cleanliness was assessed on four distinct anatomically areas of the ewe: the lumbosacral region, the lower hind legs, the udder and the perianal region. A scale from 1 to 4 was used: 1: absence of or minor soiling; 2: less than 50% of the region soiled; 3: more than 50% of the region soiled; 4 more than 50% of the region soiled with patches of dirt or presence of patches of dirt. For each region, scores 1 and 2 were considered clean (satisfactory), 3 and 4 were considered soiled. The percentage of ewes with satisfactory cleanliness of the lumbosacral region (%SCLS); ewes with satisfactory cleanliness of the hind lower legs (%SCLL); ewes with satisfactory cleanliness of the udder (%SCUD) and ewes with satisfactory cleanliness of the perianal region (%SCPA) were calculated for each farm.

2.1.9. Fleece moisture

Fleece moisture was recorded as follows: dry fleece (Grade 0): absence of humidity on the fingers; humid fleece (Grade 1): humidity felt on the fingers, deep in the fleece. The percentage of animals with a dry fleece (%DRYF) was calculated for each farm.

2.1.10. Fleece cover

Standard fleece cover of the ewe should be taken into account in order to identify abnormalities. The Lacaune breed is characterised by a low density fleece covering the upper parts of the body, except the head and neck. A ewe, with a complete fleece was considered as having a normal fleece (note 0). The fleece was considered abnormal when wooless patches or regions could be observed (note 1) (Dwyer et al., 2015). The percentage of ewes with normal fleece cover - note 0 - (%SWOOL) was calculated for each farm.

2.1.11. Skin integrity

One entire side of each ewe was observed (including the legs and the head and excluding the udder). Skin without lesions or with superficial lesions smaller than 2 cm in diameter or with scars was noted as 0 (good integrity). The presence of skin lesions and adjacent tissue, and/or a skin lesion larger than 2 cm, and/or swelling, myiasis or the presence of abscesses was noted as 1 (Jolly, 2020). The percentage of ewes with satisfactory skin integrity was calculated for each farm (%SSI).

2.1.12. Udder health

The udder was palpated starting from the attachment region. The udder and the teats were visually inspected. Udder health was classified using three grades: Grade 0: no lesion on the udder or teats; Grade 1: presence of minor or healed lesions (mammary gland fibrosis, healed lesions suggesting Parapoxvirus infection, small lumps located away from the teats); Grade 2: presence of an abscess larger than 2 cm in diameter on the mammary gland, lesions or lumps on the teats, or clinical mastitis. This scoring method was adapted from the AWIN protocol (Dwyer et al., 2015).

Grade 1 alterations were not considered as relevant from a welfare point of view "on the day of the evaluation" and were therefore deemed acceptable. The percentage of ewes with satisfactory udder health (Grade 0 + 1) was calculated for each farm (%SUDH).

2.1.13. Claw length

The claws (all four feet) were classified into one of three categories: satisfactory (Grade 0) when all claws were well balanced; mildly long (Grade 1) when at least one claw was slightly unbalanced provoking a slight alteration of the fetlock-pastern-claw angles; extremely long (Grade 2), slipper-like claws clearly causing the displacement of the animal's weight to the back of the foot (adapted from the AWIN protocol, Dwyer et al., 2015). The percentage of ewes with satisfactory claw length (%SCLAW) was calculated for each farm.

2.1.14. Locomotion

The gait of each ewe was assessed as it entered the milking parlour. If necessary, additional observations were carried out as the animal left the milking parlour. The scoring scale described by Angell et al. (2015) was simplified to three grades: Grade 0: no lameness, ewe with a balanced gait (homogeneous steps and equal weight borne on each foot). Grade 1: minor lameness with unequal steps and a difficulty to identify the lame limb OR moderate lameness with uneven steps, shortening of stride and possible identification of the lame limb. Grade 2: severe lameness with severely compromised mobility (no weight-bearing on affected limb whilst moving or standing), obvious discomfort (recumbence or reluctance to stand or move). The percentage of ewes with satisfactory locomotion (normal gait - Grade 0) (%SLOCO) was calculated for each farm.

2.1.15. Nasal secretion

Absence or presence of non-abundant transparent nasal discharge was scored as normal (Grade 0). The presence of abundant or yellowish discharge was scored as abnormal (Grade 1). The percentage of ewes with normal (Grade 0) nasal discharge (%NNAS) was calculated for each farm. The method was based on BIENE (IDELE, 2015).

2.1.16. Environment-based measurements

2.1.16.1. Allowance for a social life. Possibility for a social life: the farmer was asked about the number of animals raised alone. The indicator (ASL) was considered satisfactory when no animals were raised alone. It was not satisfactory if at least one animal was raised alone.

2.1.16.2. Freedom of movement. Freedom of movement was assessed by observing the animals and asking the farmer about the annual management of the animals. Grade 0: when none of the animals were tethered; Grade 1: when the animals could be temporarily tethered; Grade 2: when at least some animals could be systematically tethered.

2.1.16.3. Pasture access. The management of access to the pasture was also assessed as a component of the expression of normal behaviour. The duration of access in days was noted.

2.1.16.4. Available surface per animal (SPA). In each farm, the width and length of the pen were measured with a laser meter and the number of ewes counted (AWIN protocol, Dwyer et al., 2015). The surface was then divided by the number of animals to calculate the SPA for the pen (in m^2/animal). The assessor measured the surface in all the pens but only the lowest SPA was used in the analysis.

2.1.16.5. Presence of shade outdoor. Shade was noted as “absent” when no shade was available in the pasture; “hedges” when only hedges were available or “trees” when shade was provided in the interior of the paddock. It was also noted if the shade was available in all the paddocks or only in some of them.

2.1.16.6. Water supply in the building (number of troughs and cleanliness). The number and the type of functioning drinking troughs were calculated. The recommendations are either 1 m of linear drinking troughs for 35 ewes (i.e.: 2.8 cm/ewe) or 1 bowl/15 ewes (Inn’Ovin, 2019). The percentage compliance with the water supply recommendations (%CWAT) was calculated:

$$\%CWAT = (((\text{n cm of linear troughs}/2{\cdot}8) + (\text{n bowls x } 15)) / \text{n ewes present in the pen}) \text{ x } 100$$

The cleanliness of the troughs was evaluated:

- clean (clean water and drinking trough),
- partially clean (clean water with the presence of organic residue in the bottom of drinking trough),
- dirty (turbid water and/or development of algae, mould, insect larvae).

2.1.16.7. Water supply in the outdoor area. The farmer was asked about the water supply in the pasture: presence or absence; and if present, in one, some or all the paddocks. When no water was available, the farmer was asked about his management practices when the weather was hot.

2.2. Management results

2.2.1. Adult mortality

Farm registers were used to calculate adult mortality in the year prior to the visit.

$$\%AMOR = (\text{n dead adults / n adults}) \text{ x } 100$$

Depending on the type of data records, several definitions of an adult animal exist:

- all animals from weaning (%AMOR1)
- nulliparous, primiparous and multiparous females included in the last mating season (%AMOR2)
- primiparous and multiparous ewes (%AMOR3)

The three calculation methods were analysed separately.

2.2.2. Lamb mortality

Lamb mortality was calculated from farm records over the one-year period before the visit. Depending on the type of data records, the lamb mortality could be calculated from birth to 48 h (%LMOR1) or from birth to weaning (%LMOR2). These two calculation methods were used in the protocol but were analysed separately.

$$\%LMOR1 = (\text{n dead lambs from birth to 48 h / n born lambs}) \text{ x } 100\%$$

$$LMOR2 = (\text{n dead lambs from birth to weaning/ n born lambs}) \text{ x } 100$$

2.2.2.1. Number of ewes per staff member. The number of ewes per staff member (NESM) was calculated as follows:

$$NESM = \text{average number of productive ewes / n of full time manpower}$$

2.2.3. Data handling and statistics

Each animal-based and environment-based measure was first summarised into indicators, as explained in the previous section. The objective was to determine, for each farm, the level of satisfaction for each indicator. The behavioural measurements, performed in groups, were summarised in a descriptive manner, as explained in the previous section.

The population was described using the median value, the 1st and the 3rd quartiles of each indicator, using Excel (© Microsoft 2020), except for QBA. QBA was analysed by Principal Component Analysis (PCA), using R (© 2013 Rcommander.com), FactomineR and factoextra libraries.

For a more comprehensive visualisation and analysis of the results, all indicators were grouped according to the concept of the five freedoms (Brambell and Barbour, 1965).

2.3. Study 2

After Study 1 and feedback from the assessor, the steering committee selected the most consistent and meaningful measurements and proposed minor adaptations to consolidate the EBBEL protocol. The objectives of these modifications were:

1) to obtain acceptable measurement reproducibility,

2) to obtain a final assessment protocol, including at least one indicator for each of the five freedoms of animal welfare, which was sufficiently informative and simple to be used in the field.

The final assessment protocol, used in Study 2, was similar to the protocol used in the pre-experiment (Study 1) except for an improvement in measurements for claw length, udder health, teat integrity and fleece moisture because the feedback from the assessor indicated a lack of precision. In addition, the measurement of social behaviour was abandoned due to poor feasibility.

The objectives of Study 2 were:

1) Study 2.1: to create the official EBBEL reference database using data from 81 farms
2) Study 2.2: to assess the feasibility and inter-operator reproducibility of the individual animal-based measurements performed in the milking parlour. Study 2.2 was performed on the first 10 farms assessed in Study 2.1.

2.3.1. Measurement improvements

Only the measurements modified between Study 1 and Study 2 will be described in this section, all the other measurements were implemented as described in Study 1.

2.3.1.1. Claw length. The scoring scale used in Study 1 was simplified:

- Grade 0 (normal or acceptable): when the eight claws were well balanced or one or more claws were mildly long or unbalanced provoking slight joint angle alteration
- Grade 1 (not acceptable): when at least one of the claws was extremely long (slipper-like claws) clearly changing the distribution of weight towards the heels and provoking an alteration of joint angles. The percentage of ewes with satisfactory claw length (Grade 0) (%SCLAW) was calculated for each farm.

2.3.1.2. Udder health. In Study 1, a global score was provided for the udder and teats. In Study 2, the evaluation method was improved and in the final version of the EBBEL protocol the udder and teats were evaluated separately after milking.

The udder was visually inspected:

- Grade 0: absence of lesions on the udder;
- Grade 1: presence of minor (superficial, <2 cm) or healed lesions (dissymmetry of the udder, small lumps located far from the teats);
- Grade 2: presence of an abscess >2 cm or lesions >2 cm on the udder.

Grade 1 alterations were considered not to be a welfare problem “on the day of the evaluation” and were therefore acceptable. The percentage of ewes with satisfactory udder health (%SUDH) (Grade 0 + Grade 1) was calculated for each farm.

2.3.1.3. Teat integrity. A 360° visual inspection of the teats was performed:

- Grade 0: absence of lesions on both teats.
- Grade 1: presence of a lesion on the teats.

The percentage of ewes with satisfactory teat integrity (Grade 0) (%STI) was calculated for each farm.

2.3.1.4. Fleece moisture. Fleece moisture was measured:

- Grade 0: dry fleece, where the was no humidity on the fingers
- Grade 1: humid fleece, where humidity was felt on the fingers, deep in the fleece.

The percentage of animals with a dry fleece (%DRYF) was calculated for each farm.

2.3.2. The training program

The training program was modified to take into account the aforementioned improvements.

2.3.3. Study 2.1: building the EBBEL Reference database

2.3.3.1. Assessors, farms, and animals. Study 2.1 was carried out on 81 farms with herd sizes ranging from 60 to 1040 adult Lacaune lactating ewes in the Roquefort production area (France) and involved a population of 28,104 Lacaune dairy ewes (5,768 primiparous and 22,336 multiparous). This was a representative sample of farms - in this zone there are 269 farms supplying milk for the production of Roquefort cheese). The farms were assessed in July and August, 2020 or February and March, 2021.

2.3.3.2. Assessment. Individual measurements were performed in the milking parlour whereas the behavioural assessments were performed in the pen as described in Study 1.

2.3.3.3. Data handling, statistics, and descriptive analysis. Data from Study 2.1 were analysed as described in Study 1.

2.3.4. Study 2.2: feasibility and inter-operator reproducibility of individual measurements

It was necessary to demonstrate the feasibility and the reproducibility of individual measurements performed in the milking parlour and during the milking routine, in order to validate this part of the protocol.

Study 2.2 was conducted on a subset of ten farms from Study 2.1 and the assessment methodologies employed in Study 2.2 were the same as the ones used in Study 2.1. Individual measurements were performed in the milking parlour independently by two assessors on the same animals.

2.3.5. Data analysis

Inter-operator reproducibility and feasibility were studied for each individual measurement.

Feasibility: a measurement was judged to be feasible in the milking parlour if more than 95% of the animals could be observed by both assessors and if the milking routine could be continued without any major disruption.

Inter-operator reproducibility was assessed using the Cohen’s *kappa* coefficient (Cohen, 1960) and was interpreted as described by Landis and Koch (1977). The *kappa* coefficient value was calculated using Excel (method supplied by Desquilbet, 2019).

3. Results

A total of 35,481 Lacaune ewes were assessed in Studies 1 and 2 (7,096 primiparous and 28,385 multiparous) and of these, 5,889 ewes (1,118 primiparous and 4,771 multiparous) were individually assessed in the milking parlour.

3.1. Study 1

Study 1 was performed on 21 farms and involved a population of 7,377 Lacaune ewes (1,328 primiparous and 6,049 multiparous), of which 1,187 (225 primiparous and 962 multiparous) were individually assessed in the milking parlour (40 to 72 ewes per farm).

Study 1 was carried out during the winter when ewes were housed.

Table 2
A description of individual indicators and social behaviour in farms in Study 1 (21 farms) and Study 2.1 (81 farms).

Animal welfare	Measurement	Herd indicator	Study 1: pre-experiment		Study 2.1: enlarged population	
			N farms	Median (Q1; Q3)	N farms	Median (Q1; Q3)
Freedom from hunger and thirst	Body condition score	% SBCS	21	95.5 (92.9; 98.0)	81	99.0 (98.0; 100.0)
Freedom from pain, injury or disease	Fleece coverage	% SWOOL	21	88.9 (86.7; 93.2)	81	96.0 (91.0; 100)
	Skin integrity	% SSI	20*	97.2 (94.9; 98.7)	81	96.0 (93.0; 98.0)
	Udder health	% SUDH	-	-	81	97.0 (95.0; 100)
	Teat integrity	% STI	-	-	81	96.0 (93.0; 100)
	Locomotion	% SLOCO	20	100 (98.3; 100)	81	100 (97.0; 100)
	Nasal secretion	% NNAS	20	100 (98.2; 100)	81	100 (100; 100)
	Cleanliness -Perianal region	% SCPA	14*	100 (100; 100)	81	100 (100; 100)
Freedom from discomfort	Cleanliness -Lumbosacral region	% SCLS	14*	72.0 (46.1; 84.8)	81	100 (100; 100)
	Cleanliness -Udder region	% SCUD	14*	84.6 (73.7; 92.6)	81	100 (94.0; 100)
	Cleanliness - Lower hind legs	% SCLL	14*	92.3 (77.6; 99.6)	81	100 (94.0; 100)
	Claw length	% SCLAW	-	-	81	82.0 (63.0; 91.0)
	Fleece moisture	% DRYF	-	-	81	98.0 (95.0; 100)
Freedom to express normal behaviour	N of agonistic events	FAGO	19**	8.8 (0.0; 15.5)	-	-
	N of affiliative events	FAFF	19**	3.1 (0.0; 4.3)	-	-
	N of abnormal behaviour events	FABN	19**	0.0 (0.0;0.0)	-	-
Freedom from fear and distress	Flight distance	FLID	10**	2.1 (0.6; 3.0)	38 ***	1.0 (0.0; 3.0)

*Recording errors meant that the results from some farms were not included
**The standardisation of these tests was performed during the first visits in order to eliminate operational bias. The results obtained during the standardisation phase are not reported.
***This test was only performed on 38 farms because the test requires the presence of the farmer and in 43 farms the farmer responsible for the animals was not available during the visit.

The population is described in terms of median, first quartile and third quartile in Table 2. Two individual measurements (perianal cleanliness and nasal secretion) presented very low variability among farms with nearly all the farms receiving perfect scores (Table 2). Indeed, only one farm did not attain the perfect score for perianal cleanliness (score: 86.4%). Fourteen farms attained the perfect score for nasal secretion and the lowest score was 95% on one farm. The other individual indicators showed more variability. The lumbosacral cleanliness presented the highest variability between farms and the lowest level of satisfaction.

In all farms, animals were raised in groups and were not tethered. During the behavioural assessment, the animals were performing multiple activities (eating, ruminating, walking, lying and interacting). We did not observe any abnormal behaviour (FABN) and we observed a low frequency of social interactions (Table 2). The absolute maximum of agonistic + affiliative events: ≤20/100 animals/10 min observation).

The QBA allowed to distinguish between the 2 main dimensions of emotion (Fig. 1): one axis captured emotional valence ranging from the positive valence on the left side (relaxed, calm, content, assertive, sociable) to the negative valence (tense, physically uncomfortable, aggressive, defensive, frustrated, wary, apathetic, fearful) on the right side. The other axis defined arousal, ranging from high arousal on the bottom (active, alert, vigorous, agitated) to low arousal (subdued, listless) on the top of the axis. Fourteen farms (73.7%) showed a positive emotional valence. Five farms showed a negative valence, including one farm also showing low arousal. Interestingly, this farm had suffered a wolf attack a few months before the welfare assessment.

Ten farms had the Human-Animal relationship assessed (Table 2). The median distance before the first sign of avoidance was 2.1 m. In two farms the farmer could touch the ewe. In seven farms the ewes fled walking (in these farms QBA was positive for emotional valence and arousal). In the last farm the first sign of avoidance was observed at 3 m and the ewes fled by "running" away (in this farm QBA was negative for emotional state).

The environmental conditions were variable among farms as described in Table 3. The superior quartile agreed with the recommendations of Inn'Ovin (2019) for the surface per animal (1.2 m^2/ewe without lamb) and for water supply (1 m of linear trough for 35 ewes or 1 bowl/15 ewes). The median farm (Table 3) agreed with the surface per animal recommendation but not for water supply. The inferior quartile did not agree with both surface per animal and water supply recommendations.

The presence of shade and water in pasture was summarised in Table 4. In Study 1, four farmers said there were water troughs in all the paddocks, ten farmers stated that water troughs were present in some of the paddocks and seven farmers did not have water troughs in any of their paddocks. Seventeen farmers said that they provided shade (mainly a hedge bordering the field) in all the paddocks while in four farms shade was unavailable in the paddocks. The farmers who said that they did not have water or shade in any of the paddocks also said that they managed the daily length of pasture according to the weather and the quality of the pasture.

According to the management records, adult mortality was quite low in all farms and, not surprisingly, the calculation method %AMOR1 gave a slightly higher level of mortality than %AMOR2 (because of a

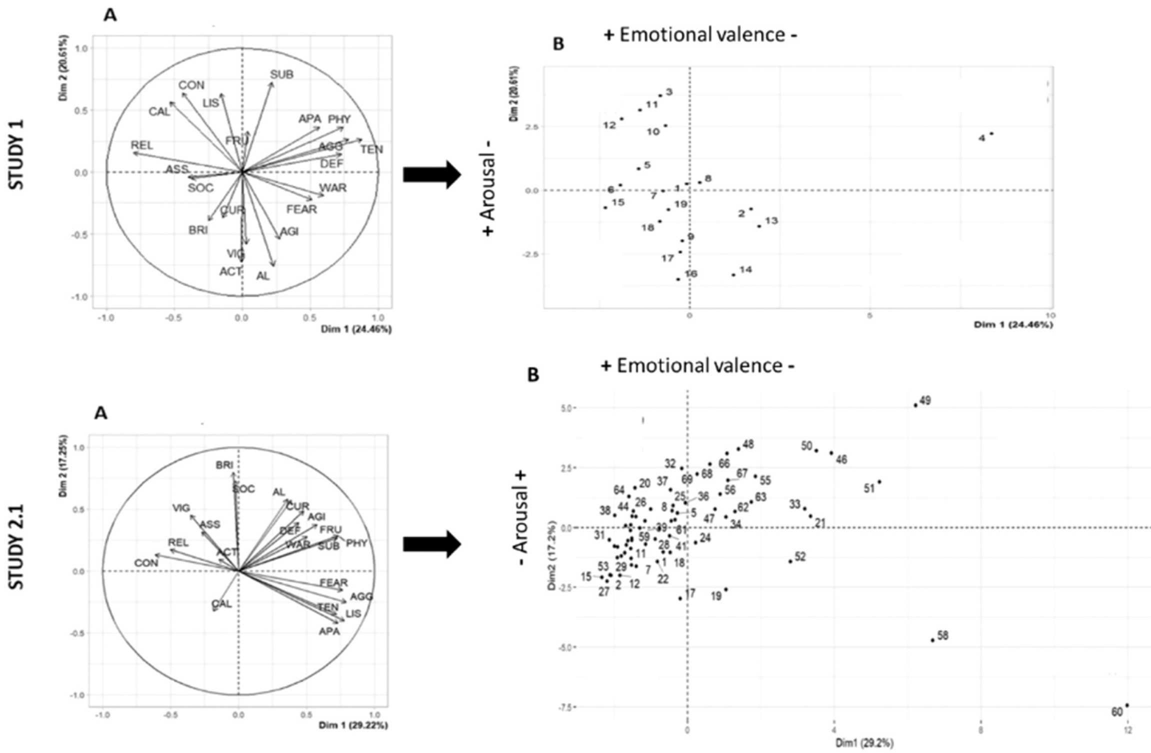

Fig. 1. Qualitative behaviour assessment of dairy ewes in intensive systems in the Roquefort region. Results from Study 1 are represented on the top panel and from Study 2.1 on the bottom panel. For each study, the PCA analysis is represented on left (A) and the farm distribution is represented on the right (B).QBA adjectives: Active (ACT); Aggressive (AGG); Agitated (AGI); Alert (AL); Apathetic (APA); Assertive (ASS); Bright (BRI); Calm (CAL); Content (CON); Curious (CUR); Defensive (DEF); Fearful (FEAR); Frustrated (FRU); Inquisitive (INQ); Listless (LIS); Physically uncomfortable (PHY); Relaxed (REL); Sociable (SOC); Subdued (SUB); Tense (TEN); Vigorous (VIG); Wary (WAR).

Table 3
Description of environment-based and management-based indicators observed in farms in Study 1 (21 farms) and Study 2.1 (81 farms).

Animal welfare	Measurement	Indicator	Study 1: Pre-experiment		Study 2.1: Enlarged population	
			N farms	Median (Q1; Q3)	N farms	Median (Q1; Q3)
Freedom from discomfort	Surface and n animals	SPA (m^2/animal)	21	1.2 (1.1; 1.3)	80*	1.3 (1.2; 1.4)
Freedom from hunger and thirst	Water troughs	%WAT	20*	80.0 (68.3; 117.5)	80*	112.0 (87.0; 159.0)
5 freedoms related	Adult mortality (1)	%AMOR1	7	4,1 (2.6; 5.6)	3***	-
	Adult mortality (2)	%AMOR2	9	3,4 (2.8; 4.2)	15	3.9 (2.3; 6.0)
	Adult mortality (3)	%AMOR3	4**	-	57	3.1 (2.1; 4.3)
	Lamb mortality (1)	%LMOR1	-	-	25****	6.0 (4.0; 8.0)
	Lamb mortality (2)	%LMOR2	21	9.1 (8.0; 16.0)	65****	7.0 (4.0; 8.0)
	n ewes and n staff member	NESM (n ewes / staff member)	21	125.0 (90.0; 142.0)	81	178.6 (133.0; 213.3)

Six farms were unable to supply adult mortality data.
* Data from one farm is missing because the assessor was not allowed to enter the pen.
** The median and quartiles were not calculated due to the low number of farms using this method of calculation. The four farm values were: 4.2%; 4.4%; 4.9%; 5.3%.
*** The median and quartiles were not calculated due to the low number of farms using this method of calculation. The three farm values were: 3.0%; 5.0%; 9.0%.
**** 9 farms supplied both %LMOR1 and %LMOR2. Both results were included in the population description.

Table 4
Presence of water and / or shade in the paddocks of the farms visited in Study 1 (21 farms) and Study 2.1 (81 farms).

Presence of	Study 1: pre-experiment				Study 2.1: enlarged population			
	Farms (n)	In all paddocks n (%)	In some paddocks n (%)	In no paddocks n (%)	Farms (n)	In all paddocks n (%)	In some paddocks n (%)	In no paddocks n (%)
Water	21	4 (19.0)	10 (47.6)	7 (33.3)	81	22 (27.2)	19 (23.5)	36 (44.4)
Shade	21	17 (81.0)	0 (0.0)	4 (19.0)	81	53 (65.4)	24 (29.6)	4 (5.0)
Water and shade	21	3 (14.3)	_*	0 (0.0)	81	22 (27.2)	_*	4** (4.9)

* the number of farms with water and shade in some of the paddocks has not been counted.
** these farmers declared that they managed the daily presence on the pasture according to the weather and pasture quality

difference in the definition of “adult”, as described in the Material and Methods). All the farms provided the %LMOR2 with the median value of 9.1% (Table 3) and the extremes varying from 0 to 18%.

The practical feedback from Study 1 highlighted the need for an improvement in the measurements for: “claw length” (%CLAW), “udder health” (%UDH) and “fleece moisture” (%DRYF). Therefore, these data (for Study 1) were not summarised in Table 2 and changes were made to the final protocol (see Material and Methods of Study 2).

3.2. Study 2.1

Study 2.1 was performed on 81 farms through the winter to the summer and involved a population of 28,104 Lacaune dairy ewes (5,768 primiparous and 22,336 multiparous), of which 4,702 (893 primiparous and 3,809 multiparous) were individually assessed in the milking parlour. From the end of winter to summer, the ewes had access to the pasture between the two daily milking and were housed during the night.

The median, first and third quartiles of this population of farms for individual indicators, Human Animal Relationship and environmental conditions were summarised in Table 2 and Table 3. Fig. 1 shows the QBA results. The presence of shade and water in the pasture were summarised in Table 4.

The first quartile is very high for all the individual indicators (Table 2) meaning that 75% of the farms in this population presented high levels of satisfaction of the different indicators. A more detailed analysis of the level of satisfaction of the individual indicators among the farms was further summarised in Supplementary data (Table 1) and allowed the grouping of individual animal welfare indicators into 4 categories:

- indicators with a high frequency of perfect scores (>85% of the farms) and low score variability (difference between the highest and lowest level of satisfaction <10%): satisfactory cleanliness of lumbosacral and perianal regions,
- indicators with an intermediary frequency of perfect scores (50 to 84.9% of the farms) and medium score variability (difference between the highest and lowest level of satisfaction between 10% and 40%): normal nasal secretion, satisfactory cleanliness of the udder and normal gait,
- indicators with a low frequency of perfect scores (<50% of the farms) and medium score variability (difference between the highest and lowest level of satisfaction between 10% and 40% of variability): satisfactory body condition score, skin and teats integrity, udder health and fleece moisture,
- indicators with a low frequency of perfect scores (<50% of the farms) and high score variability (difference between the highest and lowest level of satisfaction >40%): fleece cover and claw length. Cleanliness of lower hind legs can also be included in this category because, despite the fact that 50.6% of the farms reached the perfect score, a high variability was observed (96% difference between the farm with the highest and the farm with the lowest level of satisfaction).

The QBA was performed on 69 farms (Fig. 1) since in the remaining 12 farms the ewes were not available in the building therefore QBA was not assessed.

Once again, the QBA allowed us to capture the two main dimensions of emotion (Fig. 1): one axis captured emotional valence ranging from a positive valence on the left side (relaxed, content) to a negative valence (physically uncomfortable, aggressive, frustrated, fearful) on the right side. The other axis defined arousal, ranging from high arousal on the top (bright, alert, curious) to low arousal (apathetic, tense, listless) on the bottom of the axis.

Animals from 47 farms (69.1%) showed a positive emotional valence. In 22 farms (31.9%) they showed a negative valence, including five farms (7.2%) where the animals also showed low arousal. In these five farms, we could not identify precise factors triggering this negative emotional state.

The Human-Animal relationship was assessed in 38 farms. The median distance at the first sign of avoidance was 1 m (Table 2) but a large variability was observed (Supplementary data – Table 4): on eighteen farms (47.3%) the farmer could touch the target ewe and in twenty farms, the flight distance varied up to 3 m. However, in all these farms the ewes fled by walking, suggesting a moderate fear of humans.

The environmental conditions showed variations (Table 3). The median farm provided 1.3 m^2/animal and the first quartile provided 1.2 m^2/animal. These findings mean that 75% of the farms respected the recommendations of the French ovine industry (Inn’Ovin, 2019) in terms of surface per animal (1.2 m^2/animal). The median farm also fully satisfied the recommendations for water supply (i.e.: 1 m of linear trough for 35 ewes or 1 bowl/15ewes (Inn’Ovin, 2019). Looking in detail (Supplementary data – Table 2), 70% of the farms (n = 56) fulfilled the water supply recommendations.

All the farmers stated that access to the pasture was allowed as soon as the weather conditions were favourable. Among the eighty-one farmers, twenty-two said that they provided shade and water in all the paddocks (Table 4). However, four farmers said that they did not supply shade or water in the paddocks and managed the daily presence on the pasture according to the weather and pasture quality. The other fifty-five farmers said that either only water or only shade was available in the paddocks.

In this sample of farms, %AMOR2 and %AMOR3 were predominant, and, not surprisingly, %AMOR2 gave a slightly higher level of mortality than %AMOR3 (because of the definition of “adult”, as described in the Material and Methods). The dominant method of calculation was % AMOR3. Six farmers could not supply this information. The results were summarised in Table 3. We noted that in the median farm, adult mortality rate was close to 4% independent of the method of calculation. A detailed analysis of the raw data (Supplementary data – Table 3) showed that overall adult mortality ranged from 0.7% to 12%. In regard to lamb mortality, the annotation method %LMOR2 was more frequently observed than %LMOR1. The lamb mortality in the median farm was 6% (%LMOR1) or 7% (%LMOR2). The detailed analysis of the raw data (Supplementary data – Table 3) showed that overall lamb mortality ranged from 0.1% to 20% (from birth to 48 h - %LMOR1) and from 1.2 to 40% (from birth to weaning - %LMOR2).

3.3. Study 2.2

Study 2.2 was carried out on a sample of 610 primiparous and multiparous ewes from 10 farms participating in Study 2.1. The feasibility of individual measurements performed in the milking parlour was excellent: both assessors were able to perform all the observations for all the animals without disrupting the milking routine. Agreement between the data from the two assessors was higher than 90% for all individual measurements but the imbalanced prevalence of the scores represented a weakness for statistical analysis (Table 5). Indeed, the higher the prevalence of one of the scores for a given measurement, the weaker the statistical level of agreement. Also, the level of agreement was statistically unpredictable for the measurements exhibiting almost 100% of prevalence of one of the notes to the detriment of the other. Despite the difficulties inherent in a field study, agreement between the assessors was moderate for the majority of the measures. It was high for the measurement of normal gait and almost perfect for claw length. Udder cleanliness and teat integrity showed only fair and a slight level of agreement, respectively.

3.4. Summary of the final protocol EBBEL

After the experience obtained in Study 1, the final protocol can be summarised as follows.

Table 5
Level of agreement between the two assessors for individual measurements in the milking parlour.

Measurement	n agreement / n total (%)	Kappa value [IC95%]	P-value	Kappa value interpretation (Landis & Koch, 1977)
Body condition score (BCS) (by 0.25 point)	473/610 (77.5)	0.46 [0.39; 0.54]	<0.001	Moderate
BCS (satisfactory / not satisfactory)	610/610 (100)	NC	-	-
Fleece coverage	575/610 (94.2)	0.42 [0.27; 0.56]	<0.001	Moderate
Skin health	596/610 (97.7)	0.52 [0.31; 0.70]	<0.001	Moderate
Udder health	571/610 (93.6)	0.45 [0.31; 0.59]	<0.001	Moderate
Teat integrity	582/610 (95.4)	0.11 [0.01; 0.31]	<0.001	Slight
Locomotion	607/610 (99.5)	0.73 [0.49; 0.96]	<0.001	Substantial
Nasal secretion	610/610 (100)	NC	-	-
Claw length	578/610 (94.7)	0.84 [0.70; 0.92]	<0.001	Almost perfect
Cleanliness - Perianal region	609/610 (99.8)	NC	-	-
Cleanliness -Lumbosacral region	610/610 (100)	NC	-	-
Cleanliness - Udder region	578/610 (94.7)	0.30 [0.13; 0.47]	<0.001	Fair
Cleanliness - Lower hind legs	565/610 (92.6)	0.56 [0.45; 0.68]	<0.001	Moderate
Fleece humidity	605/610 (99.1)	NC	-	-

NC: not calculated

3.4.1. Training and measurements for the protocol

Training is strongly recommended in order to become familiar with the structure of the protocol and to illustrate the measurement method with practical examples.

3.4.2. Measurements

In summary, the measures selected cover the 5 freedoms of animal welfare and are listed in Table 6. They cover the 4 principles of welfare (corresponding to the five freedoms of animal welfare), and 11 of the 12 criteria of the Welfare Quality nomenclature (Welfare Quality®, 2009).

3.4.3. The visit

3.4.3.1. Before the visit.

- Contact the farmer to explain the visit and agree a date and time for the visit
- Ask the farmer for:
- the documents needed to calculate adult and lamb mortality results
- the number of ewes in lactation
- type of milking platform (double platform, rotary platform)
- number of milking points
- Calculate the sample size for individual measurements and the number of animals to be observed on each milking platform and in each milking session.
- Prepare the measurement annotation sheets (template provided - to be printed out or used on a tablet).

3.4.4. The visit format

For the description of the measurement, please follow links supplied

Table 6
Measures included in the EBBEL Protocol and indexed under the Welfare Quality nomenclature.

Welfare Principle		Welfare criteria	EBBEL measurement (link to the method of observation and calculations)
Good feeding (Freedom from hunger and thirst)	1	Absence of prolonged hunger	Body condition score
	2	Absence of prolonged thirst	Number of drinking troughs in the building Cleanliness of drinking troughs in the building
Good housing (Freedom from discomfort)	3	Comfort around resting	Claw length Cleanliness of the lumbosacral region Cleanliness of the udder Cleanliness of the hind lower legs Fleece moisture Available surface / animal in the building
	4	Thermal comfort	Water supply in the outdoor area Presence of shade outdoor
	5	Ease of movement	Freedom of movement
Good health (Freedom from pain, injury or disease)	6	Absence of injuries	Skin integrity Udder health Teat integrity
	7	Absence of disease	Fleece cover Locomotion Nasal secretion Cleanliness of the perianal region
	8	Absence of pain induced by management procedures	-
Appropriate behaviour (Freedom to express normal behaviour + Freedom from fear and distress)	9	Expression of social behaviours	Allowance for social life
	10	Expression of other behaviours	Pasture access
	11	Good human-animal relationship	HAR: Flight distance
	12	Positive emotional state	Qualitative Behaviour Assessment
Multiple principles		Multiple criteria	Adult mortality Lamb mortality Number of ewes / staff member

in Table 6.

1) Carry out the Qualitative Behaviour Assessment on the animals in the main pen on arrival at the farm and then the Flight distance (farmer in the same pen and assessor out of the pen).
2) Carry out individual measurements on the animals throughout milking.
3) Measure the surface area of the pens and the watering capacity in the pens.
4) Interview the farmer about the number of staff (full-time equivalent) and the presence of shade and watering in the grazing paddocks, collect the documents used to calculate adult and lamb mortality.

3.4.5. After the visit

1) Carry out the calculations as described for each of the indicators.
2) Return the result to the farmer

4. Discussion

The present work was performed:

1) to propose an animal welfare assessment protocol fully adapted to the dairy sheep context. The protocol was designed to be applicable in the field, to assess the five freedoms of animal welfare and to be objective and reproducible,
2) to test the proposed protocol on a large number of farms to generate a reference database to allow future field actions and/or scientific studies.

Both objectives were achieved and we will discuss both the methodological points of interest and the main results for field application.

4.1. The protocol design

The EBBEL protocol proposed in this study was mainly inspired by existing protocols AWIN (Dwyer et al., 2015) and Welfare Quality (Welfare Quality®, 2009).

The inclusion of different industrial stakeholders in the steering committee was essential to design a protocol with high level of feasibility, as observed in the development of the Welfare quality and the AWIN protocols. The farmers were especially interested because they were convinced of the link between animal welfare and farmer welfare. Some of them described a sort of malaise or lassitude when the flock is experiencing a problem. Our study was not designed to objectivise this relationship but the technical, scientific and social discussions during the process were necessary to build a solid basis for understanding, acceptation and future use of the results by the different stakeholders.

Also in accordance with the AWIN protocol, we proposed the calculation of the percentage satisfaction of animal welfare indicators because we believe, based on the current level of scientific knowledge, that a welfare assessment protocol intended to encourage farmers to implement higher animal welfare should inform them about the level of satisfaction of animal welfare in their farms (positive communication) rather than the severity of the problems detected (negative communication).

However, the final protocol included significant adaptations to the AWIN protocol as described in Material and Methods. The suggested visit format proposed allowed the assessment of all the individual, environmental and behavioural indicators of a farm within a maximum of 2 h. The milking parlour was defined as the place of choice to perform the individual measurements because dairy ewes enter and leave the parlour twice a day at a steady pace. In addition, the flat and hard floor of the parlour is ideal to perform locomotion scores. However, rotary parlours are not adapted to perform all the individual measurements. We estimate that 14.6 to 20% of the sheep dairy farms in France are equipped with a rotary parlour (internal unpublished statistics in 2023). On such farms, when adaptations on the parameters of the rotor cannot be considered, individual measurements must be performed in another place with a suitable surface. Despite this difficulty, the milking parlour was chosen because rotary parlours are a minority in dairy sheep farms.

4.2. Strategic decisions about behaviour measurements

4.2.1. Quantitative behaviour assessment

All behavioural observations were performed on a group of ewes in a pen, before milking and without any interactions between the assessor and the animals. In the pre-experiment, three behavioural assessments were included: quantitative social behaviour, inspired by Welfare Quality® (2009), Qualitative Behaviour Assessment and Human-Animal Relationship. The quantitative social behaviour assessment was based on filmography. This methodology had several advantages: the assessor could install the camera in a discreet place and leave the barn therefore avoiding any interference with animal behaviour; the initial minutes of the recording were ignored to avoid the period of overexpression of curiosity behaviour; and the forward/rewind functions of the recorder allowed all behaviours to be scored. To our knowledge, it is the first time such a study has reported a number of social behaviours observed during a short period (10 min) compatible with an assessment protocol. However, the quantitative social behaviour assessment was not included in the final protocol (Study 2) because it was considered to be too time consuming for field application. Further studies should be performed to improve the feasibility of these observations.

4.2.2. Human Animal Relationship

Despite the lack of consensus on the ideal method to assess the Human-Animal Relationship (Boivin et al., 2003), we decided to include this aspect in the protocol because:

1) as explained by Waiblinger et al. (2006) measuring the reaction of animals to humans would help determine how they perceive humans in general,
2) farmers from the region were interested in gaining understanding in this area.

We chose to assess the reaction of animals to a moving person (the farmer). The test included the measurement of the flight distance and the speed of reaction, previously validated by expert consensus (Phythian et al., 2011). Carrying out the test was very delicate because, as described by Rault et al. (2020) and Waiblinger et al. (2006), multiple factors are at play and can introduce bias in the results. The pre-experimental study was of fundamental importance to identify the variation factors and improve the assessment method in order to improve its reliability for further studies, such as Study 2.

4.3. Qualitative Behaviour Assessment (QBA)

The Qualitative Behaviour Assessment (QBA) was extensively discussed by the steering committee during the first meetings. Indeed, the subjectivity inherent in this methodology was questioned. However, discussions based on the literature (Fleming et al., 2016; Wickham et al., 2015; Phythian et al., 2011; Wemelsfelder, 2007; Wemelsfelder et al., 2000; Wemelsfelder, 1997) produced a consensus of opinion. QBA is currently the only valid test to assess the mental state in sheep and even if it does not tell us what the animals are feeling, it does detect their responses to their environment and is sensitive enough to detect variations in responses. Therefore, it was used in the protocol.

4.3.1. Lessons learnt from the pre-experiment

The pre-experiment (Study 1) was of great value to define operational details and improve the feasibility and reliability of the measurements. One important point for feasibility and reliability was the definition of the sampling method to produce a random selection of a maximum of four animals each time a batch of ewes entered each side of the milking parlour. This sampling procedure was consistent with the sample size recommended in the AWIN protocol (Dwyer et al., 2015) and is expected to be representative of the flock. Operational and methodological decisions were validated in Study 2.2: under these conditions, all the individual measurements could be performed (good feasibility) with a high level of agreement between the assessors (>90%). However, uncontrollability of the balance between positive and negative observations under field conditions reduced the level of agreement from a statistical point of view. Despite this methodological weakness, the reproducibility observed in the present study was comparable to the results obtained by others (Phythian et al., 2019; Angell et al., 2015; Phythian et al., 2012; Stubsjøen et al., 2011; Napolitano et al., 2009; Russel et al., 1969).

The most important methodological adaptations of the individual measurements in EBBEL were:

1) redefining of the assessment of claw length (with the aim of increasing the precision of the definition proposed in the AWIN protocol (Dwyer et al., 2015)),
2) reducing the significant skin lesion size which is considered to be problematic (from 10 cm in the longest axis in the AWIN protocol to 2 cm in present experiment). Indeed, we assumed that all lesions can cause pain and, in the dairy context, where animals are observed twice a day by the farmer, a skin lesion > 2 cm can already be treated,
3) separating the udder and teats into two distinct entities because the signs of concern are considerably different and do not necessarily have the same origins, nor the same impact on animals in terms of stress or pain. In the dairy context, where teats are mechanically milked at least once a day and where the udder can be inflamed or infected during lactation, it appeared important to pay special attention to both entities,
4) separating the cleanliness score into four different regions (lower legs, udder, lumbosacral and perianal regions). Hugues (2001) demonstrated in cows, that separating scoring areas improved the precision in interpreting the origin of the dirtiness (lying surface, pathways, bedding or diarrhoea) and to analyse risk factors. Ewes having dirty lower legs indicates inadequate building management practices and can be a risk factor for lameness (Caroprese et al., 2009) while a dirty udder is a risk factor for mastitis (Alba et al., 2019).

4.3.2. Building the reference database

The second objective of this study was to build a reference database for the set of indicators in the EBBEL protocol. We believe that the scientific community lacks objective data to support arguments to improve animal welfare. To our knowledge the EBBEL database is the largest database available in the literature for dairy ewes (81 farms – Study 2) and includes data from winter to summer, which is a good representation of the annual rearing conditions. The importance of creating this reference database was multiple:

1) to describe factually the current level of satisfaction of the indicators and obtain more transparency for any discussions or decisions related to animal welfare,
2) to identify the key points needing improvement – i.e. indicators with low level of satisfaction and/or with high variability between farms,
3) to serve as a reference to evaluate the effectiveness of an improvement plan on a farm or on a group of farms.

In summary, the database is the basic tool necessary to move from assessment to action. Periodic updates of the database will be necessary as field practices evolve.

4.3.3. Learning from the database

The information generated by the EBBEL database in the Roquefort region (Study 2) will be discussed in the next paragraphs.

Freedom from hunger and thirst (good feeding): The body condition score was considered satisfactory for the majority of the animals evaluated. The median farm had 99% of animals with a satisfactory body condition score and in the farm with the lowest score, 80% of the animals presented a satisfactory body condition. This result demonstrated that it is currently possible to feed animals correctly in a semi-intensive system to prevent extreme leanness or obesity. However, we observed high variability in compliance with water supply recommendations (Inn'Ovin, 2019). The median farm met or even exceeded the recommendations, but the farms in the inferior quartile did not meet the recommendations. Five of these farms (6.3%) supplied less than 60% of the recommendation and, according to the results of Casamassima et al. (2016), were at risk of compromising production. Water supply in the pasture was also variable. Water supply in the barn or in the pasture is a point of concern in this region, especially with climate change. The sanitary quality of the water (Schmidely et al. 2010), was not assessed in this protocol.

Freedom from pain, injury or disease (Good health): locomotion, nasal secretions, perianal cleanliness were highly satisfactory with very low variability between farms indicating good health management of the locomotor, respiratory and digestive systems. However, skin and teat integrity and udder health indicators were highly variable. Bergonier & Berthelot (2003) also observed variable prevalence of mastitis in dairy ewe farms in France. Mammary disorders do occur in lactating mammals but our results showed that high standards of udder and teat health are attainable despite the variability observed between farms. As the udder is of central importance for the dairy industry (Caroprese et al. 2016), the variability observed should raise awareness of the possibility of implementation of sanitary programs to improve udder and teat health in the farms with lower udder health and teat integrity satisfaction scores.

Freedom from discomfort due to the environment (Good housing): the majority of the environment-related indicators were variable, indicating variable environmental conditions on the farms. Claw length was the most variable indicator in our study suggesting that the claw management strategy could be improved. Indeed, according to Gelasakis et al. (2019), in semi-intensive and intensive systems, where the animals either remain housed or graze for short periods on soft and/or wet paddocks, the claws can become too long or deformed due to improper growth and this is a risk factor for lameness. Preventive claw trimming of all animals is necessary to remove overgrown claws. We also observed a certain variability in surface/animal. According to Caroprese et al. (2009) and Boe et al. (2006) a reduction in space allowance can have a negative impact on welfare and milk production. In our study, the majority of the farms (n = 59, 73.8%) agreed with the minimal recommendation of Inn'Ovin (2019). However, improvements are possible in farms not following recommendations (26.2% of the farms). In addition, the presence of trees and/or hedges which provide some protection against the sun, wind and rain varied from presence in all the paddocks to absence in all the paddocks. Together with water supply in paddocks, shade is an important means of combating heat stress (Abecia et al., 2017) and reducing the adverse behavioural and production effects of exposure to solar radiation on lactating ewes (Sevi et al., 2001).

Freedom from fear and distress and Freedom to express normal behaviour (Appropriate behaviour) were at least partly respected as the animals were housed in groups (a major behavioural requirement for sheep) and they had freedom to move in their environment in all farms. The variable reactions of the animals to the entrance of the farmer in the pen suggested different levels of HAR in the farms. Indeed, in some farms the animals did not display flight behaviour (suggesting an absence of fear) while in other farms the animals fled by walking (suggesting a moderate fear – the animals wanted to keep a distance). In none of the farms, the animals displayed panic behaviour (which would suggest strong fear).

The QBA was variable among farms. However, as explained by Fleming et al. (2016), the QBA should not be interpreted on its own and provides supplementary information for a welfare assessment.

5. Conclusion

The EBBEL protocol was designed to be specific for the sheep dairy context and especially when raw milk products are produced. In addition to assessing the compliance with all 5 freedoms, special attention was paid to udder health and teat integrity in order to maximise welfare in the dairy context. In addition, the format of the visit was designed to keep the duration of the visit below 2 h and to limit handling that is often restrictive and stressful for the animals, since farmers also prefer as little disturbance as possible. The participation of different stakeholders in the steering committee was a major component to ensure the feasibility of the protocol under conventional rearing conditions.

The database established in this study (81 dairy sheep farms) is of great value to stakeholders because it gives a clear and objective image of the farms and it helps to prioritise actions for animal welfare improvements.

In perspective, our group is enlarging the EBBEL database with data from other dairy production systems.

Funding source

The Fromageries Papillon funded the field activities (transport and a grant for the students) and provided a list of farmers for random selection and assessment. This work was supported by the Ecole nationale vtérinaire d'Alfort.

CRediT authorship contribution statement

Rouyer Elise: Data curation, Investigation, Writing – original draft. **Ponter Andrew Arthur:** Conceptualization, Writing – review & editing. **Jolly Pierre-Alexandre:** Conceptualization, Investigation, Methodology. **Patout Olivier:** Conceptualization, Formal analysis, Methodology. **de Paula Reis Alline:** Conceptualization, Formal analysis, Methodology, Project administration, Supervision, Validation, Writing – original draft, Writing – review & editing. **Foubert Gabrielle:** Investigation, Methodology. **Roman Mathilde:** Methodology, Writing – original draft. **Marin Lorie:** Investigation. **Meillac Elise:** Data curation, Investigation. **Bru Laetitia:** Funding acquisition, Resources.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank the members of the steering committee for fruitful discussions and important advice related to practical and human aspects of the protocol.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.smallrumres.2024.107209.

References

Abecia, J.A., Garcia, A., Castilho, L., Palacios, C., 2017. The effects of weather on milk production in dairy sheep vary by month of lambing and lactation phase. J. Anim. Behav. Biometeorol. 5 (2), 56–63. https://doi.org/10.31893/2318-1265jabb.v5n2p56-63.

Alba, D.F., Da Rosa, G., Hanauer, D., Saldanha, T.F., Souza, C.F., Baldissera, M.D., Dos Santos, D.S., Piovezani, A.P., Girardini, L.K., Da Silva, A.S., 2019. Subclinical mastitis in Lacaune sheep: causative agents, impacts on milk production, milk quality, oxidative profiles and treatment efficacy of ceftiofur. Microb. Pathog. 137, e103732 https://doi.org/10.1016/j.micpath.2019.103732.

Angell, J.W., Cripps, P.J., Grove-White, D.H., Duncan, J.S., 2015. A practical tool for locomotion scoring in sheep: reliability when used by veterinary surgeons and sheep farmers. Vet. Rec. 176 (20), 521. https://doi.org/10.1136/vr.102882.

Bergonier, D., Berthelot, X., 2003. New advances in epizootiology and control of ewe mastitis. Livest. Prod. Sci. 79 (1), 1–16. https://doi.org/10.1016/S0301-6226(02)00145-8.

Boe, K.E., Berg, S., Andersen, I.L., 2006. Resting behaviour and displacements in ewes—effects of reduced lying space and pen shape. Appl. Anim. Behav. Sci. 98 (3-4), 249–259. https://doi.org/10.1016/j.applanim.2005.10.001.

Boivin, X., Le Neindre, P., Boissy, A., Lensink, J., Trillat, G., Veissier, I., 2003. Eleveur et grands herbivores: une relation à entretenir. INRA Prod. Anim. 16 (2), 101–115. https://doi.org/10.20870/productions-animales.2003.16.2.3651.

Brambell, W.R., Barbour, D.S., 1965. Report of the Technical Committee to Enquire into the Welfare of Animals kept under Intensive Livestock Husbandry Systems, 4. Her Majesty's Stationery Office, p. 89.

Caroprese, M., Annicchiarico, G., Schena, L., 2009. Influence of space allowance and housing conditions on the welfare, immune response and production performance of dairy ewes. J. Dairy Res. 76 (1), 66–73. https://doi.org/10.1017/S0022029908003683.

Caroprese, M., Napolitano, F., Mattiello, S., Fthenakis, G.C., Ribo, O., Sevi, A., 2016. On-farm welfare monitoring of small ruminants. Small Rumin. Res. 135, 20–25. https://doi.org/10.1016/j.smallrumres.2015.12.010.

Casamassima, D., Vizzarri, F., Nardoia, M., Palazzo, M., 2016. The effect of water-restriction on various physiological indicators in intensively reared Lacaune ewes. Vet. Med-Czech. 61 (11), 623–634. https://doi.org/10.17221/144/2015-VETMED.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Meas. 20 (1), 37–46. https://doi.org/10.1177/001316446002000104.

Desquilbet, L., 2019. Guide pratique de validation statistique de méthodes de mesure: répétabilité, reproductibilité, et concordance. ⟨https://hal.archives-ouvertes.fr/hal-02103716⟩ (Accessed 02 June, 2021).

Dwyer, C., Ruiz, R., Beltran de Heredia, I., Canali, E., Barbieri, S., Zanella, A., 2015. AWIN Welfare assessment protocol for sheep. AWIN Anim. Welf. Indic. 72.

EFSA, 2012. Scientific opinion: " Statement on the use of animal-based measures to assess the welfare of animals ". Panel Anim. Health Welf. EFSA J. 10 (6), 2767–2796. https://doi.org/10.2903/j.efsa.2012.2767.

EFSA, 2014. Scientific opinion: " Scientific Opinion on the welfare risks related to the farming of sheep for wool, meat and milk production ". EFSA J. 12 (12), 121–128. https://doi.org/10.2903/j.efsa.2014.3933.

Eurobarometer, 2016. Report: attitudes of europeans towards animal welfare. Spec. Eur. 442, 86.

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 (5), 878–887. https://doi.org/10.1017/S1751731114003164.

Fleming, P.A., Clarke, T., Wickham, S.L., Stockman, C.A., Barnes, A.L., Collins, T., Miller, D.W., 2016. The contribution of qualitative behavioural assessment to appraisal of livestock welfare. Anim. Prod. Sci. 56 (10), 1569–1578. https://doi.org/10.1071/an15101.

Gelasakis, A.I., Kalogianni, A.I., Bossis, I., 2019. Aetiology, risk factors, diagnosis and control of foot-related lameness in dairy sheep. Animals 9 (8), 509. https://doi.org/10.3390/ani9080509.

Hughes, B.O., 1976. Behaviour as an index of welfare. Proc. 5th Eur. Poult. Conf. Hug 8, 1005–1018.

Hugues, J., 2001. A system for assessing cow cleanliness. Pract 23 (9), 517–524. https://doi.org/10.1136/inpract.23.9.517.

IDELE, 2015. BIENE: protocole d'évaluation du bien-être des brebis en condition de pâturage hivernal, test en exploitation d'élevage 2ème année., Résultats (report n°00 15 301 041), 42.

Inn'Ovin, 2019. Guide de bonnes pratiques ovines. 79p. https://www.inn-ovin.fr/le-guide-des-bonnes-pratiques-en-elevage-ovin-version-2020/.

Jolly, P., 2020. Bien-être des brebis laitières dans le rayon Roquefort: mise en place d'un protocole d'évaluation et constitution d'une base de référence en vue de la mise en place d'une charte. Thèse de médecine vétérinaire, École nationale vétérinaire d'Alfort. p. 227.

Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33 (1), 159–174.

Marcone, G., Carnovale, F., Arney, D., De Rosa, G., Napolitano, F., 2022. A simple method for on-farm evaluation of sheep welfare using animal-based indicators. Small Rumin. Res. vol.208, 106636 https://doi.org/10.1016/j.smallrumres.2022.106636.

Munoz, C.A., Campbell, A.J.D., Hemsworth, P.H., Doyle, R.E., 2019. Evaluating the welfare of extensively managed sheep. e0218603 PLoS ONE 14 (6), 14. https://doi.org/10.1371/journal.pone.0218503.

Napolitano, F., De Rosa, G., Ferrante, V., Grasso, F., Braghieri, A., 2009. Monitoring the welfare of sheep in organic and conventional farms using an ANI 35 L derived method. Small Rumin. Res. 83 (1-3), 49–57. https://doi.org/10.1016/j.smallrumres.2009.04.001.

Phythian, C.J., Michalopoulou, E., Jones, P.H., Winter, A.C., 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 (6), 943–952. https://doi.org/10.1017/S1751731110002594.

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. 193 (1), 257–263. https://doi.org/10.1016/j.tvjl.2011.12.006.

Phythian, C.J., Michalopoulou, E., Duncan, J.S., 2019. Assessing the validity of animal-based indicators of sheep health and welfare: do observers agree? Agriculture 9 (5), 88–102. https://doi.org/10.3390/agriculture9050088.

Rault, J.-L., Waiblinger, S., Boivin, X., Hemsworth, P., 2020. The power of a positive human–animal relationship for animal welfare. Front. Vet. Sci. 7, 590867 https://doi.org/10.3389/fvets.2020.590867.

Russel, A.J.F., Doney, J.M., Gunn, R.G., 1969. Subjective assessment of body fat in live sheep. J. Agric. Sci. 72 (3), 451–454. https://doi.org/10.1017/S0021859600024874.

Schmidely, P., Bayourthe, C., Enjalbert, F., Gaudré, D., Grosjean, F., Jurjanz, S., Paris, A., Pouliquen, H., Soyeux, Y., Duchemin, J., Mania, J., Montiel, A., Vialette, M., Ganiere, J.P., Vialard, J., Kammerer, M., 2010. État des lieux des pratiques et recommandations relatives à la qualité sanitaire de l'eau d'abreuvement des animaux d'élevage. ANSES édition Sci. 124.

Sevi, A., Annicchiarico, G., Albenzio, M., Taibi, L., Muscio, A., Dell'Aquila, S., 2001. Effects of solar radiation and feeding time on behavior, immune response and production of lactating ewes under high ambient temperature. J. Dairy Sci. 84 (3), 629–640. https://doi.org/10.3168/jds.S0022-0302(01)74518-3.

Stubsjøen, S., Hektoen, L., Valle, P., Janczak, A., Zanella, A., 2011. Assessment of sheep welfare using on-farm registrations and performance data. Anim. Welf. 20 (2), 239–251. ⟨https://doi:10.1017/S0962728600002724⟩.

Waiblinger, S., Boivin, X., Pedersen, V., Tosi, M.-V., Janczak, A.M., Visser, E.K., Jones, R. B., 2006. Assessing the human–animal relationship in farmed species: a critical review. Appl. Anim. Behav. Sci. 101 (3-4), 185–242. https://doi.org/10.1016/j.applanim.2006.02.001.

Welfare Quality®, 2009. Welfare Quality® assessment protocol for cattle. Welfare Quality® Consortium, Lelystad, Netherlands.

Wemelsfelder, F., 1997. The scientific validity of subjective concepts in models of animal welfare. Appl. Anim. Behav. Sci. 53 (1-2), 75–88. https://doi.org/10.1016/S0168-1591(96)01152-5.

Wemelsfelder, F., 2007. How animals communicate quality of life: the qualitative assessment of behaviour. Anim. Welf. 16, 25–36.

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 (3), 193–215. https://doi.org/10.1016/S0168-1591(99)00093-3.

Wickham, S.L., Collins, T., Barnes, A.L., Miller, D.W., Beatty, D.T., Stockman, C.A., 2015. Validating the use of qualitative behavioral assessment as a measure of the welfare of sheep during transport. J. Appl. Anim. Welf. Sci. 18 (3), 269–286. https://doi.org/10.1080/10888705.2015.1005302.