ENCAW
Working group for Area 1 – Welfare indicators

Report

Authors (by alphabetic order)
Antoni Dalmau
Ute Knierim
Luc Mounier
Isabelle Veissier (coordinator)
Antonio Velarde
Christoph Winckler

ENCAW – Working group for Area 1 – Welfare indicators Report

1. Abstract

On-farm animal welfare assessment may be predominantly based on resource based measures such as the type of housing, feeding, health plan (e.g. in the Animal Needs Index) or on animal based measures such as body condition, behaviour or health measures (e.g. in Welfare Quality®). The choice between the two types of measures seems dogmatic, some authors considering that resource based measures detect risk factors for animal welfare whereas animal based measures better reflect the true welfare state of animals. In order to review available resource vs. animal based measures, we first drew a list of properties for a measure to be considered valid for assessing animal welfare. These properties have been derived from analytical methods: selectivity, trueness, reliability (repeatability, reproducibility), stability over time, fitness for the purpose (including sensitivity), and feasibility in different systems. The properties of the measures for the 12 welfare criteria defined in Welfare Quality® were analysed. For some criteria, animal based measures are more suitable than resource based measures, this is the case for absence of prolonged hunger, absence of injuries or diseases or good human-animal-relationship for which different factors interact so that resource based measures have a low predictive value. In other cases, resource based measures are more suitable because animal based measures are not sensitive enough (for absence of thirst) or not feasible (detection of pain during some management procedures on farm). Finally, for some criteria a combination of resource and animal based measures is necessary. This is the case for comfort around resting where problems (negative aspects) can be detected on animals while positive aspects – the actual comfort – may be assessed through resource based measures.
In conclusion, there is no general rule whether animal based measures are more or less valid than resource based ones. This depends on the welfare criterion considered and to a lesser extent on the animal type. We recommend to use a mixture of animal and resource based measures (chosen according to their validity and feasibility) to assess the overall welfare of animals on farms.

2. Introduction: Context and objectives

The ENCAW-project is coordinated by a Management Team consisting of Harry Blokhuis (Chair of the Steering Group at SCAW, Swedish Centre for Animal Welfare), Mats Sjöquist (Director SCAW), Margareta Steen (Deputy Director SCAW) and Peter Svensson (Project Manager). Within the ENCAW-project a number of activities and pilot studies are carried out to further develop and define the tasks of a future European Network of reference Centres for Animal Welfare (ENCAW) and to gain relevant experience in related working processes. One such pilot study is an impact assessment on a relevant topic in order to explore and analyze the challenges of carrying out such an assessment through a 'coordinated network' effort. In the European Commission's context 'impact assessment' is understood as the process that prepares information for political decision-makers on the advantages and disadvantages of possible policy options by assessing their potential impact. This generally refers to potential economic, social and environmental consequences that new initiatives may have.
This activity aims

- to address the questions (grouped in three areas) mentioned below as part of a pilot impact assessment
- to imitate a procedure of a European Network of reference Centres for Animal Welfare

The overarching question for the impact study is : "What are the consequences of the introduction of animal based welfare assessment in legislation and official controls?"
Three more specific areas of interest were defined that should be addressed:

- Area 1: Welfare assessment (coordinator: Isabelle Veissier, INRA, France)

- Area 2: Economy (coordinator: Ruben Hoffman, SLU, Uppsala, Sweden)
- Area 3: Acceptation/Attitudes (coordinator: Karl Bruckmeier, Univ. Gothenburg, Sweden)

The present report covers Area 1. Animal based measures and resource based measures are to be compared. More specifically, this area addresses two major questions :

- What are the advantages/disadvantages of animal based measures and resource based measures in terms of the quality of the welfare assessment on farm (i.e. from a biological/fundamental point of view)? The issues of variation over time should also be addressed (animal based assessment may give more variable outcomes over time than the measurements of resources).
- If assessment is to be based on animal based measures should/could one fully rely on those or should resource based measures still be included?

3. Approach

3.1. The working group

Isabelle Veissier (Inra, France) was asked to coordinate the work package to study Area 1. She set up a working group composed of:

- L Mounier (VetAgro Sup, France)
- Christoph Winckler (Boku, Austria) and Ute Knierim (Univ. Kassel, Germany) working together
- Antonio Velarde and Antoni Dalmau (Irta, Spain) working together

The members of the working group were chosen for their expertise in assessing animal welfare. They had been involved in the Welfare Quality® project.

3.2. Welfare criteria

Welfare is a multidimensional concept. It includes physical and mental health (Hughes 1976, Dawkins 2006). Several aspects such as physical comfort, absence of hunger and disease, possibilities to perform motivated behaviour, etc. need to be covered for the welfare of an animal to be ensured (Farm Animal Welfare Council 1992, Fraser 1995, Bracke *et al* 1999, Botreau *et al* 2007). The Welfare Quality® project drew a list of 12 welfare criteria that can be checked independently from each other, the whole list is supposed to be exhaustive i.e.to cover all domains/aspects of animal welfare (Welfare Quality® 2009):

- Absence of prolonged hunger: Animals should not suffer from prolonged hunger, i.e. they should have a suitable and appropriate diet.
- Absence of prolonged thirst: Animals should not suffer from prolonged thirst, i.e. they should have a sufficient and accessible water supply.
- Comfort around resting: Animals should have comfort when they are resting.
- Thermal comfort: Animals should neither be too hot nor too cold.
- Ease of movement: Animals should have enough space to be able to move around freely.
- Absence of injuries: Animals should be free of injuries, e.g. skin damage and locomotory disorders.
- Absence of disease: Animals should be free from disease, i.e. animal unit managers should maintain high standards of hygiene and care.
- Absence of pain induced by management procedures: Animals should not suffer pain induced by inappropriate management, handling, slaughter, or surgical procedures (e.g. castration, dehorning).
- Expression of social behaviours: Animals should be able to express normal, non-harmful, social behaviours (e.g. allo-grooming).
- Expression of other behaviours: Animals should be able to express other normal behaviours, i.e. it should be possible to express species-specific natural behaviours such as foraging.

- Good human-animal relationship: Animals should be handled well in all situations, i.e. handlers should promote good human-animal relationships.
- Positive emotional state: Negative emotions such as fear, distress, frustration or apathy should be avoided whereas positive emotions such as security or contentment should be promoted.

The present task was developed from the Welfare Quality® list of criteria and the working group investigated a series of measures that potentially could be used to check each criterion.

3.3. How the working group developed the task

On-farm animal welfare assessment may be predominantly based on resources based measures such as the type of housing, feeding, health plan (as in the Animal Needs Index (Bartussek 2001)) or on output measures (i.e. on animals as in Welfare Quality®). The choice between these two types of measures seems dogmatic, some authors considering that design indicators are risk factors for animal welfare whereas output measures better reflect the true welfare state of animals.

- On which ground we compared animal vs. resource based measures?

In order to review available design or output measures, we first drew a list of properties for a measure to be considered valid for assessing animal welfare. These properties were derived from analytical methods: selectivity, trueness, reliability (repeatability, reproducibility), stability over time, fitness for the purpose (including sensitivity), and feasibility in different systems.

- What measures were reviewed?

We considered each Welfare Quality® criterion and reviewed measures (from Welfare Quality® or not) used to check each criterion. The group could not look at all species used in farming. This would have implied too much work within the time constraints of the task (6 months). For each welfare criterion, examples from several species, taken as models, were analysed. Special attention was drawn on species covered in Welfare Quality® (cattle, pigs, and poultry) because there was more available information on these species.

- What domain did we cover?

As in Welfare Quality®, we considered measures to check the welfare of animals on the farm and at slaughter. We worked in the context of a farm / slaughter visit for a certification. The total duration of the visit should be 1 day maximum. So too time consuming or too costly measures were not considered (e.g. some detailed measures specific to experiments were excluded from the analysis)

4. Characteristics to assess the quality of measures

To assess the overall quality of a welfare measure, we considered the properties defined for analytical methods (e.g. biochemical assays) and we transposed them for welfare measures:

4.1. Selectivity

For analytical methods, the selectivity is the degree to which a method can quantify the analyte accurately in the presence of interferents, i.e. other analytes (Thompson *et al* 2002)
For a welfare measure, selectivity may first refer to the degree to which a measure can quantify what we want to analyse and not something else. This property should be analysed in reference to the welfare criterion the measure is supposed to bring information about. It is often called the validity of a measure although to be valid a measure not only needs to be selective but also true, repeatable, and fit for the purpose (see below). There are several ways to check that a measure really provides good information to assess a welfare criterion (Scott *et al* 2001)(Reenen et al, 2004, presentation at Welfare Quality® meeting):

- Comparing the results produced by the method and those produced by another one already validated which would serve as a gold standard (concurrent validity). As far as animal welfare is concerned, there is generally no gold standard.
- Comparing the effects of conditions or treatments or demonstrate causal relationship between treatment and effect (predictive or construct validity).
- When the concurrent or predictive validity cannot be established objectively, at least experts should agree on the validity of a measure according to their experience (consensus or face validity)

A second issue is whether there are no confounding factors that can influence the results.
For instance, regarding the criterion Good human-animal relationship, one may ask whether the approach test (measuring at what distance we can approach an animal before it moves away) when the animal is feeding, reflects the quality of the human-animal relationship. This question was answered in calves by comparing animals that received positive versus negative contacts with humans: the former accepted to be approached at a closer distance than the latter (Lensink *et al* 2003). By this, predictive or construct validity was demonstrated. Nevertheless, lameness can affect the results of the test: lame cows can be approached by a shorter distance than healthy ones, probably because the motivation to avoid humans is counterbalanced by the pain caused by the movement (Špinka *et al* 2005). In that case lameness can be a confounding factor.

4.2. Trueness

For analytical methods, trueness refers to the closeness of agreement between a test result and the accepted reference value of what is being measures (Thompson *et al* 2002).
For welfare measures, in many cases there is no reference value. Nevertheless, we can compare the result produced by a measure according to a given method to the results of a more detailed method. For instance, regarding behaviour, one may compare the results obtained on a short period to those obtained by longer observations; and in case scan sampling is used (i.e. one snap shot observation every fixed interval of time) the results can be compared to those obtained with continuous observations.

4.3. Precision

For analytical methods, the precision is the closeness of agreement between independent test results obtained under stipulated conditions (in general expressed as standard deviation or relative standard deviation (Thompson *et al* 2002). It includes:

- Intra-assay repeatability, i.e. the precision between measures done by the same operator with the same instrument (Feinberg 1996)
- Inter-assay repeatability (sometime called reproducibility), i.e. the precision between measures done by several operators, with different instruments, at different times… (Feinberg 1996)

For welfare measures, authors ofen refer to

- Intra-observer repeatability, i.e. the precision between measures done by the same observer in similar conditions. This can be tested using video recordings for behaviours or or assessing the same animals again after a short time lapse for clinical symptoms.
- Inter-observer repeatability, i.e. the precision between measures done by several observers, in similar conditions. This may be checked by asking several observers to look at the same animals at the same time, alternatively from videos or photos.

Very often the repeatability is expressed by the correlations between observations in case of continuous data (Spearman or Pearson correlations) or concordance between observers in case of discrete data (Kendall's coefficient of concordance). In the present document we will consider that a correlation of 0.7 or concordance of 0.40 or more are acceptable (Fleiss *et al* 2003 , Martin & Bateson 2007 , Knierim & Winckler 2009) on the condition that these values are significant.

4.4. Fitness for purpose

The fitness for purpose is the extent to which the performance of a method matches the criteria, agreed between the analyst and the end-user of the data, that describe the end-user's need (Thompson *et al* 2002). The fitness for purpose can be based on:
- the detection limit, i.e. the smallest amount of concentration of analyte in the test sample that can be reliably distinguished from 0 (Thompson *et al* 2002)
- the quantification limit, i.e. the concentration below which the analytical method cannot operate with an acceptable precision (sometime defined as 10% RSD or 2 x detection limit) (Thompson *et al* 2002)
- the sensitivity, i.e. the importance of the variation of the instrumental response in relation to variations of the analyte (Feinberg 1996)
- the validated range, i.e. the interval of analyte concentration within which the method can be regarded as validated (Thompson *et al* 2002)
- the calibration and linearity, i.e. the limits between which a linear model can be used to predict correctly the concentration of an analyte (Feinberg 1996)

For welfare measures, we usually do not talk of an exact quantification but of whether we can check that a welfare criterion is fulfilled. Therefore, the distinction between the detection limit and the quantification limit is generally not done. To assess whether the detection/quantification limit of a measure is fit for the purpose, we may question whether a method can be used to detect slight welfare problems. For instance, sunken eyes is used to detect dehydration in calves however, this indicates an extreme state where this animal urgently needs to be rehydrated, often by intravenous administration of a liquid. This symptom is not sensitive enough to detect thirst that may occur on a farm if the drinking points are under the recommended number. Nevertheless, it seems appropriate to detect dehydration due to very long transports during warm weather. A further question regarding the quantification limit relates to the occurrence of rare behaviours (such as certain abnormal behaviours or play in adults), which in principle can be detected, but not reliably quantified within a given time frame (Knierim & Winckler 2009). Regarding sensitivity, one may question whether a measure allows highlighting differences between farms or slaughterplants.
We decided not to take into account the limits for linearity at this stage. As a matter of fact, what is important for a welfare measure is the relation between the result of what we measure and the level of welfare. In Welfare Quality®, it turned out that the experts consulted to assess the severity of problems never followed a linear reasoning. For instance, a 10% increase or decrease of a problem did not lead to a proportional decrement in the assessment of the related criterion: regarding lameness 0% cows (perfect situation) brought a score for absence of lameness of 100/100), 10% lame cows brought a score of 46/100 (so a reduction by 56) while 90 or 100% lame cows brought very similar scores (2 to 0). In Welfare Quality®, I-spline functions were used to address this problem (Botreau *et al* 2008).
Finally the validated range may be of importance, not only at the bottom of the scale (to detect minor problems) but also at the other extreme, to distinguish between moderate and severe welfare problems. As a matter of fact, according to Swedish law, the welfare of all animals must be ensured (Djurskyddslag (1988:534)) so it is necessary to detect severe cases that require more attention from the farmer.

4.5. Ruggedness

For analytical method, the ruggedness is the resistance to change in the result produced by an analytical method when minor deviations are made from the experimental conditions described in the procedure (Thompson *et al* 2002).
For welfare measures, this may be transposed to **stability over time**, especially day to day variations. The results of a welfare assessment should be representative of the long-term farm situation. They should not be sensitive to changes in environmental or internal conditions that are not significant for the welfare of the animals. Thus, similar recordings should be achieved at different times if no major changes on the farms occurred (Knierim & Winckler 2009). However,

there might be differences in the results obtained from some measures because the conditions for observation have slightly changed, e.g. changed weather conditions, small changes in the management or exchange of animals.
Although the overall conditions on a farm or a slaughter plant may not change during a given period and the overall welfare of animals should thus be stable, there might be differences in the results obtained from some measures because the conditions for observation have slightly changed or because of random variations between days. This is particularly true for behaviours which occurrence is multifactorial: weather conditions, occasional events like small changes in the routine management of the farm can affect the expression of behaviours.

4.6. Matrix variation

For analytical methods, the matrix is the set of constituents present in the material on which the method is applied. These constituents may have an effect on the results of the method (Thompson *et al* 2002)(Feinberg 1996). For instance the determination of hormones in different tissues or species may require specific methodologies.
For welfare measures to be applied on farms or at slaughter, it is important to check if the measure can be applied in different farming systems. For instance detecting lameness in cows in a loose barn vs. tied stalls requires a different method. In Welfare quality® lameness is assessed in loose housed cows by making them walk in a straight line and checking if the cows bears its weight equally on the four limbs and makes regular steps. For tied cows, the observer checks if the cow stands on its four limbs when undisturbed then the observer makes the cow move to the left and to the right, observing how she shifts weight from foot to foot (Welfare Quality® 2009).

4.7. Other property taken into account: feasibility

In case of observations on farms or at slaughter that must be performed on a large scale and in commercial conditions to produce an overall assessment of the farm or the slaugtherplant, the issue of feasibility of measures must be taken into account. Feasibility can be assessed by the time necessary to carry a measure and also by the need to have access to specific devices to perform a measure, or to the need to have specific skills to perform the measure. For the present study, we considered that the whole visit of a farm or a slaugtherplant should not exceed 1 day, that the observers were trained for 3-4 days before the observations but were not specialists on each welfare criterion addressed (e.g. they were not vets and could only detect clear cut clinical symptoms on animals for the criteria Absence of injuries and Absence of disease). The feasibility was assessed according to the knowledge and devices available at present. When more automatic recordings become available then the feasibility of some measures may be increased (e.g. automatic recordings of animals' movements).

5. Review of characteristics of animal based and resource based measures for each welfare criterion

For each welfare criterion, the properties of measures based on animals and on resources were analysed. The results are given in tables in Annex 1. We will discuss here the main properties of each measure in order to be able to decide whether animal based or resource based measures should be used to check a given criterion.
Within each criterion, several measures can be taken into consideration; they are not necessarily mutually exclusive but may reflect different aspects of the criterion (e.g. different disease for the criterion Absence of disease).

5.1. Absence of prolonged hunger

For the criterion Absence of prolonged hunger, the following measures were considered (Table 1):

- Animal based measures : body condition, rumen filling, vocalisation/aggression level, stereotypies, back fat thickness and stomach/gut weight (measured at slaughter)

- Resource based measures: feeder space, energy supply in the diet

Body condition score provides a reasonably accurate measure of the body reserves of an adult animal and can thus reflect the quality of feeding during a prolonged time. The scale need to be adapted for species and breeds. To avoid variations with physiological stage, only very lean animals should be considered. Body condition score has not been validated for young animals (calves, piglets).
Rumen filling (only available in ruminants) is not selective enough: it is not necessarily associated with prolonged hunger and can be confounded with health problems; in addition, it is not stable over time.
The feasibility of back fat thickness is moderate: the method is available only for pigs, it requires a special device and handling of the animals. Fitness for the purpose is low due to an inadequate detection of very lean animals whose welfare is definitively at risk.
Vocalisation or aggression levels are not selective enough; they may result from hunger but also from social interactions or the interaction of both.
Underfeeding can result in redirected oral behaviours which may develop into stereotypies (D'Eath *et al* 2009). This has been especially observed in sows (De Leeuw & Ekkel 2004), hens (Savory & Lariviere 2000), and to a lesser extent in cows (Lindstrom & Redbo 2000). However other factors influence the occurrence of such behaviours: the restriction of foraging activity the lack of activity in general or boredom (Wemelsfelder 1990, Terlouw & Lawrence 1992, Jensen *et al* 2010). Therefore stereotypies do not appear selective enough to detect absence of hunger. In addition, when the occurrence of stereotypies is rare (as in cattle) then the measure may require too long observations so feasibility is low.
In pigs, the weight of the full guts can indicate the time during which the food was removed before slaughter. This measure is probably not appropriate for birds because they do not digest during transport (Lindblad, personal communication, 2006, cited in Welfare quality® report no. 9: Assessment of animal welfare measures for layers and broilers).

Feeder space affects the access of animals to food. However other factors play a role: the social cohesion of the group (if the hierarchy is not marked then all animals can have access to food) and the quantity and the quality of the food (see below: energy supply).
Regarding the energy supply in the diet, it is necessary to take into account the nutritional quality of the feed, the amount of feed the animal is provided and the maximum it can eat, and also the needs of the animal according to its physiological stage. For this the advice of a specialist is needed; therefore assessing the balance between the energy supply and the needs of the animals appear not feasible.

5.2. Absence of prolonged thirst

For the criterion Absence of prolonged thirst, the following measures were considered (Table 2):
- Animal based measures: skin test, sunken eyes, plasma osmolarity or hematocrite
- Resource based measures: water provision including the number of water points, their functioning and cleanliness

Although often used by vets to detect dehydration especially in young animals, the skin test (that consists of observing how fast the skin returns to its normal contour) appears to be not selective: the delay needed for the skin to resume its initial position after having been pinched is linked neither to serum osmolarity, or packed cell volume which indicate the level of hydration nor to drinking behaviour (Pritchard *et al* 2006). Sunken eyes are also used by vets to detect dehydration. It seems that this sign detects only extreme cases of dehydration. Such cases do not occur in normal farm practices, except in case of a disease with strong diarrhoea. It is thus not fit for the purpose of detecting thirst due to suboptimal provision of water. It may nevertheless be used to detect dehydration after long transports.
Plasma osmolarity and hematocrite increase in case animals are deprived of water (Knowles *et al* 1995, Pritchard *et al* 2006). However, at the moment, more investigations are needed to check if these indicators can help to distinguish a moderate dehydration vs. a severe one. This measure requires a blood sample; this can be obtained easily at slaughter (from bleeding) but is more difficult on farms.

There exist recommendations on the number of water points and the total amount of water to provide to animals according to their size and/or physiological stage. It is also acknowledged that animals avoid dirty water. Therefore, the number of water points (in comparison to the number and type of animals), their water flow and cleanliness can provide an accurate estimate of whether or not the animals have sufficient access to water, although social cohesion plays also a role (limiting competition for resources between animals). In addition, the feasibility and repeatability of these measures are very high.

5.3. Comfort around resting

For the criterion Comfort around resting, the following measures were considered (Table 3):

- Animal based measures: duration of lying down and of rising up, total lying time and number of lying periods, collisions with equipment, abnormal sequences, lying outside the lying area, ruminating during lying in cattle, cleanliness of the animals, bursitis in pigs, ecto-parasites in poultry (red mites)
- Resource based measures: quality of the floor (roughness), bedding material, ergonomics of cubicles for cattle, presence of dust

Comfort around resting includes two main aspects: the quality of the flooring / bedding (more or less soft, clean, providing insulation and enough space to rest) and the design of the resting area that makes lying down / rising up movements more or less easy.
A rough lying surface can result in skin lesions: hair losses and bursitis, especially in pigs (Mouttotou *et al* 1998). This sign is selective and sensitive and thus very appropriate to assess the comfort around resting.
Animals usually avoid lying down on dirty surfaces (Phillips & Morris 2002). Therefore the cleanliness of the flanks and upper legs (cattle, pigs) can be used to assess the comfort of resting (we suppose that the animal could not lie on a clean surface and thus may feel discomfort, in addition a small lying area is more likely to be soiled). In pigs, this sign is less selective than in cattle, thermoregulation being a confounding factor (see Thermal comfort). In broilers, the cleanliness of the ventral plumage can be checked. This sign is not relevant for hens because they can rest on perches.
Animals lie down for shorter when the lying area is uncomfortable (Haley *et al* 2000, Tucker *et al* 2004). The detection of such sign requires long observations which does not seem feasible during a short visit of a farm. In case the lying area is very uncomfortable they may lie outside it. Also a too small lying area can result in animals lying partly outside this area (e.g. when cubicles for cattle are not long enough). The proportion of animals lying outside or partly outside the lying area is thus likely to be fit to detect absence of comfort of the lying area.
Difficulties in lying down and rising up can be detected visually in cattle. It is not common to perform such observations in other species although this could be meaningful in animals which are restricted, e.g. tethered sows. The duration of the movements and hits against obstacles can be measured. These measures are selective: they are linked to the design of the lying area (e.g. how the partitions between cubicles are designed, the height of the neck rails… (Veissier *et al* 2004)). Nevertheless lameness may increase the duration of lying down and rising up. Lying down and rising up are not very frequent so they need long observations of animals unless the observation is carried out at a time when there is a high probability of such movements. This is the case when one observes cows after morning milking: the animal first eat (they are generally fed at that time) then lie down.
The quality of the lying area can be assessed. It is possible to assess the roughness of the floor and the presence and quality of bedding (softness, cleanliness). However, these assessments are qualitative. More effort on the standardisation of such measures is necessary. In cattle housed in a shed with cubicles, the observer can record the type of cubicles, their dimensions and the exact place of the various bars (e.g. neck rail, presence of a brisket board). However, it is difficult to interpret these measures due to many interacting factors.

5.4. Thermal comfort

For the criterion Thermal comfort, the following measures were considered (Table 4):

- Animal based measures: signs of heat: panting, wallowing in pigs, reduced eating in cattle; signs of cold: huddling, shivering (cold);

- Resource based measures: temperature, relative humidity, air velocity

Heat induces an increase in respiratory rate. In extreme cases, this can result in panting (pigs, and probably broilers). The measure of the respiratory rate seems fit to detect heat in pigs (and maybe in poultry) but not in cattle because in this species only extreme temperatures increase respiratory rate. If they have access to a wet mixture, pigs wallow in case of heat (Huynh *et al* 2005). However, this sign cannot be observed in all farming systems (e.g. on fully slatted floors). Pigs maintained at high temperature tend to have their body soiled; this can be observed in all systems. However, the cleanliness of pigs depends also on management (cleanliness of the pens) and space so the selectivity is moderate. Cattle react to heat by reducing their feeding behaviour, especially during hot hours (Taweel *et al* 2006); however this sign requires long observations which are not compatible with a short farm visit.

In pigs, cold can result in huddling and shivering (Huynh *et al* 2005). Although huddling can be affected by social cohesion and space allowance, it seems a selective indicator and fit for the purpose (huddling frequency varies significantly with temperature). Surprisingly, the selectivity of shivering is not confirmed because this sign does not seem to vary much with temperature. It may be fit to detect only extreme cold. Cattle seem not to express cold markedly.

Measures of temperature, air humidity and velocity can in theory be applied on farms easily. However, the number of measures and the exact place where they are performed are of crucial importance and need expertise to be able to run an adequate evaluation of what the animals may perceive. The simplification of the measurements needs further work (Capdeville, personal communication).

5.5. Ease of movement

For the criterion Ease of movement, the following measures were considered (Table 5):

- Animal based measures: slipping/falling, moving,
- Resource based measures: state of the floor (e.g. inclination of a ramp), stocking density

Ease of movement depends on the possibility to move thanks to a large space allowance and to appropriate flooring (not slippery).

At slaughter, slipping and falling can be easily detected on animals. They occur quite often when animals (pigs, cattle) are downloaded from trucks. The measure is selective, feasible and fit for the purpose at slaughter. Information on the quality of a ramp to download animals can be recorded: slope of the ramp, quality of flooring, presence of stairs, lateral partitions… These factors interact with each other so that it is difficult to predict accurately if downloading will be difficult for animals. On a farm, slipping and falling are rare. Even on a slippery floor the animals may not slip or fall because they adjust their pace, e.g. making smaller steps. Hence this measure seems not fit for the purpose. The time spent walking can be recorded on a farm. However moving about depends on many factors: the possibility to move offered by the environment but also the motivation of animals to move about which can come from social behaviour, exploration, need to rest… Therefore, this measure appears not selective and not stable in time (due to variable motivations of the animals). In addition, to assess accurately the time spent walking, one must observe the animal for a very long period, e.g. during the whole daylight period, because the arousal of animals varies during the day and so their walking activity. On farm, the stocking density combined with information on how the space is structured and the quality of the floor may provide a good estimation of the ease of movement (selective, feasible, and repeatable).

5.6. Absence of injuries

For the criterion Absence of injuries, the following measures were considered (Table 6):

- Animal based measures: skin lesions, locomotion
- Resource based measures: stocking density, floor quality (too rough or too slippery), presence of holes or obstacles on floors, walls or doors;

This criterion contains two main aspects: skin lesions and lameness. Both can be observed visually on animals. A close inspection is necessary to detect skin lesions (hairless patches, wounds, scars); this can be done together with other clinical observations (e.g. cleanliness, observation of nasal or ocular rejections) so that the time necessary for the measure can be reduced. Lame animals can be detected by observing them when they are walking. This requires that some space

is available for the animal to move freely. The two measures (skin injuries and lameness detected by visual observation) are very selective. Actually they measure the disorders that are aimed to be measured. In addition walking abilities have been linked to claw lesions (Winckler & Willen 2001). The repeatability of these measures within and between observers is very good. The measures are not stable in time, reflecting the variable prevalence of these disorders with time.
Injuries on animals depend on the quality of the environment, especially the quality of the floor and the presence of obstacles. They also depend on stocking density, which affect the risk for animals to be in contact with obstacles. A high stocking density can also create social tensions and therefore injuries due to fighting / biting (pigs). Although all these elements are risk factors it is difficult to predict the amount of injuries accurately from them.

5.7. Absence of disease

For the criterion Absence of disease, the following measures were considered (Table 7):

- Animal based measures: respiratory disorders, enteric disorders, reproductive disorders, low performances (incl. low reproduction), sickness behaviour
- Resource based measures: biosecurity and control of diseases

Like injuries, clinical symptoms of diseases can be observed directly on animals or lesions produced by diseases can be detected at slaughter. Of course, during a short farm visit or a slaughter inspection, one cannot ensure that the animals are free of any disease. Nevertheless a trained observer should be able to detect the most prevalent diseases: digestive disorders, respiratory disorders, sickness behaviour. Some symptoms of respiratory or digestive disorders may not be totally specific to these disorders but can also appear in case of systemic disorders (Rebhum 1995, Straw *et al* 2006); but anyway they are to linked to the health status. Sickness behaviour combines inactivity, decreased food consumption, scruffy/oily looking of hair or feathers, all these signs being not specific to disease when taken individually but their combination appears selective to diseases (Hart 2010). Farm records may also provide information on some disorders, e.g. reproduction problems (dsytocia, metritis…), somatic cell counts in milk on dairy farms (as an early sign of masititis), mortality, decreases in performances. These disorders are not stable in time, since their incidence depends on occasional factors and these disorders are likely to disappear thanks to appropriate care and treatment from the farmer. To obtain a good picture of the health status of a farm, one should have access to data collected throughout the year, in farm records or vet records.
Biosecurity measures to protect animals from external contamination are crucial. A health plan may include prophylactic measures (e.g. vaccination, separation of animals from different species or age) and a proper management of diseases are of prime importance to improve the health of animals (Barceló & Marco 1998, Main *et al* 2001). Diseases depend on many factors including the animal “terrain”, the sanitary level of the farm and occasional factors (e.g. spreading of microbes from one farm to another, poor weather, nutritional changes such as weaning…). As a consequence, one cannot predict precisely the actual level of diseases from the various measures taken to prevent diseases or cure them.

5.8. Absence of pain induced by management procedures

For the criterion Absence of pain induced by management procedures, the following measures were considered (Table 8):

- Animal based measures: on the farm: presence of mutilations (castration, dehorning, tail docking, beak-trimming); at slaughter: signs of unconsciousness
- Resource based measures: on the farm: method for mutilation (including pain management); at slaughter: stunning resources and management

On the farm: It is possible to observe animals and detect if they have been castrated, dehorned, tail docked or beak-trimmed. This measure is selective although in rare cases, cryptorchid animals might be confused with castrated animals, tail biting with tail docking and polled animals with disbudded ones. It is not possible to observe the pain induced by these procedures on animals for feasibility reasons (the observer has very little chance to be present when the procedure is

applied). Information gathered from the farmer on how the procedure was applied is essential, especially concerning the use of pain releasers (anaesthesia, analgesia).

At slaughter, it must be ensured that stunning is efficient, i.e. induces unconsciousness in order to avoid pain during sticking and later. The absence of corneal reflex, the breathing rhythm, the loss of a righting reflex and the lack of vocalisations are signs of unconsciousness. When used in combination with each other, those signs are very selective (Holst 2001). Several aspects of the resources can affect stunning efficiency: in case of electric stunning, the intensity, the voltage and frequency of the current applied; in case of gas stunning, the gas concentration and the time of exposure; in case of mechanical stunning, the force of the cartridge and whether it is penetrative or not, and the management of the procedure (e.g. the place in applying the current or the captive bolt). If the stunning resources and management is correct, most of the animals should be unconscious (so free from pain) at the moment of sticking, however this is not always the case (Raj *et al* 1997).

5.9. Expression of social behaviours

For the criterion Expression of social behaviours, the following measures were considered (Table 9):

- Animal based measures: agonistic encounters, cohesive encounters and ratio between them, wounds/lesions
- Resource based measures: stocking density, ratio of animals to resources (e.g. feeding places, drinkers), group size, structuring of housing environment (visual barriers; presence of dead ends, narrow alleys, etc.), social management (regrouping, introduction of new animals), feeding management (availability of (attractive) feed, e.g. ad libitum or restrictive), lairage management (mixing of groups, time of resting)

The criterion Expression of social behaviour covers two aspects: whether the animals can express social behaviour; whether the interactions between animals reflect social cohesion rather than social stress.

These aspects can be assessed by direct observation of the behaviour of animals in a group: it is possible to observe a group of animals and count how many positive interactions (e.g. social licking) and negative interactions (e.g. butts, chase, and fights) they exchange. The higher the ratio between negative interactions and the total of interactions the higher the social stress in the group (Hasegawa *et al* 1997, De Groot *et al* 2001, Mounier *et al* 2005). Social interactions can be defined precisely so that the repeatability between trained observers is ensured. The frequency of social interactions can vary with time of the day, the presence of female in oestrus, recent regrouping of animals (Raussi *et al* 2005)… Recording social interactions in a group requires to observe animals during a significant duration of time (e.g. a minimum of 2 h was found in dairy and fattening cattle (Laister *et al* 2009)) and this may limit the feasibility for a short farm visit.

Aggressive interactions can lead to wounds on the body. In cattle, this is observed only between horned animals. In pigs such wounds are frequent and visible on the farm (e.g. vulva lesions) or at slaughter (e.g. bruises). In hens, aggressivity can be expressed by pecking the comb of another hen, resulting in a wound. Like other injuries, the repeatability of such measures can be very good. Aggressive social interactions are more frequent in case of a high stocking density and when access to resources is limited (too few drinking of feeding places according to the size of the group, too little attractive food) (Geverink *et al* 1996, Huzzey *et al* 2006). However the social behaviour depends largely on the composition of the group of animals, especially on how long they have been together (Bouissou & Andrieu 1977, Mounier *et al* 2006, Raussi *et al* 2008), as well as on group size (Estevez *et al* 2007), structuring of housing environment (Huzzey *et al* 2006), feeding management and at the slaughterhouse on lairage management. Therefore, it is difficult to predict the actual level of social behaviour expression in a group of animals from observation of the resources.

5.10. Expression of other behaviours

For the criterion Expression of others behaviours, the following measures were considered (Table 10):

- Animal based measures: play behaviour, exploratory behaviour, abnormal behaviour

- Resource based measures: availability of different resources (e.g. access to pasture/outdoor loafing area, rooting material, brushes, structuring of housing environment)

This criterion shall take into account the expression of behaviours the animals are motivated to express (e.g. grazing, play especially in young animals, exploration). When a normal behaviour cannot be expressed and/or a specific motivation is not fulfilled, then abnormal behaviour can occur. For instance sows subjected to low food allowance combined with few possibilities to express behaviour due to tethering perform oral stereotypies (Terlouw & Lawrence 1992) and calves denied of the possibility to suckle to obtain their milk express cross sucking (de Passillé 2001).

To detect whether the criterion is fulfilled or not, it is thus possible to observe animals and note these various behavioural patterns:

- play behaviour (locomotor play in calves for instance),
- exploration: exploration of objects (e.g. sniffing, rooting in litter in pigs and scratching the litter or soil in poultry), exploration of an area (e.g. use of outdoor range in poultry)
- abnormal behaviour: biting at the environment (bars, trough, drinker) , tongue rolling, sham chewing (sows and calves), teeth grinding (sows).

Some of these activities may be sporadic and difficult to accurately assess during a short visit (play behaviour, abnormal behaviours when they are not expressed over prolonged periods of time). Exploration occurs at a higher frequency and can more easily be detected. The time necessary to assess such activities varies between 20 min (poultry, exploration) to 90 min (calves, abnormal and play behaviour) depending on their frequency. The expression of these activities can vary during the day so it is necessary to define the time when they are observed in order to be able to compare farms. Given that the behavioural patterns are described precisely and the observers are trained, the inter-observer repeatability can be good to very good. The results are rather stable over time (Lensink *et al* 2003, Laister *et al* 2009) however some exceptions to this have been noted (Bond *et al* 2009).

The expression of many of these activities depends on the resources available: access to enrichment material (straw, objects, dust for hens, rooting material for pigs), access to nest boxes for hens, and access to pasture or an outdoor range. These resources should be in sufficient quantity according to the group size. Nevertheless, it is difficult to estimate the actual performance of behaviour from the resources provided because the expression of behaviour depends on many interacting factors. For instance, exploration in pigs increases with space allowance if bedding is present (Averós *et al* 2010). Also the quality of enrichment materials depends on many aspects (Bracke 2008).

Adult cattle do not often express play behaviour and abnormal behaviour. However they are motivated to access pasture even when provided feed and lying space in the barn (Legrand *et al* 2009). The actual time grazing could be measured by observing cattle but a more straightforward measure would be to see if they have access to pasture or not, assuming that they graze when they are on pasture.

5.11. Good human-animal relationship

For the criterion Good human-animal relationship, the following measures were considered (Table 11):

- Animal based measures: avoidance distance, vocalizations, coercion
- Resource based measures: intensity of human-animal contact (number of controls and time spent with animals), frequency of change of personnel, number of different stockpersons/animal, training level of stock people

A good human-animal relationship results in animals ready to approach people rather than avoiding them. This can be measured on farms by assessing the avoidance distance of animals, e.g. at what distance we can approach them before they step back (Lensink *et al* 2003, Waiblinger *et al* 2006). In pigs, fear of humans can also be expressed by panic responses when the observer enters the pen. Other factors may affect the way animals react to humans, e.g. a lame animal can be approached at a shorter distance (Špinka *et al* 2005). Nevertheless, the way the measure is done can be standardised in order to limit the effects of the environment or of time of the day. For instance, animals can be approached at feeding; this may help standardise the time of the day when the measure is taken, the motivation of the animal to feed, and there is generally sufficient space for the animal to move back from the trough. The inter-observer repeatability of these

measures is good and they are easy to perform on a farm. There is evidence that the results can be rather stable over time (Lensink *et al* 2003) however, this is not always the case (Bond *et al* 2009, Schulze Westerath *et al* 2009).

The behaviour of the animals towards humans depends largely on the behaviour of humans towards them: frequency of interactions, quality of interaction (gentle vs. rough), number of people working with animals… This has been highly documented in the different farm animal species (Hemsworth & Coleman 1998, Boivin *et al* 2003)(Waiblinger *et al* 2006). It is possible to ask questions to farmers on their attitude towards animals and to observe them during their routine work with animals (e.g. during feeding (Lensink *et al* 2000)). This may not be less time consuming than observing the animals avoidance distances. In addition, the quality of the human-animal-relationship depends on many interacting factors, such as the behaviour of humans but also the genetic background of animals or the group size (Raussi *et al* 2003), so that it is difficult to predict exactly the level of fear vs. attraction of animals to humans.

At slaughter, animals are subjected to extensive handling and this may cause stress. In pigs and cattle, this can be detected by recording the vocalisation, especially high pitched vocalisations in pigs. These vocalisations will reflect the level of stress at slaughter but may not be specific to reactions to handling; they may also result from aggressions between animals or the fear of the novelty of the situation. For this reason, in Welfare Quality® it was recommended considering Criterion 11 and 12 together at slaughter; this combination would reflect the level of fear at slaughter whatever the origin of the fear. To analyse whether this fear comes from poor handling of animals, one can record the level of coercion (use of electric goads, sticks, flappers, rattles…). Although it is difficult to compare the different means of coercion, this information may help identify causes of welfare problems at slaughter.

5.12. Positive emotional state

For the criterion Positive emotional state, the following measures were considered (Table 12):

- Animal based measures: qualitative behaviour assessment (QBA), novel object test, fearfulness
- Resource based measures: cognitive enrichment of the environment, management routines e.g. stroking)

This criterion shall reflect the general internal state of the animal, which cannot be measured by the observation of behaviour towards peers, humans or physical aspects of the environment (Criteria 9, 11, 10). Targeting the negative side of emotional states, several tests have been developed to assess the emotivity of an animal; these tests assess the fearfulness of animals. Reactions to novel objects can indicate whether animals are afraid of the object (do not approach it), the results being consistent with other fear tests such as tonic immobility or novel arena test (Forkman *et al* 2009). Another way to approach the whole emotional state of the animal is to observe their body language (Wemelsfelder & Lawrence 2001). This is called Qualitative Behaviour Assessment (QBA): the observers have to score a group of animals using about 20 descriptors of a supposed internal state of animals (happy, distress, apathetic, quiet…). Although this approach is based on a subjective assessment, the repeatability between observers is acceptable to high and the results correlate with physiological variables in relation to emotions (e.g. heart rate (Wemelsfelder 2009)). Therefore the method is very promising and may grasp aspects of animal welfare which cannot be measured through traditional observations of behaviour.
The general emotional state of animals on farms depends presumably on the many factors that have been listed for the various welfare criteria discussed here. It is likely that cognitive enrichment, e.g. signalling food by a sound, and providing pleasant situations to animals (stroking animals, providing brushes for animals to scratch themselves) have an impact on the emotional state of animals. However, the final state will result from the interaction of all these factors and the animal per se (its genetic background and previous experience). At present one cannot predict accurately the general emotional state of animals by taking into account the resources it is provided.
At slaughter, the general level of fear of animals is confounded with the fear due to handling. See Criterion 11 for discussion.

6. Recommendations for using animal based or resource based measures for each welfare criterion

6.1. Absence of prolonged hunger

Animal based measures seem more appropriate than resource based measures to detect absence of prolonged hunger. The most suitable measure seems body condition score as it is selective, fit for the purpose and feasible. It can allow detecting both moderate and severe problems. The method must nevertheless be refined for young animals and birds.

6.2. Absence of prolonged thirst

Tests on animals such as the skin pinch test or the observation of sunken eyes are not selective or sensitive enough to detect absence of thirst. Number of water points, their water flow and cleanliness are more valid. These latter measures can tell whether animals have access to fresh water. In case this access is limited (not enough drinkers, water flow insufficient or dirty water), it will be important to check whether animals suffer from this situation and which of them suffer. This may be done by measuring the osmolarity and hematocrite of some animals that seem in poor conditions, if not all animals.

6.3. Comfort around resting

The comfort around resting can be evaluated by different measures on animals:

- bursitis in pigs and possibly hair losses in cattle (due to frictions on a rough floor); cleanliness of flanks and upper legs are adequate signs of the comfort linked to the lying surface
- it seems possible to detect the discomfort of a lying surface or the improper design of a laying area by looking at the proportion of animals lying outside or partly outside the lying area
- the duration of lying down and difficulties in lying down (hits on obstacles) can be used to detect problems in the design of the lying area if it is limited by bars or walls.

All these measures allow detecting clear problems in relation to resting. However, they do not measure the comfort provided by the lying area (i.e. the positive aspects).
Measures on resources (roughness and cleanliness of the floor, design of the lying area) need more investigation either to standardise the way the measurements are performed or to better understand the interactions between factors that lead to problems on animals.
We suggest using animal based measures to detect problems due to the poor quality of the resting area. In case no problems are detected, some indicators of the comfort of the area may be used. For instance, it is known that some beddings are preferred by animals (Manninen *et al* 2002) and these may be scored as positive.

6.4. Thermal comfort

Animal based measures are suitable to detect responses to heat (panting, body soiled) or cold (huddling) in pigs. More information is needed to apply such measures on poultry. In cattle, animal based measures (increased respiratory rate and decreased feeding behaviour) are either not sensitive enough or not feasible on farms to detect heat and responses to cold are not visible. At the same time, resources based measures (temperature, humidity, air speed) are difficult to apply by a non expert person. Therefore, in cattle, for the moment, more work is needed to define appropriate measures to detect cold or heat.

6.5. Ease of movement

Slipping and falling are good animal based measures for the criterion Ease of movement when animals are handled, especially on a ramp when unloaded at their arrival at a slaughterhouse. On a farm it is difficult to detect problems in moving because the time spent moving varies largely during the day (due to several motivations interfering); in addition the quality of walking is also difficult to assess because animals may adapt their walking to the slipperiness of the floor of their barn. Whereas at slaughter, the quality of the ramp does not seem a more precise or feasible

measure than direct measures on animals, on the farm resource based measures seems suitable: not only the space allowance should be considered but also the quality of the space (a too structured space limits movements) and the floor.

6.6. Absence of injuries

Direct observation of injuries (skin lesions, lameness) on animals is more valid than any other measure on the environment since they measure the exact problems (so very selective and shall provide true results).

6.7. Absence of disease

We recommend using measures on animals to check the absence of diseases on farm, or rather the level of diseases on a farm. Symptoms of the most prevalent diseases can be detected on living animals or lesions can be detected at slaughter. These disorders vary with time. Therefore, to provide a good evaluation of the health status on a farm, farm records should be kept with various disorders and treatments used. At the moment the quality of such records vary greatly between farms.
Preventive and curative measures provide information on how health is managed on the farm. However, due to interactions between factors it is difficult to predict the exact health status of a farm from these indicators.

6.8. Absence of pain induced by management procedures

On the farm: We recommend that the animals are observed to detect whether mutilations have been performed (castration, dehorning, tail-docking, beak-trimming). If so we then recommend asking the farmer how mutilations were done, especially if pain release was used. If pain releasers were used, it is not possible to ensure that pain release was applied on all animals and it is assumed that the procedure used was the same for all animals. Note that even with anaesthesia and analgesia, it is not possible to ensure that no pain at all was experienced by the mutilated animals.

At slaughter, efficiency of stunning: the absence of consciousness after stunning can be detected on animals with good accuracy. Appropriate resources for the stunning of animals have a high predictive value that stunning will be efficient however difference e.g. due to management (electrodes or bolt not at the optimal place e.g. due to movements of animals) may exist and this does not ensure that all animals are appropriately stunned. We thus recommend checking the efficiency of stunning on animals.

6.9. Expression of social behaviours

Resource based measures in relation to social behaviour (e.g. stocking density, access to resources, structure of the pens, regroupings of animals) are easily feasible on farm. However, at present it does not seem possible to determine the level of social interactions accurately from such measures. More investigation in the modelling of social behaviour is needed to achieve such an aim. Animal based measures (social behaviour, wounds on the body) are essential to obtain information on the true level of social interactions in a group of animals. However, they require close observations of animals and can be time consuming. Social interactions need also to be precisely defined to ensure trueness and repeatability of the measures. It is recommended to perform the observation at a definite time of the day in order to avoid variations within the day (e.g. after the morning milking and feeding). Because occasional events can affect the expression of social behaviour (females in oestrus, recent regrouping), we recommend asking the farmer about such events. If they occur, the observations may be postponed to a further date.

6.10. Expression of other behaviours

The expression of behaviour (other than social behaviour) can be assessed by direct observation of animals: play, exploration, abnormal behaviours. Some of these activities occur at a low frequency (play, some abnormal behaviour) and may require long observations, not compatible with a short farm visit. Exploration and marked stereotypic behaviour can still be assessed on

farms and are likely to provide a good estimate of how much behavioural motivations are fulfilled on a farm. In adult cattle, positive and abnormal behaviour is rarely expressed although one cannot exclude that there exist motivations for performing certain activities. Regarding grazing behaviour we recommend checking access to pasture on farms (assessment of the resource in that case rather than observation of animals).

6.11. Good human-animal relationship

It is rather easy to collect information on resources that can affect the relation between animals and humans: behaviour of the farmers, use of coercion means at slaughter. However, collecting this information may not be less time consuming than observing directly animals (avoidance distance, vocalisations at slaughter). In addition, it is difficult, if possible at all, to determine the true quality of the human-animal relationship from resource based measures. The avoidance distance test on the farm and the behaviour of animals at slaughter (e.g. vocalisations in pigs) can be used to assess the reactions of animals to humans or handling. As for other behavioural aspects, the measures must be standardised to ensure repeatability. At slaughter, it is difficult to distinguish reactions to handling from reactions to other stressors; to do so it is necessary to assess the level of animal coercion when the animals are led to the slaughter station.

6.12. Positive emotional state

To date, there is not enough knowledge of the factors influencing the emotional state of an animal to design resource based measures that have a good predictive value. Tests can be run on animals, such as novel object test, to assess their level of fearfulness. At present, Qualitative Behaviour Assessment is a very promising method to detect aspects which cannot be measured by the traditional observation of behaviour towards peers, humans or physical aspects of the environment.

7. Conclusions

There is no general rule whether animal based measures are more or less valid than resource based ones. This depends on the welfare criterion considered and to a lesser extent on the animal type. We recommend using a mixture of animal based and resource based measures to assess the overall welfare of animals on farms. These measures should be chosen according to their validity (selectivity), feasibility, and sensitivity.
In addition, the detection of animals in a very poor state cannot be done by resource based measures. When such measures are applied and reveal a potential welfare problem then animals should be scrutinized in order to detect the ones which are in a very poor state. Some remedial measures may be taken on these animals immediately before a plan can be applied on the farm to improve the situation for all animals.

8. Additional comments on the whole process

As mentioned in the introduction of the report, one objective of Encaw is to imitate a procedure of a European Network of reference Centres for Animal Welfare (ENCAW). This activity shall help refine the process by which issues are addressed so that Scaw can propose procedures to be applied in a further ENCAW.

8.1. Choice of experts

The members of the working group were chosen for their expertise in assessing animal welfare. They had been involved in the Welfare Quality® project. They knew each other before the present project and this help developing the work in a short time.
Recommendation for ENCAW: The coordinator need to be chosen on the basis of his/her scientific expertise as well as potential to use scientific knowledge for an expertise task and to deliver. A database of potential experts that are ready to participate in an expertise should be made, including information on the specific expertise of experts (according to e.g. their previous work).

This would help ENCAW to choose a coordinator and also help the coordinator to choose members of the working group.

8.2. Development of the work

As soon as the working group was set up, a telephone meeting was organized between members (15 October 2010). The approach to develop the work was decided and the various tasks were allocated to partners:

- Definition of properties of measures: Isabelle Veissier
- Comparison of animal vs. resource based measures, Welfare Quality® Criteria 1-4: Luc Mounier
- Comparison of animal vs. resource based measures, Welfare Quality® Criteria 5-8: Antoni Dalmau and Antonio Velarde
- Comparison of animal vs. resource based measures, Welfare Quality® Criteria 9-12: Ute Knierim and Christoph Winckler

The group members planned a face to face meeting on January 19 2011 (Paris) to exchange their results on the advantages/disadvantages of animal based measures and resource based measures in terms of the quality for assessing welfare on farm and in terms of stability over time of the assessment. The question "If assessment is to be based on animal based measures should/could one fully rely on those or should resource based measures still be included?" was also addressed during the meeting and further discussed after the meeting (by e-mail).
Recommendation for ENCAW: face-to-face meetings between members of a working group are essential to exchange views and make significant progresses towards the conclusions of the work. By contrast experts are often senior scientists and it may be difficult for them to squeeze several meetings in a short period of time. Alternatively visio-conferences could be organised. Nevertheless we highly recommend that at least a face-to-face meeting is held when most information has been gathered and the group is in the phase of drawing conclusions and making recommendations.

8.3. Timing and workload

8.3.1 Timing

September 1010	Constitution of the expert group
October 2010	Telephone meeting, organisation of the work
Early November 2010	Properties of measures defined
November – February 2011	Collection of information on measures and assessment of their quality
March 2011	Synthesis of the information, conclusions, writing of the report

8.3.2 Workload

Meetings (telephone or face to face)	2 days x 5 people
Collection of information	
Properties of measures	3 days
Quality of resources vs. animal based measures	
Criterion 1-4	20 days
Criterion 5-8	20 days
Criterion 8-12	20 days
Synthesis and reporting	17 days
TOTAL man x days	90 days

8.4. Exchanges with Scaw

As soon as Scaw asked the coordinator to work on the project Encaw – task 1, she (the coordinator) discussed with Scaw to refine the task and give her views on how to develop the task (choice of experts, general approach). A workplan was attached to the contract signed between Scaw and the coordinator. The coordinator then informed Scaw after each meeting on the progress

of the work, sending the conclusions of these meetings (telephone or face to face meeting). On February, a visio conference was held between the coordinator and Scaw. The progresses made so far by the working group were discussed. Scaw asked to add two aspects in the final report:
- a discussion on the detection of the welfare level of all animals on a farm (and not only the average state of the herd), with a focus on the animals in a very poor state
- a discussion on how the work was developed in order to advise Scaw for further projects.
We suppose that in further projects, the objectives of the work will be precisely set from the beginning of the project. Nevertheless we must admit that these two additional points did not disturb very much our workplan.

8.5. Dissemination

It is important that the reports are largely disseminated. The present report can be disseminated by Scaw, e.g. in a specific book or on the Scaw website. This may not be sufficient to ensure a broad use of the results. In addition, experts have a direct interest in the dissemination of the work. It is rewarding for them to be part of an expertise. However, to be recognized by their scientific community and their hierarchy, they need to publish their work in scientific (peer-reviewed) journals. Such publications are also a guarantee that the work carried out is scientifically valid.
Recommendation: to publish the reports and allow the working group to make scientific papers out of their contribution in ENCAW task, with due acknowledgement of ENCAW.

8.6. Incentives for scientists working on consultancy basis

Incentives for scientists may be various and complementary:

- the consultancy work has to be done on top of the many other duties which form the basic missions of scientists (research i.e. producing original results, teaching, management). Financial incentive is welcome to encourage scientists, and their hierarchy, to take part in consultancy task
- Networking is also a strong incentive for scientists. This includes enlarging or consolidating their network with other scientists and enlarging their network with policy makers. This can help building transnational projects.
- As mentioned above publishing is essential for scientists. To be recognised by their peers, their hierarchy and the evaluation structures, scientists need to publish in peer reviewed journals. So it is essential that the work they do for consultancy can be used for publication.

8.7. Best conditions for the organization to enable impact assessments and research based studies

Scientists need to use the most up to date information. Most of them have access to bibliographic databases but this may not always be the case. An access to the **Scopus library** for the time of the task is likely to help scientists find the appropriate literature. This is done by scientific journals when they ask scientists to review a paper.
A clear and realistic workplan needs to be decided by the expert panel and Scaw from the start of the work. Scaw must be informed of any important deviation from the workplan as soon as possible.
During the first projects, it is important to analyze the experience gained from these projects – as Scaw asked us to do – so that the process can be further refined. Depending on the topics addressed in further projects, a general layout of project development (methods, milestones…) may be set up, as for risk assessment in Efsa. A template for presenting the report may reduce Scaw workload when the reports are put on a website or various reports are collated.

9. Annex

Tables describing each measure considered and its properties
The information contained in tables comes from Welfare quality® (Welfare quality® reports 9, 10, 11) or else in articles which references are explicitly cited.

10. References

Averós X Brossard L Dourmad J-Y de Greef KH Edge HL Edwards SA and Meunier-Salaün M-C 2010 A meta-analysis of the combined effect of housing and environmental enrichment characteristics on the behaviour and performance of pigs. *Applied Animal Behaviour Science* In Press, Corrected Proof: accepted

Barceló J and Marco E 1998 On farm biosecurity. *15 th meeting IPVS* 129-133

Bartussek H 2001 An historical account of the development of the animal needs index ANI-35L as part of the attempt to promote and regulate farm animal welfare in Austria: an example of the interaction between animal welfare science and society. *Acta Agriculturae Scandinavica, Section A, Animal Science* Supplementum 30: 34-41

Boivin X Lensink J Tallet C and Veissier I 2003 Stockmanship and Farm Animal Welfare. *Animal Welfare* 12: 479-492

Bond A Butterworth A Mullan S Temple D Dalmau Bueno A Velarde A and Gunnarsson S 2009 Final report on findings of the 'pilot study' and evaluation of the Welfare Quality assessment tool (pigs). *Welfare Quality® deliverable D4.28*

Botreau R Capdeville J Perny P and Veissier I 2008 Multicriteria evaluation of animal welfare at farm level; an application to MCDA methodologies. *Fondations of Computing and Decision Sciences* 33: 287-316

Botreau R Veissier I Butterworth A Bracke MBM and Keeling LJ 2007 Definition of criteria for overall assessment of animal welfare. *Animal Welfare* 16: 225-228

Bouissou MF and Andrieu S 1977 Etablissement des relations de dominance-soumission chez les bovins domestiques. IV. Etablissement des relations chez les jeunes. *Biologie du Comportement (Biology of Behaviour)* 2: 97-107

Bracke MBM 2008 RICHPIG: a semantic model to assess enrichment materials for pigs. *Animal Welfare* 17: 289-304

Bracke MBM Metz JHM and Spruijt BM 1999 Overall animal welfare reviewed. Part3: Welfare assessment based on needs and supported by expert opinion. *Netherlands Journal of Agricultural Science* 47: 307-322

D'Eath RB Tolkamp BJ Kyriazakis I and Lawrence AB 2009 Freedom from hunger' and preventing obesity: the animal welfare implications of reducing food quantity or quality. *Animal Behaviour* 77: 275-288

Dawkins MS 2006 A user's guide to animal welfare science. *Trends in Ecology and Evolution* 21: 77-82

De Groot J Ruis MAW Scholten JW Koolhaas JM and Boersma WJA 2001 Long-term effects of social stress on antiviral immunity in pigs. *Physiology & Behavior* 73: 145-158

De Leeuw JA and Ekkel ED 2004 Effects of feeding level and the presence of a foraging substrate on the behaviour and stress physiological response of individually housed gilts. *Applied Animal Behaviour Science* 86: 15-25

de Passillé AM 2001 Sucking motivation and related problems in calves. *Applied Animal Behaviour Science* 72: 175-187

Estevez I Andersen IL and Naevdal E 2007 Group size, density and social dynamics in farm animals. *Applied Animal Behaviour Science* 103: 185-204

Farm Animal Welfare Council 1992 FAWC updates the five freedoms. *Veterinary Record (The)* 17: 357
Feinberg M 1996 La validation des méthodes d'analyse. Une approche chimiométrique de l'assurance qualité au laboratoire. *Masson* 108-124
Fleiss JL Levin B and Paik MC 2003 *Statistical Methods for Rates and Proportions*. John Wiles & Sons: Hoboken, NJ, USA
Forkman B Heiskanen T Graml C and Waiblinger S 2009 Assessment of general fearfulness. In: Forkman B, Keeling, L. *Assessment of Animal Welfare Measures for Layers and Broilers*
Fraser D 1995 Science, values and animal welfare: exploring the 'inextricable connection'. *Animal Welfare* 4: 103-117
Geverink NA Engel B Lambooij E and Wiegant VM 1996 Observations on behaviour and skin damage of slaughter pigs and treatment during lairage. *Applied Animal Behaviour Science* 5: 1-13
Haley DB Rushen J and De Passille AM 2000 Behavioural Indicators of Cow Comfort: Activity and Resting Behaviour of Dairy Cows in Two Types of Housing. *Canadian Journal of Animal Science* 80: 257-263
Hart BL 2010 Beyond Fever: Comparative Perspectives on Sickness Behavior. In: *Encyclopedia of Animal Behavior* Elsevier Ldt:
Hasegawa N Nishiwaki A Suguwara K and Ito I 1997 The effects of social exchange between two groups of lactating primiparous heifers on milk production, dominance order, behavior and adrenocortical response. *Applied Animal Behaviour Science* 51: 15-27
Hemsworth PH and Coleman GJ (eds.) 1998 *Human-Livestock Interactions: The Stockperson and the Productivity and Welfare of Intensively Farmed Animals*. Cab International: Oxon/New-York

Holst S 2001 CO2 stunning of pigs for slaughter, practical guidelines for good animal welfare. *47th International Congress of Meat Science and Technology*
Hughes BO 1976 Behaviour as an index of welfare. *Proceedings of the 5th European Poultry Conference, Malta, 5-11 Sep. 1976* 1005-1018
Huynh TTT Aarnink AA Gerrits WJJ Heetkamp MJH Canh TT Spoolder HAM Kemp B and Verstegen MWA 2005 Thermal behaviour of growing pigs in response to high temperature and humidity. *Applied Animal Behaviour Science* 91: 1-16
Huzzey JM DeVries TJ Valois P and von Keyserlingk MAG 2006 Stocking density and feed barrier design affect the feeding and social behavior of dairy cattle. *Journal of Dairy Science* 89: 126-133
Jensen MB Studnitz M and Pedersen LJ 2010 The effect of type of rooting material and space allowance on exploration and abnormal behaviour in growing pigs. *Applied Animal Behaviour Science* 123: 87-92
Knierim U and Winckler C 2009 On-farm welfare assessment in cattle: validity, reliability and feasibility issues and future perspectives with special regard to the Welfare Quality® approach. *Animal Welfare* 18: 451-458
Knowles TG Warriss PD Brown SN Edwards JE and Mitchell MA 1995 Response of broilers to deprivation of food and water for 24 hours. *British Veterinary Journal* 151: 197-202
Laister S Hesse N, Minero, M. Lolli S Zucca D Knierim U Canali E and Winckler C 2009 Reliability of measures of socio-positive and play behaviour in dairy and beef cattle. Welfare Quality Reports No. 11, 175-188.
. In: Forkman B, Keeling, L. *Assessment of Animal Welfare Measures for Dairy Cattle, Beef Bulls and Veal Calves*
Legrand AL von Keyserlingk MAG and Weary DM 2009 Preference and usage of pasture versus free-stall housing by lactating dairy cattle. *Journal of Dairy Science* 92: 3651-3658
Lensink BJ van Reenen CG Engel B Rodenburg TB and Veissier I 2003 Repeatability and reliability of an approach test to determine calves' responsiveness to humans - a brief report. *Appl. Anim. Behav. Sci.* 83: 325-330
Lensink J Boissy A and Veissier I 2000 The relationship between farmers' attitude and behaviour towards calves, and productivity of veal units. *Annales de Zootechnie : Animal Research* 49: 313-327
Lindstrom T and Redbo I 2000 Effect of feeding duration and rumen fill on behaviour in dairy cows. *Applied Animal Behaviour Science* 70: 83-97

Main DC Webster AJF and Green LE 2001 Animal welfare assessment in farm assurance schemes. *Acta Agric. Scand. Sect A. Animal Sci* Suppl. 30: 108-113
Manninen E de Passille AM Rushen J Norring M and Saloniemi H 2002 Preferences of dairy cows kept in unheated buildings for different kind of cubicle flooring. *Applied Animal Behaviour Science* 75: 281-292
Martin P and Bateson P 2007 *Measuring Behaviour*. Cambridge University Press: Cambridge, UK
Mounier L Veissier I and Boissy A 2005 Behavior, physiology, and performance of bulls mixed at the onset of finishing to form uniform body weight groups. *J Anim Sci* 83: 1696-704
Mounier L Veissier I Andanson S and Delval EaB, A. 2006 Mixing at the beginning of fattening moderates social buffering in beef bulls. *Applied Animal Behaviour Science* 96: 185-200
Mouttotou N Hatchell FM and Green LE 1998 Adventitious bursitis of the hock in finishing pigs: prevalence, distribution and association with floor type and foot lesions. *Veterinary Record* 142: 109-114
Phillips CJC and Morris ID 2002 The ability of cattle to distinguish between, and their preference for, floors with different levels of friction, and their avoidance of floors contaminated with excreta. *Animal Welfare* 11: 21-29
Pritchard JC Barr ARS and Whay HR 2006 Validity of a behavioural measure of heat stress and a skin tent test for dehydration in working horses and donkeys. *Equine Veterinary Journal* 38: 433-438
Raj ABM Johnson SP Wotton SB and McInstry JL 1997 Welfare implications of gas stunning pigs: 3. the time toloss of somatosensory evoked potential and spontaneous electrocorticogram of pigs during exposure to gases. *The Veterinary Journal* 153: 329-339
Raussi S Lensink BJ Boissy A Pyykkönen M and Veissier I 2003 The impact of social and human contacts on calves' behaviour and stress responses. *Animal Welfare* 12: 191-203
Raussi S Boissy A Delval E Pradel P Kaihilahti J and Veissier I 2005 Does repeated regrouping alter the social behaviour of heifers? *Applied Animal Behaviour Science* 93: 1-12
Raussi S Niskanen S Veissier I Siivonen J Hänninen L Hepola H and Jauhiainen L 2008 Are long lasting social relationships significant for calves? *42nd Congress of the International Society for Applied Ethology (ISAE)* 88
Rebhum 1995 Respiratory diseases. In: *Diseases of dairy cattle* Williams and Wilkins:
Savory CJ and Lariviere JM 2000 Effects of qualitative and quantitative food restriction treatments on feeding motivational state and general activity level of growing broiler breeders. *Applied Animal Behaviour Science* 69: 135-147
Schulze Westerath H Pinent T Knierim U Kirchner M Winckler C Tessitore E and Cozzi G 2009 Final report on findings of the “pilot study” and evaluation of the Welfare Quality assessment tool (beef bulls). *Internal report, EU-Project Welfare Quality* D4.28:
Scott EM Nolan AM and Fitzpatrick JL 2001 Conceptual and methodological issues related to welfare assessment: a framework for measurement. *Acta Agriculturae Scandinavica, Section A, Animal Science* Supplementum 30: 5-10
Špinka M Dembele I Panamá J and Stìhulová I 2005 Lame dairy cows have shorter avoidance distances. *39th International Congress of the International Society for Applied Ethology* 83
Straw BE Dewey CE and Wilson MR 2006 Differential diagnosis of diseases. In: *Diseases of swine.* Blackwell Publishing: Iowa
Taweel HZ Tas BM and et al 2006 A note on eating behaviour of dairy cows at different stocking systems: diurnal rhythm and effects of ambient temperature. *Applied Animal Behaviour Science* 98: 315-322
Terlouw EMC and Lawrence AB 1992 Long-term effects of food allowance and housing on development of stereotypies in pigs. *Applied Animal Behaviour Science* 37 pp.
Thompson M Ellison SLR and Wood R 2002 Harmonized guidelines for single-laboratory validation of methods of analysis - (IUPAC technical report). *Pure and Applied Chemistry* 74: 835-855
Tucker CB Weary DM and Fraser D 2004 Free-stall dimensions: Effects on preference and stall usage. *Journal of Dairy Science* 87: 1208-1216
Veissier I Capdeville J and Delval E 2004 Cubicle housing systems for cattle: Comfort of dairy cows depends on cubicle adjustment. *J. Anim. Sci.* 82: 3321-3337

Waiblinger S Boivin X Pedersen V Tosi M-V Janczak AM Visser EK and Jones RB 2006 Assessing the human-animal relationship in farmed species: a critical review. *Applied Animal Behaviour Science* 101: 185-242
Welfare Quality® 2009 *Welfare Quality® assessment protocol for cattle (fattening cattle, dairy cows, veal calves)*. Welfare Quality® Consortium: Lelystad, The Netherlands
Wemelsfelder F 1990 Boredom and laboratory animal welfare. In: Rollin BE and Kesel ML *The Experimental Animal in Biomedical Research, 1* CRC-Press: Boca Raton, U S A
Wemelsfelder F 2009 The human perception of animal body language: A window into an animal's world? *Knowing animals* 79
Wemelsfelder F and Lawrence AB 2001 Qualitative assessment of animal behaviour as an on-farm welfare-monitoring tool. *Acta Agriculturae Scandinavica, Section A, Animal Science* Supplementum 30: 21-25.
Winckler C and Willen S 2001 The reliability and repeatability of a lameness scoring system for use as an indicator of welfare in dairy cattle. *Acta Agri Scand, Sect. A, Anim Sci* 30: 103-107