Review of the assessment of animal welfare with special emphasis on the "Welfare Quality® animal welfare assessment protocol for growing pigs"

I. Czycholl[1], K. Büttner[1], E. grosse Beilage[2], and J. Krieter[1]

[1]Institute of Animal Breeding and Husbandry, Christian-Albrechts-Universität, Olshausenstr. 40, 24098 Kiel, Germany
[2]Field Station for Epidemiology, University of Veterinary Medicine Hannover, Foundation, Buescheler Str. 9, 49456 Bakum, Germany

Correspondence to: J. Krieter (jkrieter@tierzucht.uni-kiel.de)

Abstract. This paper discusses the arising need for an objective, but feasible, reliable and valid method for assessing animal welfare on farms. Animal welfare has become especially important since the industrialisation of animal housing after the Second World War and as public awareness and concern has increased. Simultaneously, alienation of the public from agriculture has taken place, as the population has moved increasingly from rural areas to towns. This has led to a very emotional discussion concerning the welfare of farmed animals, and thus a need for not only a clear definition but also a way of objectively measuring it has arisen. It is probably best defined as a total of the different conceptions health, natural behaviour and positive affective state. In the last few years, different methods for an objective assessment have been developed; however, all of them still face great challenges in their practical implementation and acceptance. The most promising method is probably the Welfare Quality® (WQ) approach, especially as it concentrates on animal-based parameters. The development of the WQ protocols emphasised not only the different conceptions of animal welfare but also especially the feasibility, reliability and validity of the parameters to be included. One of the main challenges of these protocols remains, however, the final aggregation of the results to a welfare score. Furthermore, a thorough cost–benefit analysis has not been carried out so far. Even more importantly, only a few studies have addressed the general reliability and validity of the complete protocols, and those studies that have addressed these issues have also revealed challenges concerning the interobserver and test–retest reliability of some of the included parameters. As an example, this is discussed in detail for the "Welfare Quality® animal welfare assessment protocol for growing pigs". In conclusion, the WQ approach can be seen as promising, but it has also revealed that there are still a considerable number of challenges that need to be addressed in further studies on the WQ protocols in order to achieve constant improvement. These challenges should be borne in mind in the application of these protocols, which should not be simply referred to as a gold standard.

1 Introduction

Animal welfare has become an important topic of public and political debate in the last few decades (Hobbs et al., 2002; Fraser, 2009). However, there is a lack of agreement concerning the question of what animal welfare should comprise (Broom, 1988). Therefore, the ongoing discussion is carried out in a very emotional and subjective way. This has led to the necessity of an objective and generally accepted method for assessing the welfare situation of animals (Webster, 2005). One of the most famous methods, having gained great importance in the field of animal welfare science in the last few years, is the Welfare Quality® (WQ) approach, which promises to be objective, feasible, valid and reliable (Veissier et al., 2013). The following review discusses in Sect. 2 the definition of the term "animal welfare", how it

has become of increasing importance in the last few decades and what the public's current concerns are. Throughout this discussion, the need for a method to objectively assess animal welfare becomes obvious. Suitable parameters and existing assessment systems are introduced in the Sect. 3 and their particular advantages and disadvantages are described. Section 4 describes in detail the development of the "Welfare Quality® animal welfare assessment protocols", their construction, how they function and also the current challenges this system still faces. One of the most important challenges is that only a few studies have been carried out so far on the feasibility, reliability and validity of the complete protocols in their on-farm use. As an example, an overview of the few studies emphasising the "Welfare Quality® animal welfare assessment protocol for growing pigs" and the existing challenges revealed in these studies is provided in the Sect. 5.

2   Animal welfare

2.1   Definition

Although most people have an opinion on animal welfare and how animals should be treated (Keeling et al., 2011), there is a lack of consensus on a clear definition of the term itself (Moberg, 2000). Furthermore, in the literature, the term well-being is often used; however, according to Fraser (1998), well-being and welfare can probably be used as synonyms. One has to bear in mind that the term animal welfare is not scientifically based but rather a term that arose from a debate in society (Keeling, 2005), as described in detail later on. Animal welfare can best be described by contemplating three different conceptions: (1) basic health and biologic functioning, which includes the physical well-being of the animal; (2) natural living, which concerns the possibility to express normal behaviour; and (3) the affective state, especially concerning positive emotions (Fraser, 2008). A veterinarian, on the one hand, may well focus on the health of the animals and thus claim that indoor housing systems provide better welfare as the animals can be kept free from parasitic infections more easily. An ethologist, on the other hand, may think of good welfare more in terms of the expression of normal behaviour, thus putting more emphasis on free range systems and considering parasitic infections as minor welfare problems. Despite this disagreement, all defenders of different conceptions may agree on certain factual issues such as the mortality rate (Fraser, 2008). The challenge of a comprising definition is, therefore, to find a basis of agreement on these three conceptions, which most scientific definitions try to take into account. For example, the Brambell report (1965) emphasised the psychological aspect of animal welfare, defined as "a wide term that embraces both the physical and mental well-being of the animal. Any attempt to evaluate welfare, therefore, must take into account the scientific evidence available concerning the feelings of animals". Hughes (1976) also defined animal welfare comprehensively as "a state of complete mental and physical health, where the animal is in harmony with its environment". It is eye-catching that this definition is not far from the definition of human health given by the World Health Organization (World Health Organization, 1946), in which health is described as a "state of complete physical, mental and social well-being and not merely the absence of disease or infirmity".

2.2   Importance

In the last five decades, a considerable increase in interest, a new critical attitude and concern towards animal welfare have developed on the side of society. Nevertheless, a lack of consensus remains due to different attitudes, background knowledge and personal interpretations.

On the market, there is increasing demand from the public for products that originate from animal-friendly housing (Roex and Miele, 2005) and it is especially the public, i.e. the customers – most of them without expert background knowledge of animals – who think the affective state to be the most important part of animal welfare in general (Vanhonacker et al., 2008).

Although the truth about animal welfare is probably a compromise based on all three conceptions, i.e. basic health and biologic functioning, natural living and positive affective state, there is still an emotional discussion about what animal welfare really is. This is especially due to the lack of consensus in different groups of society defending their conception about welfare (Lassen et al., 2006). Therefore, an increasing need has arisen to overcome the highly emotional and partly also anthropomorphic discussion by finding a way of objectively describing and assessing animal welfare (Webster, 2005). This becomes even more important when one looks at the rising demand for animal-welfare-friendly products on the market described above. There is, however, still a gap between the customers' rising interest and their willingness to pay more for animal-friendly products (Bennett et al., 2002). Trustworthy labelling would help customers to find the products they really want. This can only be realised by a reliable and valid animal welfare assessment, which somehow comprises all the different aspects and understandings of animal welfare (Blokhuis et al., 2013a).

3   Methods for the assessment of animal welfare

3.1   Parameters

All parameters used for such an assessment must always provide good accuracy, validity, reliability and feasibility (Velarde and Geers, 2007). Accuracy describes how close the measured value is to its true value. The validity of a parameter refers to its extent of measuring animal welfare and to what extent it actually measures what it is supposed to measure (Velarde and Geers, 2007). For instance, positive and negative social patterns of behaviour are relevant to welfare

as positive social behaviour is known to provide a rewarding function at least for the receiver and furthermore to reinforce and stabilise social relationships (Sato, 1984; Sato et al., 1991). Negative social behaviour, on the other hand, may cause stress and injuries in the receptor animal (Tuchscherer et al., 1998; Menke et al., 1999). It can, however, be critically questioned whether the assessment of this behaviour in a relatively short interval of time with an observer effect in the stable reflects the real behaviour of the animals, thus whether the methodology as described is adequate and valid. Reliability refers to the repeatability of a measure, i.e. the relative similarity of repeated measurements on the same object. It can be divided into the interobserver, intraobserver and test–retest reliability (de Passille and Rushen, 2005). The interobserver reliability describes how well different observers agree in their findings, assessing the same objects at the same time and under the same conditions. Intraobserver reliability measures the agreement that the same observer achieves when assessing exactly the same objects at the same time. Interobserver and intraobserver reliability can be influenced by the training of the observers (Wirtz and Caspar, 2002). Intraobserver reliability in terms of animal welfare can only be measured by the assessment of pictures or video sequences due to the fact that only pictures and videos assure the assessment of exactly the same images. On-farm assessments, on the other hand, are always studies assessing the test–retest reliability of the method, because it can never be assured that truly the same objects are assessed in the practical situation of farms, due to different individuals. Even if the same individuals are assessed, there are still changes in the individuals such as weight gain or pregnancy status. The test–retest reliability thus refers to the tested method and its capability to produce consistent results despite routine procedures and minor changes in the object that are not of interest in terms of the assessment (Temple et al., 2013). Finally, feasibility means that the monitoring method produces reliable results at an affordable cost, and thus it analyses the cost–benefit ratio (Velarde and Geers, 2007).

Parameters that have been discovered as valid, reliable and feasible by the various animal welfare studies in the last few decades and are now available for the assessment of animal welfare can be divided into resource-based, management-based and animal-based parameters (Blokhuis et al., 2010). Resource-based parameters are, for instance, the evaluation of the availability of space or the measuring of the width of slats in the case of slatted flooring. These parameters are relatively easily accessible and are usually of excellent feasibility and reliability (Winckler, 2006). But simply because the resources provide good welfare does not mean that the animal itself experiences welfare (Napolitano et al., 2009). For instance, a lame animal suffering from an infectious disease will not feel well even if it lives in the best possible environment. Management-based parameters are parameters concerning the management routines, e.g. castration or tail-docking procedures as well as vaccination schemes. They,

too, are relatively easily accessible, but mainly through interviewing farm managers (Welfare Quality®, 2009). Thus, one has to rely on them telling the truth about (the farm's) management procedures. Of course, castration and tail-docking procedures can be assessed directly on the animals; however, whether pain relief has been provided can scarcely be proven. According to Rousing et al. (2001), resource- and management-based parameters represent rather an assessment of the risk of the potential welfare problems than an actual assessment of the state of welfare. In contrast, animal-based parameters are parameters concerning the animal itself, such as the evaluation of lameness or the results of lung examinations at the slaughterhouse. They reflect the actual effect of the resource- and management-based parameters on the animal. This is especially important concerning the animal's ability to cope with its environment (Smulders and Algers, 2009). These are, however, probably also those parameters that are most time-consuming to evaluate (Bartussek, 1999) and, furthermore, reliability is often questionable (Veissier et al., 2013). Moreover, interpretation of animal-based parameters may remain ambiguous when interpreted solely (Bracke, 2007); for example, laying hens in caged systems may provide better health parameters but do not necessarily show natural behaviour (Keeling, 2005). It becomes clear that, in terms of a valid, reliable and feasible assessment, the focus should be given to animal-based parameters, however not solely referring to them but rather considering a combination of these three different parameters (Hewson, 2003).

3.2   Methods for the overall assessment of welfare

In the last few decades, different methods for the overall assessment of animal welfare have been proposed and published (Johnsen et al., 2001). In the following, examples of overall welfare assessment systems taking into account all the aspects of welfare (health and biologic functioning, natural behaviour, positive affective state) are presented.

Different checklists exist which are often a combination of resource-, management- and animal-based parameters. The focus is especially directed towards a feasible assessment, and the main aim is either labelling by associations or self-checks by the farmer to detect areas of concern. Examples are the checklist of labelling for the eco-label "Bioland" (Schumacher et al., 2007) or the informative leaflet for farmers published by the German Association for Agriculture (Deutsche Landwirtschaftsgesellschaft) (Pelzer and Kaufmann, 2012).

The concept of critical control points is derived from food hygiene, in which it is used in terms of process control to identify weak points (Mortimore and Wallace, 2013). It has been adapted to animal welfare science in various approaches during the last few years. The most famous approach of critical control points is probably the control system for slaughterhouses developed by Grandin (2006). The principle is

based on checklists consisting of questions on management, environment or the animal itself. These can be answered either simply with "yes" or "no" (Grandin, 2010) or else with a traffic-light system, with green meaning that everything is in order in terms of this control point, yellow indicating minor or initial problems that should be observed further and red referring to unacceptable conditions that need to be eliminated immediately (von Borell et al., 2001). All control points have to be fulfilled in terms of quality assurance schemes; there is no distinction made concerning their importance. A good critical control point should be specific and measure many aspects (Grandin, 2006).

The animal needs index (ANI) was invented by Bartussek (1985) and was further advanced and refined in the following years. Evaluation schemes for cattle, pigs and poultry (Bartussek, 1995a, b) are now widely available. Basically, five areas influencing the animals were identified: freedom of movement, social contacts, flooring, light and air conditions, and intensity of care. For all of these areas, parameters to be measured in the stable were identified, these being mostly resource- and management-based parameters. Minimum requirements in all the influencing areas have to be fulfilled. Thereafter, the interaction between incriminating and exonerating factors is considered by allowing higher values in one area to compensate for lower values in another. This is done in order to take into account the animal's interaction with its environment and its ability to cope with its condition of life (Haiger et al., 1988). The different groups of parameters are aggregated to a total sum, thereby taking into account multifactorial, interdisciplinary and all-embracing approaches. Despite its good feasibility and broad application, especially in Austria, the ANI has been criticised since evaluation remains subjective for some parameters due to it being based on the practical experience of the assessor (von Borell, 2001) and furthermore not considering animal-based parameters to a sufficient degree (Blokhuis et al., 2013b).

Decision support systems were first introduced into animal welfare science by Bracke (2001). This overall welfare assessment system was first tested on husbandry systems for pigs (Bracke et al., 2002) and adapted to poultry (de Mol et al., 2006), cattle (Ursinus et al., 2009) and recently to salmon (Stien et al., 2013). It was invented as a tool to actually make decisions concerning animal welfare (Bracke et al., 2002). In terms of decision support systems, certain needs of the animals are identified, i.e. needs that have to be fulfilled in order to provide good welfare. For instance, for sows, needs were identified for food and water, resting, social contact, kinesis, exploration, body care, territorialism, thermal comfort, good air condition, health, safety and nest-building behaviour, and maternal care. In a further step, parameters were found to which good and bad levels were assigned according to expert opinion. Weights were attributed to these parameters and, finally, with the help of a computer model, values for the tested housing condition in terms of welfare can now be calculated (Bracke et al., 2002). The main disadvantage of this welfare

assessment tool is that parameters are mostly resource-based (Blokhuis et al., 2013c). Thus, they describe rather the potential of a housing system to provide for a good welfare state. For instance, it could be reliably proven for dairy cows that loose housing systems provide more welfare than tie stalls (Bracke et al., 2002). But just because resources provide good welfare does not mean that the animal itself indeed experiences welfare as there are many other influencing factors. Also, it is probably not suitable for assessing welfare differences between two farms of the same housing system. This can only be analysed by the use of animal-based parameters (Blokhuis et al., 2013c).

The Bristol Welfare Assurance Programme started when the initiators of "Freedom Food", a label for organic farm products in Great Britain, realised around the year 2000 that their method of certification by checklists, which only contained resource-based parameters, was no longer up to date concerning the assessment of welfare. They therefore started a project in cooperation with the University of Bristol with the aim of developing a certification tool that emphasised animal-based parameters and thus truly reflected the welfare status of a farm. Firstly, animal-based parameters with the potential capability to evaluate "Freedom Food" farms in terms of their welfare status were defined by expert opinion (Main et al., 2003; Whay et al., 2007). Existing animal welfare assessment techniques were adapted to farming and certification in the ongoing research (Leeb et al., 2004). Special attention was assigned to the relevance of the chosen parameters and to the legislative background concerning the situation of farm animals in Great Britain (Whay et al., 2003). Nowadays, evaluation sheets for the on-farm welfare assessment of laying hens, cattle and pigs are available. All concepts of animal welfare are taken into consideration by behavioural observations, scans of a sample of animals especially concerning the health status, and qualitative evaluations, the latter being based on the overall impression of the observer. Results are expressed mainly as percentages of affected animals, and these percentages are compared to threshold values. If threshold values are exceeded, guidelines of intervention are accessible, which have, however, not yet been sufficiently validated (Leeb et al., 2014). The observer needs background knowledge of the husbandry of farm animals and their behaviour, as some parameters include subjective assessment. Intensive training of the observers and strict testing of their repeatability is urgently needed in order to provide comparable results (Leeb et al., 2014). After all the conditions and requirements have been enumerated, it becomes quite obvious that reliable results in terms of the Bristol Welfare Assurance Programme still remain a great challenge especially due to the problem of subjective evaluations. Application up to now has been limited to Great Britain, as it is specifically designed for British animal protection laws (Leeb et al., 2014).

WQ was an interdisciplinary and international research project dating from the years 2004 to 2009 with the aim of

developing reliable animal welfare assessment tools on farms and at slaughterhouses, starting out with tools for poultry, cattle and pigs. In the basic definition of welfare, opinions of the public, farmers and scientists were taken into account (Blokhuis, 2008). In order to consider all aspects of welfare, animal welfare was defined in terms of four principles: good feeding, good health, good housing and appropriate behaviour. Twelve criteria were then assigned to these principles, and for each of the criteria, parameters to be measured were found in the stable. Animal-based parameters were chosen whenever possible in terms of validity, reliability and feasibility. Behavioural observations and examinations of single animals were performed using a specified sample size. The outcomes were first presented as percentages of affected animals, which were then further aggregated to grades between 0 and 100 (with 0 describing the worst and 100 the best possible condition of that parameter) by different mathematical procedures at the level of criteria and principles. Training of the observers was mandatory, but no special background knowledge was needed (Welfare Quality®, 2009). The main problem of these welfare assessment systems, however, is that they take a relatively long time for the on-farm sampling, i.e. 4 to 8 h depending on the animal species, the size of the farm and the arrangements of the buildings. Furthermore, reliability is not assured for all the parameters in their on-farm use and for the protocols themselves. However, in terms of the WQ network, the protocols are still under the continuous process of revision and improvement as research and knowledge continue to grow (Winckler and Knierim, 2014).

4   Detailed description of the WQ project

4.1   Starting point of WQ

The above-mentioned pressing need to ensure the welfare of farm animals and, what is even more important, to measure it objectively using a scientific basis, led to the WQ project. This project answered the growing societal need of consumers and citizens for high welfare standards in terms of food quality and an increased transparency of the production chain. The first approaches were determined as a response to the European Commission's call for proposals aiming at "improving animal production methods that take into account consumer demands for high standards of animal welfare, health and food quality", according to Blokhuis et al. (2013a). The research proposal containing these approaches was successful, and thus the project "Integration of animal welfare in the food quality chain: from public concern to improved welfare and transparent quality" started in May 2004 and lasted until December 2009. During that time, it developed into the largest international network of scientists and stakeholders ever having worked together in an interdisciplinary manner. About 200 scientists from 43 different institutes in 13 European and 4 Latin American countries were involved. The scientists, among whom were mathematicians as well as animal and social scientists, were integrated with farmers, processors, slaughterhouse managers, retailers, animal protection organisations (non-government organisations, NGOs) and members of the public. The project was thus a good example of science and society working together to improve the welfare of farm animals. This great number of people involved required a strong management structure, which was implemented by a supervising steering committee. The main aim became the development and establishment of a generally accepted, valid, reliable and feasible system for assessing animal welfare on-farm and at slaughterhouses (Blokhuis et al., 2003).

4.2   Definition of welfare in terms of WQ

As concerns WQ aimed at a balanced welfare assessment to satisfy public, industry, political and scientific, a holistic definition for the term was also needed. Thus, the views of customers, industrialists, farmers, legislators and scientists were considered and a dialogue between science and society was initiated. From studies based on interviewing different groups of the society, it turned out that each group considered different aspects of animal welfare to be more or less important. But, in general, it was possible to obtain good agreement concerning the basic definition of animal welfare (Bock and van Leeuwen, 2005; Bock and Van Huik, 2007; Buller and Roe, 2008). The progress of finding an all-embracing definition took place in such a way that, at the beginning, animal scientists proposed a definition of animal welfare based on three conceptions – good health and biologic functioning, natural living and positive affective state (Fraser, 2008) – as well as on the five freedoms of the Farm Animal Welfare Council (FAWC, 1979). As a result, four main principles were identified: (1) good feeding, (2) good housing, (3) good health and (4) appropriate behaviour. At this stage, 10 appending criteria were defined that formed the underlying contents of these principles. Before these elements were finally included in an animal welfare assessment scheme, this first design of a definition was assigned to citizen focus groups. Focus groups were different groups of the public (e.g. urban mothers, seniors, young singles, vegetarians, hunters, gourmets) who were given background knowledge and then encouraged to critically discuss the proposed definition of animal welfare (Blokhuis et al., 2013c). Furthermore, surveys of randomly chosen people of the population of different nations based on computer-assisted telephone interviews were carried out in order to find out their opinions on animal welfare (Kjaernes and Lavik, 2008). It was concluded that, in general, scientists thought it basically important to avoid the negative impacts of welfare such as diseases, while society in general put more emphasis on the positive aspects such as ensuring positive emotional states, thus allowing the animals to feel comfortable and content. The influence of public opinion resulted in the inclusion of two further criteria. The criteria (1) absence of prolonged hunger and (2) ab-

Table 1. Principles, criteria and parameters of the "Welfare Quality® animal welfare assessment protocol for growing pigs".

Principle		Criteria	Parameters
Good feeding	1	Absence of prolonged hunger	Body condition score
	2	Absence of prolonged thirst	Number of drinking places Functioning of drinkers Cleanliness of drinkers
Good housing	3	Comfort around resting	Bursitis Manure on the body
	4	Thermal comfort	Huddling Shivering Panting
	5	Ease of movement	Space allowance
Good health	6	Absence of injuries	Lameness Wounds on the body Tail biting
	7	Absence of disease	Mortality Coughing Sneezing Pumping Twisted snouts Rectal prolapse Scouring Skin condition Hernias Pneumonia Pleurisy Pericarditis White spots
	8	Absence of pain induced by management procedures	Castration Tail docking
Appropriate behaviour	9	Expression of social behaviours	Social behaviour
	10	Expression of other behaviours	Exploratory behaviour
	11	Good human–animal relationship	Fear of humans
	12	Positive emotional state	Qualitative behaviour assessment

sence of prolonged thirst were assigned to the principle of good feeding. For the principle of good housing, the criteria (3) comfort around resting, (4) thermal comfort and (5) ease of movement were identified. The criteria (6) absence of injuries, (7) absence of disease and (8) absence of pain induced by management procedures were allocated to the principle of good health, and the criteria (9) expression of social behaviour, (10) expression of other behaviour, (11) good human–animal relationship and (12) positive emotional state to the principle of appropriate behaviour. An overview of the principles, criteria and parameters included subsequently is presented in Table 1.

4.3 Development of welfare measures and protocols

After the final definition, measures to assess these criteria had to be found. To guarantee the feasibility and the general acceptance, it was determined that the whole assessment must be finished within one day by only one observer. As already mentioned, one of the main aims was to place the focus on animal-based parameters to truly assess the welfare state of the animals and to make the protocol internationally applicable in all different kinds of housing systems. Moreover, the overall system, and thus every single measure, had to provide good validity, reliability and feasibility. Therefore, some resource- and management-based parameters were also included if no suitable animal-based parameter could be found. The selection of appropriate measures started out with a thor-

ough review of literature in search of suitable welfare parameters (Veissier et al., 2013). Measures were considered further if they had a good validity in terms of the criterion they were supposed to assess, which was defined and decided by experts (Scott et al., 2001). Furthermore, the measures had to provide applicability on-farm in different housing systems or at slaughterhouses. During the procedure of this review of the literature, it was discovered that few welfare parameters had actually been tested accurately for reliability (Engel et al., 2003; Knierim and Winckler, 2009). The finally extracted measures to be further considered were divided into three groups regarding their already defined validity and reliability: the first group included parameters that had not been previously validated further in terms of criterion validity, i.e. the relationship of a tested measure to an already approved measure, or concept validity, i.e. the experimental proof that the measure was related to what it was supposed to measure; in our case, this considers whether it is related to the welfare state. Parameters were sorted into the second group if they had already undergone validity testing. These parameters were then tested for their reliability, which was done by observing video clips and pictures (Courboulay et al., 2009; Forkman and Keeling, 2009; Leruste et al., 2009; Schulze Westerath et al., 2009; Plesch et al., 2010). Less effort was devoted to the field of clinical examinations, which were simply expected to be of good reliability. The third group contained parameters for which validated procedures were available and had previously been considered to be reliable. Studies on the choice and extraction of parameters can be found in the WQ reports (Forkman and Keeling, 2009; Forkman and Keeling, 2009a, b).

Measures that were now found to be sufficiently valid and reliable were evaluated in terms of the information they provided in relation to all other potential parameters. Therefore, analyses of correlation and association between different animal-based parameters or else between animal-based and resource- or management-based parameters were carried out. Furthermore, before the final exclusion of a parameter, a calibration of the simplified version of the monitoring system was made against the full version. In order to simplify the use and thus provide better reliability, all parameters with more scoring possibilities than three were minimised to the following three categories: "absence", "low affection" and "high affection". This was done, for instance, in terms of the categorisation system for the parameter bursitis in pigs, which was proven to be reliable by Lyons et al. (1995) using a five-point scale. These scientifically approved versions of protocol assessment were again given to the focus groups mentioned earlier as well as to farmers, and their opinions on certain parameters were considered before a decision was made on a final version (Veissier et al., 2013).

Using this procedure, protocols were developed which promised to provide a valid, reliable and feasible animal-based animal welfare assessment which was furthermore

generally accepted, as all groups of the society were involved in the process of decision making.

4.4 Integration of data to generate an overall assessment of welfare

A scoring model was designed in order to translate the judgements at parameter level into refined and easily understandable information about the overall welfare state (Botreau et al., 2008, 2009). This model needed to be sufficiently sensitive to identify and quantify variations and differences among farms. There were many challenges for the scientists to be solved due to the multidimensional nature of animal welfare. (1) Ethical dilemmas definitely exist, such as what has to be considered more important: the good health of an animal or its appropriate behaviour. In this case, certain questions arise such as whether it is better for animals to show frightened and fearful behaviour than to be sick (Fraser, 1995), or whether one must predominantly consider the small number of animals in an extremely poor condition or the majority of the animals, i.e. at the average. Consequently, one has to decide whether a farm that houses a considerable number of animals affected by a moderate lameness is labelled as better than a farm housing only a small number of animals exhibiting severe lameness. Moreover, there is the general question of whether one aspect of welfare can be compensated for by another one. (2) Furthermore, the final protocols incorporate numerous measurements very different from each other and data are expressed at all different scale levels. Hence, completely different things need to be aggregated. (3) And even if this challenge of aggregation has been successfully mastered, there is still the question of what is to be considered as good welfare (Veissier et al., 2011).

(1) In terms of the ethical decisions, a flexible model was designed in order to most accurately represent most possible ethical decisions. Different data sets were shown to a group of scientists, and they were asked to assign a value between 0 and 100 to these theoretical data, with 0 representing the worst possible level; 20 an acceptable level, indicating that legislation requirements are met; 50 describing a welfare state of neither good nor bad; and 100 representing the best possible level. These assigned values were later discussed in juries made up of citizens and farmers. It can be concluded that more emphasis was put on animals in a poor condition than on those in a good condition, but also that the whole group was more important than individual animals (Miele et al., 2011).

(2) The calculation of scores followed a bottom-up approach. Hence, based on the measurement results, first scores were calculated for the 12 criteria and then scores for the four principles. If the results were expressed at farm level and there were a limited number of categories, a decision tree (Magerman, 1995) was applied to calculate the scores. In the case of just one measure belonging to a criterion, in which, however, several degrees were possible, the results

were expressed as percentages of affected animals for each of the degrees. To calculate the score for the criteria, a certain weight was assigned to the different degrees with the weight increasing with the severity. If the measures assigned to a criterion resulted in data expressed on different scales, these were compared to alarm thresholds and, finally, the number of alarms was valued. Hereby, scores scientists assigned in terms of the ethical decisions were used to define functions to transform the data into scores. Thereby, the scoring did not follow a linear reasoning. For instance, a farm with 10 % lame cows was judged to be of a far lower welfare standard than a farm with 0 % lame cows, but a farm with 70 % lame cows was not scored much better than a farm with 80 % lame cows. Therefore, in terms of the assignment of weights, non-linear functions, i.e. cubic I-spline functions (Curry and Schoenberg, 1966), were used. However, if many different items needed to be aggregated, the experts were incapable of considering weights. This is especially the case for the calculation of principle scores from the level of criteria. In this case, blocks of measures were first considered and Choquet integrals (Grabisch and Roubens, 2000) were used for further aggregation. Although some measures could theoretically be related to different criteria, they were only considered once in order to avoid double counting.

(3) After the calculation of the scores at criteria and principle level through these different mathematical procedures, there still remained the question of what is supposed to be good welfare. One possibility was to define as normal the average score that common farms usually achieve (Whay et al., 2003), but this approach was strongly criticised as the welfare level could be generally bad (Bekoff, 2008). It was chosen – by considering opinions of animal scientists as well as social scientists representing the opinion of the public – to define theoretical thresholds for a farm considered to be acceptable, good and excellent. As for the ethical decisions, on a given scale between 0 and 100 with 0 presenting the worst and 100 the best theoretically possible welfare state, 20 was set as a limit for acceptability representing legislative requirements to be met, 55 as an aspiration value for categorisation into enhanced, and 80 for categorisation into excellent. These four categories were chosen in order to meet all the requirements of potential implications of the protocols (e.g. compulsory or voluntary labelling, self-assessment and research). First, it was chosen to rely on unanimity; thus, these aspirational values had to be met in all four principles. However, this turned out to be too strict and unrealistic in practical use at the time, as not a single farm was scored with excellent in a study on the practical implication of the protocol. Therefore, an indifference threshold of 5 was defined, meaning that for instance a score of 50 was not regarded as significantly different from 55. Furthermore, it was determined that a farm be scored as excellent if it reached values of greater than 55 on all principles and greater than 80 on two of them. It was further scored as enhanced if all the principles exceeded a value of 20 and two of them exceeded the

value of 55. The criterion of acceptability was met if a value of greater than 10 was achieved in all principles and a value of greater than 20 on three. Therefore, it became obvious that some compensation was finally allowed. This adjustment of the model was done in order to find a balance between theoretical expectations and what can be realistically achieved at the moment, and thus in order to achieve a balance between theory and pragmatism (Botreau et al., 2007, 2009; Veissier et al., 2011).

4.5 Challenges of the protocols

Critical components of the reliability of an assessment system are, of course, the assessors themselves, who need to be credible and competent (Butterworth, 2009). Although the protocols are accessible for free and how to take the measures is well described, training is needed in order to carry out the assessments correctly (Welfare Quality®, 2009). The training programme needs to be presented in a standardised way internationally so that reliable results are obtained. Furthermore, trainers should be retrained after a certain time to assure they work correctly in terms of reliability (Velarde et al., 2010). Training has been carried out so far by members of the WQ group by travelling to the companies asking for training. However, until recently, certification of trainers had not been available, nor had there been a model to ensure that the protocols are still used correctly at some point after training (Butterworth et al., 2013).

The main concern in terms of feasibility is that a protocol assessment takes a very long time (ranging between 4 and 8 h, depending on the species, farm size and distance of farm buildings) and, consequently, the implementation is too costly. During the development of the protocols, feasibility was always taken into account when decisions were made about the inclusion or exclusion of certain measurements. All measurements currently included provide together the best possible comprehensive animal-based welfare assessment in terms of validity, reliability and feasibility. Therefore, it does not seem possible to simply exclude time-consuming measurements (Veissier et al., 2013). However, ongoing research provides the potential to automate some measures, e.g. lameness scoring (Chapinal et al., 2010). Furthermore, implementation concepts are conceivable in which not always the whole protocol assessment is carried out, but after an initial, full assessment, follow-up assessments are carried out referring only to those parts of the protocol that reveal negative or problematic issues on that particular farm (Veissier et al., 2013). However, up to now, there has been no agreement on how often protocol assessments need to be carried out at all. Furthermore, there is a lack of comprehensive cost–benefit analyses of the application of the WQ protocols (Manteca and Jones, 2013).

There also remains a problem of the general reliability of the protocols. This might seem surprising as such an enormous effort was made to evaluate the validity, reliability and

feasibility of each single measurement before it was included in the protocols. However, many parameters that were included were changed concerning, for example, their scaling in order to be more feasible in their application. For instance, the parameter bursitis as a measurement tool for comfort around resting in pigs was proven to be reliable in the study of Lyons et al. (1995), who used a five-point scale for the assessment. This five-point scale was changed without further testing and cut down to a three-point scale in terms of the "Welfare Quality® animal welfare assessment protocol for growing pigs". Moreover, most reliability studies carried out in terms of the development of the protocols were based on video clips and pictures but not on on-farm assessments. Therefore, there is a lack of studies concerning the reliability of the complete WQ protocols in on-farm use. Some studies have indicated that this may be problematic, especially concerning consistency over time (Botreau et al., 2013; Temple et al., 2013).

Although the protocols were developed with the aim of being applicable in all different housing systems worldwide, their use turned out to be problematic in extensive outdoor production systems. Behavioural observations are difficult to carry out in a large field and the animals are often not used to close observation; they need to be grouped to be observed, which requires higher input by the farmer and causes a greater disturbance to the animals (Turner and Dwyer, 2007). Furthermore, the outcomes may be strongly dependant on the weather; for instance, animals are known to change their activity patterns in particularly hot or cold weather conditions (Hahn, 1999; Tucker et al., 2007). Another challenge in terms of assessing extensive systems is that, for some parameters, validation that is obvious for intensive systems presents ambiguity such as foraging behaviour in pigs, which could be assigned to feeding or exploratory behaviour.

Potential implementations of the WQ protocols include the evaluation of whether legislative requirements on a farm are met or demands in terms of voluntary labelling are fulfilled. Moreover, they would be useful in terms of self-assessment or for the evaluation of the welfare potential of new farming systems or breeds (Botreau et al., 2013). However, the question in general is how to use the information obtained to promote and support management decisions and practices in order to improve the welfare situations on farms. Of course, the results need to be reported to the farmer and practical advice needs to be provided on what problems exist, how they are caused and how they can be tackled (Manteca and Jones, 2013). Botreau et al. (2013) reported that only a minority of farmers who had received feedback and advice actually implemented the improvement strategies, which was especially due to financial, practical and motivational problems. Therefore, improvement strategies must be practicable, robust, safe, affordable, easy to implement and in the long-term interest of the farmers. Integrated approaches should be chosen, thus including environmental, managing and genetic strategies (Boissy et al., 2005; Jones et al., 2005).

In terms of serving as a certification tool, the WQ protocols have great potential, as there is a need for homogeneous and transparent certification (Evans and Miele, 2007). But several studies have revealed that the European public had expected higher welfare standards than those achieved in terms of the WQ categories (Miele et al., 2011; Evans and Miele, 2012). The whole aggregation procedure seems to still be unsatisfactory, as only a few scientists were used to assign the weights to the single measurements, as revealed by de Vries et al. (2012). They concluded that the role of expert opinion and the type of algorithm operator used in terms of the aggregation of measures should be reconsidered.

5   Use of the "Welfare Quality® animal welfare assessment protocol for growing pigs"

The following provides an overview of studies concerning the "Welfare Quality® protocol for growing pigs" in the form published in 2009. The first study to be accomplished on the finished protocol for growing pigs was carried out in terms of the WQ project. In this study, the protocol was tested on 71 growing pig farms in different countries with varying management practices and farm sizes. It was concluded that although some adaptations of the protocol may be necessary, it can be feasibly applied under a variety of different conditions (Veissier et al., 2013). Temple et al. (2011a) applied the protocol for growing pigs to 30 conventional farms in Spain to estimate its feasibility and sensitivity. It was found that the protocol assessment was easy to perform and took about 6 h. For each animal-based parameter, confidence limits were estimated, helping to identify farms with a poor status of welfare. In general, sensitivity was proven; however, the causes of the variability and differences of the results were found to be difficult to interpret. In another study, Temple et al. (2011b) put their focus on the behaviour principle of the protocol and compared the outcomes of the measurements between intensively and extensively kept pigs. Differences between the housing systems were found in terms of the positive emotional state, which was scored higher in the extensive systems and in terms of social behaviour, which increased in intensive systems. This was interpreted as a coping strategy of the animals in intensive systems to enhance positive emotions by increased social behaviour. It becomes obvious that interpretation of the results in terms of behavioural observations of the protocol is not always straightforward. In 2012, Temple published two further studies (Temple et al., 2012a, b) in which the welfare of growing pigs in five different housing systems in two different countries were compared and factors influencing the outcomes of animal-based parameters were disclosed. These factors were the age of animals, feeding system, stocking density and type of flooring. All these studies proved that differences between farms could be detected very well by the application of the "Welfare Quality® animal welfare assessment protocol for growing pigs". However, interpretation of the outcomes was

not always straightforward. Furthermore, the studies revealed some of the challenges of the protocols mentioned already, such as the difficulties in use under extensive conditions. The revealing of the causal effects on the outcomes is of eminent importance, as these need to be considered in the further use of the protocol since it would reduce consistency over time if these aspects were not taken into account. One of the revealed effects was for instance the age of the pigs. A study on the test–retest reliability (Temple et al., 2013) dealt exactly with this problem of consistency over time. Although measures were corrected for the previously described effect of age, sufficient test–retest reliability was only found in terms of two parameters. But farms with persistent welfare problems could be safely identified. Nevertheless, test–retest reliability definitely needs enhancement in terms of future implementations of the protocols and the above-mentioned challenges have to be clarified in the future, such as the question of how often the protocol needs to be carried out. The problems already revealed in terms of the protocol also highlight the further need for studies considering the reliability of the on-farm use of the entire protocols.

6   Conclusions

The aim of the present paper was to discuss the rising need for an objective and generally accepted way of measuring animal welfare. Furthermore, existing measurement methods were described with special emphasis on the WQ approach. Thereby, in particular, the existing challenges were highlighted.

Animal welfare can be best defined considering three different concepts, i.e. health, natural behaviour and positive affective state. It can be best measured by animal-based parameters as these consider the real state of the animals, whereas resource- and management-based parameters are more a risk assessment of the potential of the environment to provide for a good welfare. The WQ protocols are based on such a broad definition and furthermore rely on animal-based parameters whenever feasible. However, there are also still a number of challenges concerning the reliability and validity of the protocols in their on-farm use and the aggregation of parameters to a final welfare score. These revealed challenges need to be addressed in further studies to allow for constant improvement and enhancement of the protocols. Even more importantly, they have to be borne in mind in the application of the protocols in terms of welfare assessment to avoid misinterpretations of the welfare situation of the animals.

Edited by: A.-E. Freifrau von Tiele-Winckler
Reviewed by: two anonymous referees

References

Bartussek, H.: Vorschlag für eine Steiermärkische Intensivtierhaltungsverordnung, Der Österreichische Freiberufstierarzt, 97, 4–15, 1985.

Bartussek, H.: Entwurf Tiergerechtheitsindex für Mastschweine: TGI35 L. Bundesanstalt für Alpenländische Landwirtschaft, Gumpenstein, Irdning, Austria, 1995a.

Bartussek, H.: Tiergerechtheitsindex fuer Legehennen: TGI 35 L. Bundesanstalt für Alpenländische Landwirtschaft, Gumpenstein, Irdning, Austria, 1995b.

Bartussek, H.: A review of the animal needs index (ANI) for the assessment of animals' well-being in the housing systems for Austrian proprietary products and legislation, Livest. Prod. Sci., 61, 179–192, 1999.

Bekoff, M.: Why "good welfare" isn't "good enough": minding animals and increasing our compassionate footprint, Annual Review of Biomedical Sciences, 10, T1–T14, 2008.

Bennett, R. M., Anderson, J., and Blaney, R. J. P.: Moral intensity and willingness to pay concerning farm animal welfare issues and the implications for agricultural policy, J. Agr. Environ. Ethic., 15, 187–202, 2002.

Blokhuis, H. J.: International cooperation in animal welfare: the Welfare Quality® project, Acta Vet. Scand., 50, S10, 2008.

Blokhuis, H. J., Jones, R. B., Geers, R., Miele, M., and Veissier, I.: Measuring and monitoring animal welfare: transparency in the food product quality chain, Anim. Welfare, 12, 445–455, 2003.

Blokhuis, H. J., Veissier, I., Miele, M., and Jones, B.: The Welfare Quality project and beyond: Safeguarding farm animal well-being, Acta Agr. Scand. A-An., 60, 129–140, 2010.

Blokhuis, H., Jones, B., Veissier, I., and Miele, M.: Introduction, in: Improving farm animal welfare, edited by: Blokhuis, H., Miele, M., Veissier, I., and Jones, B., Wageningen Academic Publishers, Wageningen, the Netherlands, 11–17, 2013a.

Blokhuis, H., Veissier, I., Jones, B., and Miele, M. Improving farm animal welfare - Science and society working together: the Welfare Quality approach, in: Improving farm animal welfare, edited by: Blokhuis, H., Miele, M., Veissier, I., and Jones, B., Wageningen Academic Publishers, Wageningen, the Netherlands, 71–85, 2013b.

Blokhuis, H., Veissier, I., Jones, B., and Miele, M.: The Welfare Quality® vision, In: Improving farm animal welfare, edited by: Blokhuis, H., Miele, M., Veissier, I., and Jones, B., Wageningen Academic Publishers, Wageningen, the Netherlands, 71–89, 2013c.

Bock, B. B. and Van Huik, M. M.: Animal welfare: the attitudes and behaviour of European pig farmers, Brit. Food J., 109, 931–944, 2007.

Bock, B. B. and van Leeuwen, F. A. C.: Review of socio-political and market developments of animal welfare schemes, in: Farm Animal Welfare Concerns. Consumers, retailers and producers (Welfare Quality Reports 1), edited by: Miele, M. and Roex, J., Cardiff University Press, Cardiff, UK, 115–150, 2005.

Boissy, A., Fisher, A. D., Bouix, J., Hinch, G. N., Le Neindre, P.: Genetics of fear in ruminant livestock, Livest. Prod. Sci., 93, 23–32, 2005.

Botreau, R., Veissier, I., Butterworth, A., Bracke, M. B. M., and Keeling, L. J.: Definition of criteria for overall assessment of animal welfare, Anim. Welfare, 16, 225–228, 2007.

Botreau, R., Veissier, I., and Perny, P.: Overall assessment of animal welfare: strategy adopted in Welfare Quality, Anim. Welfare, 18, 363–370, 2009.

Botreau, R., Winckler, C., Velarde, A., Butterworth, A., Dalmau, A., Keeling, L. J., and Veissier, I.: Integration of data collected on farms or at slaughter to generate an overall assessment of animal welfare, in: Improving farm animal welfare-Science and society working together: the Welfare Quality approach, edited by: Blokhuis, H., Miele, M., Veissier, I., and Jones, B., Wageningen Academic Publishers, Wageningen, Gelderland, the Netherlands, 147–173, 2013.

Botreau, R. L., Capdeville, J., Perny, P., and Veissier, I.: Multicriteria evaluation of animal welfare at farm level: an application of MCDA methodologies, Foundation of Computing and Decision Sciences, 33, 287–316, 2008.

Bracke, M. B. M.: Animal-based parameters are no panacea for on-farm monitoring of animal welfare, Anim. Welfare, 16, 229–231, 2007.

Bracke, M. B. M., Metz, J. H. M., and Spruijt, B. M.: Development of a decision support system to assess farm animal welfare, Acta Agr. Scand. A-An., 51, 17–20, 2001.

Bracke, M. B. M., Spruijt, B. M., Metz, J. H. M., and Schouten, W. G. P.: Decision support system for overall welfare assessment in pregnant sows A: Model structure and weighting procedure, J. Anim. Sci., 80, 1819–1834, 2002.

Brambell, F. W. R.: Report of the Technical Committee to Enquire Into the Welfare of Animals Kept Under Intensive Livestock Husbandry Systems: Presented to Parliament by the Secretary of State for Scotland and the Minister of Agriculture, Fisheries and Food by Command of Her Majesty, December, 1965, HM Stationery Office, 1965.

Broom, D. M.: The scientific assessment of animal welfare, Appl. Anim. Behav. Sci., 20, 5–19, 1988.

Buller, H. and Roe, E.: Certifying welfare: integrating welfare assessments into assurance procedures: a European perspective: 25 key points. in: Certifying Quality: Negotiating and integrating animal welfare into food assurance (Welfare Quality Reports 15), edited by: Buller, H. and Roe, E., Cardiff University Press, Cardiff, UK, 25–38, 2008.

Butterworth, A.: Animal welfare indicators and their use in society, in: Food Safety Assurance and Veterinary Public Health, edited by: Smulders, J. F. M. and Algers, B., Wageningen Academic Publishers, Wageningen, the Netherlands, 371–389, 2009.

Butterworth, A.: Relevance and implementation of Welfare Quality assessment, in: Improving farm animal welfare, edited by: Blokhuis, H., Miele, M., Veissier, I., and Jones, B., Wageningen Academic Publishers, Wageningen, the Netherlands, 201–213, 2013.

Chapinal, N., De Passille, A. M., Rushen, J., and Wagner, S.: Automated methods for detecting lameness and measuring analgesia in dairy cattle, J. Dairy Sci., 93, 2007–2013, 2010.

Courboulay, V., Meunier-Salaun, M. C., Edwards, S. A., Guy, J. H., and Scott, K.: Repeatability of Abnormal Behaviour, in: Assessment of animal welfare measures for sows, piglets and fattening pigs (Welfare Quality Reports 10), edited by: Forkman, B. and Keeling, L. J., SLU Service/Reproenheten, Uppsala, Sweden, 131–141, 2009.

Curry, H. B. and Schoenberg, I. J.: On Pólya frequency functions IV: the fundamental spline functions and their limits, J. Anal. Math., 17, 71–107, 1966.

de Mol, R. M., Schouten, W. G. P., Evers, E., Drost, H., Houwers, H. W. J., and Smits, A. C.: A computer model for welfare assessment of poultry production systems for laying hens, NJAS-Wageningen Journal of Life Sciences, 54, 157–168, 2006.

de Passille, A. M. and Rushen, J.: Can we measure human animal interactions in on-farm animal welfare assessment?: Some unresolved issues, Appl. Anim. Behav. Sci., 92, 193–209, 2005.

de Vries, M., Bokkers, E. A. M., van Schaik, G., Botreau, R., Engel, B., Dijkstra, T., and de Boer, I. J. M.: Evaluating results of the Welfare Quality multi-criteria evaluation model for classification of dairy cattle welfare at the herd level, J. Dairy Sci., 96, 6264–6273, 2012.

Engel, B., Bruin, G., Andre, G., and Buist, W.: Assessment of observer performance in a subjective scoring system: visual classification of the gait of cows, J. Agr. Sci., 140, 317–333, 2003.

Evans, A. and Miele, M. (Eds.): Consumers' Views about Farm Animal Welfare Part I: National Reports based on Focus Group Research (Welfare Quality Reports 4) Cardiff, UK, 2007.

Evans, A. and Miele, M.: Between food and flesh: how animals are made to matter (and not matter) within food consumption practices, Environ. Plann. D, 30, 298–314, 2012.

FAWC: First Press Notice, 5/12. MAFF, London, UK, 1979.

Forkman, B. and Keeling, L. J. (Eds.): Assessment of Animal Welfare Measures for Sows, Piglets and Fattening Pigs (Welfare Quality Reports 10) SLU Service/Reproenheten, Uppsala, Sweden, 2009.

Forkman, B. and Keeling, L. J. (Eds.): Assessment of Animal Welfare Measures for Dairy Cattle, Beef Bulls and Veal Calves (Welfare Quality Reports 11) SLU Service/Reproenheten, Uppsala, Sweden, 2009a.

Forkman, B. and Keeling, L. J. (Eds.): Assessment of Animal Welfare Measures for Layers and Broilers (Welfare Quality Reports 9) SLU Service/Reproenheten, Uppsala, Sweden, 2009b.

Fraser, D.: Science, values and animal welfare: exploring the "inextricable connection", Anim. Welfare, 4, 103–117, 1995.

Fraser, D.: Encyclopedia of Animal Rights and Animal Welfare, Greenwood Press, Westport, CT, USA, 1998.

Fraser, D.: Understanding animal welfare, Acta Vet. Scand., 50, S1, 2008.

Fraser, D.: Assessing Animal Welfare: Different Philosophies, Different Scientific Approaches, Zoo Biol., 28, 507–518, 2009.

Grabisch, M. and Roubens, M.: Application of the Choquet integral in multicriteria decision making, in: Fuzzy Measures and Integrals – Theory and Applications, edited by: Grabisch, M., Murofushi, T. and Sugeno, M., Physica Verlag, Heidelberg, Germany, 348–375, 2000.

Grandin, T.: Progress and challenges in animal handling and slaughter in the U.S., Appl. Anim. Behav. Sci., 100, 129–139, 2006.

Grandin, T.: Improving animal welfare: a practical approach, CABI, Wallingford, England, UK, 2010.

Hahn, G. L.: Dynamic responses of cattle to thermal heat loads, J. Anim. Science, 77, 10–20, 1999.

Haiger, A., Storhaus, R., and Bartussek, H.: Naturgemässe Viehwirtschaft, Verlag Eugen Ulmer, Stuttgart, Germany, 1988.

Hewson, C. J.: Can we assess welfare?, Canadian Vet. J., 44, 749–753, 2003.

Hobbs, A. L., Hobbs, J. E., Isaac, G. E., and Kerr, W. A.: Ethics, domestic food policy and trade law: assessing the EU animal welfare proposal to the WTO, Food Policy, 27, 437–454, 2002.

Hughes, B. O.: Behaviour as an index of welfare, 5th European Poultry Conference, Malta, 5–11 September 1976, Volume II, 1005–1018, 1976.

Johnsen, P. F., Johannesson, T., and Sandoe, P.: Assessment of farm animal welfare at herd level: many goals, many methods, Acta Agr. Scand. A-An., 51, 26–33, 2001.

Jones, T. A., Donnelly, C. A., and Dawkins, M. S.: Environmental and management factors affecting the welfare of chickens on commercial farms in the United Kingdom and Denmark stocked at five densities, Poultry Sci., 84, 1155–1165, 2005.

Keeling, L. J.: Healthy and happy: animal welfare as an integral part of sustainable agriculture, AMBIO, 34, 316–319, 2005.

Keeling, L. J., Rushen, J., and Duncan, I.: Understanding Animal Welfare, in: Anim. Welfare, edited by: Appleby, M. C., Mench, J. A., Olsson, I. A. S., and Hughes, B. O., CABI, Wallingford, England, UK, 183–199, 2011.

Kjaernes, U. and Lavik, R.: Opinions on animal welfare and food consumption in seven European countries, in: Consumption, Distribution and Production of farm animal welfare (Welfare Quality Reports 7), edited by: Kjaernes, U., Bock, B. B., Roes, E., Roe, E., and Roex, J., MPG Biddles Ltd, King's Lynn, UK, 3–126, 2008.

Knierim, U. and Winckler, C.: On-farm welfare assessment in cattle: validity, reliability and feasibility issues and future perspectives with special regard to the Welfare Quality® approach, Anim. Welfare, 18, 451–458, 2009.

Lassen, J., Sandoe, P., and Forkman, B.: Happy pigs are dirty! Conflicting perspectives on animal welfare, Livest. Sci., 103, 221–230, 2006.

Leeb, C., Whay, B., and Main, D.: Incorporation of conventional animal welfare assessment techniques into organic certification and farming, Department of Clinical Veterinary Science University of Bristol, Bristol, UK, 1–18, 2004.

Leeb, C., Whay, B., Main, D., and Webster, J.: BWAP – Bristol Welfare Assurance Programme, Tiergerechtheit bewerten, Kuratorium fuer Technik und Bauwesen in der Landwirtschaft e.V., Darmstadt, Germany, 65–72, 2014.

Leruste, H., Lensink, J., and van Reenen, C. G.: Measurement of Human-Animal Relations in Veal Calves Housed in Large Groups, in: Assessment of Animal Welfare Measures for Dairy Cattle, Beef Bulls and Veal Calves (Welfare Quality Reports 11), edited by: Forkman, B., and Keeling, L. J., SLU Service/Reproenheten, Uppsala, Sweden, 131–136, 2009.

Lyons, C., Bruce, J., Fowler, V., and English, P.: A comparison of productivity and welfare of growing pigs in four intensive systems, Livest. Prod. Sci., 43, 265–274, 1995.

Magerman, D. M.: Statistical decision-tree models for parsing, Proceedings of the 33rd annual meeting on Association for Computational Linguistics, Association for Computational Linguistics, 276–283, 1995.

Main, D. C., Whay, H. R., Green, L. E., and Webster, A. J.: Effect of the RSPCA Freedom Food Scheme on the welfare of dairy cattle, Vet. Rec., 153, 227–231, 2003.

Manteca, X. and Jones, B.: Welfare improvement strategies, in: Improving farm animal welfare, edited by: Blokhuis, H., Miele, M.,

Veissier, I., and Jones, B., Springer, Berlin, Germany, 175–200, 2013.

Menke, C., Waiblinger, S., Fölsch, D. W., and Wiepkema, P. R.: Social behaviour and injuries of horned cows in loose housing systems, Anim. Welfare, 8, 243–258, 1999.

Miele, M., Veissier, I., Evans, A., and Botreau, R.: Animal welfare: establishing a dialogue between science and society, Anim. Welfare, 20, 103–117, 2011.

Moberg, G. P.: Biological response to stress: implications for animal welfare. The biology of animal stress: basic principles and implications for animal welfare, 1–21, 2000.

Mortimore, S. and Wallace, C.: HACCP: A practical approach. Springer, Berlin, Germany, 2013.

Napolitano, F., De Rosa, G., Ferrante, V., Grasso, F., and Braghieri, A.: Monitoring the welfare of sheep in organic and conventional farms using an ANI 35 L derived method, Small Ruminant Res., 83, 49–57, 2009.

Pelzer, A. and Kaufmann, O.: Das Tier im Blick – Milchkühe, DLG e.V., Frankfurt/Main, Germany, 2012.

Plesch, G., Broerkens, N., Laister, S., Winckler, C., and Knierim, U.: Reliability and feasibility of selected measures concerning resting behaviour for the on-farm welfare assessment in dairy cows, Appl. Anim. Behav. Sci., 126, 19–26, 2010.

Roe, E. and Buller, H.: Marketing farm animal welfare, Welfare Quality factsheet, Wageningen, Gelderland, the Netherlands, 2008.

Roex, J. and Miele, M.: Farm Animal Welfare Concerns: Consumers, Retailers and Producers (Welfare Quality Reports 1), Cardiff University Press, Cardiff, UK, 2005.

Rousing, T., Bonde, M., and Sorensen, J. T.: Aggregating welfare indicators into an operational welfare assessment system: a bottom-up approach, Acta Agr. Scand. A-An., 51, 53–57, 2001.

Sato, S.: Social licking pattern and its relationships to social dominance and live weight gain in weaned calves, Appl. Anim. Behav. Sci., 12, 25–32, 1984.

Sato, S., Sako, S., and Maeda, A.: Social licking patterns in cattle (Bos taurus): influence of environmental and social factors, Appl. Anim. Behav. Sci., 32, 3–12, 1991.

Schulze Westerath, H., Brarkens, N., Laister, S., MacKintosh, N., Winckler, C., and Knierim, U.: Reliability of measures of sociopositive and play behaviour in dairy and beef cattle, in: Assessment of Animal Welfare Measures for Dairy Cattle, Beef Bulls and Veal Calves (Welfare Quality Reports 11), edited by: Forkman, B. and Keeling, L. J., SLU Service/Reproenheten, Uppsala, Sweden, 175–188, 2009.

Schumacher, U., Link, M., Volling, O., Brinkmann, J., March, S., Winckler, C., Tiergesundheitsmanagement Rinder, in: Bioland-Handbuch Tiergesundheitsmanagement, edited by: Schumacher, U., Bioland-Verlag, Mainz, Germany, 1–14, 2007.

Scott, E. M., Nolan, A. M., and Fitzpatrick, J. L.: Conceptual and methodological issues related to welfare assessment: a framework for measurement, Acta Agr. Scand. A-An., 51, 5–10, 2001.

Smulders, F. J. M. and Algers, B.: Welfare of production animals: assessment and management of risks. ECVPH Food safety assurance Vol 5, Wageningen Academic Publishers, Wageningen, the Netherlands, 2009.

Stien, L. H., Bracke, M., Folkedal, O., Nilsson, J., Oppedal, F., Torgersen, T., Kittilsen, S., Midtlyng, P. J., Vindas, M. A., and Averli, A.: Salmon Welfare Index Model (SWIM 1.0): a semantic

model for overall welfare assessment of caged Atlantic salmon: review of the selected welfare indicators and model presentation, Reviews in Aquaculture, 5, 33–57, 2013.

Temple, D., Dalmau, A., Ruiz de la Torre, J., Manteca, X., and Velarde, A.: Application of the Welfare Quality protocol to assess growing pigs kept under intensive conditions in Spain, J. Vet. Behav., 6, 138–149, 2011a.

Temple, D., Manteca, X., Velarde, A., and Dalmau, A.: Assessment of animal welfare through behavioural parameters in Iberian pigs in intensive and extensive conditions, Appl. Anim. Behav. Sci., 131, 29–39, 2011b.

Temple, D., Courboulay, V., Manteca, X., Velarde, A., and Dalmau, A.: The welfare of growing pigs in five different production systems: assessment of feeding and housing, Animal, 6, 656–667, 2012a.

Temple, D., Courboulay, V., Velarde, A., Dalmau, A., and Manteca, X.: The welfare of growing pigs in five different production systems in France and Spain: assessment of health, Anim. Welfare, 21, 257–271, 2012b.

Temple, D., Manteca, X., Dalmau, A., and Velarde, A.: Assessment of test-retest reliability of animal-based measures on growing pig farms, Livest. Sci., 151, 35–45, 2013.

Tuchscherer, M., Puppe, B., Tuchscherer, A., and Kanitz, E.: Effects of social status after mixing on immune, metabolic, and endocrine responses in pigs, Physiol. Behav., 64, 353–360, 1998.

Tucker, C. B., Rogers, A. R., Verkerk, G. A., Kendall, P. E., Webster, J. R., and Matthews, L. R.: Effects of shelter and body condition on the behaviour and physiology of dairy cattle in winter, Appl. Anim. Behav. Sci. 105, 1–13, 2007.

Turner, S. P. and Dwyer, C. M.: Welfare assessment in extensive animal production systems: challenges and opportunities, Anim. Welfare, 16, 189–192, 2007.

Ursinus, W. W., Schepers, F., de Mol, R. M., Bracke, M. B. M., Metz, J. H. M., and Groot Koerkamp, P. W. G.: COWEL: A decision support system to assess welfare of husbandry systems for dairy cattle, Anim. Welfare, 18, 545–552, 2009.

Vanhonacker, F., Verbeke, W., Van Poucke, E., and Tuyttens, F. A. M.: Do citizens and farmers interpret the concept of farm animal welfare differently?, Livest. Sci., 116, 126–136, 2008.

Veissier, I., Jensen, K. K., Botrea, R., and Sandoe, P.: Highlighting ethical decisions underlying the scoring of animal welfare in the Welfare Quality® scheme, Anim. Welfare, 20, 89–101, 2011.

Veissier, I., Winckler, C., Velarde, A., Butterworth, A., Dalmau, A., and Keeling, L. J.: Development of welfare measures and protocols for the collection of data on farms or at slaughter, in: Improving farm animal welfarem edited by: Blokhuis, H., Miele, M., Veissier, I., and Jones, B., Wageningen Academic Publishers, Wageningen, the Netherlands, 1, 115–141, 2013.

Velarde A. and Geers, R. (Eds.): On farm monitoring of pig welfare. COST ACTION 846, Wageningen Academic Publishers, Wageningen, the Netherlands, 2007.

Velarde, A., Dalmau, A., and Manteca, X.: Animal Welfare training for professional: moving from inputs to outcomes, Proceedings of the International Conference on animal welfare education: everyone is responsible, Brussels, Belgium, 1–2 October 2010, European Union-Committee of the Regions, 38–44, 2010.

von Borell, E.: An evaluation of indexing welfare in farm animals, Aus dem Institut für Tierzucht Mariensee, Mariensee, Germany, 41–46, 2001.

von Borell, E., Bockisch, F. J., Boescher, W., Hoy, S., Krieter, J., Müller, C., Parvizi, N., Richter, T., Rudovsky, A., and Sundrum, A.: Critical control points for on-farm assessment of pig housing, Livest. Prod. Sci., 72, 177–184, 2001.

Webster, J.: The assessment and implementation of animal welfare: theory into practice, Rev. Sci. Tech. OIE, 24, 723–734, 2005.

Welfare Quality®: Welfare Quality® Assessment Protocol for pigs (sows and piglets, growing and finishing pigs). Welfare Quality Consortium, Lelystad, the Netherlands, 2009.

Whay, H. R., Main, D. C. J., Green, L. E., and Webster, A. J. F.: Animal-based measures for the assessment of welfare state of dairy cattle, pigs and laying hens: consensus of expert opinion, Anim. Welfare, 12, 205–217, 2003.

Whay, H. R., Leeb, C., Main, D. C. J., Green, L. E., and Webster, A. J. F.: Preliminary assessment of finishing pig welfare using animal-based measurements, Anim. Welfare, 16, 209–211, 2007.

Winckler, C. On-farm welfare assessment in cattle from basic concepts to feasible assessment systems. Proc. 24th World Buiatrics Congr., Nice, France, 493–500, 2006.

Winckler, C. and Knierim, U.: Welfare Quality Assessment Protocol – Milchkühe und Mastrinder, Tiergerechtheit bewerten, Kuratorium für Technik und Bauwesen in der Landwirtschaft e.V., Darmstadt, Germany, 7–17, 2014.

Wirtz, M. and Caspar, F.: Beurteilerübereinstimmung und Beurteilerreliabilität, Hogrefe, Göttingen, Germany, p. 246, 2002.

World Health Organization: WHO definition of Health. Preamble to the Constitution of the World Health Organization as adopted by the International Health Conference, New York, 19–22 June, 1946; signed on 22 July 1946 by the representatives of 61 States (Official Records of the World Health Organization, no. 2, p. 100) and entered into force on 7 April 1948, 1946.