The Old School, Brewhouse Hill, Wheathampstead,
Hertfordshire AL4 8AN, UK
www.ufaw.org.uk

Animal Welfare 2021, 30: 49-59
ISSN 0962-7286
doi: 10.7120/09627286.30.1.049

Development of a list of terms in Brazilian Portuguese for the Qualitative Behaviour Assessment of broiler chickens

APO Souza*†, F Wemelsfelder‡, CA Taconeli§ and CFM Molento†

† Animal Welfare Laboratory, Federal University of Paraná, Rua dos Funcionários, 1540, 80035-050, Curitiba, Paraná, Brazil
‡ Animal and Veterinary Sciences Group, SRUC, Roslin Institute Building, Midlothian EH25 9RG, UK
§ Department of Statistics, Federal University of Paraná, Centro Politécnico, 81531-990, Curitiba, Paraná, Brazil
* Contact for correspondence: anapaulasouza.mv@gmail.com

Abstract

*Qualitative Behaviour Assessment (QBA) is a methodological approach to assess the whole animal using terms to describe and quantify the emotionally expressive qualities of behaviour and identifying larger patterns of expressivity through multi-variate statistical integration. A key condition for the success of QBA is achieving a common understanding of the meaning of descriptive terms by raters. Based on this, our study aimed to develop a list of terms in Brazilian Portuguese for the QBA of broiler chickens (*Gallus gallus domesticus*), and to test this list by studying its inter- and intra-rater reliability. Fourteen experts participated in a workshop and developed a list of 25 QBA terms, and 40 undergraduates tested this list by scoring 18 video clips using a 125-mm visual analogue scale. Principal Component Analysis was used to analyse observers' scores. Principal Component (PC) 1 ranged from disturbed/frustrated to comfortable/lively, suggesting this PC may be interpreted in terms of emotional valence. PC2 ranged from calm/dull to agitated/active, suggesting this PC indicates the level of arousal/energy of the birds. Both PC1 and PC2 clip scores showed good inter- and intra-rater reliability. This study demonstrates the importance of producing QBA term lists bottom-up as opposed to merely translating pre-existing lists from the scientific literature. Results suggest the standardised Portuguese QBA term list developed in this study is reliable in assessing the expressive qualities of broiler behaviour; therefore, a next step is to test it on-farm with experienced raters and further refine it concerning terms related to poor welfare.*

Keywords: *animal-based indicator, animal welfare, behaviour, poultry, Qualitative Behavioural Assessment, Welfare Quality® protocol*

Introduction

Brazil is the third largest producer of broiler chicken (*Gallus gallus domesticus*) meat (ABPA 2020) in the world, with a total of 5.8 billion birds slaughtered in 2019 (IBGE 2020), which means that regular assessment of broiler chicken welfare in this country is essential. Local regulations have emphasised monitoring procedures at pre-slaughter and slaughter levels. However, at farm level, no specific regulations exist for the protection of broiler chickens in this country. Some recent studies have applied the Welfare Quality® protocol (Welfare Quality® 2009) to investigate broiler chicken welfare in Brazil (Sans *et al* 2014; Souza *et al* 2015; Tuyttens *et al* 2015; Federici *et al* 2016). The Welfare Quality® protocol has been chosen because it comprises scientifically validated indicators that are predominantly animal-based (Blokhuis *et al* 2010). It also includes various indicators for behavioural assessment of the animals, something that is essential for a complete welfare evaluation in addition to housing, nutritional and health conditions (Welfare Quality® 2009).

Qualitative Behaviour Assessment (QBA) is the measure for positive emotional state in the Welfare Quality® protocol. It is a methodological approach developed to assess the whole animal, integrating information on how animals behave, and capturing it into numbers allowing for statistical analysis (Wemelsfelder *et al* 2001; Fleming *et al* 2016). QBA uses terms that describe the emotionally expressive qualities of animal behaviour, such as relaxed, agitated, scared, or comfortable. Such terms reflect an animal's experience of the situation it is facing (Wemelsfelder *et al* 2000, 2001) and provide information that is highly relevant to the animal's welfare but cannot be obtained from only measuring physical elements of behaviour. A dimensional model of valence and arousal, such as the one proposed by Russell and Bullock (1985), has been increasingly used in animal studies to classify emotions (Burn 2017), and is considered a feasible framework to study and assess affective states in animals (Mendl *et al* 2010). In studies about the expressive qualities of animal behaviour, the dimensional model helps to align and interpret QBA dimensions, delivering information about animal mood and energy, which are relevant to animal welfare assessment (de Boyer des Roches *et al* 2018).

The expression of an animal may provide important information regarding its welfare state. Some advocate that welfare could be improved by understanding how animals feel, since animal welfare does not just concern the absence of negative emotions, but also includes the presence of positive ones (Boissy *et al* 2007). In such an approach, QBA provides information about an animal's affective state, contributing to a complete welfare assessment and going beyond the traditional assessment of health, nutrition and housing. As for chickens, there are scientific findings evidencing that they experience complex positive and negative emotions combined with cognition and sociability (Marino 2017); thus, it seems interesting to further study tools to assess the affective states of these animals.

There are two QBA approaches, one allowing the raters to create their own list of terms through the Free Choice Profiling (FCP) method (Wemelsfelder *et al* 2000, 2001), and the other using a standardised list of previously validated terms. The standardised list is more practical for on-farm assessments and, provided that raters are adequately trained, is more feasible (Fleming *et al* 2016). QBA has been tested using both FCP and pre-fixed term lists in different scenarios and species, such as dairy and beef cattle, dairy buffalo, dogs, horses, pigs and sheep (Fleming *et al* 2016); donkeys (Minero *et al* 2016) and dairy goats (Grosso *et al* 2016; Battini *et al* 2018). For broiler chickens, standardised lists were used generally at group level to test correlation with other indicators (Bassler *et al* 2013; Muri *et al* 2019), or as part of application of the Welfare Quality® protocol in broiler chicken farms (Sans *et al* 2014; De Jong *et al* 2015; Souza *et al* 2015; Buijs *et al* 2016; Federici *et al* 2016). More specifically, for broiler chickens the standardised list has been useful to understand fear of humans (Muri *et al* 2019) and the effect of a dark period (Bassler *et al* 2013), while its correlation to other animal-based measures, such as contact dermatitis, lameness and mortality remains unclear; however, QBA results provide information on the whole animal's welfare state that should be regarded as complementary to the information provided by other measures (Muri *et al* 2019). The QBA for broiler chickens was initially developed by Wang (2004) using the FCP method, and further refined by Wemelsfelder *et al* (2009) into a standardised list of terms. Even though QBA is part of the Welfare Quality® protocol for broiler chickens, there is still a need for inter- and intra-rater reliability tests to further validate this method (De Jong *et al* 2014; Muri *et al* 2019).

The success of QBA is dependent on recognition and common understanding of the meaning of QBA terms by raters. The list proposed in the Welfare Quality® protocol was developed in English and has required translation to Portuguese to be applied in Brazil. According to Meagher (2009), terms used in an assessment scale are expected to be clear and understood by the raters, implying that providing QBA descriptors in an assessor's own language is more appropriate. However, the translation of English QBA terms to Brazilian Portuguese does not seem to be the best approach, since the translation may not properly address the regional usage of language and understanding of terms by native speakers, especially in big countries, such as Brazil. The need to develop chicken QBA terms in Portuguese has already been proposed in a previous study (Federici *et al* 2016). Based on this, our study aimed to develop a list of terms in Brazilian Portuguese for the QBA of broiler chickens, and to test it by studying its inter- and intra-rater reliability.

Materials and methods

Ethical statement

This project was approved by the Human Research Ethics Committee of the Health Science Sector (n 1,958,250; March 10th, 2017) and by the Animal Use Ethics Committee of the Agricultural Campus (n 122/2016; December 7th, 2016), both of the Federal University of Paraná, Southern Brazil

Development of the list of descriptive terms

Study animals and video recording

Ten commercial (25°17'49.1"S, 54°05'41.7"W) and one experimental (25°23'11.3"S 49°07'35.7"W) poultry farms were visited to make video recordings in January and April 2017, respectively. Commercial poultry barns had sidewalls with wire mesh, one covered by blackout curtains working as dark house and nine covered by yellow curtains with natural lighting, all equipped with automatic feeders, nipple drinkers, sprinklers, exhaust fans and wood-shaving litter. The mean (± SD) commercial broiler chicken barn area was 1,540 (± 187) m^2 and the number of birds per house was 18,904 (± 2,604). Birds were male and female Cobb 500®, 41.3 (± 2.0) days of age. The experimental barn had 560 m^2 of area divided into 116 floor pens of 2.06 m^2, with 21 birds each. The experimental barn had sidewalls with wire mesh covered by blue curtains with natural lighting, manual feeders, cup drinkers, brooders, exhaust fans and wood-shaving litter. Broilers were male and female Ross 308, 16 days of age.

Video-recording sessions were aimed at gathering examples of emotionally expressive qualities of broiler chicken behaviour, recording birds in situations associated with high and low energetic intensities of positive and negative emotional states. Birds were recorded in groups during regular situations inside the poultry barns ranging from good to low animal welfare states. Clips presented images of birds while they were resting, sleeping, walking, standing, scratching the litter, feeding, drinking, being disturbed, interacting with the environment and with each other, in the presence of a familiar and a non-familiar human being, in thermal comfort and discomfort, and performing comfort behaviours, such as preening and dust-bathing. On the experimental farm, items such as straw, wooden platforms and pieces of coloured plastic were added to the floor pens to encourage exploratory behaviours. Additionally, two birds were recorded in situations of reduced welfare, one experiencing fear during the isolation test and another that was severely lame. A total of 21 videos were selected to be representative of the four quadrants in a two-dimensional model of arousal and valence (Russell &

Bullock 1985). All the video clips were recorded in high definition using a Sony Cyber-shot DSC-W320 camera (Sony, China) with stock Carl Zeiss lens (Sony) on a tripod, without filters and preserving the surrounding sound.

Term generation session

In August 2018, 24 experts were invited to participate in a session of 4-h duration, to develop descriptive terms for broiler chicken expression. Participants were selected based on their academic or professional knowledge of animal welfare or broiler chicken production, and included post-graduate students in veterinary and animal sciences, as well professionals from the government and the broiler chicken meat industry, all in the State of Paraná, Southern Brazil.

The session began with a brief introduction about QBA and two practice videos to discuss any questions presented by the participants and the type of terms that were expected to be developed. Participants were instructed to write down terms to describe *how* birds behaved rather than *what* birds were doing. The session comprised of four steps, identified as S1, S2, S3 and S4, and are described in the following text.

(S1) Individual term generation: Participants watched 21 video clips of 1 min each. Following Phythian *et al* (2013), video clips were ordered to contrast in valence or, in the same valence but contrasting in intensity, to prompt participants to observe the differences and stimulate the generation of terms. More extreme videos, such as birds experiencing fear, pain, or a playful situation were put further to the back. Based on the first phase of the FCP method (Wemelsfelder *et al* 2001), at the end of each clip, participants had 2 min to write down, individually, as many terms as they wanted to describe observed expressive qualities of bird behaviour. In cases where video clips showed birds in groups, participants were instructed to write down terms describing expressions observed for the group as a whole.

(S2) Individual term list refinement: After a brief explanation about the four quadrants in a two-dimensional model of arousal and valence (Russell & Bullock 1985), each participant had 20 min to select, for each quadrant, a minimum of three terms of their own list of terms generated in S1.

(S3) Term list refinement in groups: Three groups were formed with participants equally distributed according to their level of academic knowledge and professional competence in animal welfare, broiler chicken production and disease. They had 30 min to discuss their personal term lists refined in S2 and to build a single list per group, divided into positive and negative valence, including terms representative of low and high arousal.

(S4) Final term list definition: This step consisted of an open session where all participants discussed the three lists built in S3, and on that basis came to an agreement on a final list of 18 terms.

After this session, all the terms of each step were typed into an Excel® file and evaluated by the lead experimenter. Any term that was not representative of an emotionally expressive quality of animal behaviour was marked and removed from the list afterwards, such as terms expressing what the bird was physically doing (eg drinking, panting, foraging, preening). Terms were counted using the Insite Website (http://linguistica.insite.com.br/corpus.php).

Having evaluated the distribution of the list of terms created by the participants across the four quadrants, the experimenters added seven more terms to the final list, adding up to a total of 25 terms. Presentation of terms on the scoring form was ordered by the researchers so that terms that were similar in meaning were not placed close to each other to avoid contagion during assessment.

Testing of the list of QBA descriptive terms

In October 2018, 36 undergraduates of animal sciences (seven males and 29 females), who had enrolled in the ethology course, were invited to participate in a 2-h classroom instruction about broiler chicken behaviour and QBA, with the goal of testing the inter- and intra-observer reliability of the fixed term list. After the classroom instruction, participants discussed the meaning of the terms for 30 min, so that there was a common understanding of the terms within the group (Meagher 2009). They also practiced the application of the terms with three video clips. Participants then watched 18 video clips of 1 min each, followed by 2 min to score each video using a scoring form with a 125-mm visual analogue scale (VAS) added to each term, anchored with 'minimum' and 'maximum' at each end. The video clips selected were those developed for the term generation session, following the same rule to be ordered. A total of ten out of 12 video clips presented birds in the last week of life in regular situations inside the commercial poultry barns, and six of them were horizontally mirrored and repeated to test intra-rater reliability. The other two video clips were of the younger birds interacting with environmental enrichment, a wooden platform and straw. Participants were instructed to score video clips of animals in groups considering the group as a unit, assessing the total intensity of different expressive qualities in the interactive dynamics of the group's movements. Following this session, participants discussed their views on the applicability of the terms. VAS values for each term were determined measuring the distance in mm with a ruler, from the minimum point of the scale to the point where the participant marked the VAS. These values were entered into a Microsoft Excel® worksheet to be analysed.

Statistical analysis

Principal Component Analysis (PCA), using a correlation matrix and applying no rotation, was conducted (Johnson & Wichern 2007). Parallel analysis (Franklin *et al* 1995), based on simulated datasets under independence structure, was used to choose how many components to retain. The PC scores attributed to the 18 video clips on the first four principal components were evaluated for inter- and intra-rater reliability using intraclass correlation coefficient (ICC) (Bartko 1966). Intra- and inter-rater agreement for separate terms were assessed through the Variance Components method, by fitting linear mixed models (McCulloch & Searle 2004). Agreement was also assessed by means of ICC, computed from the variances associated to videos, subjects, and error.

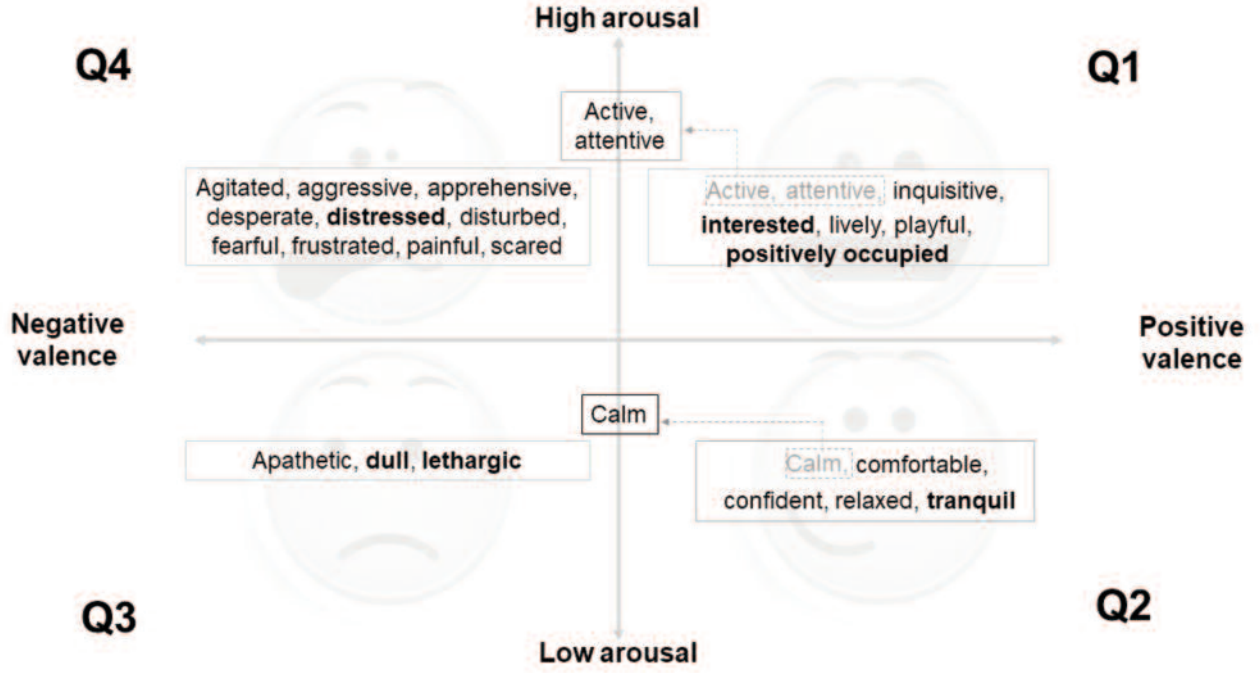
Table 1 Descriptive terms obtained in a workshop with 14 experts for the development of a Qualitative Behaviour Assessment term list in Brazilian Portuguese (English translation supplied for reference) to assess the expressive qualities of broiler chicken behaviour; results of discussion in groups (step 3) and consensus in an open session (step 4).

Step 3						Step 4	
Group 1		Group 2		Group 3		All participants	
Portuguese	**English**	**Portuguese**	**English**	**Portuguese**	**English**	**Portuguese**	**English**
Positive terms							
Ativos	Active	Atentos	Attentive	Ageis	Agile	Atentos	Attentive
Calmos	Calm	Brincalhões	Playful	Altivos	Proud	Ativos	Active
Confortáveis	Comfortable	Calmos	Calm	Atentos	Attentive	Brincalhões	Playful
Curiosos	Inquisitive	Confiantes	Confident	Ativos	Active	Calmos	Calm
Indiferentes	Indifferent	Confortáveis	Comfortable	Calmos	Calm	Confiantes	Confident
Ocupados positivamente	Positively occupied	Curiosos	Inquisitive	Com vitalidade	Lively	Confortáveis	Comfortable
Relaxados	Relaxed	Relaxados	Relaxed	Confortáveis	Comfortable	Curiosos	Inquisitive
		Sociáveis	Sociable	Corajosos	Brave	Relaxados	Relaxed
		Tranquilos	Tranquil	Curiosos	Inquisitive		
				Exploradores	Explorer		
				Relaxados	Relaxed		
				Tranquilos	Tranquil		
Negative terms							
Agitados	Agitated	Agitados	Agitated	Agitados	Agitated	Agitados	Agitated
Amedrontados	Fearful	Alertas	Vigilant	Agressivos	Aggressive	Agressivos	Aggressive
Apáticos	Apathetic	Apáticos	Apathetic	Amedrontados	Fearful	Apáticos	Apathetic
Apreensivos	Apprehensive	Assustados	Scared	Angustiados	Distressed	Apreensivos	Apprehensive
Desconfortáveis	Uncomfortable	Com medo	Fearful	Assustados	Scared	Assustados	Scared
Incomodados	Disturbed	Frustados	Frustrated	Desconfortáveis	Uncomfortable	Com medo	Fearful
		Inquietos	Restless	Desesperados	Desperate	Desesperados	Desperate
		Preocupados	Concerned	Estressados	Stressed	Com dor	Painful
				Inquietos	Restless	Frustrados	Frustrated
				Prostrados	Prostrate	Incomodados	Disturbed
				Tensos	Tense		

For intra-rater analysis, only videos that were shown twice were considered. For analysing inter-rater agreement of these videos only the ratings provided in the first exhibition were considered. The order in which videos were presented (first or second exhibition) was adjusted when analysing intra-rater agreement. Since some features presented a degree of skewness, and it is known that agreement is underestimated in such situations (Carrasco *et al* 2007), bootstrap bias corrected point estimates and confidence intervals were obtained (Efron & Tibshirani 1994; Karlsson 2009). For this purpose, a total of 5,000 simulations were performed for each analysed feature. As a general guide, the ICC reliability coefficient was considered poor when below 0.40, fair when between 0.40 and 0.59, good when between 0.60 and 0.74, and excellent when higher than 0.75 (Cicchetti 1994).

Analyses were conducted in R statistical environment, version 3.5.1 (R Core Team 2018) packages lme4 (Bates *et al* 2015) for linear mixed models, boot (Canty & Ripley 2017) for bootstrap resampling, and psych (Revelle 2017) for PCA.

Figure 1



Descriptive terms of Qualitative Behaviour Assessment for broiler chickens developed in Brazilian Portuguese. Terms are listed in alphabetical order, terms in bold were added by the researchers, dashed line indicates the valence initially considered by the experts.

Results

Development of the list of QBA descriptive terms

A total of 14 participants accepted the invitation (five male and nine female). All participants were experienced in broiler chicken production, except one PhD student in swine welfare. Experience ranged from one to 17 years. Three participants had previous experience with QBA.

In S1, participants wrote down a total of 970 terms, including repeated terms and terms not representative of expressive qualities of animal behaviour. Of these, 136 were distinctive terms, of which 88 could be considered to describe expressive qualities of behaviour (QBA terms). The median number of QBA terms generated per participant was 21 (10–30). In S2, participants chose a total of 91 distinctive terms from their lists generated in S1, of which 73 were considered proper QBA terms. When participants worked in groups in S3, group 1, 2 and 3 provided a list with 13, 17 and 23 QBA terms, respectively, divided into positive and negative valence (Table 1). In S4, eight positive and ten negative QBA terms were selected by participants. Due to different interpretations, participants could not come to an agreement as to whether painful or desperate would be the better term to describe low mood in broiler chickens. Both terms were therefore included in the final list by participants as indicative of poor chicken welfare.

This list of QBA descriptive terms as put together by the workshop participants was evaluated by the researchers, in light of their distribution across the four quadrants of emotional expression. The decision was taken to add some terms to relatively under-populated quadrants, and also in order to more clearly specify the valence of certain terms on the list, such as 'active', 'attentive', and 'calm', which can have both positive or negative meanings (eg an animal can be active or attentive in both a happy and a distressed way). Thus, the terms 'positively occupied', 'lively', 'interested', 'tranquil', 'lethargic', 'dull', and 'distressed' were added to the participants' list of 18 terms, creating a list of a total of 25 terms (Figure 1, Table 2).

Table 2 Descriptive terms in Brazilian Portuguese to assess the expressive qualities of broiler chicken behaviour, ordered to be applied using a visual analogue scale.

Language	Terms
Portuguese	Assustados, Curiosos, Com dor, Relaxados, Agressivos, Ocupados positivamente, Letárgicos, Confortáveis, Com medo, Ativos, Entediados, Confiantes, Agitados, Interessados, Apáticos, Brincalhões, Desesperados, Apreensivos, Atentos, Perturbados, Calmos, Frustrados, Com vitalidade, Incomodados, Tranquilos
English translation (provided for reference)	Scared, Inquisitive, Painful, Relaxed, Aggressive, Positively occupied, Lethargic, Comfortable, Fearful, Active, Dull, Confident, Agitated, Interested, Apathetic, Playful, Desperate, Apprehensive, Attentive, Distressed, Calm, Frustrated, Lively, Disturbed, Tranquil

Table 3 Descriptor loadings on the first four Principal Components (PC).

Terms	PC1	PC2	PC3	PC4
Scared	–0.251	0.157	–0.078	–0.275
Inquisitive	0.163	0.307	–0.133	0.164
Painful	–0.257	0.055	**–0.061**	–0.171
Relaxed	0.183	–0.190	**–0.323**	–0.246
Aggressive	–0.071	0.177	–0.090	0.250
Positively occupied	0.213	0.216	–0.083	–0.009
Lethargic	–0.136	–0.189	–0.293	0.193
Comfortable	**0.242**	–0.072	–0.249	**–0.284**
Fearful	–0.259	0.144	–0.119	–0.277
Active	0.163	**0.330**	**–0.051**	0.066
Dull	–0.115	**–0.202**	–0.316	**0.435**
Confident	0.216	0.122	–0.173	–0.131
Agitated	–0.034	**0.349**	–0.085	0.037
Interested	0.189	0.304	–0.099	0.133
Apathetic	–0.167	–0.150	**–0.361**	**0.323**
Playful	0.144	0.268	–0.109	0.112
Desperate	–0.256	0.161	–0.096	**–0.329**
Apprehensive	–0.231	0.114	–0.220	–0.033
Attentive	0.050	0.198	–0.230	0.064
Distressed	–0.246	0.135	–0.157	–0.142
Calm	0.170	**–0.251**	–0.307	–0.152
Frustrated	**–0.265**	0.043	–0.196	0.031
Lively	**0.223**	0.149	–0.177	–0.034
Disturbed	**–0.269**	0.107	–0.096	0.061
Tranquil	0.205	–0.188	–0.303	–0.205
Eigenvalue	8.5	4.4	2.6	1.3
% of variance explained	36.8	18.2	9.4	5.0
% cumulative variance explained	36.8	55.0	64.4	69.4

Highest and lowest loadings for each PC are in bold.

Testing of the list of QBA descriptive terms

PCA identified four main components with Eigenvalues greater than 1, together explaining 69.4% of the variance (Table 3). Principal Component (PC) 1 ranged from disturbed and frustrated to comfortable and lively, whereas PC2 ranged from calm and dull to agitated and active. PC3 presented no positive loadings; however, higher and lower loading terms ranged from apathetic and relaxed to active and painful. PC4 ranged from desperate and comfortable to dull and apathetic.

The inter- and intra-rater reliability for scores on PC1 and PC2 were good, while PC3 presented good intra- but not inter-rater reliability (Table 4). As to inter- and intra-rater reliability for separate QBA descriptors, five terms achieved good/excellent inter-rater reliability, eight showed fair, and 12 poor agreement. Considering the intra-rater reliability of separate terms, nine terms were classified as good, 15 as fair and one as poor (Table 5).

Discussion

Development of a list of QBA terms for broiler chickens

This study aimed to develop and test a list of descriptors in Brazilian Portuguese for the QBA of broiler chickens. This is the first study which has investigated the development of QBA terms in a non-English language, Brazilian Portuguese, with the aim of creating a QBA assessment tool that is meaningful to Brazilian animal welfare stakeholders and enables a greater level of agreement between raters than when a language foreign to these raters would be used. The study invited 24 Brazilian experts with various professional backgrounds to create and refine their own list of terms, engaging in a process of discussion in which participants could explain their point of view and hear those from their colleagues. This led to an agreed list of 18 terms, to which the project leaders added several terms in order to ensure there was sufficient balance in the distribution of terms across the four emotion quadrants (Russell & Bullock 1985).

Adding terms was especially required for quadrant Q3, in which the only term mentioned by experts was apathetic (Figure 1). This quadrant is recognised as reflecting experience indicative of low mood and energy, often leading to reduced behavioural engagement and activity. For example, the reduced behavioural repertoire of broiler chickens in the last week of life, likely caused by factors, such as barren environment, low light intensity, lameness and high stocking density (European Food Safety Authority [EFSA] 2010), can be considered as natural by people who work with fast-growing breeds. Such people have grown used to seeing the birds in this state and may not easily recognise the birds' passivity as lethargy or boredom, particularly because such states tend not to encompass easily observable overt signals, such as tail or ear movements (Burn 2017). Inactive behaviour is generally considered challenging to record during animal welfare assessment, as it could potentially indicate either positive or negative states, such as post-consummatory inactivity or inactivity due to under-stimulation, respectively (Fureix & Meagher 2015). However, QBA might facilitate judging the meaning of inactive behaviour for welfare, since it does not measure physical inactivity, but the expressive quality of such inactivity which could, for example, be tense and agitated, or alert and relaxed.

Most of the list's terms were located in quadrant Q4 (Figure 1), most likely because the experts recognised these terms as reflecting serious and readily identifiable disturbances of broiler chicken welfare, with undesirable consequences for bird health and fitness that companies want to avoid, such as birds piling on top of each other (Jones 1996), reduced feed intake and reduced resting behaviour (EFSA 2012).

Table 4 Inter- and intra-rater reliability of Qualitative Behaviour Assessment terms developed in Brazilian Portuguese for broiler chickens, analysed using intra-class correlation coefficient (ICC) for the first four Principal Components (PC).

PC	Inter-rater reliability		Intra-rater reliability	
	ICC	Confidence interval (95%)	ICC	Confidence interval (95%)
1	**0.70**	0.43–0.82	**0.73**	0.63–0.80
2	**0.65**	0.39–0.79	**0.63**	0.48–0.77
3	0.05	0.01–0.11	**0.76**	0.65–0.84
4	0.28	0.12–0.46	0.54	0.40–0.65

Good agreement is shown in bold.

Table 5 Mean (± SD) values of Qualitative Behaviour Assessment terms for broiler chickens developed in Brazilian Portuguese (English translation supplied for reference), measured by 36 undergraduates using a 125-mm visual analogue scale (VAS); inter- and intra-rater reliability of terms using intra-class correlation coefficient (ICC).

QBA term	VAS mean (± SD) (mm)	Inter-rater		Intra-rater	
		ICC	Confidence interval (95%)	ICC	Confidence interval (95%)
Desperate	13.99 (± 28.82)	**0.81**	0.59–0.89	*0.54*	0.42–0.64
Fearful	15.12 (± 26.74)	**0.70**	0.44–0.82	*0.52*	0.40–0.62
Painful	15.82 (± 29.13)	**0.68**	0.44–0.81	**0.62**	0.49–0.71
Playful	11.47 (± 22.74)	**0.66**	0.41–0.79	0.39	0.25–0.51
Scared	14.51 (± 26.09)	**0.65**	0.39–0.78	*0.53*	0.41–0.64
Inquisitive	29.45 (± 32.85)	*0.59*	0.34–0.75	*0.58*	0.46–0.67
Interested	36.46 (± 35.52)	*0.57*	0.32–0.73	0.45	0.32–0.58
Active	44.19 (± 35.75)	*0.56*	0.32–0.72	*0.54*	0.36–0.70
Distressed	20.23 (± 33.89)	*0.46*	0.23–0.64	**0.70**	0.60–0.78
Positively occupied	39.32 (± 37.65)	*0.45*	0.23–0.63	*0.58*	0.44–0.69
Frustrated	27.42 (± 36.95)	*0.44*	0.21–0.63	**0.71**	0.59–0.79
Disturbed	34.39 (± 40.24)	*0.43*	0.22–0.62	**0.72**	0.62–0.79
Agitated	38.42 (± 35.14)	*0.40*	0.19–0.59	*0.51*	0.31–0.69
Comfortable	48.23 (± 37.39)	0.35	0.16–0.54	*0.57*	0.44–0.68
Lively	50.51 (± 36.60)	0.35	0.16–0.55	**0.64**	0.50–0.73
Apprehensive	24.47 (± 32.44)	0.33	0.15–0.52	*0.59*	0.45–0.68
Calm	53.29 (± 37.27)	0.32	0.14–0.51	*0.53*	0.38–0.66
Tranquil	58.14 (± 38.32)	0.32	0.15–0.51	*0.50*	0.35–0.62
Relaxed	47.52 (± 36.67)	0.28	0.11–0.46	*0.48*	0.34–0.61
Confident	33.17 (± 31.05)	0.25	0.11–0.45	**0.65**	0.53–0.73
Dull	37.29 (± 36.84)	0.19	0.07–0.34	**0.70**	0.59–0.78
Lethargic	32.16 (± 36.74)	0.17	0.06–0.32	**0.63**	0.50–0.73
Apathetic	30.37 (± 37.02)	0.16	0.06–0.32	*0.58*	0.44–0.69
Aggressive	7.46 (± 15.10)	0.08	0.02–0.18	**0.61**	0.48–0.71
Attentive	44.14 (± 30.66)	0.06	0.01–0.14	*0.50*	0.35–0.61

Good agreement is shown in bold and fair agreement is in italics.

The Brazilian Portuguese list of terms proposed here has nine terms in common with the existing Welfare Quality® term list for broilers (Wemelsfelder *et al* 2009): active, agitated, calm, comfortable, confident, fearful, frustrated, relaxed and scared. Terms related to comfort, agitation and fear are reportedly common expressions used by professionals who work in Brazilian broiler chicken production and appear to be key terms for assessing broiler emotional expressivity (Souza & Molento 2015). In the final phase, S4, ten further Portuguese terms for broiler expression were added to the list, so that it included both local Brazilian Portuguese and more international terms for broiler welfare. The final list of terms consisted of 40% local Brazilian terms, which supports the importance of producing lists bottom-up by the people working with the animals on the ground, as opposed to translating pre-existing lists from the scientific literature. This method may provide a valuable contribution to increase the use of the QBA in countries where English is not the native language.

Investigating the reliability of the QBA term list

The QBA dimensions identified in this study align closely with the primary valence (PC1) and arousal (PC2) components of emotion proposed by Russell and Bullock (1985). Similar outcomes have been reported in QBA studies with broiler chickens (Bassler *et al* 2013; Muri *et al* 2019), dairy goats, donkeys and sheep (Phythian *et al* 2013; Grosso *et al* 2016; Minero *et al* 2016). Video sampling in the present study aimed to present observers with a variety of emotional expressions in broiler chickens, and not for showing them a representative sample of Brazilian broiler chicken farms. However, ten out of 12 video clips originated from regular commercial broiler chicken farms, and so finding these QBA dimensions within such a sample indicates that QBA can be used to differentiate between the different levels at which birds cope with the environments on those farms.

Good inter- and intra-rater concordance in scoring the 25 QBA terms was observed for the first two QBA dimensions which suggests that these terms can serve as a reliable assessment tool for broiler chickens. This is a similar finding to studies with beef cattle, broiler chickens, donkeys, pigs, sheep and veal calves (Forkman & Keeling 2009; Wemelsfelder *et al* 2009; Phythian *et al* 2013, Minero *et al* 2016, 2018; Muri & Stubsjøen 2017). In addition, our study found good intra-rater agreement for the first three QBA dimensions, indicating the raters' ability to score terms consistently in different situations, even when they lack experience with the species.

When inter- and intra-rater reliability were analysed for individual terms separately, few terms presented good agreement level, which is also in accordance with other studies (eg Grosso *et al* 2016; Minero *et al* 2016). What this indicates is that QBA is robust primarily at the level of expressive patterns that emerge from integrated analysis of the individual terms' scores, which aligns with the dynamic whole-animal principle upon which it is based (Wemelsfelder *et al* 2000, 2001). Studying individual terms is useful to optimise them and improve the reliability of the standardised list. As observed by Grosso *et al* (2016) and Muri and Stubsjøen (2017), improving reliability of individual terms will optimise the robustness of QBA dimensions.

All terms located in the 'low arousal' quadrants (Q2 and Q3; Figure 1) showed poor concordance among raters, indicating that although, as suggested above, QBA could potentially be helpful in judging the meaning of inactivity for welfare, in reality assessing the expressive qualities of inactivity still proves difficult even for experienced assessors. By contrast, terms such as desperate, fearful, painful, playful and scared had good inter-rater reliability. These terms are consistent with the basic emotions described by Panksepp and Watt (2011). Our results show that recognition of broiler chicken body expressions related to such terms was easier than the recognition of expressions related to terms describing low arousal states. Since it is important to consider the appropriateness of the descriptive terms (Meagher 2009; Fleming *et al* 2016), there is a need to further study terms in Brazilian Portuguese for poor welfare situations when the bird is vulnerable or helpless. As discussed by Muri and Stubsjøen (2017), our findings indicate that the process of including or excluding terms in a standardised QBA scoring tool is complex and depends on a combination of discussions and testing.

Some factors may have affected inter-rater reliability of the terms in this study. Raters were given descriptions of individual terms in the list in Brazilian Portuguese, which had been prepared by an experienced researcher in broiler chicken welfare. They had the opportunity to discuss these prior to assessment of video clips, but they did not have these descriptions with them during the assessment. The need for extra material to help raters was also observed by the fact that some raters marked a sign in the first page of the standardised list of terms to identify each term as positive or negative. Good intra-rater reliability was observed for nine out of 25 terms, including some terms of quadrants Q2 and Q3, such as confident, dull and lethargic. So, even if these terms were interpreted differently by the raters when observing birds, once each rater had formed an idea of the terms' meanings, they scored them consistently throughout the video session. To date, no descriptions of QBA terms have been published for broiler chickens, but further development and refinement by experts of descriptions for QBA terms for broilers may help optimise the reliability of the proposed terms list in this study (Minero *et al* 2016).

Training and the level of knowledge raters have of a species to be assessed are important factors when considering using QBA. According to Fleming *et al* (2016), raters with little experience with the animal species in question can contribute to QBA because they are encouraged to observe how an animal is behaving, and lack of experience does not seem to be a constraint to their ability to perceive animals' expressivity. Previous studies comparing groups with different levels of knowledge during the assessment of pig, dairy buffalo and dairy cattle support this rationale (Bokkers *et al* 2012; Napolitano *et al* 2012; Wemelsfelder *et al* 2012). For broiler chickens it has been argued that scoring birds using QBA terms is more difficult for non-

experienced people (De Jong *et al* 2014) and raters need to have sufficient knowledge about broiler chicken production and behaviour to obtain reliable results (Muri *et al* 2019). Our results suggest lack of experience did not prevent participants from using QBA to assess the birds in a coherent manner. However, there is a need to improve the reliability of using individual terms by training the raters to develop the observational skills required to perform robust animal welfare assessment. More specifically, it would be beneficial to invest more time in training raters to distinguish between, and correctly identify, low arousal emotions, such as indicated by descriptors in quadrants Q2 and Q3 of Figure 1 (eg tranquil and apathetic). In this study, most video clips presented birds in groups. The use of QBA at group level for broiler chickens is common practice, as specified by the Welfare Quality® protocol (2009). Difficulty of scoring broiler chickens at group level has been discussed by De Jong *et al* (2014). Birds are normally performing different types of behaviour at the same time in a poultry house, such as resting, feeding and walking. Since the group is assessed as a unit (Fleming *et al* 2016), raters are expected to observe the atmosphere in the entire group and score it accordingly. Depending on the situation, it is difficult to control exactly what the raters are observing, and they may look to different animals and different situations when observing large groups or assessing welfare at farm level (Phythian *et al* 2013; Muri & Stubsjøen 2017). As observed in this study, the terms calm, tranquil and relaxed had poor inter-rater concordance, perhaps due to difficulty in balancing these states while animals are moving around, panting and resting at the same time. In this case, the intensity of each term may be perceived differently by the raters, with consequences for inter-rater reliability. This possibility could be tested by comparing the reliability of assessing broiler chickens in groups and individually, giving an idea of any added difficulty presented by assessing chickens in groups.

Animal welfare implications and conclusion

This study reflects a first step for the application of QBA in welfare assessment of broiler chicken farms in Brazil using terms created in Brazilian Portuguese. It demonstrates the importance of producing lists bottom-up, working with people on the ground, as opposed to merely translating pre-existing lists from the scientific literature. Aligning with valence and arousal components of animal emotion, the QBA terms included in the proposed list allow comprehensive assessment of the broilers' affective states. Our results indicate the proposed term list to be reliable in assessing the expressive qualities of broiler chicken behaviour and, as such, should be readily able to be tested on-farm and by experienced raters, as well as being further developed with particular regard to terms describing poor broiler welfare. Expanding the studies to different regions in Brazil is also advisable. There seems to be a challenge in recognising emotions of low arousal in broiler chickens; thus, providing training on these specific terms seems important to improve general inter-rater reliability. Given the power to differentiate between positive and negative mood, as well as between high and low energy qualities of animal behaviour, the use of the Brazilian Portuguese QBA term list developed in this study could serve as a useful tool, adding valuable complementary information to welfare assessment in Brazilian broiler chickens.

Acknowledgements

The authors wish to thank Professor Alex Maiorka, the partner company and the farmers; the experts and the undergraduates who participated in this research; Vanessa Souza Soriano, Maria Alice Schnaider, Santiago Rucinque and Elaine Sans for video recording support; the undergraduate, Julia de Paula Soares Valente for workshop organisation, Elsa Helena Barreto for mirroring the videos and Luiza Mangerona Fukuzaki and Manuela Cosenza for data organisation. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

References

ABPA 2020 *Relatório anual 2020*. ABPA: São Paulo, Brazil. [Title translation: Annual report]. https://abpa-br.org/wp-content/uploads/2020/05/abpa_ relatorio_anual_2020_portugues_web.pdf

Bartko JJ 1966 The intraclass correlation coefficient as a measure of reliability. *Psychological Reports 19*: 3-11. https://doi.org/10.2466/pr0.1966.19.1.3

Bassler AW, Arnould C, Butterworth A, Colin L, De Jong IC, Ferrante V, Ferrari P, Haslam S, Wemelsfelder F and Blokhuis HJ 2013 Potential risk factors associated with contact dermatitis, lameness, negative emotional state, and fear of humans in broiler chicken flocks. *Poultry Science 92*: 2811-2826. https://doi.org/10.3382/ps.2013-03208

Bates D, Mächler M, Bolker B and Walker S 2015 Fitting linear mixed-effects models using lme4. *Journal of Statistical Software 67*: 1-48. https://doi.org/10.18637/jss.v067.i01

Battini M, Barbieri S, Vieira A, Can E, Stilwell G and Mattiello S 2018 The use of Qualitative Behaviour Assessment for the on-farm welfare assessment of dairy goats. *Animals 8*: 13. https://doi.org/10.3390/ani8070123

Blokhuis HJ, Veissier I, Miele M and Jones B 2010 The Welfare Quality® project and beyond: safeguarding farm animal well-being. *Acta Agriculturae Scandinavica 60*: 129-140. https://doi.org/10.1080/09064702.2010.523480

Boissy A, Manteuffel G, Jensen MB, Moe RO, Spruijt B, Keeling LJ, Winckler C, Forkman B, Dimitrov I, Langbein J, Bakken M, Veissier I and Aubert A 2007 Assessment of positive emotions in animals to improve their welfare. *Physiology & behavior 92*: 375-397. https://doi.org/10.1016/j.physbeh.2007.02.003

Bokkers EAM, De Vries M, Antonissen ICMA and De Boer IJM 2012 Inter- and intra-observer reliability of experienced and inexperienced observers for the Qualitative Behaviour Assessment in dairy cattle. *Animal Welfare 21*: 307-318. https://doi.org/10.7120/09627286.21.3.307

de Boyer des Roches A, Lussert A, Faure M, Herry V, Rainard P, Durand D, Wemelsfelder F and Foucras G 2018 Dairy cows under experimentally induced *Escherichia coli* mastitis show negative emotional states assessed through Qualitative Behaviour Assessment. *Applied Animal Behaviour Science 206*: 1-11. https://doi.org/10.1016/j.applanim.2018.06.004

Buijs S, Ampe B and Tuyttens FAM 2016 Sensitivity of the Welfare Quality® broiler chicken protocol to differences between intensively reared indoor flocks: which factors explain overall classification? *Animal 11*: 244-253. https://doi.org/10.1017/S1751731116001476

Burn CC 2017 Bestial boredom: a biological perspective on animal boredom and suggestions for its scientific investigation. *Animal Behaviour 130*: 141-151. https://doi.org/10.1016/j.anbehav.2017.06.006

Canty A and Ripley B 2017 *boot: Bootstrap R (S-Plus) Functions*. https://cran.r-project.org/web/packages/boot/index.html

Carrasco JL, Jover L, King TS and Chinchilli VM 2007 Comparison of concordance correlation coefficient estimating approaches with skewed data. *Journal of Biopharmaceutical Statistics 17*: 673-684. https://doi.org/10.1080/10543400701329463

Cicchetti DV 1994 Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment 6*: 284-290. https://doi.org/10.1037/1040-3590.6.4.284

De Jong IC, Gunnink H and Hindle V 2014 *Implementation of the Welfare Quality® broiler assessment protocol – final report. Livestock Report 833*. Wageningen UR Livestock Research: Wageningen, The Netherlands. https://edepot.wur.nl/339412

De Jong IC, Hindle VA, Butterworth A, Engel B, Ferrari P, Gunnink H, Perez Moya T, Tuyttens FAM and van Reenen CG 2015 Simplifying the Welfare Quality® assessment protocol for broiler chicken welfare. *Animal: An International Journal of Animal Bioscience 10*: 117-127. https://doi.org/10.1017/S1751731115001706

Efron B and Tibshirani RJ 1994 *An Introduction to the Bootstrap*. CRC Press: Florida, USA. https://doi.org/10.1201/9780429246593

EFSA 2010 Scientific Opinion on the influence of genetic parameters on the welfare and the resistance to stress of commercial broilers. *EFSA Journal 8*: 1-82. https://doi.org/10.2903/j.efsa.2010.1666

EFSA 2012 Scientific report updating the EFSA opinions on the welfare of broilers and broiler breeders. *EFSA Supporting Publication 9*: 116. https://doi.org/10.2903/sp.efsa.2012.EN-295

Federici JF, Vanderhasselt R, Sans ECO, Tuyttens FAM, Souza APO and Molento CFM 2016 Assessment of broiler chicken welfare in Southern Brazil. *Brazilian Journal of Poultry Science 18*: 133-140. https://doi.org/10.1590/18069061-2015-0022

Fleming PA, Clarke T, Wickham SL, Stockman CA, Barnes AL, Collins T and Miller DW 2016 The contribution of qualitative behavioural assessment to appraisal of livestock welfare. *Animal Production Science 56*: 1569-1578. https://doi.org/10.1071/AN15101

Forkman B and Keeling L 2009 *Assessment of animal welfare measures for dairy cattle, beef bulls and veal calves*, *Welfare Quality® reports n11*. Cardiff University: Cardiff, UK

Franklin SB, Gibson DJ, Robertson PA, Pohlmann JT and Fralish JS 1995 Parallel analysis: a method for determining significant principal components. *Journal of Vegetation Science 6*: 99-106. https://doi.org/10.2307/3236261

Fureix C and Meagher RK 2015 What can inactivity (in its various forms) reveal about affective states in non-human animals? A review. *Applied Animal Behaviour Science 171*: 8-24. https://doi.org/10.1016/j.applanim.2015.08.036

Grosso L, Battini M, Wemelsfelder F, Barbieri S, Minero M, Dalla Costa E and Mattiello S 2016 On-farm Qualitative Behaviour Assessment of dairy goats in different housing conditions. *Applied Animal Behaviour Science 180*: 51-57. https://doi.org/10.1016/j.applanim.2016.04.013

IBGE 2020 *Pesquisa trimestral do abate de animais*. [Title translation: Animal slaughter database]. https://sidra.ibge.gov.br/tabela/1094

Johnson RA and Wichern DW 2007 *Applied Multivariate Statistical Analysis, Sixth Edition*. Pearson Education: New Jersey, USA

Jones RB 1996 Fear and adaptability in poultry: insights, implications and imperatives. *World's Poultry Science Journal 52*: 131-174. https://doi.org/10.1079/WPS19960013

Karlsson A 2009 Bootstrap methods for bias correction and confidence interval estimation for nonlinear quantile regression of longitudinal data. *Journal of Statistical Computation and Simulation 79*: 1205-1218. https://doi.org/10.1080/00949650802221180

Marino L 2017 Thinking chickens: a review of cognition, emotion, and behavior in the domestic chicken. *Animal Cognition 20*: 127-147. https://doi.org/10.1007/s10071-016-1064-4

McCulloch CE and Searle SR 2004 *Generalised, Linear, and Mixed Models*. John Wiley & Sons: London, UK

Meagher RK 2009 Observer ratings: Validity and value as a tool for animal welfare research. *Applied Animal Behaviour Science 119*: 1-14. https://doi.org/10.1016/j.applanim.2009.02.026

Mendl M, Oliver HP and Paul ES 2010 An integrative and functional framework for the study of animal emotion and mood. *Proceedings of the Royal Society B*: 2895-2904. https://doi.org/10.1098/rspb.2010.0303

Minero M, Dalla E, Dai F, Anne L, Murray M, Canali E and Wemelsfelder F 2016 Use of Qualitative Behaviour Assessment as an indicator of welfare in donkeys. *Applied Animal Behaviour Science 174*: 147-153. https://doi.org/10.1016/j.applanim.2015.10.010

Minero M, Dalla E, Dai F, Canali E, Barbieri S, Zanella A, Pascuzzo R and Wemelsfelder F 2018 Using qualitative behaviour assessment (QBA) to explore the emotional state of horses and its association with human-animal relationship. *Applied Animal Behaviour Science 204*: 53-59. https://doi.org/10.1016/j.applanim.2018.04.008

Muri K, Marie S, Vasdal G, Oppermann R and Georg E 2019 Associations between qualitative behaviour assessments and measures of leg health , fear and mortality in Norwegian broiler chicken flocks. *Applied Animal Behaviour Science 211*: 47-53. https://doi.org/10.1016/j.applanim.2018.12.010

Muri K and Stubsjøen SM 2017 Inter-observer reliability of Qualitative Behavioural Assessments (QBA) of housed sheep in Norway using fixed lists of descriptors. *Animal Welfare 26*: 427-435. https://doi.org/10.7120/09627286.26.4.427

Napolitano F, De Rosa G, Grasso F and Wemelsfelder F 2012 Qualitative behaviour assessment of dairy buffaloes (*Bubalus bubalis*). *Applied Animal Behaviour Science 141*: 91-100. https://doi.org/10.1016/j.applanim.2012.08.002

Panksepp J and Watt D 2011 What is basic about basic emotions? Lasting lessons from affective neuroscience. *Emotion Review 3*: 1-10. https://doi.org/10.1177/1754073911410741

Phythian C, Michalopoulou E, Duncan J and Wemelsfelder F 2013 Inter-observer reliability of Qualitative Behavioural Assessments of sheep. *Applied Animal Behaviour Science 144*: 73-79. https://doi.org/10.1016/j.applanim.2012.11.011

R Core Team 2018 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria

Revelle W 2017 *Psych: Procedures for Personality and Psychological Research*. Northwestern University: Illinois, USA

Russell JA and Bullock M 1985 Multidimensional scaling of emotional facial expressions: Similarity from pre-schoolers to adults. *Journal of Personality and Social Psychology 48*: 1290-1298. https://doi.org/10.1037/0022-3514.48.5.1290

Sans ECO, Federici JF, Dahlke F and Molento CFM 2014 Avaliação de grau de bem-estar de frango de corte tipo caipira pelo Welfare Quality®. *Revista Brasileira de Ciencia Avicola 16*: 297-306. https://doi.org/10.1590/1516-635x1603297-306. [Title translation: Evaluation of free-range broilers using the Welfare Quality® protocol]

Souza APO and Molento CFM 2015 Good agricultural practices in broiler chicken production in the state of Paraná: focus on animal welfare. *Ciência Rural 45*: 2239-2244. https://doi.org/10.1590/0103-8478cr20141877

Souza APO, Sans ECO, Müller BR and Molento CFM 2015 Broiler chicken welfare assessment in GLOBALGAP certified and non- certified farms in Brazil. *Animal Welfare 24*: 45-54. https://doi.org/10.7120/09627286.24.1.045

Tuyttens FAM, Federici JF, Vanderhasselt RF, Goethals K, Duchateau L, Sans ECO and Molento CFM 2015 Assessment of welfare of Brazilian and Belgian broiler flocks using the Welfare Quality® protocol. *Poultry Science 94*: 1758-1766. https://doi.org/10.3382/ps/pev167

Wang PY 2004 *Investigating the validity of qualitative assessments of behaviour using a Free Choice Profiling approach in chickens*. University of Edinburgh: Edinburgh, UK

Welfare Quality® 2009 *Welfare Quality® Assessment protocol for poultry (broilers, laying hens)* p 116. Welfare Quality® Consortium: Lelystad, The Netherlands

Wemelsfelder F, Hunter AE and Lawrence AB 2012 Assessing pig body language: Agreement and consistency between pig farmers, veterinarians, and animal activists. *Journal of Animal Science 90*: 3652-3665. https://doi.org/10.2527/jas2011-4691

Wemelsfelder F, Hunter EA, Mendl MT and Lawrence AB 2000 The spontaneous qualitative assessment of behavioural expressions in pigs: First explorations of a novel methodology for integrative animal welfare measurement. *Applied Animal Behaviour Science 67*: 193-215. https://doi.org/10.1016/S0168-1591(99)00093-3

Wemelsfelder F, Hunter TEA, Mendl MT and Lawrence AB 2001 Assessing the 'whole animal': a free choice profiling approach. *Animal Behaviour 62*: 209-220. https://doi.org/10.1006/anbe.2001.1741

Wemelsfelder F, Knierim U, Lentfer T, Staack M and Sandilands V 2009 Qualitative behaviour assessment. In: Forkman B and Keeling L (eds) *Welfare Quality® Reports N9 - Assessment of Animal Welfare Measures for Layers and Broilers* pp 113-119. Welfare Quality® Consortium: Uppsala, Sweden