Possible improvement of measures within the principle 'appropriate behaviour' of the Welfare Quality® boiler assessment protocol

Ingrid de Jong, Johan van Riel, Tosca Hoevenaar, Thea van Niekerk

Report 1192

WAGENINGEN
UNIVERSITY & RESEARCH

Possible improvement of measures within the principle ‘appropriate behaviour’ of the Welfare Quality® boiler assessment protocol

Ingrid de Jong, Johan van Riel, Tosca Hoevenaar, Thea van Niekerk

This research was conducted by Wageningen Livestock Research, commissioned and funded by the Welfare Quality Network

Wageningen Livestock Research
Wageningen, September 2019

Report 1192

De Jong, I.C., Van Riel, J., Hoevenaar, T., Van Niekerk, T. 2019. *Possible improvement of measures within the principle 'appropriate behaviour' of the Welfare Quality® boiler assessment protocol.* Wageningen Livestock Research, Report 1192

Summary The aim of the current study was two-fold. First, we tested whether or not a scan sampling technique can be used as a relatively quick, simple scoring of the behaviour in commercial-size broiler chicken flocks, and could thus be included as a measure of appropriate behaviour in the Welfare Quality® broiler assessment protocol. Four existing datasets were analysed, these were collected for different purposes but were based on the same sampling technique, i.e. counting the number of broilers engaged in different behaviours using direct observations, at different time intervals and at different locations in the house. Analysis showed that various factors affected the outcome. Whereas those such as observer, time of the day and location could be taken into account when designing an observation protocol (e.g. by training, by selection of different observation locations and times), it must first be determined how to take into account the effect of subsequent scans, the apparent difference between flocks and genetic strains, and the effects of different behaviours. This is important when e.g. thresholds need to be set or scores reliably assigned regarding the prevalence of birds showing certain behaviours during a farm visit. Second, it was tested whether alternative measures (rather than distance to observer or a novel object) could be used as indicators of fear in broiler chickens. Previous research suggested a possible relationship between walking ability and the number of birds within arm's reach in the touch test, thus implying that the touch test may be confounded by impaired walking ability and might thus be a suboptimal method of assessing fear of humans in broilers. The present study was carried out at an experimental farm and reported in detail in a student thesis. An extended summary is included in this report. The results showed that both distance to observer or novel object and behaviour of the bird (alert body posture, neck posture, body position towards object or observer) seemed to be valid indicators of fear, however, when a relatively short time is available for testing, the birds' behaviour should always be measured as older broilers need more time to move away.

This report can be downloaded for free at https://doi.org/10.18174/499603 or at www.wur.nl/livestock-research (under Wageningen Livestock Research publications).

P.O. Box 338, 6700 AH Wageningen, The Netherlands, T +31 (0)317 48 39 53,
E info.livestockresearch@wur.nl, www.wur.nl/livestock-research. Wageningen Livestock Research is part of Wageningen University & Research.

Wageningen Livestock Research Report 1192

Table of contents

Acknowledgements

Dr Gerrit Gort, Biometris, Wageningen University and Research, is greatly acknowledged for performing part of the statistical analyses and dr Jorine Rommers, Wageningen University and Research, is acknowledged for preparing the data sets. Tosca Hoevenaar, Malou van der Sluis, Bas Rodenburg and Britt de Klerk are acknowledged for their work to improve the fear tests. Brian Jones is acknowledged for reviewing the draft report.

This study was financially supported by the Welfare Quality Network.

Summary

The current measures under the principle 'appropriate behaviour' in the Welfare Quality® (WQ) broiler assessment protocol are criticised as possibly being invalid and very limited. Therefore, it was proposed to add a simple assessment of broiler behaviour as a measure to the existing WQ broiler assessment protocol. The current study aimed to test whether or not the scan sampling technique can be used as a relatively quick and simple method of scoring the behaviour in commercial-size broiler chicken flocks, based on existing data sets. If the results were positive further steps could be taken to develop this method into a measure that can be included under principle 4, appropriate behaviour. Datasets from previous experiments in which behaviour was observed in broiler flocks at commercial farms were used. These datasets consisted of the same type of behavioural observations, i.e. scan sampling of group behaviour by counting the number of broiler chickens engaged in different behaviours, in series of scans at fixed time intervals (further referred to as timereps) and at different locations in the broiler house. These datasets were analysed in order to determine whether or not reducing the number of scans per location and/or the number of locations observed can still predict the outcome of the total set of observations. If so, a reduced set of observations could be used in the WQ broiler assessment protocol. In addition we determined the effects of observer, time of day, breed (fast vs slow growing broilers) and location in the house on the outcomes.

All four datasets contained results of observations during 1-5 days before depopulation; this is the time frame in which the Welfare Quality® assessment protocol is carried out. Some also included observations at earlier ages. The following behaviours were analysed: 'total active', foraging, dustbathing, comfort behaviour or a combination of dustbathing and foraging behaviour. The trait 'total active' was originally composed of all active behaviours (excluding feeding and drinking), and in addition analyses were performed including a reduced set of active behaviours. As the datasets were collected in studies with differing aims, there were differences between datasets (e.g. in the total number of scans per timerep) so each dataset was analysed separately. Using logistic regression analysis, the effect of individual scans was determined as well as the optimum number of scans. Further, effects of age, time of the day, observer and location in the house were determined, for one or more datasets (dependent on the variable recorded per dataset).

To assess behaviour measurement with the scan sampling technique the observers usually observed one location for various periods (minutes). During that time the number of birds performing a behaviour is scored several times; this is referred to as the number of scans. Dataset 1, with fast growing birds, showed a large scan-effect on all behaviours, indicating that the proportion of each behaviour increased with increasing scan number. Dataset 3, also with fast growing birds showed a similar significant scan effect for 'Active' behaviour and combined 'Dustbathing-comfort', but not for foraging. For slow growing birds (dataset 2 and 4) the scan effect was not seen for most behaviours. Different compositions (including different behaviours) of the trait 'active' were tested, to determine whether scan effects could be reduced if different behaviours were included. It appeared that a maximum of 4 scans for 'active', and active consisting of 'standing, walking, running, foraging' or 'standing, walking running' would be the best option for both fast and slower growing broilers, although especially for slower growing strains it would be good to further test this in commercial flocks.

With increasing age, the proportion of broilers showing the different behaviours decreased ($P<0.05$ at least), although this decrease was not always significant for comfort and dustbathing. However, due to the very low frequency of these behaviours, analysis was less reliable. A significant effect of timerep (time of the day) ($P<0.05$) was found for active, dustbathing and foraging in the different datasets. The proportion of broilers active reduced with increasing timereps, (thus, between morning and afternoon), whereas dusbathing-comfort seemed to peak around midday. Effects differed between datasets, due to different observation times. One dataset was suited to analysis of the effect of location in the house, and this showed a highly significant effect of location, with broilers close to the

walls being less active than in the central area (P<0.001). Another dataset revealed observer effects on all behaviours (P<0.05) and particularly on dustbathing, comfort and foraging behaviour.

In conclusion, analysis of existing datasets with scan sampling data of direct observations in commercial broiler flocks showed that various factors affected the outcome. Whereas factors such as observer, time of the day and location in the house could be taken into account when designing an observation protocol (e.g. by training, by selection of different observation locations), it must first be determined further how to take into account the effects of subsequent scans and of the apparent differences between flocks and genetic strain. This is important when setting thresholds for a different scoring system for the prevalence of birds showing certain behaviours, as in that case a reliable picture of the actual prevalence of a behaviour is critical.

1 The scan sampling technique as a possible method of measuring broiler behaviour

1.1 Introduction

The European Welfare Quality® project developed a standardized method of assessing the welfare of pigs, poultry (laying hens and broilers) and cattle (Blokhuis et al., 2013). One of the innovations of the Welfare Quality® (WQ) animal welfare assessment system was that it focused on animal-based measures (e.g. directly related to animal body condition, health aspects, injuries, behaviour, etc.) instead of only resource- or management-based measures. A common, standardized approach across animal species was chosen, based on 4 welfare principles: good feeding, good housing, good health and appropriate behaviour. These four principles then were subdivided into various criteria, again equal for all species. The actual measurements per criteria are species specific, but also have a common, standardized approach with similar measurements where possible. For instance, stocking density and mortality are measured in each protocol although thresholds for acceptable levels may differ per species. After the release of the first editions of the Welfare Quality® protocols the protocols have been applied to many herds and flocks and the collected data and experiences have led to new insights and ideas for further improvement of the protocols. This was foreseen in the first edition which stated: "It is important to remember that research is continuing to identify new and better measures and that Welfare Quality® protocols will be updated in the light of new knowledge" (Welfare Quality, 2009).

Experiences with the broiler assessment protocol led to the suggestion that the principle 'appropriate behaviour' for broiler chickens consisted of a rather limited set of measures that can suffer a lack of validity. Firstly for example, the validity of the Qualitative Behaviour Assessment has been studied for a number of species, (e.g., Bokkers et al., 2012; Phythian et al., 2016; Czycholl et al., 2017) but not for broiler chickens. Secondly, the touch test score as a measure of fear of humans was claimed to be related to lameness of the birds (Vasdal et al., 2018a). Thirdly, researchers would prefer animal-based measures rather than the current resource-based measure 'cover on range' (de Jong et al., 2015). Finally, the protocol has no measure of social behaviour in broiler chickens.

As 'appropriate behaviour' is an important aspect of animal welfare, the underlying criteria should have sound and valid measurements and cover the various aspects of 'appropriate behaviour'. Therefore, the 'broiler working group' (WG) of the Welfare Quality Network (WQN) proposed including measures of species-specific behaviours, to better substantiate the 'appropriate behaviour' principle (de Jong et al., 2015). They suggested including measures of behaviours considered essential like foraging, dust bathing, preening and exploration, and which also reflect the extent to which production systems meet the needs of broilers to perform their natural behaviours. The WG therefore proposed to study whether or not these behaviours could be observed and scored during a farm visit. After some discussion in the WQN it was determined that the proportion of broilers showing these behaviours, the time spent on each one and the quality of the behaviour (i.e., whether it can be fully performed) should all be included to correctly evaluate to what extent housing systems meet the behavioural needs of the birds. Alternatively, a series of simple behavioural observations could be added to the broiler assessment protocol to improve the measures under the principle 'appropriate behaviour'.

New measurements for the WQ-protocol in general should meet some requirements:

1. They should be simple and straight forward, so that they can be carried out by non-scientists;
2. The measurements should be reliable and robust, meaning that they should have a high repeatability and inter-observer reliability;
3. There should be clear cut-off points for the various scoring classes, so scoring can be done with a high repeatability;

4. Preferably the measurements should be animal based, as this is the most direct way to assess bird welfare;
5. Preferably the scoring should not take too long, as the current protocols are criticised because of their long duration (de Jong et al., 2016).

Various techniques can be used to asses behaviour in broilers. A robust technique in commercial flocks seems to be scan sampling (e.g., Bailie et al., 2018b; De Jong and Gunnink, 2019; Vasdal et al., 2018b). For this technique observers usually observe one location for apredetermined period either by direct or video observations. During this period the number of birds performing a behaviour is scored several times. Depending on the complexity of the environment this procedure can be repeated at various spots in the house. Also the assessment can be repeated several times per day. Although the assessment can be carried out by one observer, in complex situations more observers can be used.

Our aim was to test whether the scan sampling technique can be used as a relatively quick and simple scoring of the behaviour in commercial-size broiler chicken flocks, based on existing data sets. These data sets were explored with the following aims:

1. How robust are the behavioural measurements (repeatability and inter-observer reliability)?
2. How many scans are needed to get a reliable outcome?
3. Are the measurements applicable for all type of birds (fast and slow growing broiler flocks)?
4. What other aspects should be taken into account that could affect the results?

If a limited number of scans can provide a reliable picture of broiler flock behaviour, further steps can be taken to include such a measure in the WQ broiler assessment protocol.

The following behaviours were selected: foraging (scratching/pecking at the litter), dustbathing, explorative pecking at enrichment objects, perching, preening and/or the total level of active behaviour. The latter measure was included because apathy or inactivity is a welfare concern in broiler chickens (EFSA, 2010) and because management conditions that are considered to stimulate natural behaviour and positive emotions, such as environmental enrichment (including outdoor range or veranda) and the provision of natural light are often reported to increase the activity level of the broilers in general (Riber et al., 2018). The other behaviours were selected because their expression is considered essential for chickens, e.g. maintaining plumage in good conditions (preening, dustbathing), searching for food (foraging) and to acquiring information about the environment (explorative pecking).

1.2 Material and methods

1.2.1 Data selection and analysed behaviours

In the present study, datasets from previous experiments in which behaviour was observed in broiler flocks at commercial farms were used. These datasets consisted of the same type of behavioural observations, i.e. scan sampling by counting the number of broiler chickens engaged in different behaviours, in series of scans at fixed time intervals (further referred to as timereps) and at different locations in the broiler house. Our analysis aimed to determine if a reduced set of observations (i.e., fewer scans per location and/or fewer locations) is still predictive of the outcome of the total set of observations; if so fewer observations can be used in the WQ broiler assessment protocol. Potential treatment effects were not interested studies here. Four datasets were selected, because of the following reasons: conventional fast growing broiler flocks and slower growing broiler flocks (2 datasets per broiler type), different observers included (1 dataset), different times of the day included (2 datasets), exact location in the house registered (one dataset), and at least four successive scans per location present (all datasets). All observations were made during a period of 1-5 days before depopulation since this is the time frame in which the Welfare Quality® assessment protocol is carried out (Welfare Quality, 2009).

The following behaviours were analysed: 'total active', foraging, dustbathing, comfort behaviour or a combination of dustbathing and foraging behaviour. The trait 'total active' was originally composed of all active behaviours (excluding feeding and drinking), but different compositions of these traits were studied, see table 2.1 for definitions. The other behaviours were defined as; foraging: pecking and/or scratching the litter; dustbathing: all elements of dustbathing behaviour according to Van Liere (1991); comfort behaviour: preening, wing flapping, wing stretching, feather ruffling or feather shaking (see (De Jong and Gunnink, 2019) for the ethogram of all behaviours).

1.2.2 Statistical analysis

Because of the varying conditions and research questions behavioural sampling was not always done in exactly the same way for all datasets, i.e. the number of subsequent scans on one location and the ethograms could (slightly) differ (e.g., sometimes dustbathing and comfort behaviour were merged into one category). Therefore the datasets were analysed separately. Although the exact behaviours differed per dataset, all datasets were analysed using the same 3 models.

1.2.2.1 Analysis of 'active behaviour'

For all datasets, an analysis was done on the fraction "active behaviour". As the ethogram used in the various studies differed, the fraction "active behaviour" was not always composed of exactly the same behaviours. We analysed various compositions of the "active" trait (see Appendix 1 for full table with composition per dataset). Active 0 contains all active behaviours, Active 1 - 4 are composed by subsequent omission of one or more behaviours.

Table 1 *Various compositions of the trait "active". Study specific features refer to specific aspects of a study, such as busy with enrichments. For the ethogram, we refer to (De Jong and Gunnink, 2018).*

	Composition of the trait "Active"
Active 0	Study specific features
	Walking/standing/running
	Foraging
	Comfort/dustbathing
	Disturbance
	Aggression/feather pecking
	Other behaviour
Active 1	Walking/standing/running
	Foraging
	Comfort/dustbathing
	Disturbance
Active 2	Walking/standing/running
	Foraging
	Comfort/dustbathing
Active 3	Walking/standing/running
	Foraging
Active 4	Walking/standing/running

In addition, the behaviours "foraging", "dustbathing", "comfort behaviour" and the combined "dustbathing and comfort behaviour (dust-comf)" were analysed separately. All behaviours were analysed as dependent variables and expressed as proportion of birds performing a specific behaviour.

We used logistic regression with the binomial distribution for the counts (of a specific behaviour), using the total count of behaviours as binomial totals. Multiple random effects entered the model, depending on the type of study, to model the effects of all combinations of housing, number and age/time (of the day) of measurement (HT-effects in model equation 1,2 and 3). Because extra-binomial variation could not be ruled out, we added an observation-level random effect (corresponding to location, nested within other factors).

In model 2 we analysed the original counts per scan, introducing a linear trend for scan (as fixed effect), and an extra random effect for scan (which now is at the observation level). Model 3 was the same as model 2, but scan was added as a class variable, to estimate the effect for each scan (to check non-linearity of the trend for scan). Binomial variation was corrected for over-dispersion, $\text{variance}(Y \mid p) = \varphi np(1-p)$. In the models the relation between the chance of p ($0 < p < 1$) on the presence of a certain type of behaviour of the observed birds and the explanatory variables are described, using the logit-link function.

$$Logit(p) = Ln(p/1-p) = \alpha_{fixed\_effects} + \varepsilon_{HT} + \varepsilon_{Location:HT}$$
model equation 1

$$Logit(p) = Ln(p/1-p) = \alpha_{fixed\_effects} + \beta_1 * X(scan) + \varepsilon_{HT} + \varepsilon_{Location:HT} + \varepsilon_{Scan:Location:HT}$$
model equation 2

$$Logit(p) = Ln(p/1-p) = \alpha_{fixed\_effects} + \lambda_{scan\_i} + \varepsilon_{HT} + \varepsilon_{Location:HT} + \varepsilon_{Scan:Location:HT}$$
model equation 3

Fixed effects in the models were dependent on the specific study, e.g. treatment (of a specific study), age, time of the day and interactions.

As the analysis showed a scan effect, with consecutive scans showing an increase in behavioural frequency in some behaviours and a decrease in others, we tried to minimize this effect by finding the optimum number of scans. By removing the last scans one by one we studied the effect of reduction of the number of scans on the variance components for the different sources of variation. We also report average standard errors of the means (for the treatment groups which are relevant here, such as location or time of the day).

1.2.3 Description of datasets and list of analyses per dataset

1.2.3.1 **Dataset 1 – fast growing broilers**
This dataset contained records of conventional fast growing broilers (one genetic strain). Two treatments were applied, but for the current study the number of repeated scans per location and the observer effect are specifically interesting:
- 7 farms
- 2 flocks per farm (6 farms) and 4 flocks per farm (1 farm)
- 2 production cycles
- 2 or 4 houses per farm
- 2 ages per flock
- 2 treatments
- 3-5 observation days per farm
- 6 observations locations per house
- 8 scans per location
- 8 observers

<u>Analysed behaviours:</u>
- Active = walking + standing + litter pecking + preening/comfort behaviour + dustbathing + aggressive behaviour + disturbed + feather pecking + other
- Active 0-4: see Appendix 1
- Foraging
- Dustbathing
- Comfort behaviour (comfortbehav)
- Combined dustbathing and comfort behaviour (dustcomf)

1.2.3.2 **Dataset 2 – slower growing broilers**
This dataset contained records of commercial flocks of slower growing broilers (all flocks in the same housing system). For this study variation caused by observation time (on the day) (time replicates) is specifically interesting:
- 10 farms, each farm observed on a separate day
- 1 house per farm observed
- 3 time replications (timereps) per day (=farm=stable), corresponding to time of day (morning, around midday, afternoon)
- 8 observation locations per house
- 4 scans per location
- no treatments applied

Analysed behaviours:

- Active = on enrichment + close to enrichment + busy with enrichment + dustbathing + comfort behaviour + foraging + disturbed during resting + walking/running/flapping + other
- Active 0-4: see Appendix 1
- Foraging
- Dustbathing
- Comfort behaviour (comfortbehav)
- Combined dustbathing and comfort behaviour (dustcomf)

1.2.3.3 **Dataset 3 – fast growing broilers**

This dataset contained behavioural observations of one farm with fast growing broilers. There were 3 treatments (each treatment in one house), but for this report the observation time (morning or afternoon) and the location in the stable (close to wall or centre area (area between feeder and drinker lines)) are interesting:

- single farm
- 5 production cycles
- 10 observation days (2 per cycle)
- 2 observation ages per flock/house
- 3 houses per observation day; observation date and stable are confounded factors
- 3 treatments
– 2 timereps per stable per day (morning and late afternoon)
- 6 locations per stable per day (2 close to walls, 4 in the central area between feeders and drinkers)
– 5 scans per location

Analysed behaviours:

- Active = running/walking + standing + scratching + dustbathing/comfort behaviour + aggressive behaviour + disturbed + other
- Active 0-4: see Appendix 1
- Foraging
- Combined dustbathing and comfort behaviour (dustcomf)

1.2.3.4 **Dataset 4 – slower growing broilers**

Dataset 4 contains behavioural observations of one farm with slower growing broilers. There were no treatments, but 4 locations in the stable. For this report the number of scans is an interesting factor:

- single farm
- 2 production cycles
- 8 dates = observation days (every 14 days in each production cycle)
- 4 locations per stable per day
– 4 timereps per day (two observations in the morning, two in the afternoon)
– 4 scans per location
- Additional location information

Analysed behaviours:

- Active = running/walking + standing + scratching + dustbathing/comfort behaviour + aggressive behaviour + disturbed + other
- Active 0-4: see Appendix 1
- Foraging
- Combined dustbathing and comfort behaviour (dustcomf)

1.3 Results

1.3.1 Age of the birds

In general behaviours were influenced by age. Dataset 1 showed that younger birds were significantly more active and showed more foraging and dustbathing behaviour than older birds (table 2). This was confirmed in dataset 3 and 4 for active behaviour. However, in general the fractions for 'foraging' and 'dustbathing' were very low, indicating that the analysis was less reliable than for the other behaviours.

Table 2 *Effect of age on various behaviours. The estimate indicates the difference between the oldest age and the youngest, i.e. a negative estimate means that the number of broilers showing that particular behaviour decreased with increasing age.*

DATASET 1:

	Estimate	SE	P
Active	-0.749	0.085	<0.001
Foraging	-1.227	0.144	<0.001
Dustbathing	-0.479	0.275	<0.10
Comfort behaviour	-0.283	0.090	<0.010
Dust-comf	-0.376	0.080	<0.001

DATASET 3:

	Estimate	SE	P
Active	-0.585	0.205	0.004
Foraging	-0.756	0.326	0.020
Dust-comf	-0.162	0.111	0.143

DATASET 4:

	w2-w4			w2-w6			w2-w7		
	Estimate	SE	P	Estimate	SE	P	Estimate	SE	P
Active	0.396	0.223	<0.10	-0.109	0.216	<0.001	-1.459	0.215	<0.001
Foraging	0.108	0.326	n.s.	-1.334	0.331	<0.001	-1.893	0.335	<0.001
Dust-comf	0.101	0.350	n.s.	0.626	0.348	<0.10	0.514	0.348	n.s.

As the fractions dustbathing and comfort behaviour were very low in all datasets a combined analysis was carried out, but the combined fractions dustbathing-comfort behaviour were still low (overall mean 0.035, 0.033, 0.044 and 0.077 resp. in dataset 1-4). In the first dataset (fast growing birds) there was an age effect, where older birds performed less dustbathing-comfort behaviour. In the 3rd and 4th dataset (slower growing birds) there was no significant age effect for dustbathing-comfort. The 2nd dataset could not be analysed on this aspect (only one age included in the dataset).

1.3.2 Time of the day

From both the second and third dataset it appeared that 'active' was influenced by time of the day (timerep). In the second dataset three observations were done per day. There was a significant time effect (table 3), showing that the broilers became less active during day. In the third dataset one observation was done in the morning and one in the afternoon. Again the birds were significantly less active in the afternoon than in the morning.

Table 3 *Analysis of dataset 2 and 3 for fraction 'Active behaviour': the estimate indicates the effect of timerep, and a negative estimate means a decreased proportion of birds active with increasing timerep number (from morning to afternoon) (fixed effects).*

DATASET 2
Fixed effects:

	Estimate	SE	P
(Intercept)	-0.595	0.236	0.012
timerep2	-0.495	0.152	0.001
timerep3	-0.881	0.152	< 0.001
Scan	-0.010	0.0217	0.655

DATASET 3
Fixed effects:

	Estimate	SE	P
(Intercept)	-1.644	0.168	< 0.001
Timerep	0.216	0.077	0.005
Scan	0.084	0.012	< 0.001

The second dataset showed no significant difference in dustbathing frequency over the day, but the fraction dustbathing was very low with an overall mean of 0.003. The 3rd and 4^{th} dataset could not be analysed on this aspect. For the slow growing broiler flocks from the second dataset there was a slight difference in comfort behaviour frequency over day, where the fraction scored in the second timerep was higher than in the first and third timerep (P=0.0446). For the second dataset with slower growing broilers the combined fractions dustbathing-comfort behaviour differed slightly over time, with the second timerep showing slightly higher fractions than the first and third timerep (p=0.0195). In the third dataset significant differences were found between observation periods, with fractions being higher in the afternoon compared to the morning (p<0.001). The 4^{th} dataset also showed higher combined fractions dustbathing-comfort behaviour in the afternoon. Especially the first timerep (second observation moment) was lower than the other three timereps (Fig. 1).

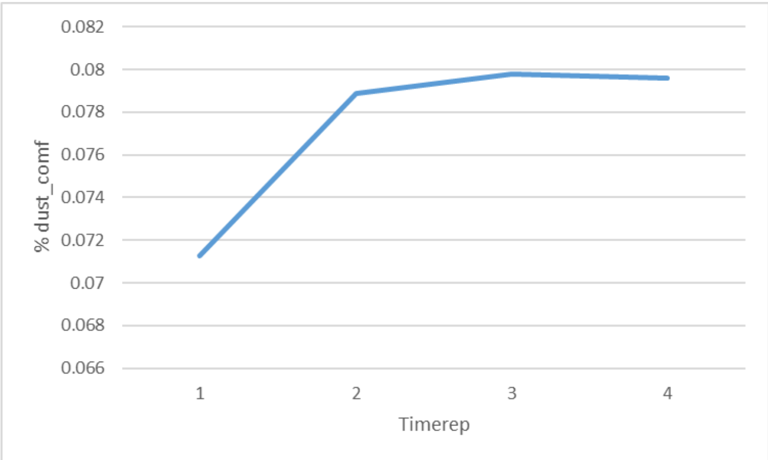

Figure 1 *Combined fractions dustbathing-comfort behaviour (% dust_comf) for 4 timereps in dataset 4 (slower growing broiler flocks).*

Average observation periods: timerep 1: 9.30 - 10.30 h; timerep 2: 11.30 – 12.30 h; timerep 3: 13.30 – 14.30 h; timerep 4: 15.00 – 16.00 h

1.3.3 Location

Dataset number 3 showed highly significant effects of location in the house (table 4). Birds close to the wall were less active than birds in the central area of the house (figure 2). Also, the combined fractions dustbathing-comfort was higher in the centre of the house than close to the wall (figure 3). The other 3 datasets contained no location information so could not be analysed on this aspect.

Table 4 *Analysis of dataset 3 for location in the house (close to the wall or central area in the house): The estimate shows the effect of the wall compared to the central area; a negative estimate indicates a lower proportion of birds showing a particular behaviour when being close to the wall as compared to the central area in the house.*

	Estimate	SE	P
Active-0	-0.403	0.360	0.004
Foraging	-1.043	0.205	0.004
Dust-comf	-0.545	0.084	<0.001

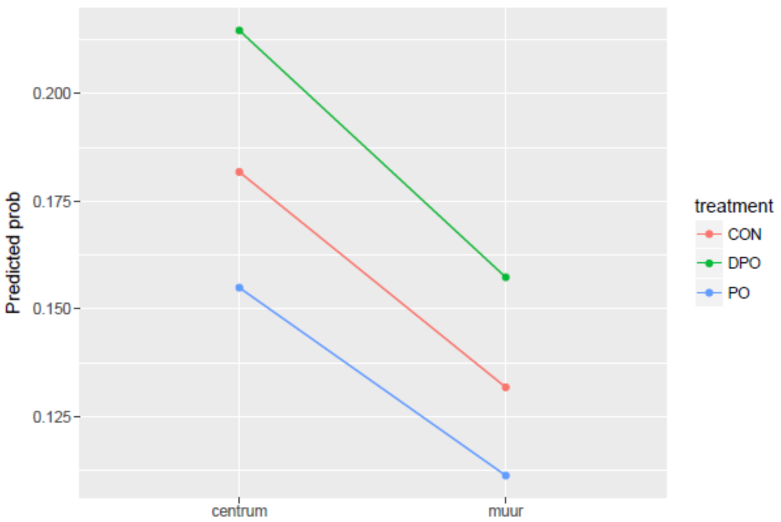

Figure 2 *Location effect on behaviour 'active'. Each line represents one treatment in the dataset; all treatments showed a reduction in proportion of broilers being active near the walls as compared to the central area (centrum=middle of the stable; muur= close to the wall).*

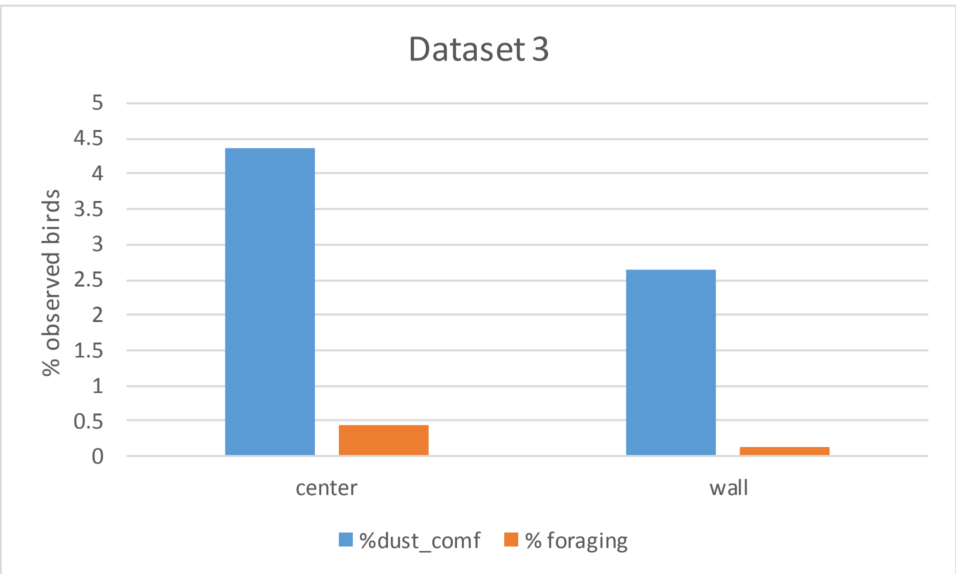

Figure 3 *Fraction Foraging and combined fractions dustbathing-comfort behaviour (% dust_comf) close to the wall and in the center of the house (resp. p=0.006 and p<0.001).*

1.3.4 Observer effect

From the analysis of dataset 1 (fast growing birds) it appeared there was a large observer effect for the combined trait “active behaviour” and for the fractions “foraging”, “dustbathing”, “comfort behaviour” and the combined fraction “dustbathing-comfort behaviour”(Dust-comf) (table 5). The other datasets could not be analysed on this aspect (only one observer per experiment).

Table 5 *Analysis of dataset 1 for observer effect. Variance components of the effect of observer for each of the behaviours are shown, indicating that the variance was highest for comfort behaviour and the combined trait dust-comf. These traits also showed a high standard deviation.*

Random effects:

Behaviour	Variance	SD
Active-0	0.152	0.096
Foraging	2.170	1.220
Dustbathing	2.202	2.275
Comfort behaviour	4.874	2.719
Dust-comf	4.722	2.650

1.3.5 Scan effect

To assess the behaviour with the scan sampling technique the observers usually observed one spot for various minutes. During that period the number of birds performing a behaviour was scored several times. This procedure was then repeated a number of times per day. The latter we call timereps, the first is referred to as the number of scans.

Dataset 1, with fast growing birds, showed a large scan-effect on all behaviours, indicating that the proportion of each behaviour increased with increasing scan number (Table 6). Dataset 3, also with fast growing birds showed a similar significant scan effect for ‘Active’ Behaviour and ‘Dust-comf’, but not for foraging. For slow growing birds (dataset 2 and 4) the scan effect was not present in most behaviours. Dataset 2 only had a significant scan effect for Comfort behaviour and Dust-comf, dataset 4 had no significant scan effect.

Table 6 *Analysis for scan effect. The estimate shows the effect of scan, a negative estimate indicating a decrease with increasing scan and a positive estimate indicating an increase with increasing scan number.*

Dataset 1	Estimate	SE	P
Active	0.052	0.006	<0.001
Foraging	0.115	0.008	<0.001
Dustbathing	0.178	No convergency	
Comfort behaviour	0.180	0.009	<0.001
Dust-comf	0.127	0.007	<0.001

Dataset 2	Estimate	SE	P
Active	-0.010	0.022	0.655
Foraging	0.158	0.107	0.140
Dustbathing	-0.010	0.083	0.904
Comfort behaviour	0.105	0.028	<0.001
Dust-comf	0.095	0.026	<0.001

Dataset 3	Estimate	SE	P
Active	0.084	0.012	<0.001
Foraging	0.078	0.068	0.250
Dust-comf	0.143	0.013	<0.001

Dataset 4	Estimate	SE	P
Active	-0.006	0.026	0.808
Foraging	-0.006	0.026	0.420
Dust-comf	0.001	0.017	0.962

1.3.6 Number of scans

Given the significant scan effect on many behaviours, the question is how many scans are needed to achieve a reliable outcome. Dataset 1 has 8 scans per behaviour. For foraging behaviour, dustbathing, comfort and dust-comf dataset 1 shows a slight reduction in overall means if less scans are carried out (Table 7). For foraging 5 or less scans gives a large reduction in the standard error, although the average does not follow this. As foraging did not occur a lot, this will have influenced the standard error as well. With respect to dustbathing the standard error gets extremely high if 3 or less scans are carried out and for comfort and dust-comf similar effects on the standard error are visible for 2 or less scans. These data should be interpreted with care, as these are probably affected by a very low prevalence of the behaviours.

Table 7 *Analysis of dataset 1 for active behaviour (Active-0), foraging, dustbathing, comfort behaviour and combined trait dust-comf; relation between number of scans and overall means and average SE for treatment (within age)**

	Active		Foraging		Dustbathing		Comfort		Dust-comf	
# scans	Overall means	SE	Overall means	SE	Overall means	SE	Overall means	SE	Overall means	SE
8	0.189	0.199	0.026	0.524	0.001	1.117	0.033	0.690	0.034	0.682
7	0.185	0.202	0.024	0.530	0.001	1.087	0.031	0.682	0.032	0.676
6	0.182	0.204	0.023	0.514	0.001	1.199	0.029	0.673	0.030	0.672
5	0.179	0.210	0.022	n.d.	0.001	1.345	0.028	0.646	0.029	0.652
4	0.178	0.221	0.021	n.d.	0.001	1.461	0.026	0.689	0.027	0.693
3	0.176	0.213	0.019	n.d.	0.001	n.d.	0.024	0.641	0.024	0.649
2	0.173	0.206	0.018	0.572	0.001	n.d.	0.021	n.d.	0.021	n.d.
1	0.165	0.213	0.016	0.555	0.001	n.d.	0.018	0.548	0.018	n.d.

* Extremely large or small standard errors are the result of the fact that the statistical program cannot find convergency. The values therefore are no real values and thus replaced by n.d. = not determined.

Dataset 2 has 4 scans per behaviour and only shows significant scan effects for Comfort and Dust-comf behaviour (Table 8). A reduction from 4 to 3 scans results in an almost double SE for foraging behaviour. For none of the other behaviours was there a large effect on the means and SE if the number of scans was reduced, although only one scan seems to be less reliable for a good mean and SE.

Table 8 *Analysis of dataset 2 for active behaviour, foraging, dustbathing, comfort behaviour and combined trait dust-comf; relation between number of scans and overall means and SE for treatment (within age).*

	Active		Foraging		Dustbathing		Comfort		Dust-comf	
# scans	Overall means	SE	Overall means	SE	Overall means	SE	Overall means	SE	Overall means	SE
4	0.307	0.220	0.005	0.692	0.003	1.250	0.030	0.181	0.033	0.189
3	0.306	0.217	0.005	1.094	0.003	1.406	0.028	0.191	0.031	0.204
2	0.309	0.204	0.004	1.225	0.003	1.450	0.028	0.179	0.032	0.198
1	0.301	0.215	0.003	1.958	0.004	1.434	0.023	0.202	0.027	0.226

Dataset 3 has 5 scans per behaviour and shows a decline in active behaviour and dust-comf with fewer scans, but an increase in foraging behaviour (Table 9). However, the average for foraging behaviour is very low, indicating that this behaviour was not frequently observed so drawing any conclusions is difficult.

Table 9 *Analysis of dataset 3 for active behaviour, foraging and combined trait dust-comf; relation between number of scans and overall means and SE for treatment (within age).*

	Active		Foraging		Dust-comf	
# scans	Overall means	SE	Overall means	SE	Overall means	SE
5	0.194	0.116	0.004	0.364	0.044	0.116
4	0.188	0.120	0.004	0.414	0.041	0.125
3	0.186	0.127	0.004	0.531	0.040	0.151
2	0.181	0.132	0.004	0.598	0.036	0.167
1	0.175	0.147	0.004	1.317	0.033	0.224

The red indicated SEM values are much higher than the SEM for 2 or more scans and therefore are an indication of reduced reliability

Dataset 4 shows a scan effect on the combined fractions dustbathing-comfort behaviour, where the fraction is increasing in consecutive scans. Also there is an effect on the fraction of foraging birds, which is quite variable, but on average increases with consecutive scans. Figure 4 shows the percentages birds performing dustbathing-comfort behaviour and foraging. In general there is a scan effect on the combined behaviour dustbathing-comfort behaviour, where the percentage increases with increasing scans. For foraging this is less obvious: in some flocks a scan effect is observed but not in others.

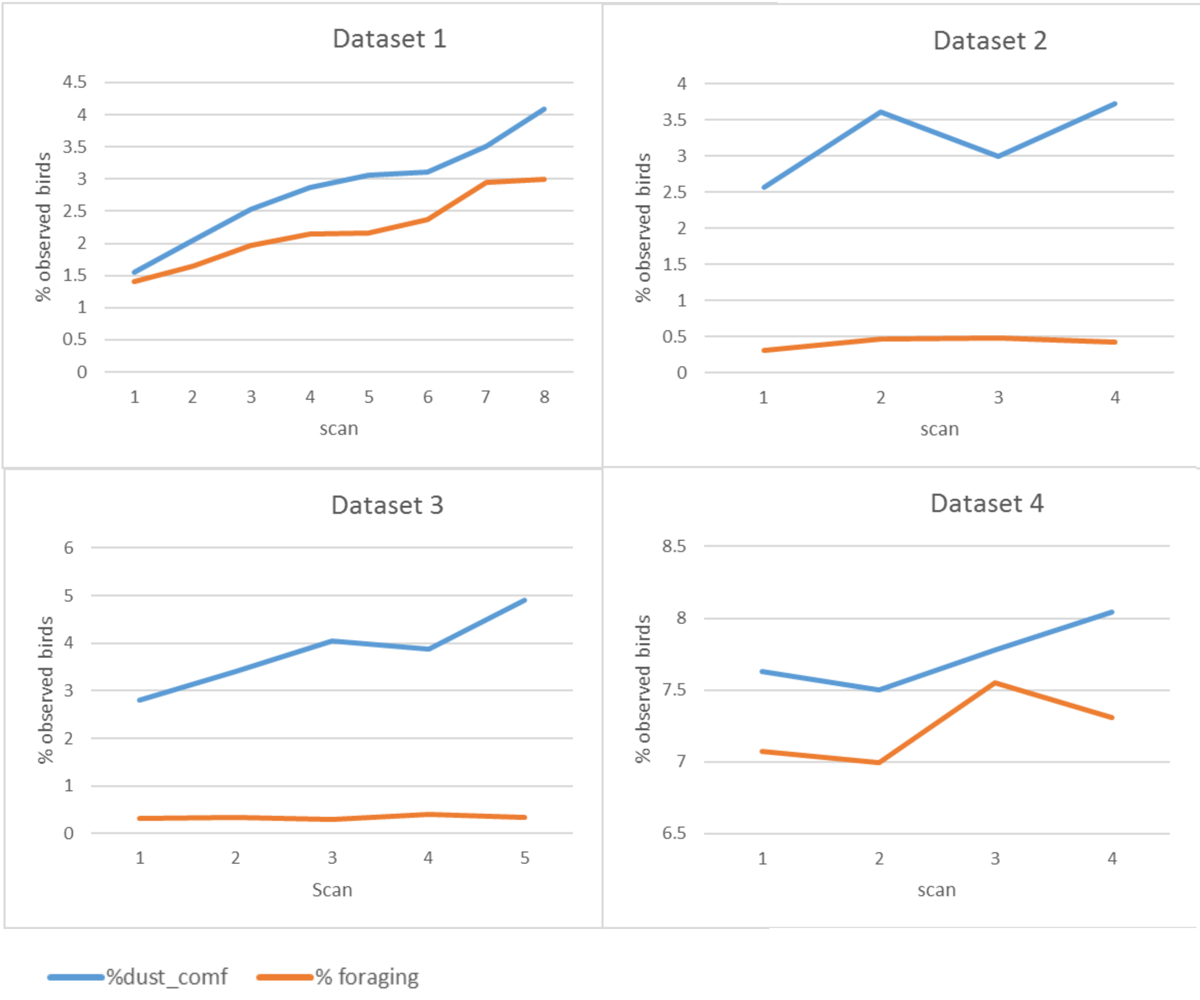

Figure 4 *Results for consecutive scans of combined fractions dustbathing-comfort behaviour (% dust_comf) and % foraging. The number of the scan is presented on the X-axis and differs between the datasets.*

1.3.7 Composition of Active behaviour

As for most datasets there was a significant observer and scan effect for active behaviour, the behaviour is not sufficiently robust to serve as a reliable measurement for inclusion in the WQ protocol. An additional analysis was done to see if changing the composition of 'Active' behaviour would result in a more robust measurement.

Figure 5 provides the fraction "Active" for different compositions of this trait. Active-0 contains all active traits and by consecutively leaving out one or more behaviours the fractions Active-1 to Active-4 are composed (see also Appendix 1). A clear scan effect is visible, but this decreased when the trait Active is reduced to walking/standing/running and foraging. For Active-4 the scan effect is absent and for Active-3 the scan effect is mainly present only when more than 5 scans are carried out.

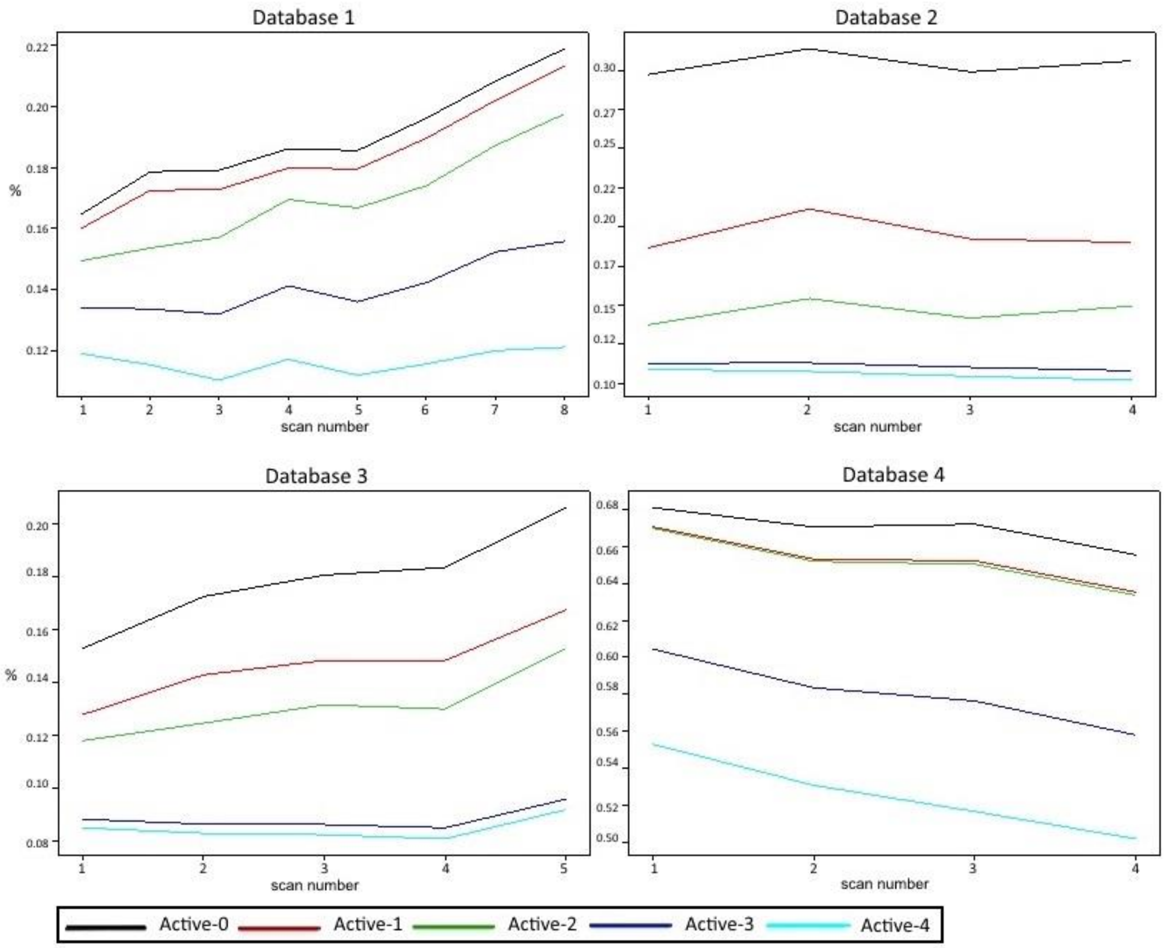

Figure 5 *Fraction for Active 0 – Active 4 for consecutive scans, presented as proportion of birds performing the behaviour in each scan. Results are presented for each dataset with the consecutive scan number on the X-axis. For explanation of Active 0-4, see Table 2.1. The figures show that with active 3 and 4, for the first 4 scans, the proportion of broilers being active is relatively stable except for dataset 4.*

Databases 1 and 3 are fast growing broilers, databases 2 and 4 are slow growing broilers

Table 10 presents the estimated variance components for observer effect for the different compositions of the trait 'Active' (only dataset 1 as the other datasets contained only one observer). For Active-3 the variance is the lowest, meaning that the observer effect is the smallest.

Table 10 *Estimated variance components for Observer effect (component of variance, s.e.; Dataset 1)*

	Variance	SE
Active-0	0.152	0.096
Active-1	0.145	0.092
Active-2	0.137	0.087
Active-3	0.134	0.087
Active-4	0.151	0.098

Taking the outcome of figure 5 and table 10 together the best option for scoring Active behaviour would be using Active-3 and 4 scans per measurement.

1.4 Discussion

In the present analysis we used existing datasets to study whether scan sampling of broiler behaviour, using direct observations, could be a possible measure to be included in welfare assessment protocols such as the Welfare Quality® broiler protocol (Welfare Quality, 2009). The data had been previously collected for various research projects. Because of that, there were some limitations in the analysis. As an example, datasets were not completely similar with respect to the behaviours recorded and the numbers of scans and observers. Therefore, the datasets were analysed separately instead of being combined (which would have provided a more powerful analysis). Despite these limitations, some general conclusions and indications were found that can be taken into account when considering if such a method would be useful to include in welfare assessment protocols.

The traits that were analysed, were measured at various ages and at various times of the day but for the Welfare Quality® protocol it would be more efficient if the traits could be scored at one particular moment on the same day as the rest of the on-farm part of the protocol is carried out. To get an idea about possible effects of age and time of day on the traits and the need for standardisation this has been analysed whenever the dataset allowed. The effect of age for the traits 'active' and foraging is in agreement with literature, showing that the activity of broiler chickens decreases with increasing age (e.g., fast growing broilers: Shields et al., 2005; Bailie et al., 2013; Bailie et al., 2018b; Baxter et al., 2018; De Jong and Gunnink, 2019); slower growing broilers: Bokkers and Koene, 2003; Wallenbeck et al., 2016). A consistent age effect was not found for comfort behaviour and dustbathing but both behaviours were only seen infrequently. It may be that comfort behaviour and dustbathing either increase or stay the same with age, or that factors such as the quality of the bedding (for dustbathing) and stocking density played a role. The Welfare Quality® assessment protocol is usually applied 1-5 days before flock depopulation (Welfare Quality, 2009) so thresholds or scores should be representative for that particular age.

A significant effect of time of the day was found for the proportion of broilers being active, dustbathing and foraging. The proportion of active broilers decreased between morning and afternoon, whereas dustbathing and comfort behaviour seemed to peak around midday. This also confirms earlier findings (e.g., Bailie et al., 2018a; De Jong and Gunnink, 2018). Because the frequencies of these behaviours seems dependent on the time of the day then related thresholds or scores must take into account the daily variation or the assessment should be performed at a particular time of the day.

Another important effect is the location in the house, broilers close to walls are less active than those in the centre of the house. It is known that broiler chickens prefer to rest close to the wall, probably because this provides shelter, can be a little more dark, and prevents disturbance by other birds (Buijs et al., 2010; Ben Sassi et al., 2019). To gain a reliable impression of the activity level of a flock, both the areas close to the wall and in the centre use should be included in an assessment. Similar findings and hence requirements were found for other welfare indicators, e.g. lameness and sick birds (Ben Sassi et al., 2019). We could not analyse the effect of enrichments on the activity level close to and away from the enrichment, but it is also known that, dependent on the type of enrichment, these may stimulate certain behaviours such as exploration and dustbathing (Kells et al., 2001; Bailie et al., 2013; Bailie and O'Connell, 2014; Bailie et al., 2018b). So, enrichment locations should also be taken into account in an observation protocol. Both aspects of location in the house could be solved by designing a protocol in which birds between the wall and feeder or drinker lines or within 1.5 m of the enrichment are recorded as well as birds more than 1.5 m away from the enrichment and between feeder and drinker lines in the central area of the house.

One dataset was open to analysis of observer effect. This effect was found for all behaviours despite previous observer training, and was largest for dustbathing, comfort and foraging behaviour. Several factors could play a role: either a) training was insufficient, b) broilers respond differently to the presence of different observer, e.g. due to variation in observer behaviour or posture, or c) it may have been difficult to score these behaviours using direct observations. Whereas standardisation can be done for factors like time of the day, location in the house and age, the prevention of intra-observer differences (and maybe also for inter-observer differences) might be more difficult even after intensive training. Indeed, a significant observer effect on the prevalence of welfare issues was found when applying the AWIN broiler assessment protocol which contains relatively simple measures of welfare (Ben Sassi et al., 2019). The relatively large observer effects on dustbathing, comfort and foraging behaviour might reflect difficulties in recognising these behaviours, e.g. because very specific behavioural elements, such as side-rubbing when dustbathing, are alternated with lying in the litter, which is sometimes very difficult to distinguish from lying with direct observations, even for trained observers. Also preening bouts are often interrupted by sitting or standing idle, and pecking and scratching, and foraging can be interrupted by walking. Moreover, foraging, dustbathing and comfort behaviours were generally observed at low frequencies and a confounding with observation location could have been present. A possible alternative could be to simply score whether or not these behaviours are observed at a particular location rather than including the proportions of birds showing these behaviours, but, the difficulty in recognizing the individual behaviours remains. Also, a yes/no scoring scale does not provide specific information about prevalence in a flock and thus will only provide limited information.

Increased activity levels in broiler flocks is generally considered positive with respect to broiler welfare. Being active itself is considered positive and it may stimulate bone strength and thus reduce the risk of locomotion problems (Reiter and Bessei, 1998). For the Welfare Quality® and other assessment protocols it might therefore be important to record the activity of the birds, and we created the trait 'Active'. The variation in recorded behaviours in each dataset caused this trait to differ slightly between datasets. We could have just used walking and standing as active behaviours and left out all other activities, but differences in experimental design and/or the presence of enrichment would likely be influential. Indeed, in dataset 2 a behaviour called 'busy with enrichment' was scored. Excluding these birds from the prevalence of the trait 'active' could lead to an underestimation of the activity level. Therefore we chose to incorporate any kind of activity into the trait 'active'.

However, an observer effect was again found for the trait 'active'. One may thus conclude that the trait active is not easy to score consistently. Another explanation may be that some parts of 'active' behaviour (particularly those of short duration) may have been missed due to the scan sampling technique. To determine if different compositions of the trait 'active' would influence the observer effect, various options were analysed, starting with 'Active-0' containing all possible active behaviours from the given datasets, to 'Active-4' being the most reduced version, only containing walking, standing and wing flapping (see appendix 1). The observer effect was the lowest in Active-3, being walking, standing, wing flapping and foraging. In Active-4, in which foraging was excluded, the variation again increased. This indicates that Active-3 was the best combination in terms of reducing the observer effect. However, as only database 1 obtained more than one observer and these results were thus based on a single study, a follow up study should be carried out.

Finally, in addition to all other effects as discussed above, we found a significant effect of scan number. To our knowledge, such a scan effect has only been reported once (Bokkers and Koene, 2003). Observers stand at a certain spot until the birds resume their 'normal' behaviour (predefined habituation period) and then start the scan sampling, and after doing several scans they move to a new observation location. The general idea is, that by performing a number of scans per location we get a more reliable picture of the behaviour as compared to a single scan only. As recording starts when the birds are supposed to resume their normal behaviour, one would not expect any scan effect. However, especially in the fast growing flocks activity increased with increasing scan numbers. This was actually opposite to our expectation. One would expect the birds to be more active with an unfamiliar person around, and calm down as soon as they are habituated. Especially the trait 'Active-0', 'Active-1' and 'Active-2' were much affected by scan number. These traits contain comfort behaviour and dustbathing, and this might explain the increase in the trait 'Active' in consecutive scans, as these are typically behaviours that birds perform when they feel safer. This is also shown in dataset 4, where the % dustbathing and comfort behaviour increased in consecutive scans. To overcome the scan effect, when monitoring fast growing broilers, based on the present analysis we suggest to limit the trait 'Active' to the elements in 'Active-3'and 'Active-4' and only perform a maximum of 4 scans in flocks with fast growing broiler chickens. Surprisingly, in slow growing broilers the effect was not seen in dataset 2 and birds became less active in consecutive scans dataset 4. Because of this contradictory effect in at least one of the datasets with slower growing chickens, this scan effect needs to be explored further before designing an assessment protocol. Limiting sampling to one scan may not provide a representative picture of flock behaviour, because we do not know which series of scans represent best the behaviour of a flock. Comparing video observations and live observations in flocks of broilers could determine if the scan sampling technique using direct observation is suitable for observing broiler flock behaviour and if so, how many scans should be included or how much habituation time should be allowed before starting a scan. This, in addition to the other drawbacks of scan sampling with direct observation, illustrates the pressing need for the application of sensor or video imaging techniques for measuring flock behaviour in welfare assessment in welfare assessment (Ben Sassi et al., 2016).

A scan effect was also found for dustbathing-comfort behaviour and foraging behaviour, where the use of at least 4 scans seemed to be the best option. However, frequencies were again low and a large observer effect was found. So, it is questionable if these behaviours should be scored using the scan sampling technique with direct observations.

To conclude, the present study illustrated that scan sampling of behaviour using direct observations in commercial broiler flocks has several limitations that affect the reliability of the results, and thus the possibilities to include such a measure in welfare assessment protocols. Whereas time of the day, age, and location effects could be incorporated in the assessment protocol, the observer and scan effects that were found merit further study. On the other hand, incorporating reliable behavioural measures would greatly improve the Welfare Quality® protocol and as there are currently no affordable techniques for automated monitoring of behaviour, it is worth to further study ways of overcoming the above limitations.

2 Alternative variables to measure fear in broiler chickens

2.1 Guidance for reading

Here we present an extended summary of the master thesis of Tosca Hoevenaar, entitled 'Development of a fear test for the Welfare Quality® protocol using different genetic lines of broilers'. This work was carried out as part of the project reported in chapter 1. The thesis work was supervised by ir Malou van der Sluis, dr ir Ingrid de Jong and prof. Bas Rodenburg (Wageningen University and Research) and dr Britt de Klerk (Cobb Europe).

2.2 Extended summary

Introduction and aim

The Welfare Quality® assessment protocol for poultry contains all measurements that can be used to determine overall welfare and is used to assess the welfare status of broiler chickens on commercial farms. In order to assess broilers' fear of humans, the touch test is currently included; this measures the number of animals that withdraw from the observer when they reach an arm out towards the birds – avoidance would indicate fear of the observer. However, as many broilers may show impaired walking ability, especially in the week before depopulation when the assessment is carried out, the birds might be less able or unable to withdraw from the observer thereby affecting the results. For example, Vasdal et al. (2018a) found a significant relationship between gait score and the number of broilers within arm's reach of the observer, which may indicate a confounding effect of walking ability.

Thus, the touch test should either be validated, which has not yet been done for broiler chickens, replaced by another measure of fearfulness, or, alternative variables related to fearfulness and unaffected by walking ability should be identified. Furthermore, the expression of fear can differ between genetic lines (Nielsen, 2012; Mignon-Grasteau et al., 2017). So, it is crucial that the results represent valid levels of overall fear for all lines.

This study aimed to determine if we could find alternative variables indicative of the level of fear in broilers that are not affected by walking ability. We studied two fear tests: the novel object test (measuring general fearfulness) and the touch test (measuring fear of humans).

Methods

The responses of habituated (to human contact and novel objects) and non-habituated groups of broiler chickens were compared in both the novel object and the human approach test. We considered that the habituation treatment would decrease fearfulness and thereby result in contrasting levels of fear between the two groups (lower in the habituated groups). Furthermore, behavioural responses were compared between different genetic lines of fast growing broiler strains, between young and older birds (1.5 versus 5 weeks of age) and between sexes. Two experiments were conducted in August 2018 and November 2018. In the first experiment, the touch test and the novel object test were conducted at 1.5 and 5 weeks of age and each group of broiler chickens was subjected twice to the same test. Six pens were habituated before testing and 6 pens were not; two genetic lines were equally present in the habituated and non-habituated pens (each pen contained one line).

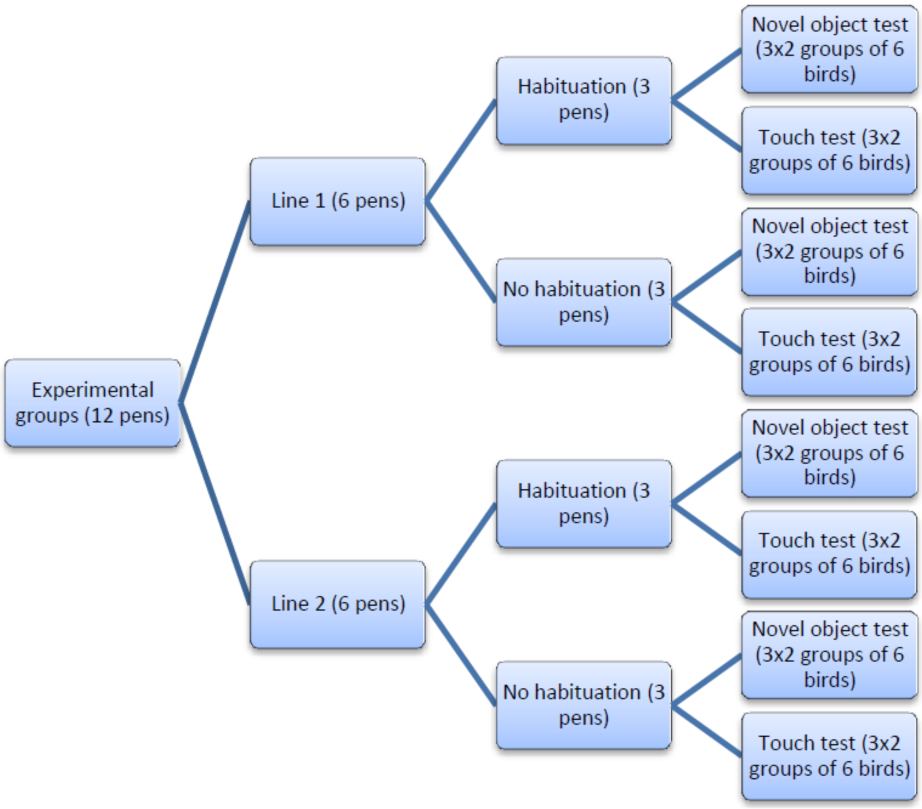

Figure 6 *Schematic overview of the design of experiment 1.*

From each pen, 4 groups of 6 birds were randomly selected and marked, of which half were subjected to the novel object test and half to the human approach test. Testing was done in a test arena located in the same room as their home pen. Figure 6 shows the experimental setup of experiment 1. On the day of testing, a randomly selected group of 6 birds was brought to the test pen and either subjected to the novel object test (10 min) or to the touch test (10 min), after a habituation period of 10 minutes in the test pen. In the novel object test, a small coloured plastic block was placed in the middle of the pen and in the touch test, the observer stretched her arm out and measured the birds' responses. The birds' behaviour during testing was recorded on video and analysed later. Behaviour was scored using the Observer XT 10 software (Noldus Information Technology B.V., Wageningen, The Netherlands). During both tests six birds were observed continuously for 30 seconds after the start of the test. Thereafter, instantaneous sampling was done every 30 seconds. Latency to approach, the number of birds within arm's length and the number of birds that approached were scored in Excel. Behaviours were scored for individual birds, but statistical analysis was performed on group level.

In experiment 2, the touch test and novel object test were conducted using groups of 60-66 broilers in their home pens, in order to create a more comparable situation to commercial flocks. Twenty-four pens with 4 different lines were used. In total, 6 pens of each genetic line were used (3 pens with females and 3 pens males). Like experiment 1, half of the birds were habituated and half were not. The pens that received habituation were randomly selected. The number of pens that were habituated per sex was also randomized. The birds were tested in their home pen so no habituation period was applied. The novel object was placed in the middle of the pen, and for the human approach test the observer first squatted down for 30 seconds and then stretched her arm out (Figure 7). Behaviour was recorded on video and analysed afterwards. Only the behaviour of broilers present in an imaginary circle of 1.4 m around the object or arm was recorded (Figure 7). Behaviour was scored using Excel. Instantaneous sampling was used for all tests. In total, 14 samples per test were taken, every 15 seconds for 3.5 minutes. During both tests the number of alert birds, body positions and proximity to the object or experimenter were scored. The duration of alertness was tested by comparing the proportions of alert birds per sampling moment.

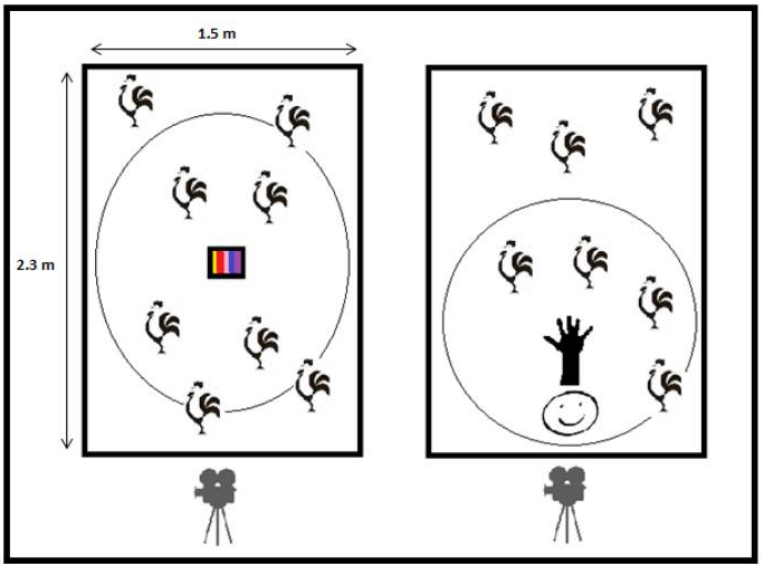

Figure 7 *Schematic drawing of the position of the novel object or observer and the imaginary circle in which the behaviour of the chickens was recorded.*

All statistical analyses were performed using SAS. Normality of the data was checked first and transformation was applied in the event of a non-normal distribution. Mixed model ANOVA's were used with genetic line, habituation and age included as fixed effects and pen as random effect in Experiment 1 whereas genetic line, time (of sampling), habituation and sex were included as fixed effects and pen as random effect (experiment 2). Correlations between variables within a test were also analysed.

Results

Experiment 1
During the novel object test of experiment 1, behaviours performed during the test were significantly affected by habituation and age ($p < 0.05$). During the touch test, behaviour was furthermore affected by genetic line ($p < 0.05$). During both tests, broiler chickens of 5 weeks of age walked less and spent more time lying during the test than chickens of 1.5 weeks of age. During the novel object test, birds of 1.5 week of age seemed more alert than those of 5 weeks, whereas during the touch test birds of 5 weeks of age seemed more alert than younger ones. This may imply that different types of fear were measured in the different tests. Birds of line 1 seemed more fearful during the touch test than those of line 2, indicating line differences in fear of humans. Habituation significantly affected behaviour during both tests. In the novel object test, habituated chickens showed more walking ($p<0.02$), were closer to the object (<0.01; within a bird length), less frequent erect neck postures ($p<0.01$) and less frequent alert body postures ($p<0.01$; standing motionless without performing other behaviours). Continuous sampling showed that habituated broilers spent: less time in alert body postures (and thus showed more other postures), more time exploring, and more time with a low neck position (see Figure 8).

In the touch test instantaneous sampling showed that habituated broilers were closer to the observer ($p=0.01$), showed fewer erect neck postures ($p=0.01$), and were less alert ($P<0.001$) than non-habituated chickens. Further, habituated chickens had a shorter latency to approach the observer ($p<0.05$) and more chickens approached the observer ($p<0.01$) or were observed within arm's length ($p<0.001$). Analysis of the continuous sampling showed generally the same trends (see Figure 9).

In order to determine which behaviours might be predictive of the behaviour 'alert' and thus might indicate fear, alertness was correlated with all other behaviours that were measured. The results showed that body positions probably did depend on the level of fearfulness, i.e. alert birds were more oriented towards the object or observer. Furthermore, while the head is oriented towards the stimulus fewer other behaviours, such as foraging, were shown.

Experiment 2
In Experiment 2 an adjusted ethogram was used based on the results of experiment 1 which included the number of alert birds, body positions and proximity to the object or experimenter. The Experiment 2 results showed significant differences between genetic lines in their responses during both tests ($P<0.05$). Furthermore, males seemed less fearful than females ($P<0.05$). In the novel object test, habituated broilers were closer to the object, with a body position oriented away, and less frequently alert ($P<0.0001$ for all variables). Generally the same results were found for the touch test. Figure 10 shows the proportion of broilers being alert in both tests.

Conclusions
The touch test and the novel object test measure different types of fear and therefore behavioural responses can differ accordingly, although in general several variables can be included in both tests as a measure of fear (of novelty, or of humans). As expected, habituation affected the various behaviours and postures observed. In addition, distance from the inanimate or human stimuli seemed positively associated with, and hence, an indicator of fear. As broilers at slaughter age often show less movement a measure based solely on distance from the stimulus might yield misleading results, especially if relatively short observation sessions are used. **We therefore suggest that either other variables, such as 'alertness', 'erect neck', or 'body position towards the object', should be used as measures of fear, or that these are combined with the variable 'distance to the stimulus'.** An adjusted protocol to be included in the Welfare Quality broiler assessment protocol (Welfare Quality, 2009) is shown in Appendix 3. We strongly advise that this modified protocol should be tested in commercial farms before applying it on a large scale.

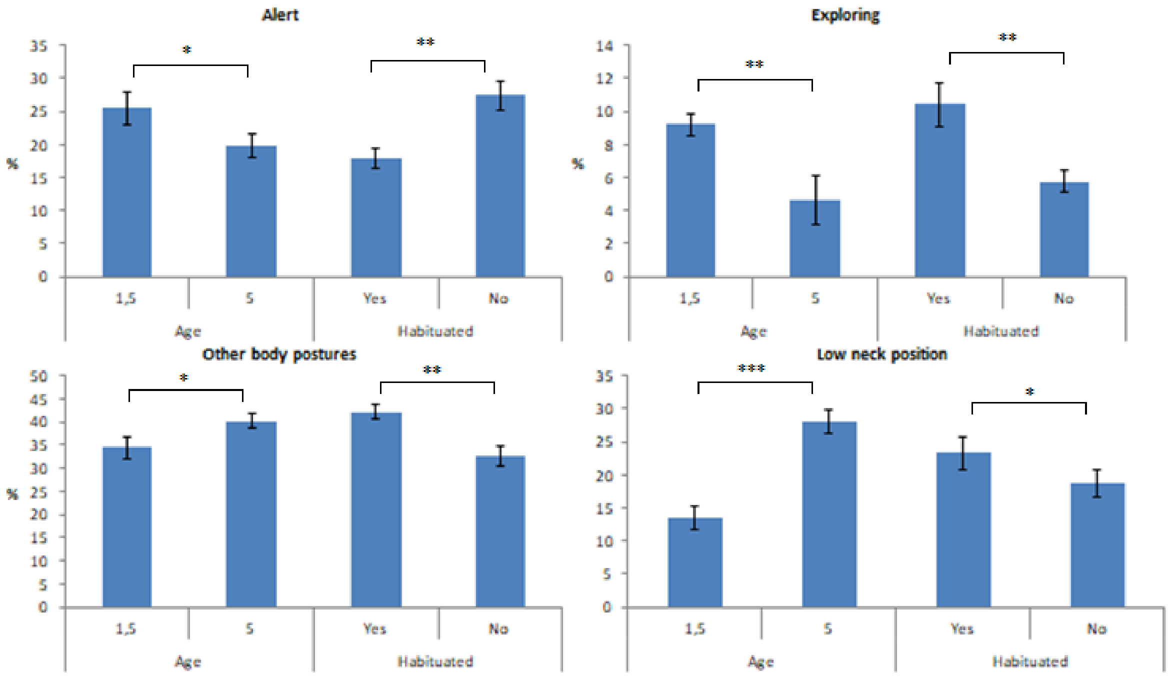

Figure 8 *Behavioural activities (proportions of the total observation time) alertness, exploratory behaviour, other body postures, and a low neck position of habituated and non-habituated groups and of broilers of 1.5 and 5 weeks of age recorded during the novel object test. Data are expressed as means ± SEM. Alert: stands motionless without performing active behaviours; Exploring: exploring the novel object; Other body postures: all other postures than alert; Low neck position: neck not erected.*

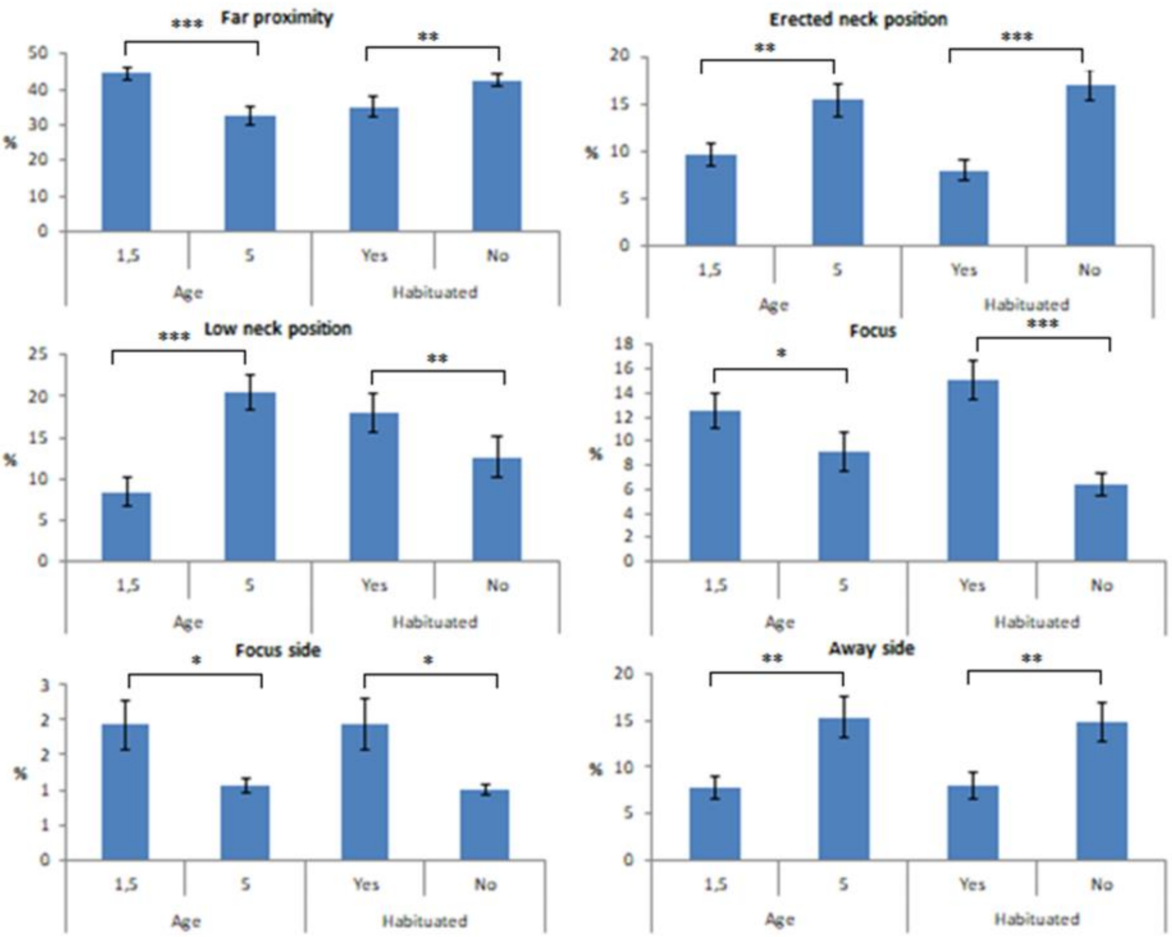

Figure 9 Behavioural activities (proportions of the total observations): far proximity of the experimenter, an erected neck position, a low neck position, focussed body position, a focussed body position combined with a side wards head position and a away oriented body position combined with a side ward head position of habituated and non-habituated groups and of broilers of 1.5 and 5 weeks of age recorded during the touch test. Data are expressed as means ± SEM.

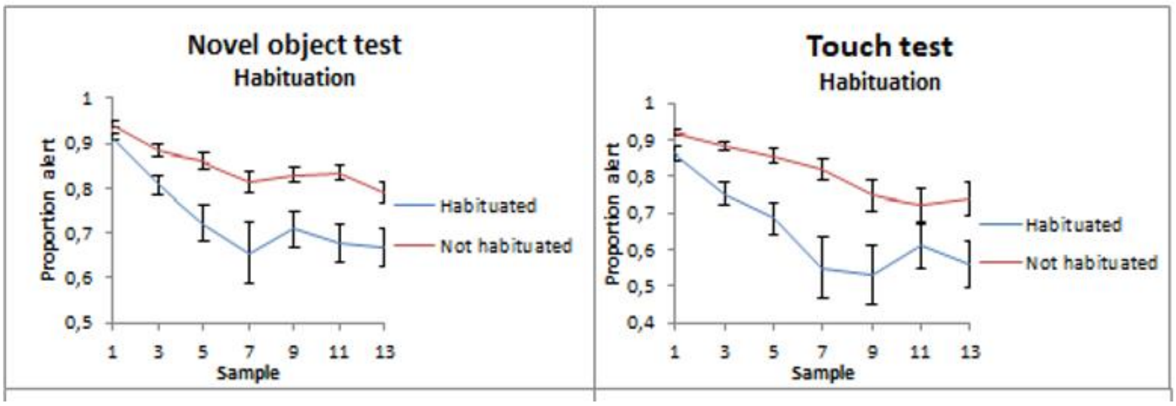

Figure 10 Duration of alertness of habituated and non-habituated groups recorded during the touch test and novel object test. Data are expressed as means ± SEM.

3 References

Bailie, C.L., Ball, M.E.E., O'Connell, N.E., 2013. Influence of the provision of natural light and straw bales on activity levels and leg health in commercial broiler chickens. Animal 7, 618-626.
Bailie, C.L., Baxter, M., O'Connell, N.E., 2018a. Exploring perch provision options for commercial broiler chickens. Applied Animal Behaviour Science 200, 114-122.
Bailie, C.L., Ijichi, C., O'Connell, N.E., 2018b. Effects of stocking density and string provision on welfare-related measures in commercial broiler chickens in windowed houses. Poultry Science 97, 1503-1510.
Bailie, C.L., O'Connell, N.E., 2014. The effect of level of straw bale provision on the behaviour and leg health of commercial broiler chickens. Animal 8, 1715-1721.
Baxter, M., Bailie, C.L., O'Connell, N.E., 2018. Evaluation of dustbathing substrate and straw bales as environmental enrichments in commercial broiler housing. Applied Animal Behaviour Science 200, 78-85.
Ben Sassi, N., Averos, X., Estevez, I., 2016. Technology and Poultry Welfare. Animals : an open access journal from MDPI, 6.
Ben Sassi, N., Averos, X., Estevez, I., 2019. The potential of the transect method for early detection of welfare problems in broiler chickens. Poultry Science 98, 522-532.
Blokhuis, H., Miele, M., Veissier, I., Jones, B., 2013. Improving farm animal welfare science and society working together: The welfare quality approach. Doi 10.3920/978-90-8686-770-7
Bokkers, E.A.M., de Vries, M., Antonissen, I., de Boer, I.J.M., 2012. Inter- and intra-observer reliability of experienced and inexperienced observers for the Qualitative Behaviour Assessment in dairy cattle. Anim. Welf. 21, 307-318.
Bokkers, E.A.M., Koene, P., 2003. Behaviour of fast- and slow growing broilers to 12 weeks of age and the physical consequences. Appl. Anim. Behav. Sci. 81: 59-72.
Buijs, S., Keeling, L.J., Vangestel, C., Baert, J., Vangeyte, J., Tuyttens, F.A.M., 2010. Resting or hiding? Why broiler chickens stay near walls and how density affects this. Applied Animal Behaviour Science 124, 97-103.
Czycholl, I., Beilage, E.G., Henning, C., Krieter, J., 2017. Reliability of the qualitative behavior assessment as included in the Welfare Quality Assessment protocol for growing pigs. Journal of Animal Science 95, 3445-3454.
De Jong, I.C., Gunnink, H., 2019. Effects of a commercial broiler enrichment programme with or without natural light on behaviour and other welfare indicators 13: 1751-7311.
de Jong, I.C., Hindle, V.A., Butterworth, A., Engel, B., Ferrari, P., Gunnink, H., Moya, T.P., Tuyttens, F.A.M., van Reenen, C.G., 2016. Simplifying the Welfare Quality((R)) assessment protocol for broiler chicken welfare. Animal 10, 117-127.
De Jong, I.C., Butterworth, A., Tuyttens, F.A.M., Buijs, S., 2015. Suggestions for modifications of the WQ broiler chicken welfare assessment protocol, Welfare Quality Network meeting, Welfare Quality Network, Barcelona.
EFSA, 2010. Scientific Opinion on the influence of genetic parameters on the welfare and the resistance to stress of commercial broilers. EFSA journal 8, 1666 -.
Kells, A., Dawkins, M.S., Borja, M.C., 2001. The effect of a 'freedom food' enrichment on the behaviour of broilers on commercial farms. Anim. Welf. 10, 347-356.
Mignon-Grasteau, S., Chantry-Darmon, C., Boscher, M.-Y., Sellier, N., Le Bihan-Duval, E., Bertin, A., 2017. Genetic Determinism of Fearfulness, General Activity and Feeding Behavior in Chickens and Its Relationship with Digestive Efficiency. Behav. Genet. 47, 114-124.
Nielsen, B.L., 2012. Effects of ambient temperature and early open-field response on the behaviour, feed intake and growth of fast- and slow-growing broiler strains. Animal 6, 1460-1468.
Phythian, C.J., Michalopoulou, E., Cripps, P.J., Duncan, J.S., Wemelsfelder, F., 2016. On-farm qualitative behaviour assessment in sheep: Repeated measurements across time, and association with physical indicators of flock health and welfare. Applied Animal Behaviour Science 175, 23-31.
Welfare Quality, 2009. The Welfare Quality Assessment Protocol for Broiler Chickens and Laying Hens. The Welfare Quality Consortium, Lelystad.

Reiter, K., Bessei, W., 1998. Effect of locomotor activity on bone development and leg disorders in broilers. Arch. Geflugelkd. 62, 247-253.
Riber, A.B., van de Weerd, H.A., de Jong, I.C., Steenfeldt, S., 2018. Review of environmental enrichment for broiler chickens. Poultry Science 97, 378-396.
Shields, S.J., Garner, J.P., Mench, J.A., 2005. Effect of sand and wood-shavings bedding on the behavior of broiler chickens. Poultry Science 84, 1816-1824.
Van Liere, D.W., 1991. Function and organization of dustbathing in laying hens, Landbouwuniversiteit Wageningen, Wageningen.
Vasdal, G., Moe, R.O., de Jong, I.C., Granquist, E.G., 2018a. The relationship between measures of fear of humans and lameness in broiler chicken flocks. Animal 12, 334-339.
Vasdal, G., Vas, J., Newberry, R.C., Moe, R.O., 2018b. Effects of environmental enrichment on activity and lameness in commercial broiler production. Journal of applied animal welfare science : JAAWS, 1-9.
Wallenbeck, A., Wilhelmsson, S., Jonsson, L., Gunnarsson, S., Yngvesson, J., 2016. Behaviour in one fast-growing and one slower-growing broiler (Gallus gallus domesticus) hybrid fed a high-or low-protein diet during a 10-week rearing period. Acta Agric. Scand. Sect. A-Anim. Sci. 66, 168-176.

Appendix 1 Composition of the trait “Active” per database

	General setup	Dataset 1	Dataset 2	Dataset 3	Dataset 4
Active 0	**Study specific features**		on.enrichment+ close.to.enrichment+ busy.with.enrichment		
	Walking/standing/flapping	lopen+ staan+	walking.running.flapping	ren.loop+ sta	ren/loop+ staan
	Foraging	bodempik+	foraging+	scharrel	foraging
	Comfort/dustbathing	poets_comfort+ stofbad+	dust.bathing+ comfort.behaviour+	dust.comf	dustcomf
	Disturbance	verstoring+	disturbed.during.resting	verstoring	verstoring
	Aggression/featherpecking	agressie+ verenpikken+		aggres	aggres
	Other behaviour	overig	other.behaviour	overig	overig
Active 1	**Walking/standing/flapping**	lopen+ staan+	walking.running.flapping	ren.loop+ sta	ren/loop+ staan
	Foraging	bodempik+	foraging+	scharrel	foraging
	Comfort/dustbathing	poets_comfort+ stofbad+	dust.bathing+ comfort.behaviour+	dust.comf	dustcomf
	Disturbance	verstoring+	disturbed.during.resting	verstoring	verstoring
Active 2	**Walking/standing/flapping**	lopen+ staan+	walking.running.flapping	ren.loop+ sta	ren/loop+ staan
	Foraging	bodempik+	foraging+	scharrel	foraging
	Comfort/dustbathing	poets_comfort+ stofbad+	dust.bathing+ comfort.behaviour+	dust.comf	dustcomf
Active 3	**Walking/standing/flapping**	lopen+ staan+	walking.running.flapping	ren.loop+ sta	ren/loop+ staan
	Foraging	bodempik+	foraging+	Scharrel	foraging
Active 4	**Walking/standing/flapping**	lopen+ staan+	walking.running.flapping	ren.loop+ sta	ren/loop+ staan

Appendix 2 Details of models used per dataset

Dataset 1

Models used:

Counts for active, foraging, dustbathing, comfort are calculated as counts and totals:

G1 <- summaryBy(active + foraging + dustbathing + comfortbehav + dustcomf + N ~ study + farm + flock + stable + observer + treatment + age + date + observer + location2)

Analysis per behaviour without taking scan into account. Analysis done per behaviour for active, foraging, dustbathing, comfortbehav, dustcomf. Used model (where BH stands for each analysed behaviour):

Formula: cbind(BH.sum, N.sum - BH.sum) ~ treatment + age + treatment:age + (1 | observer) + (1 | farm/flock/stable/date/location2)

Analysis per behaviour including the effect of scan. Analysis done per behaviour for active, foraging, dustbathing, comfortbehav, dustcomf. Used model (where BH stands for each analysed behaviour):

Formula: cbind(BH, N - BH) ~ treatment + age + treatment:age + scan + (1 | observer) + (1 | farm/flock/stable/date/location2/scan)

Dataset 2

Models used:

Counts for active, foraging, dustbathing, comfort are calculated as counts and totals:

G1 <- summaryBy(active + foraging + dustbathing + comfortbehav + dustcomf + N ~ study + farm + farmID + stable + date + timerep + location2)

Analysis per behaviour without taking scan into account. Analysis done per behaviour for active, foraging, dustbathing, comfortbehav, dustcomf. Used model (where BH stands for each analysed behaviour):

Formula: cbind(BH.sum, N.sum - BH.sum) ~ timerep + (1 | farm/timerep/location2)

Analysis per behaviour including the effect of scan. Analysis done per behaviour for active, foraging, dustbathing, comfortbehav, dustcomf. Used model (where BH stands for each analysed behaviour):

Formula: cbind(BH, N - BH) ~ timerep + scan + (1 | farm/timerep/location2/scan)

Dataset 3

Models used:

Counts for active, foraging, dustcomf are calculated as counts and totals:

G1 <- summaryBy(active + foraging + dustcomf + N ~ study + farm + stable + date + stabledate + age2 + timerep + locinstable + location2 + treatment + time + period)

Analysis per behaviour without taking scan into account. Analysis done per behaviour for active, foraging, dustcomf. Used model (where BH stands for each analysed behaviour):

Formula: cbind(BH.sum, N.sum - BH.sum) ~ treatment + period + locinstable + age2 + (1 | stable) + (1 | date) + (1 | stabledate/timerep/location2)

Analysis per behaviour including the effect of scan. Analysis done per behaviour for active, foraging, dustbathing, comfortbehav, dustcomf. Used model (where BH stands for each analysed behaviour):

> *Formula: cbind(BH, N - BH) ~ treatment + period + locinstable + age2 + scan + (1 | stable) + (1 | date) + (1 | stabledate/timerep/location2/scan)*

Dataset 4

<u>Models used:</u>

Counts for active, foraging, dustcomf are calculated as counts and totals:

> *G1 <- summaryBy(active + foraging + dustcomf + N ~ study + farm + stable + date + stabledate + age2 + timerep + locinstable + location2 + treatment + time + period)*

Analysis per behaviour including the effect of scan. Analysis done per behaviour for active, foraging, dustbathing, comfortbehav, dustcomf. Used model (where BH stands for each analysed behaviour):

> *Formula: cbind(BH, N - BH) ~ treatment + period + locinstable + age2 + scan + (1 | stable) + (1 | date) + (1 | stabledate/timerep/location2/scan)*

Appendix 3 Adjusted protocol fear test

Here, an adjusted protocol for the touch test is suggested which can be included in the Welfare Quality broiler protocol.

It is recommended to observe the animals for at least 3.5 minutes as broilers at slaughter age probably need more time to move away from the stimulus. Alertness seems to be an indicator of fear in broilers at slaughter age. Alert animals express several behaviours while being alert, however, behavioural responses depend on the degree of fearfulness. Therefore, the degree of alertness characterized by specific behaviours might be a valid indicator of fear in broilers as suggested in the table below. Furthermore, proximity to the observer still seemed an indicator of fear in broilers at slaughter age and should therefore included in the protocol as well.

It is recommended to perform instantaneous scan sampling. The birds should be observed for at least 3.5 minutes and at least 7 samples should be taken of the number of birds within certain distance of the object or observer and the number of birds with different degree of alertness (once every 30 seconds).

1 - Not alert	**2 - Slightly alert**	**3 - Alert**	**4 - Very alert**
Animal does not pay attention to the stimulus and might show activity behaviours as preening and foraging. Muscles are not tensed, the bird shows a normal neck position and the bird continues with behaviours relatively fast after introducing the stimulus. The bird has a low motivation to move away from the stimulus and might approach the stimulus.	The bird pays attention to the stimulus several times during the observation but does show activity behaviours as preening and foraging as well. While being alert, the animals might show slightly tensed muscles, the head is oriented towards the stimulus and the neck might be in an erected position. The animal does not feel comfortable in close proximity of the stimulus but is also not highly motivated to move away from the stimulus. The bird might eventually approach the stimulus.	The bird is focussed on the stimulus with the head oriented towards the stimulus. Muscles are tensed and the neck is often erected. The animal has a high motivation to move away from the stimulus and would not approach the stimulus. Activity behaviours as foraging and preening are hardly shown during the observations.	The bird seems very afraid of the stimulus and might show freezing behaviour. Muscles are tensed, neck is erected and head is oriented towards the stimulus. The animal has a high motivation to move away or flight from the stimulus. The animal focusses on the stimulus during the whole observation and does not perform other activity behaviours as foraging.

To explore the potential of nature to improve the quality of life

Wageningen Livestock Research
P.O. Box 338
6700 AH Wageningen
The Netherlands
T +31 (0)317 48 39 53
E info.livestockresearch@wur.nl
www.wur.nl/livestock-research

Wageningen Livestock Research creates science based solutions for a sustainable and profitable livestock sector. Together with our clients, we integrate scientific knowledge and practical experience to develop livestock concepts for future generations.

Wageningen Livestock Research is part of Wageningen University & Research. Together we work on the mission: 'To explore the potential of nature to improve the quality of life'. A staff of 6,500 and 10,000 students from over 100 countries are working worldwide in the domain of healthy food and living environment for governments and the business community-at-large. The strength of Wageningen University & Research lies in its ability to join the forces of specialised research institutes and the university. It also lies in the combined efforts of the various fields of natural and social sciences. This union of expertise leads to scientific breakthroughs that can quickly be put into practice and be incorporated into education. This is the Wageningen Approach.