Modules for Welfare Quality® laying hen protocol

T.G.C.M. van Niekerk, H. Gunnink, M. Wijhe-Kiezebrink en M. Pickhard

WAGENINGEN
UNIVERSITY & RESEARCH

Modules for Welfare Quality® laying hen protocol

Modules voor het Welfare Quality® protocol voor leghennen

T.G.C.M. van Niekerk, H. Gunnink, M. Wijhe-Kiezebrink en M. Pickhard

This research was conducted by Wageningen Livestock Research, commissioned and funded by the Ministry of Agriculture, Nature and Food Quality, within the framework of Policy Support Research theme 'Animal Welfare' (project number BO-20-008-033)

Wageningen Livestock Research
Wageningen, June 2018

Report 1104

T.G.C.M. van Niekerk, H. Gunnink, M. Wijhe-Kiezebrink en M. Pickhard, 2017. *Modules for Welfare Quality® laying hen protocol; Modules voor het Welfare Quality® protocol voor leghennen.* Wageningen Livestock Research, Report 1104.

Samenvatting NL Om meer informatie uit het Welfare Quality protocol te kunnen halen, zijn drie modules ontwikkeld, die op drie deelgebieden gedetailleerdere informatie geven. Deze detailscores passen op de originele WQ-scores, zodat een overal-berekening van het welzijn volgens de WQ-methode mogelijk is. Twee van de drie modules (borstbenen en snavels) gaven een goede herhaalbaarheid en verstrekten tevens zinvolle informatie. De module voor kleine veerbeschadigingen is niet verder ontwikkeld, omdat deze beschadigingen te weinig relatie leken te hebben met verenpikkerij en omdat de werkelijke oorzaken van de schade niet duidelijk waren.

Summary UK To get more information out of the Welfare Quality assessment, three modules are developed, to get more detailed information on three aspects. These detailed scores fit on the original WQ-scores, so that the overall welfare assessment according to the WQ-methodology can be done. Two modules (keel bones and beaks) had a good repeatability and provided useful information. The module for subtle feather damage was not further developed, because this damage didn't appear to have much relation with feather pecking and the actual causes of this damage were unclear.

This report can be downloaded for free at https://doi.org/10.18174/451287 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 is NEN-EN-ISO 9001:2015 certified.
All our research commissions are in line with the Terms and Conditions of the Animal Sciences Group. These are filed with the District Court of Zwolle.

Wageningen Livestock Research Report 1104

Table of contents

Summary

The Welfare Quality laying hen protocol is time consuming, hampering its feasibility to it in practice. Simplification, to make it less time consuming, was not possible. To overcome the problems of the Welfare Quality® Laying hen Protocol a study was conducted addressing the following questions:

- If simplification and time reduction is not possible, is there a way to get more information out of the protocol, so that in the same time span more information is obtained and thus the time spent on it is paid back. This could be done by making modules aiming at specific problems, that provide a more detailed insight.

- Is it possible to apply the Welfare Quality® protocol earlier than 50 weeks of age by focussing some important measures more on the situation at younger age?

The idea was to make three modules, that can be applied if wanted. The modules aimed at the following aspects:

- Beaks: when keeping hens with intact beaks a number of variations and abnormalities of the beak tip occur, that are not or hardly distinguished by the Welfare Quality® protocol. This applies to beak tip damage and extent of abrasion (e.g. due to the use of pecking blocks). Details of the measurement could give more insight.

- Feather cover: with the standard 3-point scale for plumage condition subtle feather damage can't be detected. The measure therefore can't be used to detect early signs of feather pecking and thus can't be used as an 'early warning signal'.

- Keel bone: the standard 3-point scale is not distinctive enough to provide a detailed idea of the type and extent of keel bone deviations and fractures.

In several steps and tests the modules were developed and the repeatability and reliability was tested. Due to an outbreak of bird flu in the Netherlands, visits to commercial farms were restricted and part of the work was carried out at Carus, the experimental facility in Wageningen.

After some testing the module for feather damage was not further developed, because it was unclear what the cause of the detected feather damage was. The aim was to develop a protocol for early detection of feather pecking problems, but the test results gave reason to doubt a strait-forward relationship between the detected feather damage and feather pecking. First research has to reveal the causation and interpretation of various types of feather damage and feather irregularities, then a further development of a feather scoring module can be started.

With the developed module for keel bone damage experienced observers were able to get a good reliability of scoring the deviations on a 4-point scale and the fractures on a 3-point scale, as was shown in the trial on commercial farms. This means that this module can only be used successfully after thorough training of the observers. The scoring on eight commercial farms indicated that part of the scored deviations in fact may originate from fractures as well, which is in accordance with findings of Scholz et al. (2008).

After testing various options, the final module for beaks focussed on the beak tip rim and sharpness. The reliability of the scoring was good for the beak tip and reasonable for sharpness.

In conclusion one can state that two of the three modules that were developed actually can be applied in a meaningful way to get more detailed information on the status of the exteriour of the birds. For feather cover more research on small feather deformities is needed before continuing the development of this module. The initial goal to score at an earlier age than 50 weeks was especially aimed at feather cover. For beaks it is not known how these develop over time. For keel bone fractures it is known that these increase over time, although the majority of fractures have occurred at 50 weeks of age (Petrik et al. 2015). This makes an assessment at early age not feasible.

1 Introduction

The Welfare Quality® protocols are designed to make welfare of livestock measureable through a protocol that is applicable under commercial conditions and by non-scientists and that meets international scientific quality standards. For laying hens this protocol is very time consuming, making it less attractive for application in practice. A study financed by the Dutch government was aiming at simplification of the protocol and thus reducing the time needed. For this simplification a test was done to see what measurements were highly related with others. In that case only one of these related measurements would be needed to still get a good result. However, for laying hens it appeared that simplification in this way was not possible. Also, the most time consuming part of the protocol is the clinical scoring of 100 birds. Reducing the number of birds results in a less reliable outcome and therefore is no option.
Another problem of the protocol is the fact that it can only be applied to older flocks (at least 50 weeks of age), because the clinical scoring otherwise is not sufficiently discriminative. This applies mainly for the scoring of the feather cover, but probably also for other measurements (e.g. foot pads, keel bones, wounds).
Finally the Welfare Quality® protocol uses simple measurements on a 3-point scale, that works well as a general estimation, but does not always providing sufficient insight in the specific situation.

To overcome the problems of the Welfare Quality® Laying hen Protocol a study was conducted addressing the following questions:

- If simplification and time reduction is not possible, is there a way to get more information out of the protocol, so that in the same time span more information is obtained and thus the time spent is paid back. This could be done by making modules aiming at specific problems, that provide a more detailed insight.

- Is it possible to apply the Welfare Quality® protocol earlier than 50 weeks of age by focussing on some important measures more on the situation at younger age?

The idea is to make 3 modules, that can be applied if wanted. The modules need to meet the following demands:

- Each module replaces the original measurement of the Welfare Quality® protocol (is in fact a refining of it), but still fits into the systematics of Welfare Quality®, making it possible to do the calculation for the final aggregation and scoring.

- Each module can be applied or omitted according to the demand or goal of the application of the Welfare Quality® protocol.

- The modules don’t or hardly take extra time compared to the original measurement.

- Because of the extra information the modules allow to obtain through the Welfare Quality® protocol, the fairly long timespan needed to apply the protocol is less of a problem, because in this time much more information is obtained than with the standard protocol.

The modules aim at the following aspects:

- Beaks: when keeping hens with intact beaks a number of variations and abnormalities of the beak tip occur, that are not or hardly distinguished by the Welfare Quality® protocol. This applies to beak tip abnormalities and extent of abrasion (e.g. due to the use of pecking blocks). Details of the measurement could give more insight in for instance: 1. Any malfuncton of the system, causing damage to the beaks (e.g. traps); 2. the effectivity of pecking stones; 3. Risk for injuries when pecking starts (sharpness of the beak).

- Feather cover: with the standard 3-point scale for plumage condition small spots of damage can’t be detected. The measure therefore can’t be used to detect early signs of feather pecking and thus can’t be used as an 'early warning'. By zooming in on small signals of feather damage, indicating the early onset of feather pecking, management measures can be put in place in time to prevent any further development of this unwanted behaviour.

- Keel bone: the standard 3-point scale is not distinctive enough to provide a detailed idea of the type and extend of keel bone deviations and fractures. Scientists mostly agree that deviations most likely are not really a welfare problem, whereas fractures are a welfare problem. Distinction between the two therefore is important to be able to indicate welfare issues related to keel bones. There is some discussion among scientists about the extent of impact on welfare of small and larger fractures, making a distinction useful. As more and more retail is asking proof of good welfare, farmers will be in need of a clear determination of the state of the art of the keel bones. re, farmers will be in need of a clear determination of the state of the art of the keel bones.

1.1 Possible use of the modules

The protocol is applicable for more purposes and therefore can be used by various groups or people. For certifying bodies the protocol can be used to monitor the welfare of a flock of laying hens in a scientifically sound and underpinned way. The protocol including the modules is targeting use by the industry. With the help of this new protocol a good insight can be obtained in both general welfare of laying hens, but also in detailed welfare aspects. As the assessment can be done at a young age, the protocol can be used as an 'early warning' for upcoming welfare issues.

1.2 Requirements for the modules

The modules should be made in such a way that they still allow the aggregation of the scores into a final welfare score according to the systematics used for the Welfare Quality® protocols. This means that the detailed scores in the modules are in such a way related to the original scoring, that they do fit into the original scorings.
Therefore, apart from designing the modules, also directions should be given how to fit the scores of the modules to the scores of the original protocol.

2 Materials and methods

The idea of the modules is that they are more detailed than the existing WQ-scoring and thus giving more detailed information. For the calculation of the final score of the WQ-protocol these detailed scores can't be used. They have to be brought back to the original, existing scores. In the development of the modules this is taken into account and realized, because for each module one can combine one or more scores of the module to one of the categories of the original, existing scoring system. As a consequence of this the modules could not be too much different from the original scoring, it could not score other features, but could only score more detail of the in the WQ-protocol defined features.

The modules were developed in six steps. The first four steps were part of a master's thesis, carried out by Melvin Pickhard, a student of Wageningen University. The last two steps were carried out by personnel of Wageningen Livestock research.

<u>Step 1:</u>
In this step the various scoring methods available in the literature were investigated.

<u>Step 2:</u>
New modules were made based on the old modules. This was done for the criteria plumage condition, keel bone damage and beak damage. For plumage condition, a whole new module was made. The aim was to look for the first signs of feather pecking. Therefore, we looked at the extent of damage to the feathers. For keel bone and beak damage we split the existing scores up into more detailed scores.

<u>Step 3:</u>
The third step was about practicing the handling of the hens and the WQ assessment protocol. This was done at the Carus facility in Wageningen (white hens) and at a commercial farm (brown hens). During those visits, new ideas and improvements were investigated. On the commercial farm the modules were tested for their practical usefulness.

<u>Step 4:</u>
Due to the bird flu in November 2016, we had to further test the protocols at Carus (white hens). Here we also did an assessment of the inter- and intra-observer repeatability and reliability.
The modules for plumage condition were also tested on pullets of 8 weeks old (white hens). With this, we investigated whether the new module for feather damage can also be used for rearing pullets.

<u>Step 5:</u>
Based on step 4 the modules were slightly modified and once more tested at Carus (white hens) for their repeatability and reliability. This was also done at two commercial farms, one with white hens and one with brown hens.

<u>Step 6:</u>
With the newly developed modules eight commercial flocks were scored. From each flock 100 birds were scored by two observers.

2.1 Animals

2.1.1 Carus

At Carus and commercial farms we tested the new modules. During the first three visits, fourteen pens at Carus were available. In each pen, between four and nine hens were housed. The hens were part of another research focusing on feather pecking. In each pen either hens from a low feather pecking line or a high feather pecking line were housed (White Leghorn line, divergently selected for high or low

feather pecking behaviour; Kjaer et al., 2001). During the first visit on 12 September 2016, the hens were 40 weeks old. Eight pens were available with 48 hens in total (batch 1). During the second visit on 28 September 2016, five pens were available with 33 hens in total (batch 2). During the third visit 16 hens were assessed with the draft version of the new modules. During a fourth visit, repeatability and reliability were investigated. From both the high feather pecking line and low feather pecking line, nine hens were assessed two times. This was done by three observers.
We used also pullets for plumage condition assessment. Those pullets were the next generation of the selection lines used in this study. The pullets were 8 weeks old and divided over ten pens. In each pen, 12 pullets were housed of either the low feather pecking line or the high feather pecking line.

2.1.2 Commercial flocks

For testing the protocol (step 3), one farm was visited. Here hens were housed in a “Rondeel” barn. The barn contained ten compartments. A total of 36.000 hens of the breed Novogen Brown Lite were housed from which in total, 69 hens of around 50 weeks old were assessed for the protocol.

For step 5 two commercial flocks were visited. One was the same flock as in step 3. The second flock was a white flock housed in a conventional aviary henhouse.

For step 6 a number of eight flocks was visited, 4 white genotypes and 4 brown genotypes and varying in age. In Table 2.1 the details of the flocks are given.

Table 2.1: *Details of the investigated flocks in step 6*

Number of birds on site (at the time of the visit)	Age (w) at day of inspection	Genotype
38321	74	Dekalb White
34200	49	Novogen Light brown
10720	78	Novogen White
9402	29	Dekalb White
9000	33	Novogen Brown Light
73000	50	Lohman Brown Classic
73000	30	Lohman Brown
32800	88	Dekalb White

2.1.3 Measurements commercial flocks

For step 3 and 5 the measurements on the commercial flocks only focussed on the WQ-modules. In step 6 more measurements were done. This was done to see if the WQ-measurements have any relation with housing conditions and behaviour.
The measurements done were:
- WQ-module keel bone
- WQ-module beak
- Possibilities for and actual movement through the housing system
- Specific lay-out of the system (presence or absence of side partitions, free range, covered veranda, etc.)

In appendix 2 the recording protocol is given.

2.1.4 Statistics

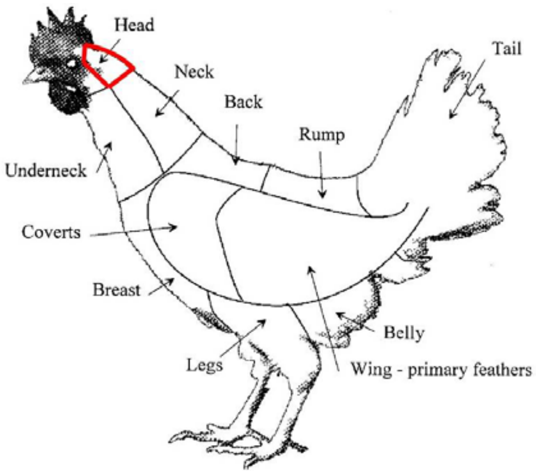
The observations of the first four steps were analysed with SPSS (IBM SPSS Statistic 23). To calculate the inter-observer reliability across all three observers the Kendall’s coefficient of concordance (W) was used (with significance set at $p=0.05$), as the data were not normally distributed. Each module was analysed per pen, round and for all observations. Pen was divided in a high feather pecking line and a low feather pecking line and round was divided in round 1 and round 2 (two different batches). For step 5 and 6 the data were analysed with the excel software program, using the correlation formula.

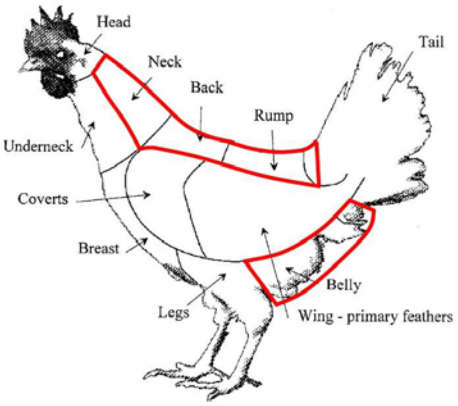
3 Results

3.1 Module Plumage damage

In the WQ-protocol plumage damage is scored twice: once as plumage damage on the back of the head and once as plumage damage due to feather pecking. For the latter specific areas are scored: back and rump, tail basis, around the cloaca (belly) and back of the neck.
Plumage damage on back of the head is scored separately from plumage damage on other parts of the body, as damage on the back of the head is caused by aggressive behaviour, not by feather pecking.
In the WQ protocol both types of feather damage are mentioned under the principle Appropriate behaviour. Plumage damage on the back of the head is scored under the criterion Expression of social behaviours and plumage damage caused by feather pecking is mentioned under the criterion Expression of other behaviours.

3.1.1 Original WQ-protocols

Title	**Plumage damage on the back of head**
Scope	Animal-based measure: Laying hens
Sample size	Sample size according to § 5.1.5
Method description	The feathers of normal birds should be smooth with no signs of disturbance. All feather shafts then usually point in one direction resulting in a protective and insulating cover to the skin. Aggressive pecking is usually directed downwards to the head region. Plumage damage in this area is an indicator of aggressive behaviour. Birds are visually inspected individually. Score each animal according to the indicated body part marked with an orange line in the drawing. For each bird a score is given on a 3-point scale, taking into consideration the indicated bodypart: **0** – no or slight wear, (nearly) complete feathering (only single feathers lacking); **1** – moderate wear, i.e. damaged feathers (worn, deformed) or one or more featherless areas < 5 cm in diameter at the largest extent; **2** – at least one featherless area ≥ 5 cm in diameter at the largest extent  © Bilcik, B. & L.J. Keeling, 1999

Title	Plumage damage
Classification	**Percentage** of birds with scoring categories 0,1,2
Scope	Animal-based measure: Laying hens
Sample size	Sample size according to § 5.1.5
Method description	The feathers of normal birds should be smooth with no signs of disturbance. All feather shafts then usually point in one direction resulting in a protective and insulating cover to the skin. Due to abrasion against wire, feather shafts can be broken. Due to pecking behaviour feathers can be disturbed, broken or even torn out. Areas where feather damage usually starts are the tail, neck and cloacal region. Feather damage at the back of the head is an indication of aggressive behaviour and is recorded separately (see par. 6.1.4.1) Birds are visually inspected individually. Score each animal according to three individual body parts (see photographic reference). For each bird 3 scores are given (i.e. 1 for each body part): being the back and rump together, around the cloacae (belly) and head and neck together. The 3 body parts are chosen to give information regarding the cause of feather damage: damage to feathers of the back and rump usually indicate feather pecking, damage to the feathers of head and neck can be caused by abrasion, and feather damage to the belly can be seen in highly productive animals. (However, the latter can also be caused by vent pecking.). For each bird a score is given on a 3-point scale, taking into consideration the 3 indicated bodyparts: **0** – no or slight wear, (nearly) complete feathering (only single feathers lacking); **1** – moderate wear, i.e. damaged feathers (worn, deformed) or one or more featherless areas < 5 cm in diameter at the largest extent; **2** – at least one featherless area ≥ 5 cm in diameter at the largest extent  © Bilcik, B. & L.J. Keeling, 1999
Classification	**Percentage** of birds with scoring categories 0,1,2

3.1.2 Module Plumage condition

To get a more precise measurement for the situation in which there is only little feather damage an additional module was designed with the help of the student project. From the original WQ-protocol the scoring was used for assessing denuded areas. An additional module was created to detect minor feather damage. As a sign of early stages of feather pecking the number of feathers with missing pieces was used. Those missing pieces are the gaps in the feathers, mostly found towards the tip of the feather.
The number of feathers that had at least one missing piece were recorded on a 4-point scale. If the feathers in the neck were difficult to see, they were stroked upwards or a paper was placed under the feathers. The number of damaged feathers in a given region determined the score, where limits of the scale were set at 3 and 10 damaged feathers. When a region was assigned at least a score 1 for denuded areas, feather damage was assigned a score 3.
The denuded areas were scored on a four point scale, with limits of the denuded area of larger or smaller than 2.5 cm and 5 cm.

Table 3.1 *WQ module Plumage condition*

Feather damage	Denuded area
0: No damage	0: No damage
1: ≤3 feathers damaged	1: ≤2.5cm denuded
2: 4≤10 feathers damaged	2: 2.5≤5cm denuded
3: >10 feathers damaged	3: >5cm denuded

The same body regions as in the original protocol were assessed as shown in the figure in paragraph 3.1.1 for plumage damage. The head and comb region was not assessed. For the back, denuded areas were measured from the neck/back until the end of the rump where the tail begins. For the vent, the feathers above the cloaca and under the tail feathers were assessed. These were feathers that easily can be assessed for missing pieces. For denuded areas, the whole vent was assessed.

For the vent, we included an extra measurement on a binomial scale. If the vent was denuded we also measured the number of wounds and small lesions due to pecking. Feather pecking at the vent can become vent pecking. This is a cannibalistic form of pecking and can increase the mortality of the flock.
0: No wounds or pecks
1: One wound or peck

3.1.3 Combined WQ-protocol with modules

As the modules should be designed to be used on top of the original WQ-protocol, Table 3.2 indicates how this is established.

Table 3.2 *Combination of WQ-protocol and WQ module feather damage*

WQ Protocol	WQ module	
	Feather damage	*Denuded areas*
0 – no or slight wear, (nearly) complete feathering (only single feathers lacking)	0a: No damage	0: No damage
	0b: ≤3 feathers damaged	
1 – moderate wear, i.e. damaged feathers (worn, deformed) or one or more featherless areas < 5 cm in diameter at the largest extent	1a: 4≤10 feathers damaged	
		1b: ≤2.5cm denuded
	1c: >10 feathers damaged	1c: 2.5≤5cm denuded
2 – at least one featherless area ≥ 5 cm in diameter at the largest extent		1c: >5cm denuded

3.1.4 Reliability of scores

As the data were not normally distributed the Kendall's W, which is a nonparametric test, was used to calculate the agreement of three or more observers with ranks. The scores are between 0 and 1. Following Martin and Bateson (2007), a reliability of 0.4 is defined as acceptable and a reliability of 0.7 is defined as good.
Overall reliability of the feather cover scores was not high (Table 3.2). Acceptable reliability was obtained for denuded spots in the neck, back damage in round 2 and in the HFP-line and denuded spots around the vent in the HFP-line. A good reliability was only obtained for feather damage in the neck in the LFP-line.

Table 3.3: *Kendall's W test for scores for feather cover and wounds, obtained at Carus*

	Neck		Back		Vent	
Pair	**Denuded**	**Damage**	**Denuded**	**Damage**	**Denuded**	**Wounds**
All	0.432*	0.153*	0.169*	0.280*	0.258*	0.076
Round 1	0.496*	0.209*	0.231*	0.104	0.249*	0.081
Round 2	0.372*	0.115	0.124	0.512*	0.294*	0.083
High FP-line	0.276*	0.09	0.242*	0.478*	0.420*	0.152
Low-FP line	0.616*	0.718*	0.111	0.358*	0.097	-

* Significance with p=0.05. – Not calculated

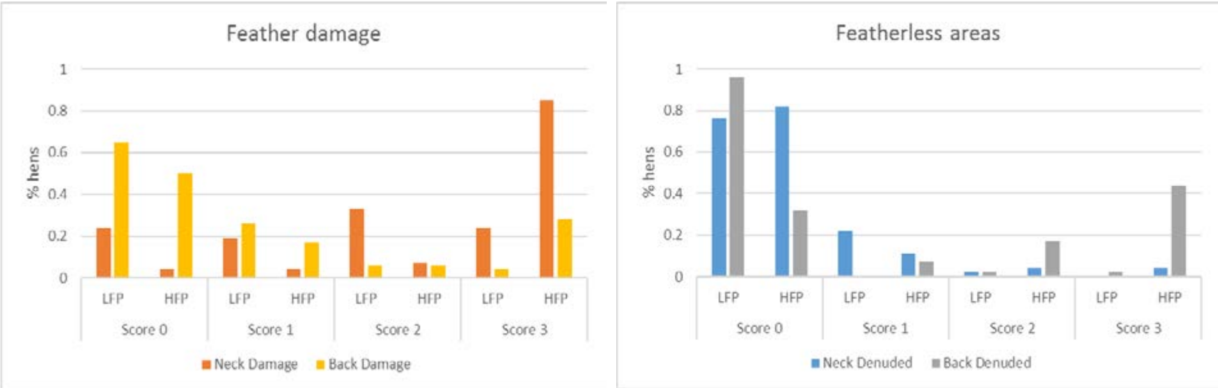


Figure 3.1: *Percentage of birds assigned to scores for feather damage and featherless areas*

There was a clear line difference for score 3 (severe damage), where the HFP-line had a higher percentage of birds falling into that category than the LFP-line. This can be expected, as the HFP-line is performing more severe feather pecking, resulting in more feather damage, as has also been confirmed by behavioural studies (carried out on the same birds, but in other projects). As can be expected the percentage of birds in with score 0 is larger for the LFP-line compared to the HFP-line. For score 2 and 3 however, the results are not in line with the expectations. For featherless areas (denuded areas) the results for score 0 and 3 are again as expected, but the result for scores 1 and 2 for feather damage are not according to expectation and score 1 for denuded areas also is not in line with the expectation. Especially when taking into account that the score for feather damage of the neck had a good reliability for the LFP-line, the results are odd. Score 1 and 2 are focussing on subtle signs of damage of the feathers. These signs often are little missing pieces of the feather, that have broken exactly on a so called stressline in the feather (see figure 3.2 for example). Apparently these occurred more in the LFP-lines than in the HFP-line. This result then suggests that the subtle signs of feather damage may not have a strong relation with feather pecking.

Figure 3.2: *Feather damage on stress-lines: pieces of feathers are broken off on stress-lines*

3.2 Module Keel bones

3.2.1 Original WQ-protocol

Originally the WQ-protocol had only 2 classes for keel bones: 0 - completely straight; 2 - deviation, deformation or thickened section. Later a class 1 was added to make a distinction between deviations and fractures. Fractures are painful and thus a welfare problem, whereas this is unclear for deviations.

Title	**Keel bone deformation**
Scope	Animal-based measure: Laying hens
Sample size	Sample size according to § 5.1.5
Method description	Keel bones are normally straight without dips, bulbs, deviation or other palpable abnormality. Abnormalities can be caused by badly shaped perches, healed breaks (fractures) or by decalcification of the keel. A keel bone deformation is any abnormality from the normal straight shape of the keel. The majority of keel bone deformations are caused by breaks and thus represent a major welfare issue. Healed breaks usually have thickened sections due to extra calcification. Some minor deviations are not caused by breaks, but originate from decalcification. They may represent a welfare risk, but are of a lesser order than actually broken keel bones. Examine the breast of the hen by looking at it (in case of a featherless breast) and by running your fingers alongside and over the keel bone. Make sure to check the keel ends for deviations. Collection of birds for assessment can be made by either penning (corralling) birds or by picking up individual birds in several areas of the henhouse. The number of places to collect hens is dependent on the housing system and the number of compartments. Pick up a bird from within the penned group or from the litter or slatted floor, inspect the keel area visually and palpate the area. In cage systems take the birds from different areas of the house and from different tier levels. Compare to the photographic reference and assess according to the following: 0 - No deviations, deformations or thickened sections, keel bone completely straight 1 - Deviations (flattening, s-shape, bending) present in very slight form 2 - Deviation or deformation of keel bone (including thickened sections)
Classification	**Flock level:** **Average keel bone score** **Percentage** of birds in the flock in category 2

Examples: score 0 score 1 score 2

3.2.2 Module Keel bone

The additional module for keel bones tries to differentiate the extent of the deviations and fractures and thus the extent of the problem. By palpating the keel bones any signs of fractures are determined. Thickened areas and extreme deviations are regarded as an indication of a fracture. Deviations in keel bones are defined as all not-strait keel bones that have no fractures. The module that was developed is presented in Table 3.4.

Table 3.4 *WQ module keel bone*

Deviations		Fractures
0: No deviation	*Completely straight*	0: No fractures
1: Slight deviation	*Slightly bent*	1: One fracture
2: Moderate deviation	*Sharply bent in one direction or slightly in two directions*	2: Multiple fractures
3: Severe deviation	*S-shaped, sharply bent in two or more directions*	

3.2.3 Combined WQ-protocol with modules

The new additional module basically makes a distinction in the extent of the deviation by dividing it into three classes. For fractures a distinction is made between one or more fractures. In Table 3.5 it is indicated how to use the new module on top of the original WQ-protocol.

Table 3.5 *Combination of WQ-protocol and WQ module keel bone*

WQ Protocol	WQ module	
	Deviations	*Fractures*
0 - No deviations, deformations or thickened sections, keel bone completely straight	0: No deviation	0: No fractures
1 - Deviations (flattening, s-shape, bending) present in very slight form	1a: Slight deviation	
	1b: Moderate deviation	
	1c: Severe deviation	
2 - Deviation or deformation of keel bone (including thickened sections)		2a: One fracture
		2b: Multiple fractures

3.2.4 Reliability of scores

In Table 3.6 the outcome of the Kendall W test for reliability between all observers is presented. Overall the reliability for deviations is very low and mostly not significant. For fractures it is higher and significant, but only in round 2 and in the LFP-line the reliabilities are reasonable.

Table 3.6 *Kendall's W Keel bone.*

Dataset	Deviations	fractures
All	0.085	0.466*
Round 1	0.06	0.335*
Round 2	0.115	0.632*
High FP-line	0.286*	0.388*
Low-FP line	0.088	0.623*

* Significance with p=0.05.

3.3 Module Beaks

The original WQ protocol for beaks focussed on beak trimmed beaks. Later this was modified to also fit intact beaks, but this merely aimed at the normality of the shape and possible abnormalities. The protocol does not take variation in beak shape and sharpness of the beak tip into account.

3.3.1 Original WQ protocol

Title	**Beak damage and abnormalities (not caused by trimming)**
Scope	Animal-based measure: Laying hens
Sample size	Sample size according to § 5.1.5
Method description	Due to poor equipment design beaks can be trapped and injured. Apart from that, selection in the rearing may not have been strict enough, leaving birds with misshaped beaks in the flock. 100 birds are examined and information from these birds will be included in the final score. The final score is based on both the inspection of 100 birds and visual observations during other work in the hen house and the number of birds with beak damage or abnormalities assessed. The classification reflects the number of birds with beak damage or abnormalities.
Classification	**Flock level:** **0** – No evidence of beak damage or abnormalities **1** – Fewer than 3 birds with beak damage or abnormalities **2** – 3 or more birds with beak damage or abnormalities

1: Normal untrimmed beak

2: Broken beak

3: broken and cloven tip of beak

4: broken tip of beak

© van Niekerk, WUR

Title	**Beak trimming**
Scope	Animal-based measure: Laying hens
Sample size	Sample size according to § 5.1.5
Method description	Beak trimming is painful for bird. Severe trimming leads to abnormalities of the beak and a higher risk for chronic pain. Abnormally shaped beaks may impair birds in their foraging, drinking and preening behaviour. Score 100 birds - pick up the birds from within the penned group or from the litter or slatted floor. In cage systems take the birds from different areas of the house and from different tier levels: Examine the beak on both sides according to classification presented in the photographic reference.
Classification	**Individual level:** **0** – No trimming **1** – Moderate to light trimming with moderate to no abnormalities; lower beak should not be longer than upper beak **2** – Severe abnormalities or severe trimming, with clear abnormalities or lower beak is longer than upper beak


Score=0


Score=1


Score=2


Score=0


Score=1


Score=2

© White hens: Gunnink, WUR; © Brown hens: van Niekerk, WUR

3.3.2 Module Beaks – first draft

In Table 3.7 the first draft of a beak module is given. This draft made a distinction between the upper and lower beak and was developed to evaluate intact beaks. For trimmed beaks the part about abnormalities can be used. With this part also broken beaks and deformed beaks can be scored. As it is difficult to make distinctions in severity of abnormalities, a 2-point scale was used, where all abnormalities were scored the same.

Table 3.7 *WQ module beaks first draft*

Tip	Lower beak	Abnormalities
0: No damage	0: No damage	0: No abnormalities
1: Blunt	1: Blunt	1: Abnormal
2: Sharp and/or overgrowth	2: Sharp and/or overgrowth	
3: Broken and/or split	3: Broken and/or split	

All scores were visually determined, except for the determination of sharp/blunt for which the observer ran his finger along the beak to feel weather it was sharp or blunt.
The module was tested in the experimental unit Carus. Lower beak condition and beak abnormalities were in all cases assessed as no damage. In Table 3.9 the outcome of the Kendall W test for reliability between all observers is presented. For round 1 and the HFP-line the reliability is reasonable, but for round 2 and the LFP-line it is very low.

Table 3.9. *Kendall's W for beak tip*

Dataset	Tip
All	0.244*
Round 1	0.422*
Round 2	0.144
High FP-line	0.389*
Low-FP line	0.120

* Significance with p=0.05

3.3.3 Module Beaks – Second draft

As the first draft of the beak module didn't give very reliable results a new testing round was done at Carus (white flock) with one observer and in 2 commercial flocks (white and brown) with two observers (the same in both flocks). The beaks were scored using the scores of Table 3.7. Additionally a micrometer was used to measure the horny rim of the beak tip and lower beak. This is the part of the beak starting where the living tissue ends up to the very end of the beak, as indicated in figure 3.3 for the beak tip rim. Also the amount of overgrowth of the upper beak over the lower beak was measured with the micrometer, as indicated in figure 3.4 The beak was held tight and the part of the upper beak extending the lower beak was measured.

Although the first measurements at Carus indicated that it is possible to measure the overgrowth of the upper beak over the lower beak, the comparison of the outcomes of the two observers on the first commercial farm indicated a very low correlation (0.37) between the measurements. The correlation between the measurements of the 2 observers for the horny rim of the lower and upper beak was good (0.71 for the tip, 0.84 for the lower beak). However, there was no correlation between the measurements of the lower beak and the upper beak and measuring the lower beak was time consuming. As damage due to pecking is mainly caused by the upper beak, it was decided to stop the measurements on the lower beak. On the second commercial farm therefore only measurements on the upper beak were carried out.
The correlations between the measurements of the two assessors on the first commercial farm are presented in Table 3.8. As can be seen the correlation between the measurement of lower beak and upper beak is very low for both assessors. The correlation between tip rim length and tip sharpness score was fairly high for both assessors, indicating that longer tips rims tended to be sharper. However, when repeated this measure in a white flock, the correlation was much lower for both assessors. This indicates could be the effect of the different shape of the beaks of white hens, but could also indicate that the scoring method is not very reliable.
As using the micrometer is difficult under dark conditions, on the second commercial farm a test was done with a 3-point scale for length of beak tip rim, where 0 = hardly any, 1 = < 2.5 mm and 2 = >2.5 mm. In Table 3.9 the correlations between the various scores is given for both assessors. The correlation between tip rim in mm and tip rim score was quite good, but the rim score had no correlation with the sharpness score.

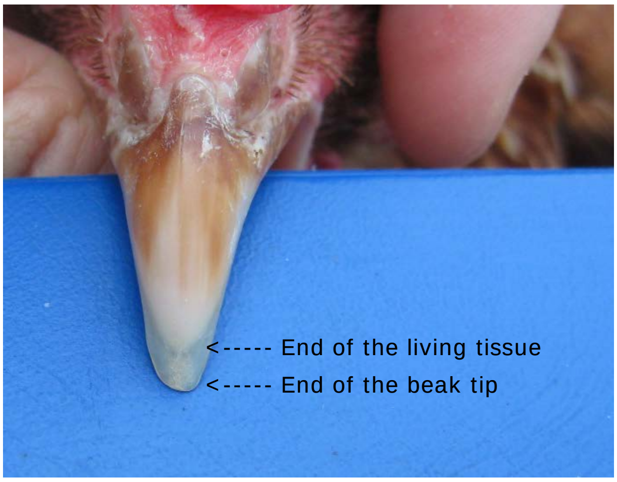


Figure 3.3: Beak tip rim, being the part between the end of the living tissue and the end of the beak tip

Figure 3.4: *Overgrowth of the upper beak over the lower beak*

Table 3.8 *Correlations beak measurements in a brown commercial flock with the second draft module*

	Observer 1	Observer 2
mm upper beak - mm lower beak	0.07	0.24
mm tip rim – sharpness score*	**0.53**	**0.66**

* Sharpness score: 0=blunt tip, 1=moderate sharp tip, 2=sharp tip

Table 3.9 *Correlations beak measurements* in a white commercial flock with the second draft module*

	Observer 1	Observer 2
mm tip rim – sharpness score	0.44	0.12
mm tip rim – tip rim score	0.46	0.60
tip sharpness score – tip rim score	0.09	0.16

* Sharpness score: 0=blunt tip, 1=moderate sharp tip, 2=sharp tip
  Tip rim score: 0 = hardly any, 1 = < 2.5 mm and 2 = >2.5 mm

3.3.4 Module Beaks – Third draft

Based on the results of the first and second draft module the third and final draft was reduced to two measurements:

- Beak tip rim: visually the size of the rim was estimated on a 3-point scale
- Sharpness of the beak: by running a finger across the beak tip the sharpness of the beak was determined.

The classes are given in Table 3.10.

Table 3.10 *WQ module beaks third draft*

Beak tip rim	Beak sharpness	Abnormalities
0 = hardly any	0 = blunt	0 = no abnormalities
1 = < 2.5 mm	1 = moderate	2 = abnormal
2 = ≥ 2.5 mm	2 = sharp	

3.3.5 Combined WQ-protocol with modules

In Table 3.11 the combined scores of the modules and the WQ-protocol are given, showing how the scores of the module can be brought back to scores of the original WQ-protocol.

Table 3.11 *Combination of WQ protocol and WQ module beaks*

WQ Protocol		WQ module		
Beak damage and abnormalities (not caused by trimming)	Beak trimming	Tip rim	Beak sharpness	Abnormalities
0 – No evidence of beak damage or abnormalities	0 - No trimming	0: hardly any	0: blunt	0: No abnormalities
		1: < 2.5 mm	1: moderate	
		2: ≥ 2.5 mm	2: sharp	
1 - Fewer than 3 birds with beak damage or abnormalities	1 - Moderate to light trimming with moderate to no abnormalities; lower beak should not be longer than upper beak			2a: Abnormal
2 - 3 or more birds with beak damage or abnormalities	2 - Severe abnormalities or severe trimming, with clear abnormalities or lower beak is longer than upper beak			

3.4 Results 10 commercial farms

3.4.1 Farm results

In Table 3.12 some characteristics of the flocks are given, that have been monitored. Selection of the flocks was done on availability and willingness of the farmer to co-operate, but also the aim has been to get an equal number of white and brown flocks. This has been achieved. All flocks were housed in aviaries. If a covered veranda wasn't used, it is considered not present.

Table 3.12 *Characteristics of the 8 commercial farms*

Farm code	Number of birds on site	Age (w)	Genotype	System	Free range	Covered veranda	sections in the house	Type of section division
1W	38321	74	Decalb White	aviary	yes	yes	7	wire
2B	34200	49	Novogen Light brown	aviary	no	yes	10	wire
3W	10720	78 [1)]	Novogen White	aviary	yes	yes	4	closed
4W	9402	29	Decalb White	aviary	no	no	2	wire
5B	9000	33	Novogen Brown Light	aviary	yes	yes	3	closed
6B	73000	50	Lohman Brown Classic	aviary	no	no	4	wire
7B	73000	30	Lohman Brown	aviary [2)]	no	no	4	wire
8W	32800	88	Decalb White	aviary	no	no	3	wire

[1)] organic, moultred, back in lay

[2)] 2-level house

Birds work the dirt in a free range area with their beaks, making it possible that the soil caused some wear on the beak tip. Also the stocking density in the house is lower, reducing the chance for collisions to the furniture, at the moments a part of the birds is in the free range. However, the presence of free range or a covered veranda, the number of sections and flock size had no high correlation with beak scores or keel bone scores. Only the age of the flock had a high correlation with keel bone deviations and keel bone fractures (Table 3.13).

Table 3.13 *Correlations between beak and keel bone scores and characteristics of the henhouse*

	Beak tip	Beak sharpness	Keel deviation	Keel fractures
Free range	-0.537	-0.522	0.063	0.021
Veranda	**-0.622**	**-0.752**	-0.013	-0.069
No sections	-0.445	-0.239	-0.043	-0.036
Flock size	0.160	0.489	0.015	-0.021
Age (w)	0.302	0.119	**0.850**	**0.910**

Pecking at mineral blocks can make the beaks blunt. Therefore this aspect was recorded. As can be seen in Table 3.14 all flocks had pecking stones. Mostly they are used well. One flock didn't use the blocks. This flock didn't have access to pecking blocks in the rearing period, which could be the reason for not using the blocks. Experiences on commercial farms indicate that hens don't use pecking blocks well if they haven't been provided during rearing. The other flock that didn't use the pecking blocks well had a different type of blocks compared to the other farms. The farmer indicated that the blocks were harder than the commonly used blocks and therefore less used. Surprisingly the flocks that didn't use the pecking blocks had the least sharp beaks. Also the beak tip rim was not different in length compared to the flocks that did use the pecking blocks.

Table 3.14 *Pecking blocks and beak scores in 8 commercial flocks*

Farm code	Pecking blocks in rear	Pecking blocks in lay	Type of pecking blocks in lay	Use of pecking blocks	Beak tip	Beak sharpness
1W	yes	yes	Sand-lime brick [1)]	much	1.20	1.51
2B	yes	yes	Mineral blocks [2)]	much	1.46	1.41
3W	no	yes	Aerated blocks	no	1.59	1.39
4W	yes	yes	Aerated blocks	much	1.50	1.51
5B	yes	yes	Aerated blocks [3)]	little	1.43	1.38
6B	yes	yes	Aerated blocks	much	1.66	1.50
7B	yes	yes	Aerated blocks	much	1.56	1.50
8W	yes	yes	Aerated blocks	much	1.92	1.53

1) Up to 30 w mineral blocks

2) +alfalfa +loose straw

3) Aerated blocks are bought from builder and are harder than the usual ones. Therefore used less.

The incidence of keel bone fractures can be influenced by collisions hens have with furniture. This can be related to the necessity to jump through the system and the risk of falls (Heerkens et al. 2016). Also, light intensity can influence the number of collisions with perches (Moinard et al. 2004). In Table 3.15 the smallest and largest width of aisles of the aviaries are given as well as the light intensity. As can be seen there is a very large variation in aisle width both between and within systems. This has to do with the type of system and how to fit it into existing buildings. Light intensity also varies a lot, ranging from almost dark to way above 20 lux, the often advised minimum light intensity. Average keel bone fractures and deviations are also given. Deviations mostly range between Slight (hardly visible) and Moderate (max. 0.5 cm), with one exception of hens with almost severe deviations (> 0.5 cm). No strong correlations were found between aisle width and keel bone deviations or fractures. Also light intensity had no correlations with keel bone deviations or fractures. All correlations are given in appendix 1.

Table 3.15 *Aisle width, light intensity and average keel bone deviations and fractures*

Farm code	Aisle (smallest; cm)	Aisle (widest; cm)	Light (lux)	Keel deviation	Keel fractures
1W	120	120		1.67	0.39
2B	85	260	2.2	1.57	0.27
3W	190	365	7.2	1.90	0.44
4W	105	145	32.5	1.11	0.13
5B	96	327	26.2	1.59	0.21
6B	140	225	19.2	1.79	0.33
7B	141	282	30.7	1.35	0.18
8W	125	310	24	2.53	0.77

In Table 3.16 the behavioural observations are given. Given are the total number of observed bird movements in 2x 5 min observing (5 min per observer). Although the numbers are not very informative, the comparison between flocks does give some information.
In flock 1 the movements are mainly horizontal and crossing an aisle. The latter is a risk for keel bone fractures. The keel bone damage in flock 1 (0.39; Table 3.15) is slightly above the average, being 0.34.
Flock 2 has a lot of vertical movements, especially downwards. Theoretically this would increase the risk for keel bone damage, especially with low light intensities. Flock 2 also has many horizontal movements that don't cross an aisle, which theoretically are positive in preventing keel bone damage. The actual level of keel bone damage in flock 2 (0.27) is slightly below the average, being 0.34.

Flock 3 and 4 are quite similar in the movements of the birds, but they differ in keel bone damage. The same accounts for flock 6 and 7, who also have similar movement patterns, but differ in keel bone damage. Looking at the light intensities of these 4 flocks, the flocks with the higher light intensities have less severe keel bone damage, which is in compliance with the literature (Moinard et al. 2004). Flock 8 had an average number of vertical movements, but a lot of horizontal movements crossing an aisle. Also it has the highest number of birds falling or clashing in obstacles. This is in accordance with the relatively high score for keel bone fractures. The high number of keel bone deviations may well contain some fractured bones as well, as more extreme deviations are hard to distinguish from fractures.
Table 3.17 provides the correlations between the behavioural observations and the beak and keel bone scores. Fairly high correlations are found between keel bone fracture scores and crossing an aisle or falling/clashing into furniture, which is in accordance with the literature. Surprisingly these two behaviours also have a high correlation with beak sharpness. None of the other correlations are high enough to be considered relevant.

Table 3.16 *Behavioural observations movements through the aviaries*

Farm code	Flying vertical down	Flying vertical up	Jumping platform-perch same side	Jumping platform-perch other side	Falling / clashing
1W	18	18	2	34	1
2B	78	26	62	0	1
3W	21	38	15	4	0
4W	43	45	0	6	1
5B	30	25	11	2	0
6B	38	42	25	19	1
7B	68	35	39	22	0
8W	28	30	0	49	2

Table 3.17 *Correlations between behavioural observations and beak characteristics and keel bones*

	Beak tip	Beak sharpness	Keel deviation	Keel fractures
Flying vertical down	0.007	-0.068	-0.448	-0.465
Flying vertical up	0.477	0.194	-0.232	-0.226
Jumping platform-perch same side	-0.072	-0.360	-0.248	-0.340
Jumping platform-perch other side	0.374	**0.753**	**0.625**	**0.725**
Falling / clashing	0.415	**0.629**	0.514	**0.622**

Table 3.18 provides the correlations between beak characteristics and keel bone scores. The strong correlation between keel bone deviations and keel bone fractures may be caused by the difficulty to distinguish the two if the deviation is more severe. A surprising correlation is found between keel deviation and beak tip rim.

Table 3.18 *Correlations between beak and keel bone characteristics*

	Beak tip	Beak sharpness	Keel deviation	Keel fractures
Beak tip		0.234	**0.633**	0.597
Beak sharpness	0.234		0.086	0.245
Keel deviation	**0.633**	0.086		**0.965**
Keel fractures	0.597	0.245	**0.965**	

3.4.2 Inter-observer reliability

Per flock 100 birds were scored individually. Each bird was scored by two observers, making it possible to determine the inter-observer reliability and thus the reliability of the scoring method.
In Table 3.19 the correlations are given between the scores of both observers. In general the correlations were very high, except for the scoring of the beak sharpness, that was slightly lower. The genotype of bird (white or brown) did not affect the scoring.

Table 3.19 *Correlations between scores of two observers*

Bird type Scoring	Brown	White	All
Beak tip	0.885	0.844	0.865
Beak sharpness	0.633	0.650	0.642
Keel deviation	0.888	0.886	0.889
Keel fractures	0.905	0.855	0.878

In Figure 3.5 the beak and keel bone scores of both observers are given for white and brown hens. The variation in the scores is small and both observers scored quite equally. Beak scores were on average slightly higher for white genotypes, which was consistent for both observers.

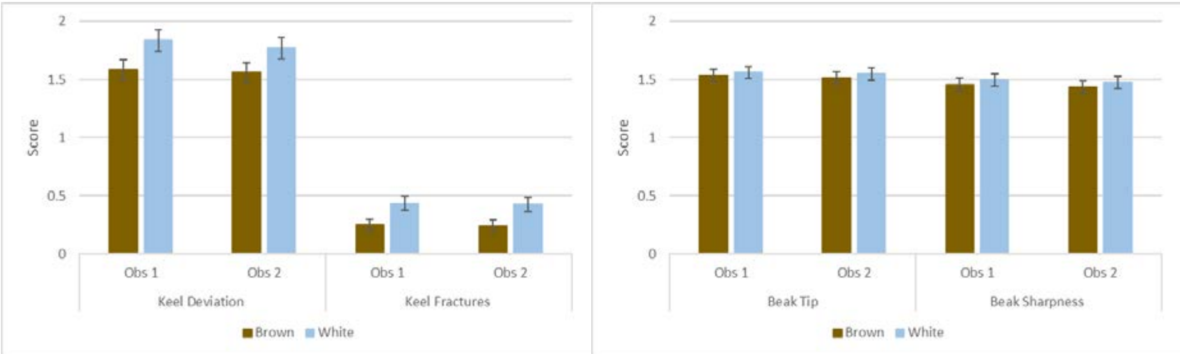


Figure 3.5 *Beak and keel bone scores of two observers for two genotypes (+SEM)*

In Figure 3.6 the beak scores per flock and per observer are given, including the standard error of means (SEM). White flocks are on the left side of the graph, brown flocks are on the right side. The scoring of the beak tip showed more variation between flocks for white flocks compared to brown flocks. For the sharpness score this effect was not present. Observer 2 had slightly lower beak tip scores then observer 1 for flock 2B. For all other flocks the beak tip and beak sharpness scores of both observers were equal.
In figure 3.7 the keel bone scores per flock and per observer are given. For keel deviations observer 2 scored lower than observer 2 in flock 1. For all other flocks both observers scored equal.

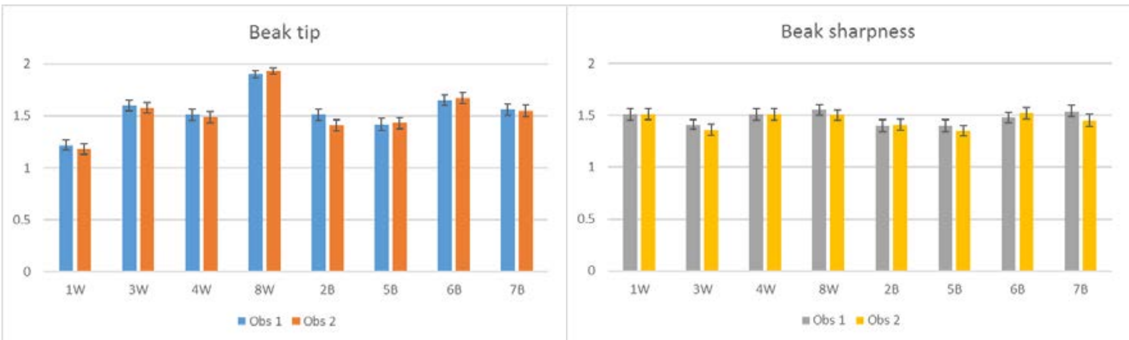


Figure 3.6 *Beak scores of two observers for eight flocks (+ SEM)*

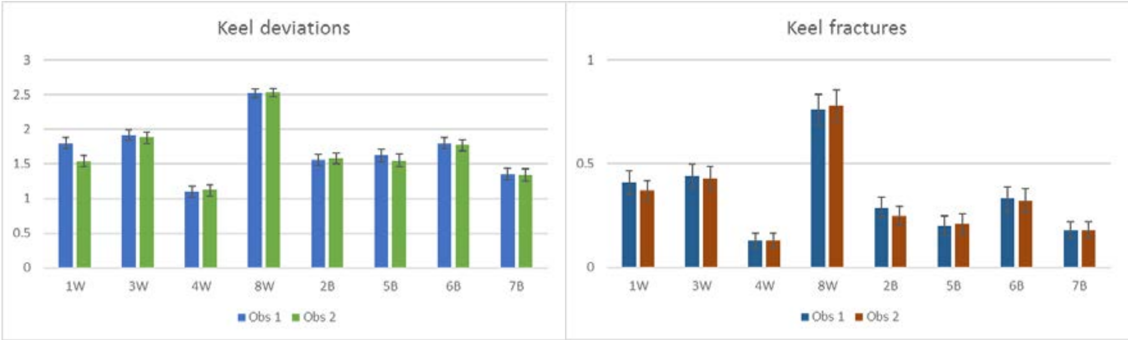


Figure 3.7 *Keel bone scores of two observers for eight flocks (+ SEM)*

4 Discussion and conclusions

Modules on top of the WQ-protocol are developed with the aim to get more or more specific information out of the assessment without losing the possibilities to get an original WQ-assessment. The development has been done in several steps. The first steps have been carried out by a MSc student of Wageningen University. He first developed 3 modules for 3 areas of concern: feather damage, beaks and keel bones. It appeared difficult to develop a detailed protocol with sufficient reliability and repeatability. After some tests the classification had to be made more simple, because the differences between the classes were not clear enough, making the scoring unreliable. The repeatability increases if the classifications are clear and thus scoring is simple. This is one of the difficult aspects of developing a scoring system: on the one hand one wants to have more classes to be able to detect small differences, but on the other hand it is hard to define clear cutting points in the classification, making it often necessary to reduce the number of classes. Protocols used by certifying bodies therefore often have few, but clear defined classes. Protocols used for research purposes often have more classes, but then it requires training to get a consistent result. Protocols that give some more detail, but are still applicable without too much training would be the ideal for commercial farms to get insight in specific issues.

After some testing the module for feather damage was not further developed, because it was unclear what the meaning of the detected feather damage was. The aim was to develop a protocol for early detection of feather pecking problems, but the test results gave reason to doubt a strait forward relation between the detected feather damage and feather pecking. Often parts of the feather shafts were broken on so-called stress lines. However, this mainly was found in the LFP-line, in a flock were earlier research has found hardly any feather pecking. Also, in other research projects similar effects were seen and when zooming in on various feather deformations for which the origin was unclear (Janssen et al. 2017). Therefore it was concluded that before continuing the development of a feather damage module, various questions need to be answered: what causes these stress lines? Do these lines have any relation with feather pecking or stressors inducing feather pecking? What causes parts of the shaft to break off? Is it pecking or collision into furniture? After research has revealed the causation and meaning of various types of feather damage and feather irregularities the further development of a feather scoring module can be started.

The module for keel bone assessment started with more categories to distinguish between various extents of deviations and fractures. It appeared difficult to get clear distinctions for the different categories and at first the reliability of scoring was low, as calculated by the student. With experienced observers it was possible to get a good reliability of scoring the deviations on a 4-point scale and the fractures on a 3-point scale, as was shown in the trial on commercial farms. This means that this module can only be used successfully after thorough training of the observers.

The module for beaks aimed at scoring intact beaks, as the original WQ-protocol was more aiming at abnormalities caused by beak trimming. Intact beaks can differ in sharpness and thus in effect when pecking at conspecifics. Often the ‘hook’ is mentioned as the causation of injuries. Also, beaks differ in sharpness. Finally intact beaks can get caught in feed chains and be injured (broken/split). As this was seen in earlier monitoring projects, this aspect was incorporated in the protocol. However, as the frequency appeared to be very low, this was not further scored. Any broken or split beaks can be scored under abnormalities. In the first draft module the sharpness and overgrowth was taken into one category. The reliability of the scoring was low, so the module was modified. Sharpness and overgrowth of the beak were then scored separately. As it appeared very difficult to score the overgrowth of the beak, it later was excluded from the module. The final module focusses on the beak tip rim and sharpness. The reliability of the scoring was good for the beak tip and reasonable for sharpness.

The scoring done on eight farms had an additional goal, i.e. to get an idea of relations with management and lay-out of the housing system. In relation with the beak tip and sharpness the idea was to compare flocks with and without pecking blocks. As all flocks had pecking blocks, a comparison could not be made. For the two flocks that didn't use the pecking blocks well, the outcome was contrary to what was expected. As pecking blocks are causing wear of the beaks one would expect blunter beaks with less tip rim, which was not the case. We have no explanation for this. The two flocks with low pecking block use both had outdoor access, so pecking in the dirt may have caused some wear of the beaks. Further, the results of only two flocks are not representative and chance may very well have played a role.

As for correlations between lay-out of the system and behaviour, no strong correlations could be found with keel bone deviations and fractures. Only jumping from a platform to a perch on the other side of an aisle and falling/clashing into furniture had relationships with increased deviations and fractures, which is according to the expectation. When birds have hard landings when jumping, this can cause fractures. The relation with keel deviations is not that clear, as deviations are thought to originate from resting with the keel bone on perches. However, it appeared difficult to distinguish well between more extreme deviations and fractures, as was also shown by the high correlation between the two. Also in the literature it is indicated that larger deviations mostly are caused by fractures (Scholz et al. 2008). Although from the work of Moinard et al. (2004) a relation between keel bone damage and aisle width or light intensity could be expected, this didn't come out of our research. This may have been caused by the limited number of flocks and the large variation in lay-out of systems. We tried to catch the important features of each system with aisle width and jumping possibilities, but probably a more detailed system or different characterisation needs to be made.
A surprising outcome was the relation between beak sharpness and falling andclashing. An explanation could be that birds that are pecked with a sharp beak are more likely to suddenly react. This might impair their judgement of good landings, causing them to fall. This could also be the explanation for the relation between beak tip rim and keel deviations, if longer beak tips would be sharper. However the relation between beak tip rim and sharpness was very low. We therefore have no explanation why longer beaks are correlated with keel deviations. The limited number of flocks and measurements make it hard to draw hard conclusions.

In conclusion one can state that two of the three modules that were developed actually can be applied in a meaningful way to get more detailed information. For feather cover more research on small feather deformities is needed before continuing the development of this module. The initial goal to score at an earlier age than 50 weeks especially was mainly aimed at feather cover. For beaks it is not known how these develop over time. For keel bone damage it is known that this increases over time, although the majority of fractures have occurred by 50 weeks of age (Petrik et al. 2015). This makes an assessment at early age not useful.

References

Heerkens, J. L. T., E. Delezie, B. Ampe, T. B. Rodenburg, and F. A. M. Tuyttens. 2016. Ramps and hybrid effects on keel bone and foot pad disorders in modified aviaries for laying hens. *Poultry Science* 95 (11): 2479-2488. http://dx.doi.org/10.3382/ps/pew157.

Janssen, T., J. v. Rooij, T. G. C. M. v. Niekerk, and N. d. Bruijn. 2017. Veren, wat vertellen ze ons? *Pluimveehouderij* 47: 22-23.

Martin, P., and P. Bateson. 2007. *Measuring Behaviour, an introductory guide*. Cambridge: Cambridge University press.

Moinard, C., P. Statham, M. J. Haskell, C. McCorquodale, R. B. Jones, and P. R. Green. 2004. Accuracy of laying hens in jumping upwards and downwards between perches in different light environments. *Applied Animal Behaviour Science* 85 (1-2): 77-92. http://dx.doi.org/10.1016/j.applanim.2003.08.008.

Petrik, M. T., M. T. Guerin, and T. M. Widowski. 2015. On-farm comparison of keel fracture prevalence and other welfare indicators in conventional cage and floor-housed laying hens in Ontario, Canada. *Poultry Science* 94 (4): 579-585. http://dx.doi.org/10.3382/ps/pev039.

Scholz, B., S. Rönchen, H. Hamann, M. Hewicker-Trautwein, and O. Distl. 2008. Keel bone condition in laying hens: A histological evaluation of macroscopically assessed keel bones. *Berliner und Munchener Tierarztliche Wochenschrift* 121 (3-4): 89-94. http://dx.doi.org/10.2376/0005-9366-121-89.

Appendix 1 Correlations

	Beak tip	Beak sharp-ness	Keel deviation	Keel fractures	Flying vertical down	Flying vertical up	Jumping platform-perch same side	Jumping platform-perch other side	Falling / clashing	Free range	Veranda	No sections	Flock size	Age (w)	Aisle smallest (cm)	Aisle widest (cm)	Light
Beak tip		0.234	**0.633**	0.597	0.007	0.477	-0.072	0.374	0.415	-0.537	**-0.622**	-0.445	0.160	0.302	0.303	0.520	0.091
Beak sharpness	0.234		0.086	0.245	-0.068	0.194	-0.360	**0.753**	**0.629**	-0.522	**-0.752**	-0.239	0.489	0.119	-0.021	**-0.617**	0.586
Keel deviation	**0.633**	0.086		**0.965**	-0.448	-0.232	-0.248	**0.625**	0.514	0.063	-0.013	-0.043	0.015	**0.850**	0.301	0.454	-0.282
Keel fractures	0.597	0.245	**0.965**		-0.465	-0.226	-0.340	**0.725**	**0.622**	0.021	-0.069	-0.036	-0.021	**0.910**	0.300	0.297	-0.223
Flying vertical down	0.007	-0.068	-0.448	-0.465		0.114	**0.860**	-0.338	-0.093	**-0.665**	-0.184	0.440	0.388	-0.568	-0.398	0.014	-0.142
Flying vertical up	0.477	0.194	-0.232	-0.226	0.114		-0.055	-0.248	-0.071	-0.482	**-0.651**	-0.573	0.094	-0.307	0.395	0.046	0.297
Jumping platform-perch same side	-0.072	-0.360	-0.248	-0.340	**0.860**	-0.055		-0.422	-0.244	-0.374	0.158	**0.674**	0.429	-0.323	-0.173	0.218	-0.558
Jumping platform-perch other side	0.374	**0.753**	**0.625**	**0.725**	-0.338	-0.248	-0.422		**0.635**	-0.173	-0.427	-0.144	0.389	0.581	0.128	-0.174	0.350
Falling / clashing	0.415	**0.629**	0.514	0.622	-0.093	-0.071	-0.244	**0.635**		-0.488	-0.378	0.097	0.078	0.474	-0.316	-0.364	0.009
Free range	-0.537	-0.522	0.063	0.021	**-0.665**	-0.482	-0.374	-0.173	-0.488		**0.775**	0.013	-0.495	0.276	0.254	0.157	-0.211
Veranda	**-0.622**	**-0.752**	-0.013	-0.069	-0.184	**-0.651**	0.158	-0.427	-0.378	**0.775**		0.562	-0.488	0.212	-0.081	0.170	**-0.680**
No sections	-0.445	-0.239	-0.043	-0.036	0.440	-0.573	**0.674**	-0.144	0.097	0.013	0.562		0.175	0.158	-0.311	-0.176	**-0.780**
Flock size	0.160	0.489	0.015	-0.021	0.388	0.094	0.429	0.389	0.078	-0.495	-0.488	0.175		-0.108	0.146	-0.152	0.108
Age (w)	0.302	0.119	**0.850**	**0.910**	-0.568	-0.307	-0.323	0.581	0.474	0.276	0.212	0.158	-0.108		0.433	0.177	-0.470
Aisle smallest (cm)	0.303	-0.021	0.301	0.300	-0.398	0.395	-0.173	0.128	-0.316	0.254	-0.081	-0.311	0.146	0.433		0.388	-0.159
Aisle widest (cm)	0.520	**-0.617**	0.454	0.297	0.014	0.046	0.218	-0.174	-0.364	0.157	0.170	-0.176	-0.152	0.177	0.388		-0.382
Light	0.091	0.586	-0.282	-0.223	-0.142	0.297	-0.558	0.350	0.009	-0.211	**-0.680**	**-0.780**	0.108	-0.470	-0.159	-0.382	

Appendix 2 Scoring protocol 8 commercial farms

Horizontal and vertical movement

Horizontal movement	0 = no clear obstacles 1 = obstacles, but birds can negotiate them fairly easy or wide aisles (>1 m) without obstacles 2 = obstacles prevent birds from moving freely or wide aisles (> 1 m) with vertical obstacles (e.g. wires, lights)
Vertical movement	0 = birds to go up and down easily (= they don't need to fly, except to litter) 1 = obstacles, but birds can negotiate them fairly easy (= they can walk to other, but not all floors or there are easy to take steps) 2 = birds have difficulty to go up and down (= hardly any in-between to go up or down) 3 = birds are obstructed to go up and down (= no in-between to go up or down)

System:

Aviary system	0 = internal stairway
	1 = extended perches
	2 = external platforms
	3 = external stairways
Aisle	Width of aisle in cm (extended perches not calculated)
Light	Light intensity (lux) (measure 30 cm above litter in middle of internal aisle, Between lights, sensor horizontal)
Visibility of jumps	0 = hens can see perches/platforms on other side of aisle good
	1 = hens can see perches/platforms on other side of aisle less good due to obstacles blocking (part of) the sight or due to very dim light

Behavioural observations:

Record during 5 minutes in one middle aisle the number of birds:	
Flying vertical	Down
	Up
Jumping from platform/perch to platform/perch	On same side
Falling / clashing	To other side

Clinical Scoring:

Beak tip rim	0= hardly any 1=< 2.5 mm 2= ≥ 2.5 mm
Beak sharpness	0=blunt 1=moderate 2= sharp
Keel deviation	0=no deviation 1=slight 2=moderate 3=severe
Keel fractures	0=no fracture 1=1 fracture 2= more fractures

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.