Wageningen UR Livestock Research

Partner in livestock innovations

Report 589

Welfare Quality® assessment protocol for laying hens

Results assessment of 122 flocks

May 2012

LIVESTOCK RESEARCH
WAGENINGEN UR

Colophon

Publisher
Wageningen UR Livestock Research
P.O. Box 65, 8200 AB Lelystad
Telephone +31 320 - 238238
Fax +31 320 - 238050
E-mail info.livestockresearch@wur.nl
Internet http://www.livestockresearch.wur.nl

Editing
Communication Services

Wageningen UR Livestock Research and Central Veterinary Institute of Wageningen UR, both part of Stichting Dienst Landbouwkundig Onderzoek (DLO Foundation), together with the Department of Animal Sciences of Wageningen University comprises the Animal Sciences Group of Wageningen UR (University & Research centre).

Single numbers can be obtained from the website.

ISO 9001 certification by DNV emphasizes our quality level. All our research projects are subject to the General Conditions of the Animal Sciences Group, which have been filed with the District Court Zwolle.

Abstract
Results of a study on the Welfare Quality® assessment protocol for laying hens. It reports the development of the integration of welfare assessment as scores per criteria as well as simplification of the Welfare Quality® assessment protocol. Results are given from assessment of 122 farms.

Keywords
laying hens, Welfare Quality, welfare assessment

Reference
ISSN 1570 - 8616

Author(s)
T.G.C.M. van Niekerk
H. Gunnink
K. van Reenen

Title
Welfare Quality® assessment protocol for laying hens

Report 589

LIVESTOCK RESEARCH
WAGENINGEN UR

Report 589

Welfare Quality® assessment protocol for laying hens

Welfare Quality® assessment protocol voor leghennen

T.G.C.M. van Niekerk
H. Gunnink
K. van Reenen

May 2012

This study was financed by the Ministry of Economic Affairs, Agriculture and Innovation.

Ministerie van Economische Zaken, Landbouw en Innovatie

Preface

Between 2004 and 2009 a method was developed within the Welfare Quality® project to assess animal welfare on cattle, pig and poultry farms. The resulting Welfare Quality® assessment protocols, published in 2009, provide a detailed account of the necessary measurements, and how these can be combined to provide a single overall statement of the state of welfare on the farm assessed. The method has attracted a lot of interest from European and national policy makers, NGO's and the farming community, but has to date not been adopted in any commercial scheme nor used by individual farmers to improve animal welfare on their farm. The main drawback seems to be the amount of time required to carry out the measurements. In 2010 the Dutch ministry of Economic Affairs, Agriculture and Innovation commissioned Wageningen UR Livestock Research to do a series of studies aiming to simplify and shorten the original protocols. In collaboration with former Welfare Quality® partners and the Dutch laying hen sector, farm visits were organised and data collected between April and August 2011. This report describes the findings based on analyses of the collated data, along with an attempt to find simpler alternatives for the time consuming measurements. The results will be presented to the international Welfare Quality® Network, which is working on further improvement of the protocols. They will also be recommended to the Dutch Ministry, who can introduce them for improvement of farm animal welfare in collaboration with the Dutch laying hen sector.

Without these three stakeholder groups, the Dutch Ministry, representatives from the laying hen industry and the Welfare Quality® Network this work would not have been possible. On behalf of the project team I would like to thank Bart Crijns, Amanda Manten, and Léon Arnts (Ministry of Economic Affairs, Agriculture and Innovation) and Henk Hulsbergen (PPE) for their contributions to this work.

Paul Vriesekoop
April 2012

Summary

A protocol for the assessment the welfare of laying hens has been described in the Welfare Quality® assessment protocol for poultry (Welfare Quality®, 2009). This protocol describes measurements indicative of laying hen welfare on-farm. Although individual measurements are described, extrapolation of the scores to flock level is lacking. In this report we present an extrapolation. The fundamental calculations are described in a separate report (Van Niekerk et al., 2012).
On average, it takes approximately 7 hours per flock to perform the laying hen welfare assessment protocol (Welfare Quality®, 2009). For practical use and acceptance a reduction in time is demanded. Therefore, on-farm time saving methods have been investigated.
The aim of the project was twofold: 1. to integrate individual measurementss to allow extrapolation of scores for each criterion at flock level; 2. An investigation of the potential for time saving operations aimed at a simplification of measurement performance. This in an attempt to improve applicability of the assessment protocol.
In order to determine whether or not simplification of the laying hen assessment protocol is possible, the protocol for laying hens (Welfare Quality®, 2009) was used to assess a large number (122) of flocks. A flock was defined as birds housed together in a single shed on a farm. Flocks were kept in various housing systems, ranging from traditional cages, with limited space and no furniture, to organic systems with ample space, litter, nest boxes, perches and a free range area.
Data were provided from 2 different countries: 74 Dutch flocks (29 assessed in 2007-2008 and 45 flocks assessed in 2011) and 48 Swedish flocks (assessed in 2007). It was possible that each farm could have more than one layer house, implying that more flocks could be assessed per farm. In addition, as assessment of the Dutch flocks was performed over different years, some farms were revisited later and thus in that case more flocks were assessed from the same house. In total 5 assessors (4 in the Netherlands and 1 in Sweden), all trained in the theory and practice by experts, performed the data collection. Assessment of the flocks was performed on flocks of at least 50 weeks of age according to the Welfare Quality® laying hen assessment protocol (Welfare Quality®, 2009).
Results from all measurements are presented. Because the laying hens were housed in different systems, the results are presented in four classes of housing systems: regular (non organic) aviaries and floor systems, organic aviaries and floor systems, conventional cages and furnished cages (floor systems refer to systems with partly slatted floor and partly litter floor, manure pit underneath the slatted floor; furnished cages is the scientific correct name for enriched cages). Almost all measurements show considerable variation between flocks, as well as within housing systems.

Unfortunately, 51 flocks (43.2%) could not be classified due to missing information. The Welfare Quality® protocol is not very flexible on this point: a missing measurement restricts further calculation of one or more criteria and thus prohibits calculation of scores for all principles. These flocks should not be confused with the column Not Classified. The latter were classified, but ' Not classified' indicates that their scores were not high enough to fall into the category acceptable (or higher).
From all flocks 47 (39.8%) were classified as 'Acceptable', 15 (12.7%) were classified as 'Not classified' and 5 (4.2%) were classified as 'Enhanced'. No flocks were classified as 'Excellent'.
From the enhanced flocks 4 out of 5 were housed in furnished cages and no flocks from organic systems could be classified as 'Enhanced'. On the other hand no flocks from organic systems were classified as 'Not classified', which indicated a higher minimum welfare level than in the other systems.

Furnished cage flocks received high scores because first and foremost, all 4 enhanced furnished cage flocks were Swedish with good health status and intact beaks. This gave them a high score for principle 3. They also had a reasonable amount of litter and nest boxes available, resulting in a score above 20 for principle 4. Although space is limited in furnished cages, the presence of perches and a good climate in the house resulted in a score above 20 for principle 2. This together with ample feeder and drinker space, resulted in a classification as 'enhanced'.
Although the results are supported by findings of other researchers, some measurements may need improvement. The measurements in question lack refinement under certain circumstances, resulting in an over exaggeration of some negative aspects. To overcome this it is suggested that measurement of: feeder space, drinker space, keel bone deformities, aggression, toe damage and litter use should all be modified. Also the comparability of the Novel Object Test in cages and non-cage systems is questioned.

Welfare Quality® protocols are often seen as classification systems resulting in an estimation of overall flock welfare. Indeed this is the outcome. However, this only becomes relevant when certain

valuations are taken in relation to the welfare classification. For example, when a higher price or product labelling is allocated to eggs from flocks with a higher qualification. However, the Welfare Quality® protocol offers much more than just a qualification and can in fact be utilized as an on-farm management tool to improve animal welfare .
The Welfare Quality® protocol provides a good tool for:

- Identifying areas of concern
- Identifying aspects which can lead to improving animal welfare
- Comparing farms on overall welfare and on individual welfare criteria and measurements
- Comparing flocks on overall welfare and on individual welfare criteria and measurements

Examination of the results from assessments of various criteria can help farmers identify where possibilities for improvement lie. At measurement level, the protocol can be used as a management tool.

In order to examine possibilities for time saving, correlations were calculated, and a principal component analysis was performed to investigate whether or not various measurements in the laying hen protocol were correlated. These analyses clearly showed that no meaningful correlations exist between laying hen protocol measurements. Therefore, simplification strategies based on prediction equations were not an option.
Reduction of sample size was analysed in a simulation study involving random sampling of half and 25% of total sample size (100 birds), respectively, and repeating this procedure 100 times.
The results of this simulation study clearly show that reduction of the sample size from 100 to 50 or 25 hens is not a viable alternative. Reduction of sample size compromises specificity and/or sensitivity .
Although reduction of protocol performance time is not possible, this is not necessarily too great a problem if frequency of monitoring can be limited even more (normal frequency is already low, being once every flock and thus less than once per year). In order to reduce this frequency possibilities should be investigated for selective usage of measurements as signals to indicate when full protocol assessment is advisable. Examples of such indicators include: mortality and culling levels. Also measurements could be taken into account that are no part of the full protocol, e.g. numbers (type) of second grade eggs.

Samenvatting

In het Welfare Quality® assessment protocol for poultry (Welfare Quality®, 2009) is een protocol beschreven om het welzijn van leghennen vast te stellen. Dit protocol beschrijft metingen die indicatief zijn voor het welzijn van leghennen op het legbedrijf. hoewel de individuele metingen beschreven staan, mist de extrapolatie naar koppelscores. In het onderliggende rapport wordt deze extrapolatie gegeven. het rekenmodel dat hieraan ten grondslag ligt, is beschreven in een apart rapport (Van Niekerk et al., 2012).
Gemiddeld kost het protocol 7 uur per koppel om uit te voeren (Welfare Quality®, 2009). Vanwege de praktische uitvoerbaarheid en acceptatie is een reductie in tijd gewenst. Daarom zijn enkele opties doorgerekend die tijdsbesparing op het legbedrijf zouden kunnen betekenen.
Het doel van het project was tweeledig: 1. integratie van individuele metingen, zodat extrapolatie van de scores voor elk criterium op koppelniveau verkregen kan worden; 2. inventarisatie van tijdbesparende mogelijkheden, gericht op het versimpelen van de uit te voeren metingen. Dit zou de praktische toepasbaarheid van het protocol verhogen.
Om te onderzoeken of tijdsbesparing mogelijk is, was een grote database nodig. Hiertoe zijn 122 koppels met het protocol doorgemeten. Een koppel is gedefinieerd als hennen die op een bedrijf in dezelfde legstal gehuisvest zijn. De koppels weden gehuisvest in deiverse houderijsystemen, varierend van conventionele kooien, met beperkte ruimte en zonder extra faciliteiten, tot biologische houderijsystemen met veel ruimte, strooisel, nesten, zitstokken en buitenuitloop.
De gegevens werden verzameld in twee landen: 74 Nederlandse koppels: (29 beoordeeld in 2007-2008 en 45 beoordeeld in 2011) en 48 Zweedse koppels (beoordeeld in 2007). In een aantal gevallen zin meerdere koppels per bedrijf beoordeeld. Dit kon zijn, omdat er meerdere legstallen aanwezig waren. Ook zijn de beoordelingen in Nederland over meerdere jaren verspreid, waardoor een stal later opnieuw bezocht kan zijn, zodat meerdere koppels uit dezelfde stal beoordeeld zijn. In totaal hebben 5 beoordelaars (4 in Nederland en 1 in Zweden) de koppels bezocht. Zij waren allen door experts getraind in de theorie en praktijk van het protocol. Alle koppels die beoordeeld zijn, waren minimaal 50 weken oud, zoals voorgeschreven in het Welfare Quality® laying hen assessment protocol (Welfare Quality®, 2009).
De resultaten van alle metingen worden gepresenteerd. Omdat leghennen in diverse houderijsystemen gehuisvest zijn, worden de resultaten gepresenteerd voor 4 groepen van houderijsystemen: reguliere (niet biologische) volières en strooisel/rooster-systemen, conventionele en verrijkte kooien (incl. koloniehuisvesting). Alle metingen tonen flink wat variatie, zowel tussen koppels als binnen houderijsystemen.

Helaas konden 51 koppels (43,2%) niet geclassificeerd worden als gevolg van missende informatie. Het Welfare Quality® protocol is niet eg flexibel op dit punt: een missende meting verhindert verdere berekening van een of meer criteria en maakt berekening van de scores voor alle principes onmogelijk. Deze koppels dienen niet verward te worden met de groep 'not classified', hetgeen de classificatie-term is voor de koppels die niet hoog genoeg scoorden om in de klasses 'acceptable' of hoger te komen.
Van alle koppels werden er 47 (39,8%) geclassificeerd als 'Acceptable', 15 (12,7%) werd geclassificeerd als 'Not classified' en 5 (4,2%) bereikte de classificering 'Enhanced' . Geen enkel koppel viel in de klase 'Excellent' . Van de 5 'enhanced' koppels waren er 4 gehuisvest in verrijkte kooien. Geen enkel biologisch koppel bereikte de scoren 'enhanced', maar daarentegen waren ze allemaal hoger geclassificeerd als ' Not classified', hetgeen een hoger minimum welzijnsniveau suggereert als in de andere houderijsystemen.

De verrijkte kooi koppels die hoge scores bereikten waren alle 4 Zweedse koppels met goede gezondheid en niet-snavelbehandeld. Dit gaf ze een hoge score voor principe 3. Er waren daarnaast een redelijke hoeveelheid strooisel en legnesten aanwezig, waardoor de score voor principe 4 boven 20 uit kwam. Hoewel de ruimte in verrijkte kooien beperkt is, zorgde de aanwezigheid van een zitstok en een goed klimaat ervoor dat ook voor principe 2 een score verkregen werd die boven 20 lag. Samen met ruime voer- en watervoorzieningen resulteerde in de classificatie 'enhanced'.
Hoewel de resultaten overeenkomen met bevindingen van andere onderzoekers, kunnen enkele metingen verbeterd worden. De betreffende metingen zijn te grof onder bepaalde omstandigheden, waardoor ze een overdrijving van enkele negatieve aspecten geven. Om dit te verbeteren, wordt aanpassing van de volgende metingen aangeraden: metingen aan voer- en drinkwater, borstbeenvervormingen, agressie, teenbeschadigingen en gebruik van strooisel. Ook is het de vraag of de Novel Object Test vergelijkbare resultaten geeft in kooien en niet-kooi-systemen.

Welfare Quality® protocols worden vaak gezien als classificatiesystemen, die een overall score voor welzijn van een koppel geven. Dit komt inderdaad uit de berekeningen. Echter, dit is alleen relevant als er waarde wordt toegekend aan bepaalde classificaties, bijvoorbeeld een hogere eierprijs of een bepaalde product-labelling aan eieren van koppels met een hogere kwalificatie.
Het Welfare Quality® protocol bidet echter meer dan alleen een kwalificatie en kan goed gebruikt worden als bedrijfsmanagement systeem om het welzijn van de hennen te verbeteren.
Het Welfare Quality® protocol bidet geode handvatten om:

- probleemgebieden met betrekking tot dierenwelzijn te identificeren
- aspecten te identificeren die kunnen leiden tot verbetering van dierenwelzijn
- bedrijven te vergelijken op overall dieren welzijn en op individuele welzijnscriteria en -metingen
- koppels te vergelijken op overall dieren welzijn en op individuele welzijnscriteria en -metingen

het bestuderen van de resultaten van metingen van diverse criteria kan pluimveehouders helpen de mogelijkheden tot verbetering te bepalen. Op het niveau van metingen kan het protocol gebruikt worden als management tool.

Bij de berekeningen ter vereenvoudiging van het protocol zijn correlaties berekend en is een principle component analysis uitgevoerd om te zien of bepaalde metingen in het protocol gecorreleerd waren. Hieruit kwam naar voren dat er geen betekenisvolle correlaties tussen de metingen zijn. Vereenvoudiging door de ene meting te laten voorspellen uit de andere is daarom niet mogelijk.
Een andere optie tot reductie van benodigde tijd voor het protocol was het verkleinen van het aantal individueel te beoordelen dieren tot de helft of zelfs 25% van het totale aantal van 100 hennen. Dit werd onderzocht met behulp van een simulatiestudie, waarbij uit de bestaande dataset ad random de helft of 25% van de resultaten gekozen werd. Dit werd 100 keer herhaald. De resultaten geven duidelijk aan dat reductie van het aantal te beoordelen dieren niet mogelijk is, omdat de specificiteit en sensitiviteit te sterk aangetast wordt.
Hoewel reductie van de tijd benodigd voor het protocol dus niet mogelijk blijkt, hoeft dit niet noodzakelijk een groot probleem te zijn, als de frequentie waarmee het protocol toegepast wordt nog verder gereduceerd kan worden (de normale frequentie is slechts eens per koppel, hetgeen minder dan een keer per jaar is). Hiertoe zou gekeken moeten worden naar mogelijkheden tot het gebruik van individuele metingen die als signaal kunnen dienen om aan te geven wanneer een volledige beoordeling dient te worden uitgevoerd. Voorbeelden van dergelijke metingen zijn: uitval en selectie van dieren. Ook zou gekeken kunnen worden naar metingen die nu geen onderdeel uitmaken van het volledige protocol, bijvoorbeeld aantal (type) tweede soort eieren.

Index

1 Introduction

1.1 Welfare Quality® assessment protocols

The European Welfare Quality® project developed standard assessment protocols for broiler chickens and laying hens, sows, growing pigs, veal calves and dairy cattle. Measurements for determination of laying hen welfare, have been described in the Welfare Quality® assessment protocol for poultry (Welfare Quality®, 2009). One of the key characteristics of the Welfare Quality® assessment protocols is that they focus on animal based measurement (e.g. injuries or behaviour) than on design or management criteria (e.g. flock size) (Blokhuis et al., 2010).

Welfare Quality® assessment protocols are based on the concept that welfare is multidimensional, addressing both physical and mental health. The basic framework of the Welfare Quality® project has been used to measure welfare of various types of animals. Different welfare measurements, e.g. for laying hens the number of keel bone deformations, can be integrated into a score for twelve independent welfare criteria. These criteria are integrated into four principle scores, that are subsequently extrapolated as an overall flock score. Table 1.1 contains a list of the twelve welfare criteria and the four principles .

Table 1.1 The basic principles and criteria of the Welfare Quality® assessment protocols (Blokhuis et al., 2010).

Welfare Quality® Principles	Welfare Quality® Criteria
Good feeding	1 Absence of prolonged hunger
	2 Absence of prolonged thirst
Good housing	3 Comfort around resting
	4 Thermal comfort
	5 Ease of movement
Good health	6 Absence of injuries
	7 Absence of disease
	8 Absence of pain induced by management procedures
Appropriate behaviour	9 Expression of social behaviours
	10 Expression of other behaviours
	11 Good human-animal relationship
	12 Positive emotional state

The welfare assessment protocol for laying hens (Welfare Quality®, 2009)describes individual measurements indicative of laying hen welfare on-farm, but calculations for extrapolation to an overall flock score are missing. In this report we present this integration. The underlying calculations are presented in a separate report (Van Niekerk et al., 2012).
On average, performance of the laying hen welfare assessment protocol takes about 7 hours per flock to complete (Welfare Quality®, 2009). In practice this protocol takes too long to perform. Therefore, on-farm time saving methods have been investigated .
Dutch stakeholders have expressed interest in the assessment protocols for the different types of farm animals, but they also stated that a reduction in performance time would improve the chance of acceptability of the assessment protocol by stakeholders (Manten and De Jong, 2011).
In addition to the requests of the stakeholders for performance time reduction, there was little practical experience with the methodology. The Dutch Ministry of Economic Affairs, Agriculture and Innovation along with certain parties within the poultry sector requested that small scale robust testing be performed prior to wider implementation of welfare protocols.

1.2 Aim of the project

The aim of the project was twofold: 1. development of integration of individual measurements as a score per criterion for extrapolation as an overall flock score; 2. determination of the possibilities for simplification of the assessment protocol for laying hens with a view to time saving in performance of

measurement. This will improve applicability and the potential for practical acceptance of the assessment protocol.
In order to determine whether or not simplification of the laying hen assessment protocol was possible, the welfare assessment protocol for laying hens (Welfare Quality®, 2009) was applied using 122 layer flocks. These flocks, differed in housing conditions and breeds and were monitored to provide a reliable database for statistical analysis. Two approaches were considered as possible ways of simplifying the laying hen welfare assessment protocol:

a) Use of predictors. If significant and meaningful correlations can be found between individual measurements within the assessment protocol, the value of one measurement could then be predicted using the value of other (related) measurement. A simplified assessment protocol may then consist of a reduced set of measurements, that will enable value prediction of related measurements. The structure of the assessment protocol (measurement – criterion scores – principle scores) remains unchanged. The final outcome of any simplified protocol would then be compared with the final outcome of a full assessment protocol (the 'golden standard');
b) Reduction in number hens scored. The largest time consumer is the time required to score 100 individual hens per flock (Welfare Quality®, 2009). If fewer birds could be used, the time taken for on-farm assessment would be reduced significantly. The method described under a) use of predictor measurements was compared with the final outcome of a full assessment protocol (the 'golden standard', with 100 birds scored).

1.3 Content of this report

In the current project a large number of data of laying hen flocks were collected. These were added to data of flocks that were monitored in 2007 and 2008. This report contains a general overview of the data, in terms of variability in individual measurements and differences between various housing systems., we present an integration of the individual measurements into criteria scores and provide the findings of an analysis of the possibilities for simplification of the laying hen assessment protocol. In addition, adaptations in the assessment protocol due to practical constraints and some practical experiences are presented.

2 Methods

2.1 Laying hen flocks

A total of 122 flocks were assessed for this project. A flock was defined as birds housed together in a single house at a particular farm. Flocks were kept in various housing systems, ranging from traditional cages, with limited space and no furniture, to organic systems with ample space, litter, nest boxes, perches and a free-range area (table 2.1).
Data were provided from 2 different countries: 74 Dutch flocks: (29 assessed in 2007-2008 and 45 assessed in 2011) and 48 Swedish flocks (assessed in 2007). As some farms may have more than one layer house, it was the case that more flocks could be assessed at a particular farm. Also, as some years had elapsed between the first and later batches of flocks, some farms were visited more than once and thus more flocks from the same house may have been assessed.

Table 2.1: Flocks per housing system, country and year of assessment

		2007&08-NL	2007&08-S	2011-NL	Totaal
Regular	Aviary	13	10	19	42
	Rondeel	0	0	2	2
	Floor	6	22	10	38
	Other *	1	0	0	1
Organic	Aviary	0	0	7	7
	Floor	1	5	3	9
	Other *	1	0	0	1
Cage	Traditional	6	0	4	10
	Furnished	1	11	0	12
Total		29	48	45	122

NL =The Netherlands; S = Sweden;
* systems that have aviaries and slatted floor or litter floor in the same house.

2.2 Visits

Assessments were performed between July 2007 and April 2008 or between April and August 2011. A total of 5 assessors (4 in The Netherlands and 1 in Sweden), all trained by experts in the theory and practice , performed the data collection. Assessment of the flocks was performed on flocks of at least 50 weeks of age according to the Welfare Quality® laying hen assessment protocol (Welfare Quality®, 2009).

2.3 Training of assessors

The assessments in 2007-2008 were performed by one assessor in Sweden and one assessor in The Netherlands. These two were involved in the Welfare Quality® project and helped to develop the laying hen protocol. Besides several meetings to discuss the measurements and an additional day for the theory, they also performed training-assessments on 2 farms in The Netherlands and 2 farms in Sweden. After several months this was repeated, both in Sweden and in The Netherlands when flock assessments in Sweden and a flock assessment in The Netherlands were performed by the two assessors together. This allowed the possibility for calibration of the assessments.
In 2011, three new assessors received training for measurements in accordance with the assessment protocol. Training consisted of one day of theory, followed by a day of on-farm training. The Dutch assessor that had performed the assessments in 2007-2008 was one of the two teachers involved in this training. The other teacher was the researcher involved in the development of the protocol in 2007, who also gave laying hen protocol training sessions in Chile, UK and Italy.
At the end of the on-farm training, the measurements of the assessors were compared to those of the experienced assessor, the 'golden standard'. A minimum of 75% agreement between the assessor

and the ‘golden standard’ was achieved. After this training the assessors received instructions on how to use Personal Digital Assistant (PDA) software for registration of data on-farm.
For the first visit two assessors were sent to a particular farm where they performed all measurements together in order to further 'calibrate' their assessments.

2.4 Measurements

Table 2.2 contains an overview of the measurements per criteria, together with a brief description.. For detailed descriptions we refer to the Welfare Quality® laying hen assessment protocol (Welfare Quality®, 2009).
Specialised software to register the data on a PDA was developed. This data was then loaded into an access database immediately after the farm visit. The uploaded data were subsequently checked by a researcher for any missing or incorrect entries. Any discrepancies were corrected in cooperation with the assessor and/or the farmer.
Part of the data could not be recorded directly on-farm, but had to be added to the database later. These were data taken from official poultry house documentation, containing the exact number of feeders, drinkers, nest boxes and any other relevant information(either a so-called CPE report or a KAT-registration). A photo was made of these documents at the farm and the data added to the database later.

Table 2.2 Measurements per principle and criteria for the Welfare Quality® assessment protocol for laying hens (Blokhuis et al., 2010).

Welfare Quality® Principles	Welfare Quality® Criteria	Measurements
Good feeding	1 Absence of prolonged hunger	Feeder space
	2 Absence of prolonged thirst	Drinker space
Good housing	3 Comfort around resting	Perch shape, total perch length Red mites, dust
	4 Thermal comfort	Panting, huddling
	5 Ease of movement	Stocking density Slatted floors
Good health	6 Absence of injuries	Keel bone deformations, skin lesions, foot pad lesions, toe damage
	7 Absence of disease	Mortality, culls, diseases, various clinical issues
	8 Absence of pain induced by management procedures	Beak treatment
Appropriate behaviour	9 Expression of social behaviours	Aggressive behaviour, plumage condition, comb pecks
	10 Expression of other behaviours	use of nest boxes, litter, environmental enrichments, free range, cover on range, covered veranda
	11 Good human-animal relationship	Avoidance Distance test (ADT)
	12 Positive emotional state	Novel Object test (NOT) Qualitative Behavioural assessment (QBA)

2.4.1 General flock information

Each visit started with a short questionnaire for the farmer. The following details were registered: name and address of the farmer, the number of the house assessed, number of birds on site, number of birds in the house at placement, number of birds in the house on day of visit, date of placement, age of the birds, genotype, moulted flock or not, type of housing system, free range or not, covered veranda or not, mortality numbers and numbers culled.

The number of drinkers, total feed trough length, number and dimensions of the nests, dimensions of litter area, slatted area and total perch length were also recorded. These data were copied from the official document farmers receive for new or re-furnished houses (CPE-report/KAT-registration).
The assessor had to enter in his or her name, the starting time of visit i.e. when entering the house, and had to register the end time of the visit when all assessments had been completed.

2.4.2 Absence of prolonged hunger

The total feed trough length in the house was noted. This enables calculation of feed trough length per bird.
When calculating feeder space per bird the following aspects need to be taken into account.

1. Linear tracks are sometimes accessible from two sides and occasionally only from one side (e.g. in cages, where the feed trough is outside the cage). Depending on the accessibility of the trough only one side or both sides were counted.

   Total Feeder space trough =
   2*(total trough length with access from both sides) +
   total trough length with access from one side

2. Apart from feed troughs, feeder pans are also used for laying hens. Circular feeders are always only accessible from one side.

   Total feeder space pans = number of pans * circumference of pan

3. The total feeder space for circular feeders should be interpreted differently than feeder space for linear tracks. Calculations were carried out to convert total feeder space of circular feeders into feeder space of linear tracks. Considering European legislation for laying hens (EU, 1999) minimum feeder space is set on 10 cm for linear feeders and 4 cm for circular feeders. Therefore the following conversion was used, where 4 cm feeder pan equals 10 cm feed trough.

   Total Feeder Space (total cm feeder space) =
   (10*Total feeder space pans/4) + Total feeder space trough

   Feeder space/bird (cm feeder space per bird) =
   Total Feeder Space (total cm feeder space) / Total number of birds present

2.4.3 Absence of prolonged thirst

The type of drinker was noted, together with the number of drinkers in each house. From this the number of birds per type of drinker was calculated.
When calculating drinker space per bird the following aspects need to be taken into account.

1. Drinker space for nipple drinkers and cups were calculated in the same way:

   Total number of nipple drinkers =
   total number of nipples +
   total number of cups

2. In addition to nipples and cups, circular drinkers (bell drinkers) are also used for laying hens.

   Total drinker space bell drinkers (in cm) =
   number of bell drinkers * circumference bell drinkers

3. The total number of nipple drinkers should be interpreted differently from drinker space at circular drinkers. Calculations were performed to convert total drinker space at circular drinkers into number of drinking spots. Considering European legislation for laying hens (EU, 1999) minimum drinking space for circular drinkers is set at 1 cm per bird. Therefore the

following conversion was used, where 1 cm bell drinker per hen is required or 1 nipple drinker per 10 birds, implying that 10 cm bell drinker equals 1 nipple drinker.

Total drinker space = total number of nipples =
Total drinker space bell drinkers/10 +
Total drinker space nippels + total drinker space cups

Drinker space/bird = birds/nipple=
Total drinker space (# nipples) / total number of birds present

2.4.4 Comfort around resting

2.4.4.1 Perches

With regards to perches 3 different measures were recorded:

- Shape of perches: presence of any sharp edges is recorded as yes orno. Wooden rectangular perches are considered to have sharp edges if the edges are not rounded.
- Positioning of the perches: more than 50% of the perches positioned in a resting zone is recorded as yes or no. This can be a top level of an aviary or a positioning of perches in A-frames without food troughs in between.
- Perch length per bird: Total length of available perches divided by the total number of birds present (in cm).

2.4.4.2 Red mites

The presence of red mites is checked during the stay in the henhouse and by asking the farmer. The worst of these two findings was recorded in three classes:

- no red mites detectable
- small numbers of red mites found
- large quantities of red mites found.

2.4.4.3 Dust

The WQ-protocol uses an indicative measurement for dust using a black plate or piece of black paper. This is placed in the hen house at the start of all measurements, situated out of reach of hens and so that feeders, manure belts, egg belts and ventilation equipment have the least influence on it. At the end of all measurements the plate or black paper is compared with a clean plate or paper. Three categories of dust are identified:

- clean
- dusty, but black still visible
- black no longer visible.

2.4.5 Thermal comfort

2.4.5.1 Panting

An estimation of the percentage of birds panting is made prior to, halfway through and at the end of all measurement sessions.

2.4.5.2 Huddling

An estimation of the percentage of birds huddling is made prior to, halfway through and at the end of all measurement sessions.

2.4.6 Ease of movement

2.4.6.1 Stocking density

The total available space is recorded using measurements contained in the official document farmers receive for new or re-furnished houses (CPE-report). The total available space comprises both litter and perforated floors. Free range areas are not included and covered veranda's are only included if they are permanently available to the birds (thus access never prohibited).
Stocking density is given as cm space per bird and is calculated by dividing total available space by the number of hens present.

2.4.6.2 Perforated floors

The percentage of perforated floors is calculated using measurements contained in the official document farmers receive for new or re-furnished houses (CPE-report). The percentage perforated floor is calculated by dividing total available perforated floor area by the total available space (as used for calculating stocking density) and multiplying the outcome by 100.

2.4.7 Absence of injuries

For the absence of injuries 100 individual hens were examined and the following injuries scored:
- Keel bone deformations: scored as yes or no
- Skin lesions: scored in three classes: no lesions, minor lesions or large lesions
- Foot pad lesions: scored in three classes: no lesions, minor lesions, large lesions (dorsally visible)
- Toe damage (wounds, missing toes): scored as yes or no

Flock percentage can be calculated based on the outcome of the 100 hens.

2.4.8 Absence of disease

Absence of diseases can be established using the following measurements:
- Mortality numbers and number of cullings : these numbers come from the farm records.
- Diseases and various clinical issues: based on examination of 100 individual birds and visual clinical issues when carrying out the protocol in the hen house six issues can be scored on a 3-point scale (not present, fewer than 3 birds, more than 3 birds). These six issues are: enlarged crops, eye pathologies, respiratory infections, enteritis, parasites (excl. red mites) and comb abnormalities.

2.4.9 Absence of pain induced by management procedures

Beak treatment includes beak trimming or more gentle procedures to shorten/blunt the tip of the beak of laying hen chicks. Examination of the beaks of 100 individual hens is based on scoring according tothe following 3-point scale:
- no beak treatment and no abnormalities
- moderate beak treatment, fairly normal beak shape
- severe beak trimming or beak abnormalities

Flock percentage is estimated based on the outcome of the 100 hens examined.

2.4.10 Expression of social behaviours

Aggressive behaviour is scored on a 2-point scale (yes or no) involving registration of any aggressive behaviour observed during the time spent in the hen house.

Plumage condition is scored for 100 individual hens. Only three zones (head-neck, back-rump, cloacal region, indicated with different colours in figure 2.1) are scored, as these may bare relation to welfare, whereas plumage damage in other zones may only reflect system effects (Bilcik and Keeling, 1999). Total plumage condition is scored with one score, for the three indicated zones together, according to the following 3-point scale:

- no or very limited plumage damage
- moderate plumage damage and/or minor featherless spots
- larger featherless spots

Flock percentage is estimated based on the outcome of the 100 hens examined.

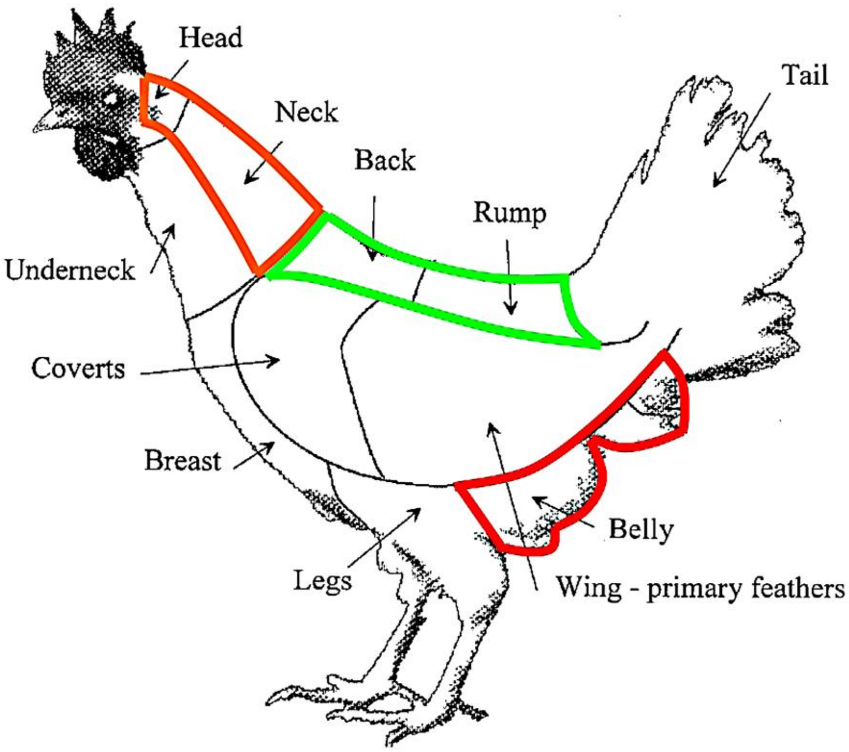


Figure 2.1 Schematic indication of (coloured) zones scored for plumage damage **(**orange: head-neck; green: back, rump; red: cloacal region)

Comb pecks are recorded for 100 individual hens and scored in the following 3 classes:

- no comb pecks present
- less than 3 comb pecks
- 3 or more comb pecks

Flock percentage can be estimated based on the outcome of the 100 hens examined.

2.4.11 Expression of other behaviours

Use of nest boxes is determined with the following 4 measurements:

- availability nest boxes (yes / no)
- distribution of nest boxes over the hen house (even / not even)
- distribution of eggs between nest rows (even / not even)
- distribution of eggs within nest rows (even / not even)

In addition to usage the available nest space per bird is also calculated. For single nests the number of hens per nest is taken from the official document farmers receive for new or re-furnished houses (CPE-report). If group nest are available, the cm^2 nest space per bird is taken from the same document.
For the final calculations of nest space area (cm^2) per bird the figure for single nests is converted to cm^2 nest space per bird. The conversion used is based on the European Directive (Council Directive 1999/74) in which minimum nest space is set at 1 single nest per 7 hens or in case of group nests 120 hens per m^2 group nest.

- For group nests the space per bird is 100*100cm/120 = 83.33 cm^2/hen
- The average size of a single nest (width*depth) is 30*35 = 1050 cm^2
  Thus, average space/hen in single nest = 1050/7=150 cm^2
- (space single nest/150) = (space group nest/83.33)
  space single nest/1.8=space group nest

Use of litter is scored on the following 3-point scale:

- birds have been observed dust bathing
- birds have been observed scratching in the litter
- no litter available or litter not used

Environmental enrichments, a free range area and covered veranda are all scored according to the following a3-point scale:

- 50-100% of the birds use it
- less than 50% of the birds use it
- not available or not used

For cover on range an estimate is made of the area of range covered. The cover can be artificial (e.g. roofs) or natural (e.g. bushes).

2.4.12 Good human-animal relationship

2.4.12.1 Avoidance Distance test (ADT)

The Avoidance Distance Test (ADT) is a measurement used to determine how close the observer can approach a hen. The observer therefore chooses a hen sitting on the edge of an elevated slatted floor, with the head directed towards the observer. The observer then slowly approaches the bird from a distance of 1.5 m. As soon as the bird replaces one foot, the observer stops and measures the distance between him and the place the hen was standing. The ADT is performed with 21 hens per flock, well distributed throughout the hen house.
In cage houses the observer walks along an aisle and chooses a hen with the head out of the cage. The observer approaches the hen from a distance of 1.5 m. As soon as the hen pulls the head into the cage, the observer stops and measures the distance between him and the cage front.
The ADT is designed to measure fear of humans. The level of fear for humans is indicated by how close the observer is able to approach the birds.

2.4.13 Positive emotional state

2.4.13.1 Novel Object Test (NOT)

The Novel Object Test (NOT) provides a measurement for fearfulness for unknown objects, which is supposed to reflect general fearfulness.
During the two minutes of the NOT every 10 seconds the number of hens is recorded that are within 30 cm of the object. Per flock this test is performed at 4 different locations in the litter area or, in case of cages, in or above the feed trough. The novel object used is a tube of 2.5 cm diameter and 50 cm length, with coloured bands (figure 2.2). The more birds counted, the less fearful the birds are.

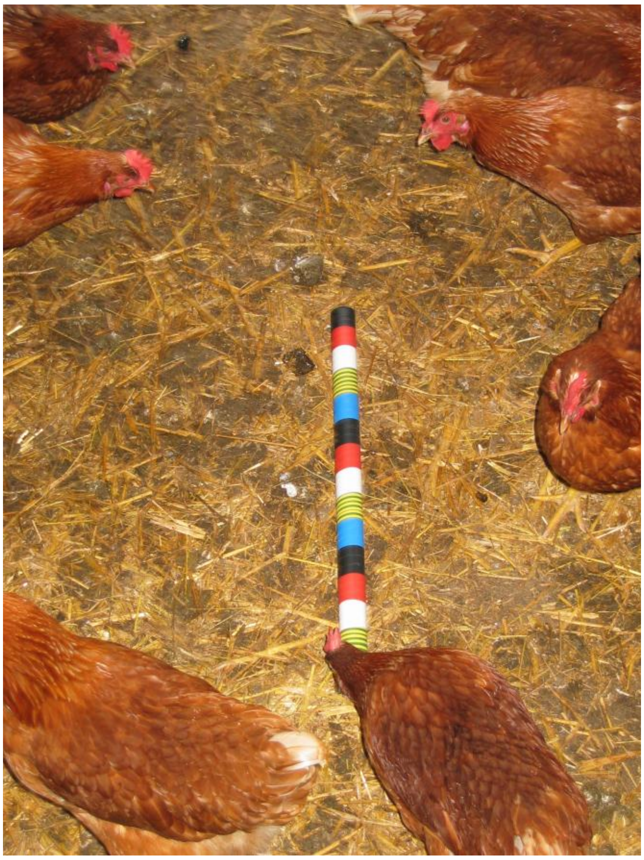


Figure 2.2 Novel Object Test

2.4.13.2 Qualitative Behavioural Assessment (QBA)

Qualitative Behavioural Assessment (QBA) is a method to estimate the state of mind of the flock. It is different from traditional methods, as it allows a certain interpretation of what is observed. The observer uses his expertise of the type of animal in combination with his observations to estimate the state of mind of the animals according to 20 characteristics. Before scoring the observer stands for 20 minutes in 2 - 4 places in the house and gains an impression of how the birds behave and what their motivations are. Despite the somewhat subjective character of the test, research has showed a high repeatability of the assessment and results agree with the expectations of experts ((Main et al., 2003; Wemelsfelder, 2007; Brscic et al., 2009; Stockman et al., 2011).
The 20 characteristics that have to be scored can be divided into 10 positive terms (e.g. active, content) and 10 negative terms (e.g. fearful, frustrated). All characteristics are scored on a scale of 120 mm long, without any numbering, only two marks are given to indicate the minimum and maximum of the scale (figure 2.3). The observer marks a place on the scale and later measures the distance (in mm) from the minimum mark .

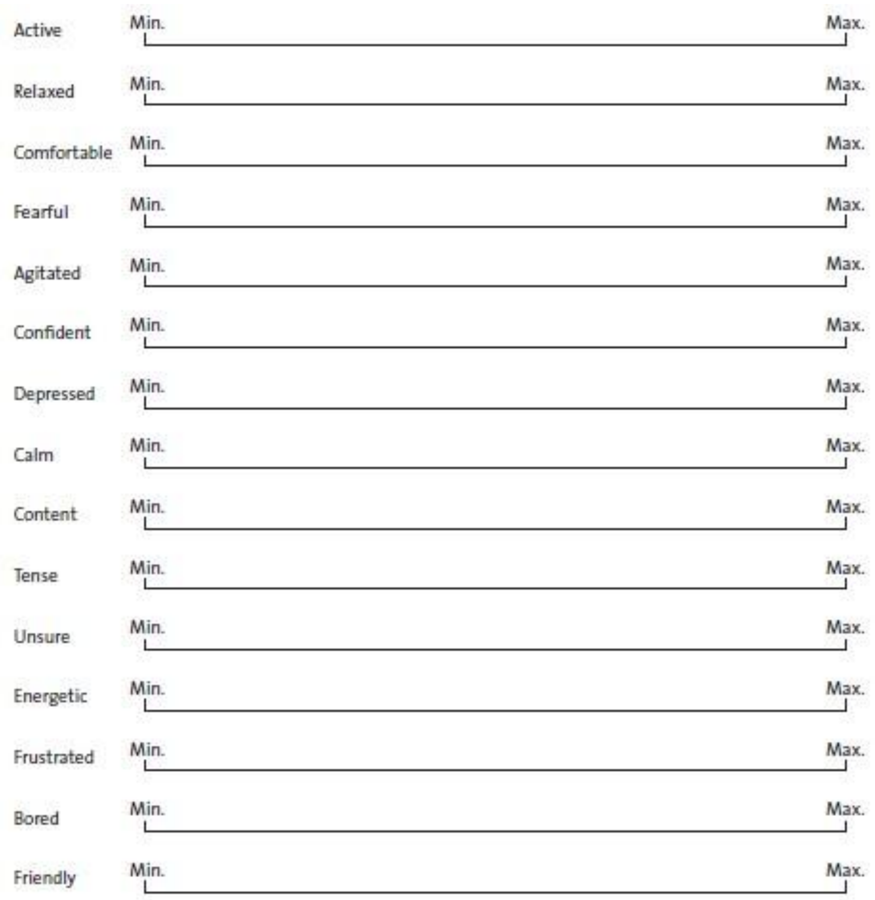


Figure 2.3 Example of scoring list for QBA with unnumbered scale.

3 Results Welfare Quality®-measurements

The results of the Welfare Quality®-measurements are presented in the sequence of the protocol. Some characteristics of the flocks are given in table 3.1. Most flocks monitored were regular (i.e. not organic) flocks, distributed over aviaries and floor systems. All Swedish flocks were non-beak trimmed, while of the Dutch flocks only organic and the Rondeel flocks were not beak trimmed.
Age of inspection was intended to be at least 50 weeks of age, but in practice averaged 65 weeks. Rondeel houses were monitored at a younger age than the other flocks.

Table 3.1 numbers (#) of flocks, flock size and age at inspection

		Total # flocks	# flocks with free range**	# flocks without free range	flock size	Age at inspection (in weeks)
Regular	Aviary	42	10	32	22,193	66
	Rondeel	2	0	2	30,000	53
	Floor	38	4	34	9,247	65
	Other *	1	1	0	5,590	64
Organic	Aviary	7	7	0	11,467	67
	Floor	9	7	2	6,007	62
	Other *	1	1	0	13,775	60
Cage	Conventional	10	0	10	38,741	68
	Furnished	12	0	12	24,749	65
Total/ Average		122	30	92	17,882	65

* systems that have aviaries and slatted floor or litter floor in the same house.
** free range with official dimensions according to European legislation

In the following paragraphs results are presented per welfare principle. In general, large variation was found between flocks for individual measurements. Variation for individual measurements are illustrated in appendix 1.
The following paragraphs contain the average results per housing system used for laying hens. The Rondeel and the two “other” systems investigated had too few flocks for monitoring and are not mentioned in the next paragraphs.

3.1 Good feeding

Good feeding comprises both feeder and water space. Figure 3.1 gives a view of the average feeding trough space (cm) and numbers of hens per nipple drinker. Feeder space in regular aviaries and organic floor systems is high. In furnished cages the variation in feeder space is low. The variation in numbers of birds per nipple drinker is similar between systems, with conventional caged systems displaying the least variation on numbers per drinker. On average organic floor systems provide the most feeder space and have low numbers of hens per nipple drinker. Also the number of hens per nipple drinker in cages is low. No statistical analysis has been done on these figures.

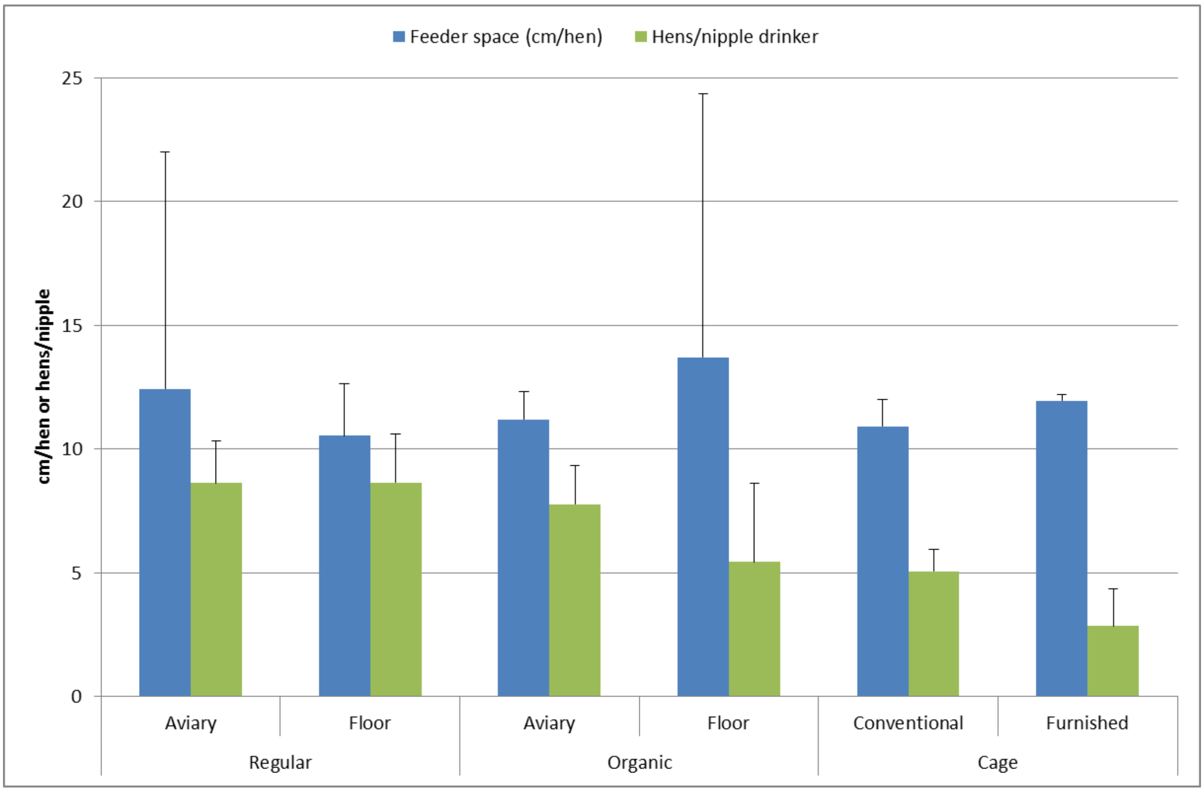


Figure 3.1: Average feeder space and number of hens per nipple drinker per housing system

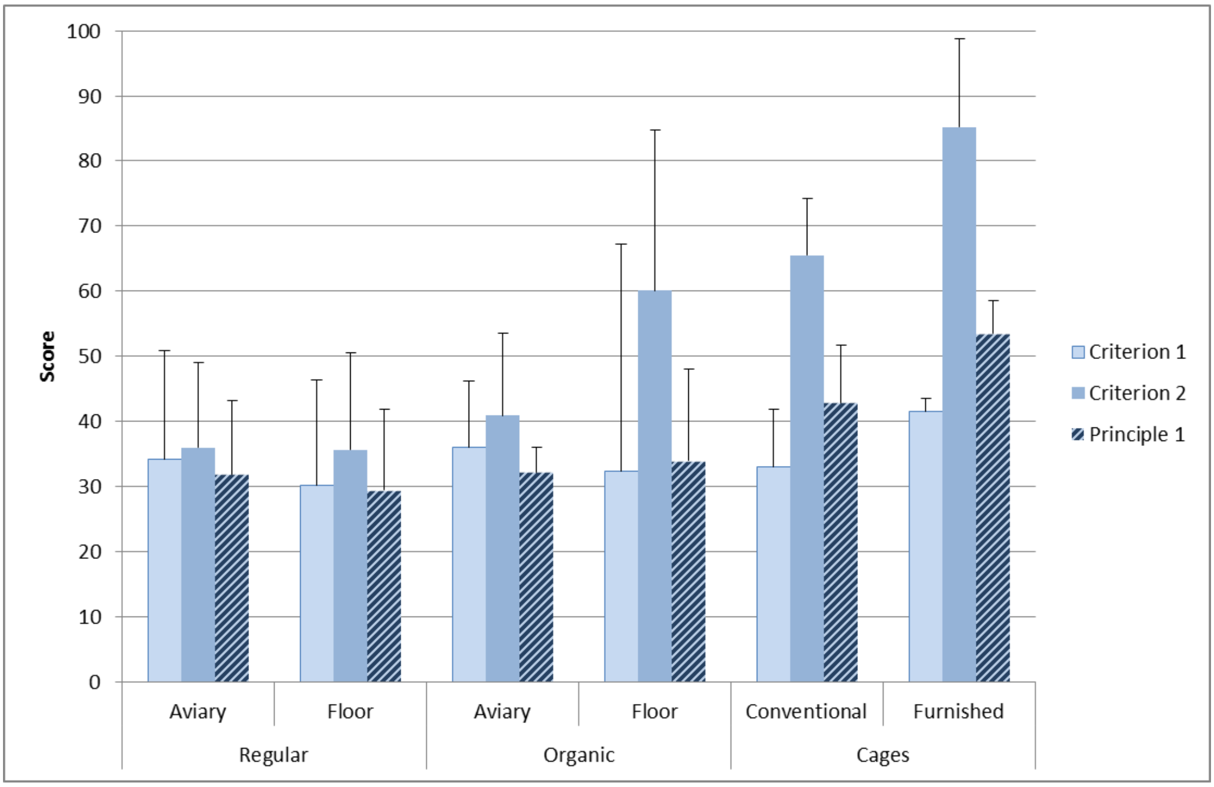


Figure 3.2: Scores for criterion 1 (absence of prolonged hunger), criterion 2 (absence of prolonged thirst) and principle 1 (Good feeding) per housing system

3.1.1 *Criteria and scores for principle 1*

In figure 3.2 the data for feeder space and number of hens per nipple have been transformed into criterion scores for absence of prolonged hunger and absence of prolonged thirst. This transformation not only involves scoring on a scale from 0 to 100, but also a weighting of the implications for animal welfare. This was performed according to the methodology as described by Van Niekerk & Van Reenen (2012). The two criterion scores are later transformed into an overall score for Principle 1 'Good feeding'. The transformation is unique for each species, but the method is in accordance with previously developed mathematical models use for other species (Welfare Quality®, 2009).
Because numbers of hens per nipple drinker is lower in cages, criterion 2 is higher and consequently the score for principle 1 is highest in organic floor and cage housing systems.

3.2 Good housing

This principle concerns welfare around resting, thermal comfort and ease of movement.

3.2.1 *Comfort around resting*

Figure 3.3 presents the average perch length per hen in various housing systems. Variation in perch length per hen is largest in floor systems. In aviaries the variation is much lower, probably as a result of the fixed lay-out of the system. In furnished cages there is hardly any variation in perch length per bird.

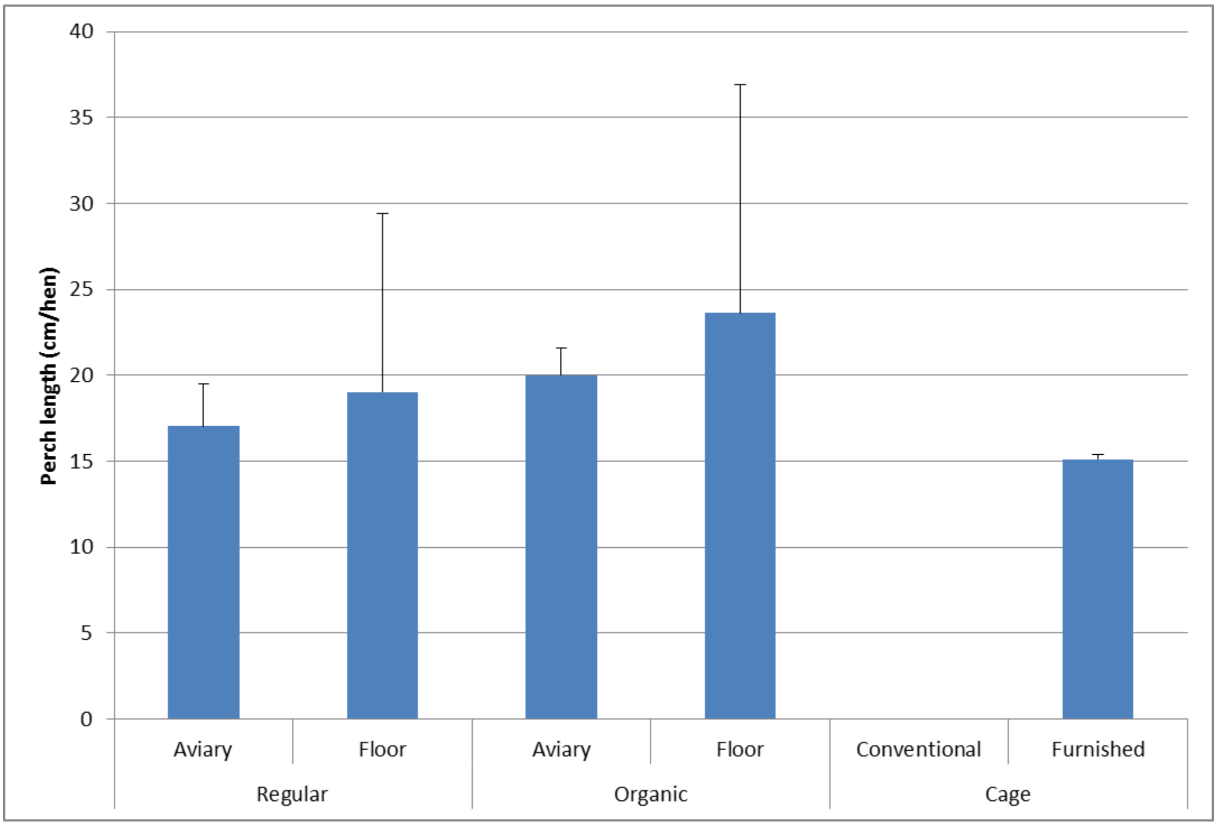


Figure 3.3: Average perch length (cm/hen) per housing system

Although there is considerable variation in red mite infestations, the level of infestation seem to be lowest on floor systems (figure 3.4). This is according to expectation, as floor systems usually have a manure pit underneath the slatted floor. Experiences on Dutch commercial farms have indicated that predator mites and other predators can easily survive in manure pits, resulting in more predation on mites and thus lower levels of red mites (Van Niekerk, personal communications).

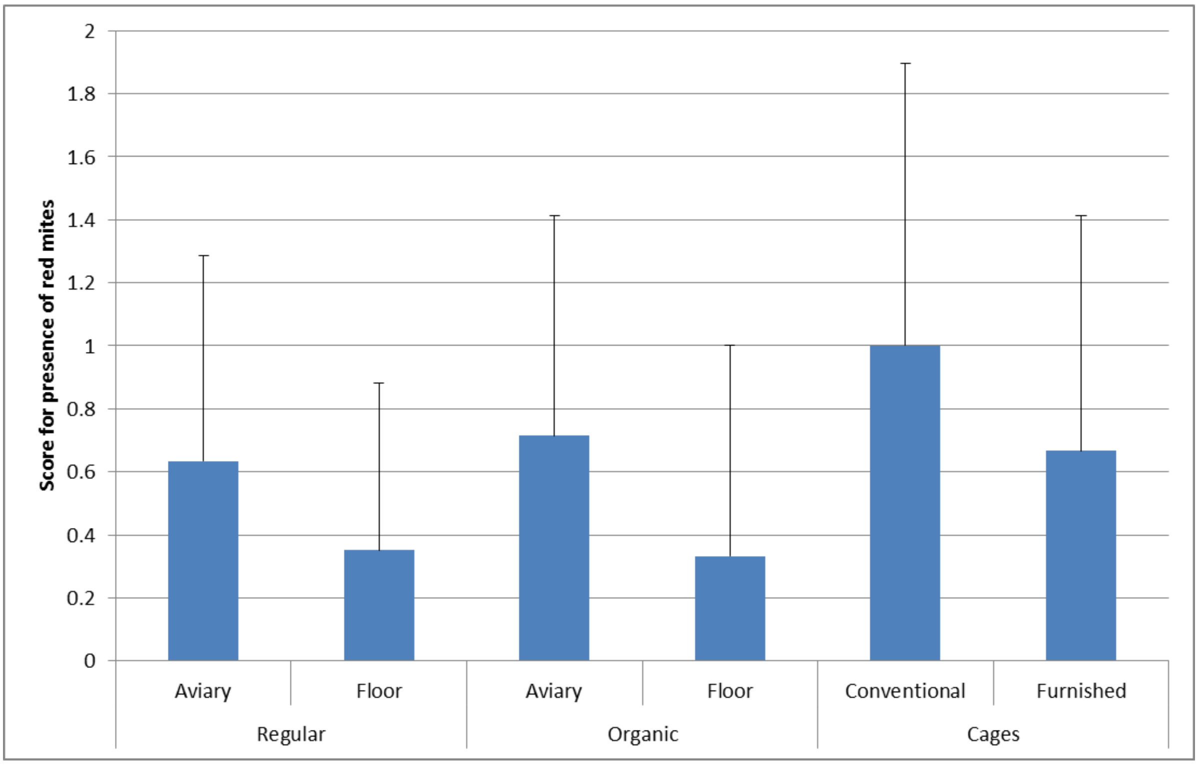


Figure 3.4: Average red mite score per housing system (0= no red mites, 2= heavy infestation)

Last measurement for determination of comfort around resting was the level of dust. This measurement was the same for all poultry houses (i.e. moderate level score 1).

3.2.2 Thermal comfort

Thermal comfort was measured by scoring the characteristics panting and huddling. Apparently all farms were scored during a period without extremes in climate , therefore it is not surprising that variation was low in these measurements.

3.2.3 Ease of movement

Ease of movement is expressed by stocking density and percentage area of perforated floor space. Stocking density is expressed as available space per hen (cm^2). As shown in figure 3.5, available space per bird was lowest in cage systems. Space per hen in furnished cages was also low, but for Welfare Quality® litter and nesting areas are excluded, whereas this is included in the calculation of space according to European legislation (EU, 1999). Space per bird is lower in regular than in organic systems, as was expected based on the legal requirements for both types of system.

Figure 3.6 shows the percentage of slatted floor, in relation to total area (excluding nests), per housing system. In the Netherlands, furnished cages often do not have a litter area, which means almost 100% perforated floor in cage systems.

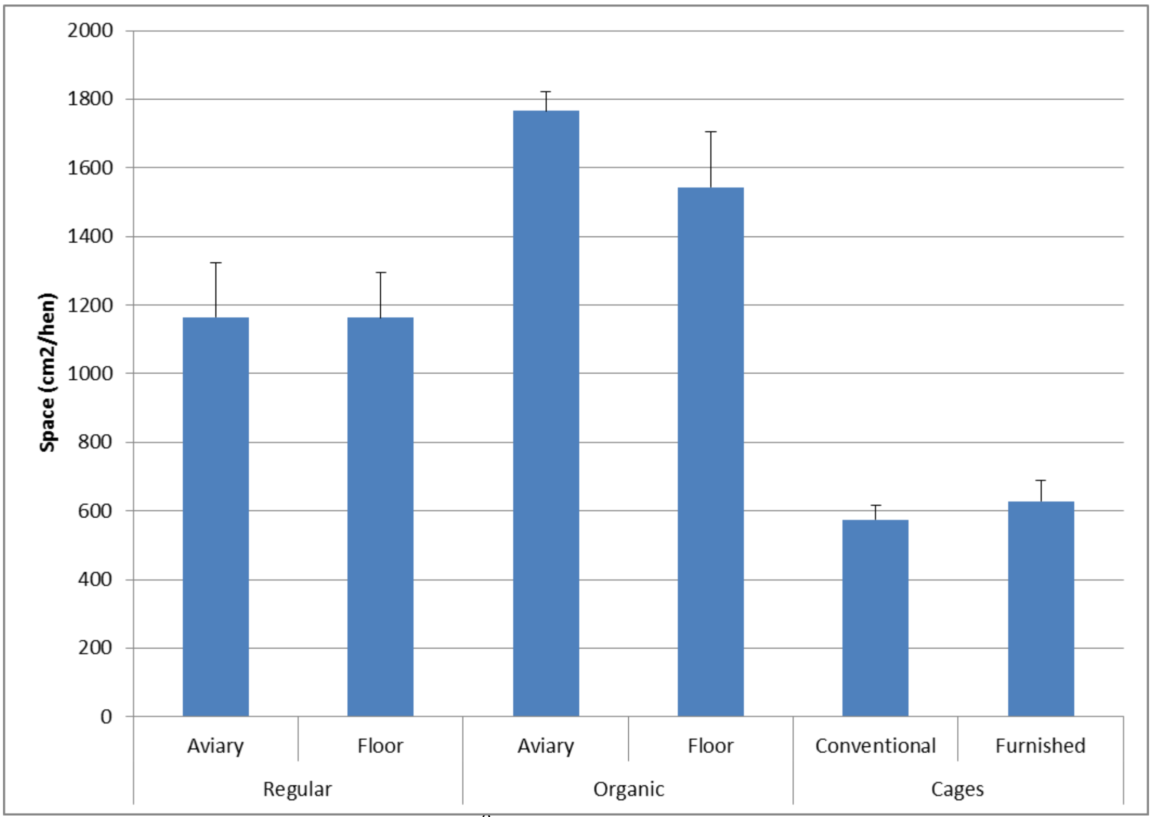


Figure 3.5: Average available space (cm^2/hen) per housing system

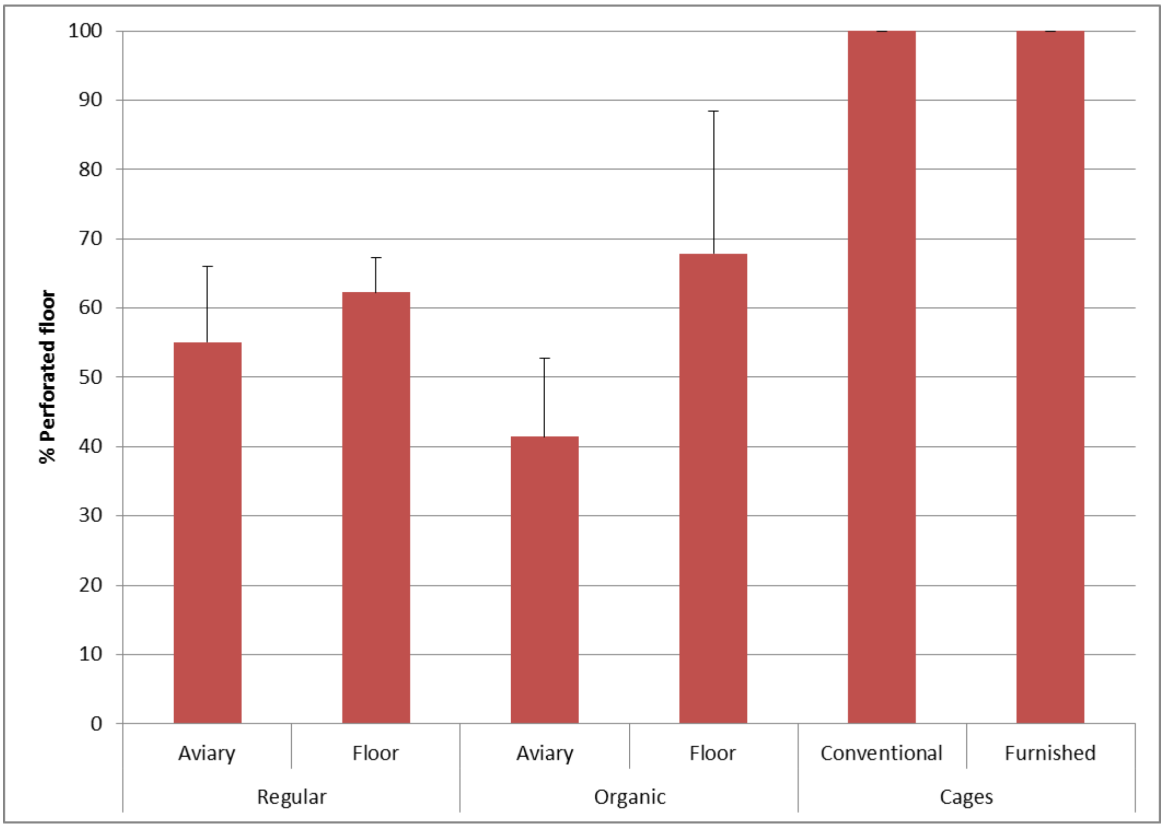


Figure 3.6: Average percentage of slatted floor per housing system

3.2.4 Criteria and scores for principle 2

Figure 3.7 displays scores for criterion 3 (Comfort around resting place), criterion 4 (Thermal comfort), criterion 5 (Ease of movement) and results of integration of these scores into an overall score for principle 2 (Good housing) per housing system. As explained in paragraph 3.2.2, thermal comfort was equal for all flocks, resulting in equal criterion scores for all systems. Therefore, any variation in principle scores is completely due to differences measured in criteria 3 and 5.
As no perches are available in conventional cages the score for criterion 3 is lower than for other systems. Limited space per bird in both conventional and furnished cages results in low scores for criterion 5 for these systems. Therefore the scores for good housing are lower for cage systems than for non-cage systems.

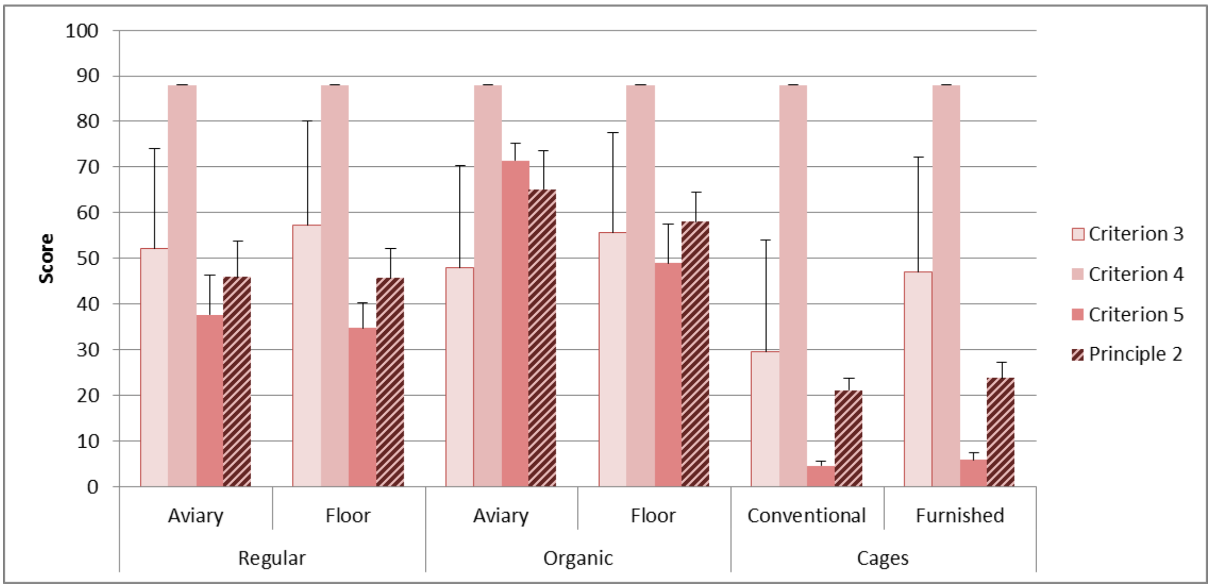


Figure 3.7: Scores for criterion 3 (Comfort around resting), criterion 4 (Thermal comfort), criterion 5 (Ease of movement) and principle 2 (Good housing) per housing system

3.3 Good health

Good health is divided into 3 criteria: absence of injuries, absence of disease and absence of pain induced by management procedures.

3.3.1 Absence of injuries

This criterion is determined by measuring keel bone deformations, skin lesions, foot pad lesions and toe damage.

Keel bone deformation was determined for 100 hens examined individually. Palpation and, when possible, visual appraisal of the keel bone enabled scoring as straight or not straight. For keel bones that were not straight no distinction was made between severe deformations or minor irregularities (e.g. slight bending). Results from the various housing systems are presented in figure 3.8. As keel bone deformation has a strong relationship with the presence of perches, it is not surprising that conventional cages scored lower for keel bone deformations (Schotz et al., 2008, Picket et al., 2010, Kapelli et al., 2011).

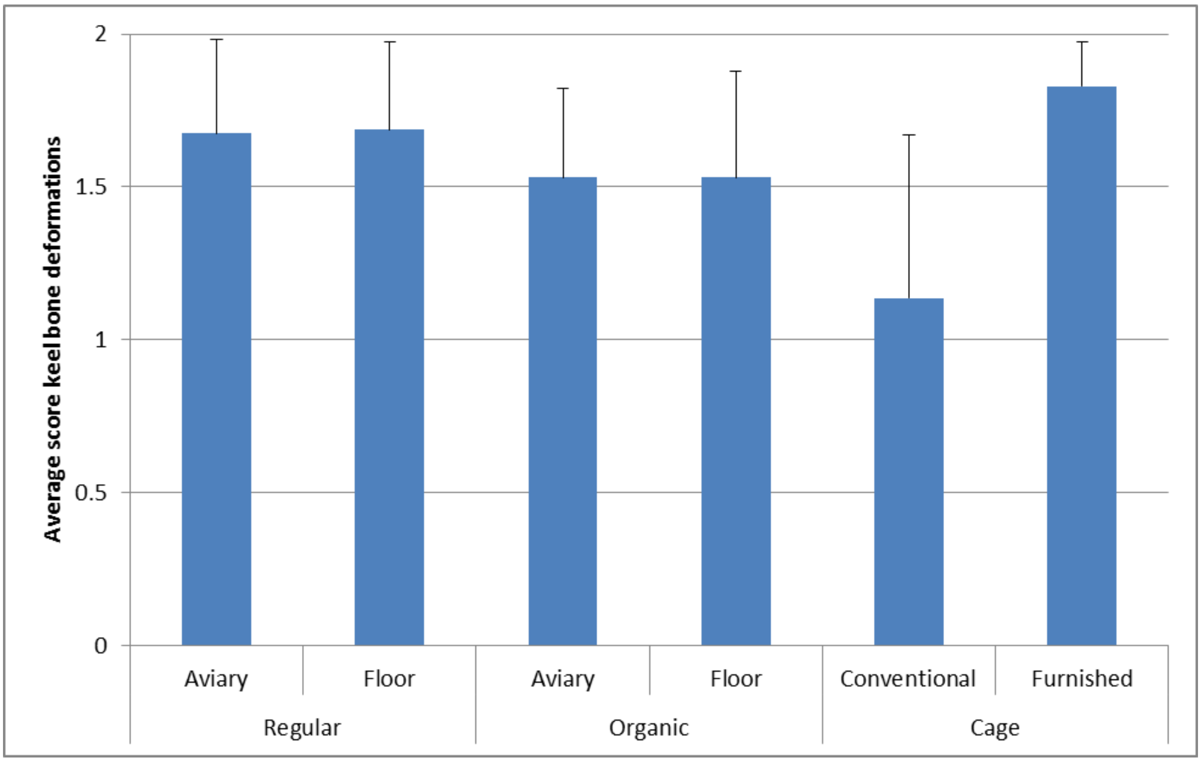


Figure 3.8: Average keel bone score per housing system (0 = no deformaton; 2 = deformation)

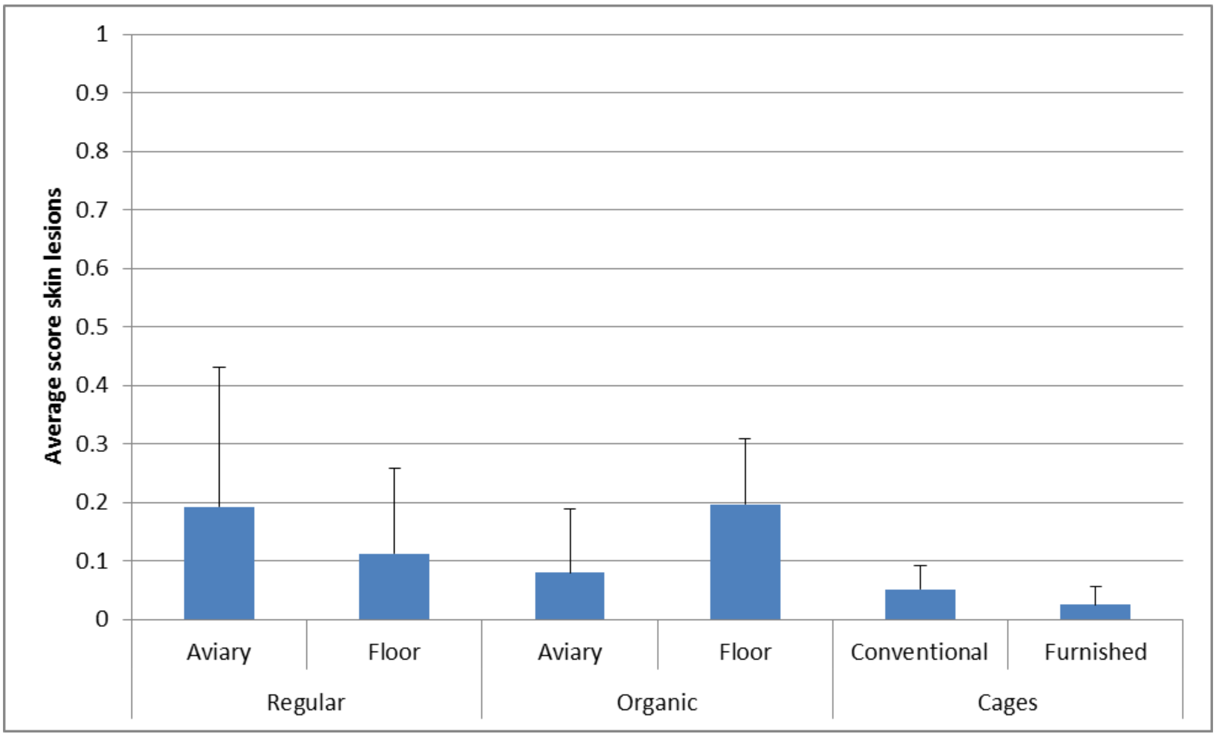


Figure 3.9: Average score for skin lesions per housing system (per flock 100 birds are scored on a 3-point scale: 0 = no lesions; 1 = minor lesions; 2= more/larger lesions)

Skin lesions were also scored individually for 100 birds per flock. On average skin lesions were very minor (figure 3.9). Approximately 2% of birds housed in regular aviaries attained a score of 2. In regular floor systems only 1.1 % of the birds received score 2. For all other systems less than one percent of the birds received score 2. The average percentage of birds receiving score 1 was highest in organic floor systems (17.8%), followed by regular aviaries (15.1%), regular floor systems (9.1%) and organic aviaries (7.1%). The other systems scored below 5% birds with score 1.

Foot pad lesions were scored for 100 hens per flock examined individually. Both feet were examined, but the score used was that for the worst foot. Minor lesions (score 1) usually refer to hyperkeratosis or small skin wounds. Severe lesions (score 2) refer to bumble foot, where the foot pads are swollen to such an extent that it is dorsally visible.
Especially foot pad scores in furnished cages were very low, with 98% score 0 reflecting hardly any abnormality. On average scores for the other housing systems were also very low, with regular floor systems having the best foot pad scores for non-cage systems (figure 3.10).
The average percentage of hens with score 2 was highest in organic floor systems (2.6%), followed by regular aviaries (1.9%) and regular floor systems (1.5%). The percentage of birds with score 2 in the other systems approached zero.

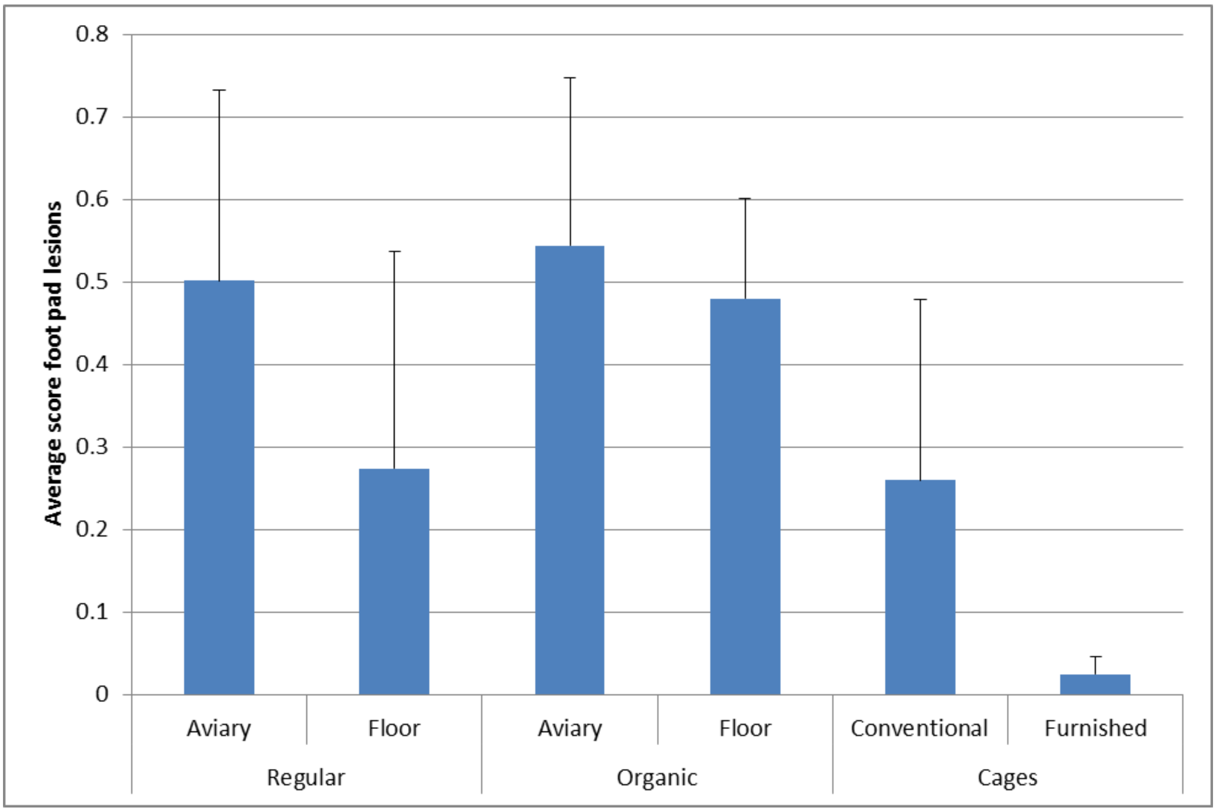


Figure 3.10: Average score for foot pad lesions per housing system (per flock 100 birds are scored on a 3-point scale: 0 = no lesions; 1 = minor lesions; 2= more/larger lesions)

Toe damage refers to missing parts of toes and toe wounds. This was scored in two ways: individual examination of 100 birds and general observations during the work in the hen house. Scores from the 100 birds provided an indication of the percentage of hens with toe damage. Observational scores were divided into 3 classes: no damage, less than 3 birds with damage and 3 or more hens with toe damage. Although there was a clear correlation between the two scoring methods, individual examination of 100 birds (clinical scoring) appeared to provide more precise information and thus this measurement was used for the Welfare Quality® protocol.
In general, average toe damage was low with organic aviaries being the highest with almost 1.6% (figure 3.11). Damage consisted mainly of toes missing through accidents. It is not clear if these accidents occurred during the rearing or laying period.

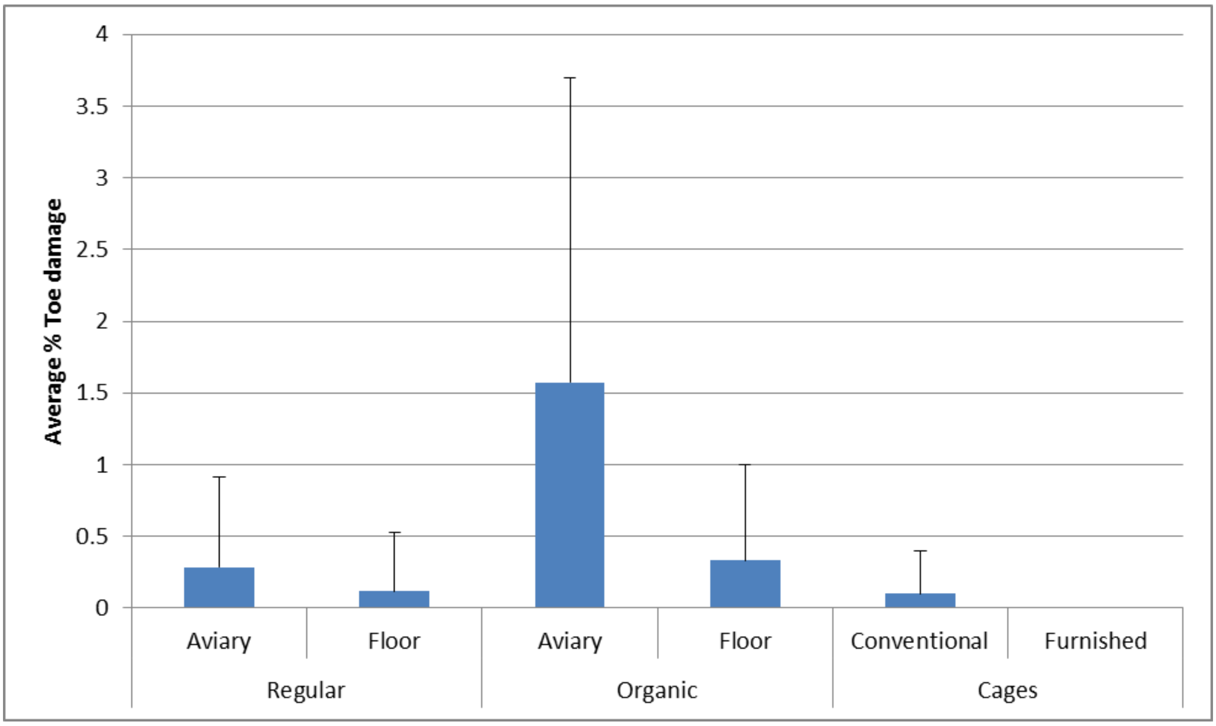


Figure 3.11: Average % toe damage per housing system (individual examination of 100 birds per flock)

3.3.2 Absence of disease

Under the criterion absence of disease not only diseases and various clinical issues are recorded, but also mortality numbers and birds culled. As almost none of the farms recorded culls separately from mortality, both figures have been combined and therefore only the figures for total mortality (including culls) is presented (figure 3.12). On average mortality was between 3.1% (furnished cages) and 7.4% (organic aviaries), but there was a large variation, especially in non-cage flocks.

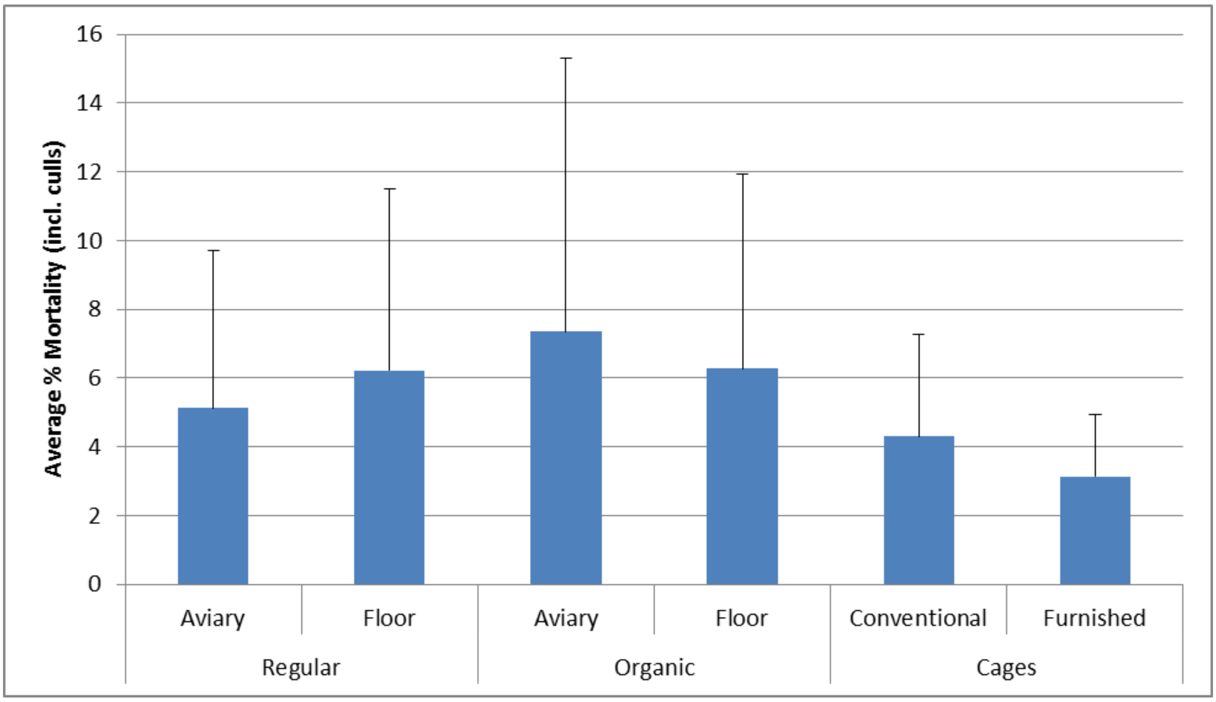


Figure 3.12: Average percentages of mortalities (including culls) per housing system

Diseases and other clinical issues were scored after all the work in the hen house had been completed. Based on the 100 individually screened birds and a general impression of the flock 5 clinical indicators of health problems were scored on a 3-point scale, where 0 = none, 1 = less than 3 birds and 2 = 3 or more birds. These clinical indicators included: enlarged crops, eye pathology, respiratory infections, enteritis and comb abnormalities. No respiratory infections were found. Only low incidences were observed for the remaining indicators (figure 3.13) Enteritis seems to occur more often in floor systems and furnished cages. Comb abnormalities were predominantly seen in conventional cages. However, overall levels remained below 1, implying that on average less than 3 birds per flock displayed health problems.

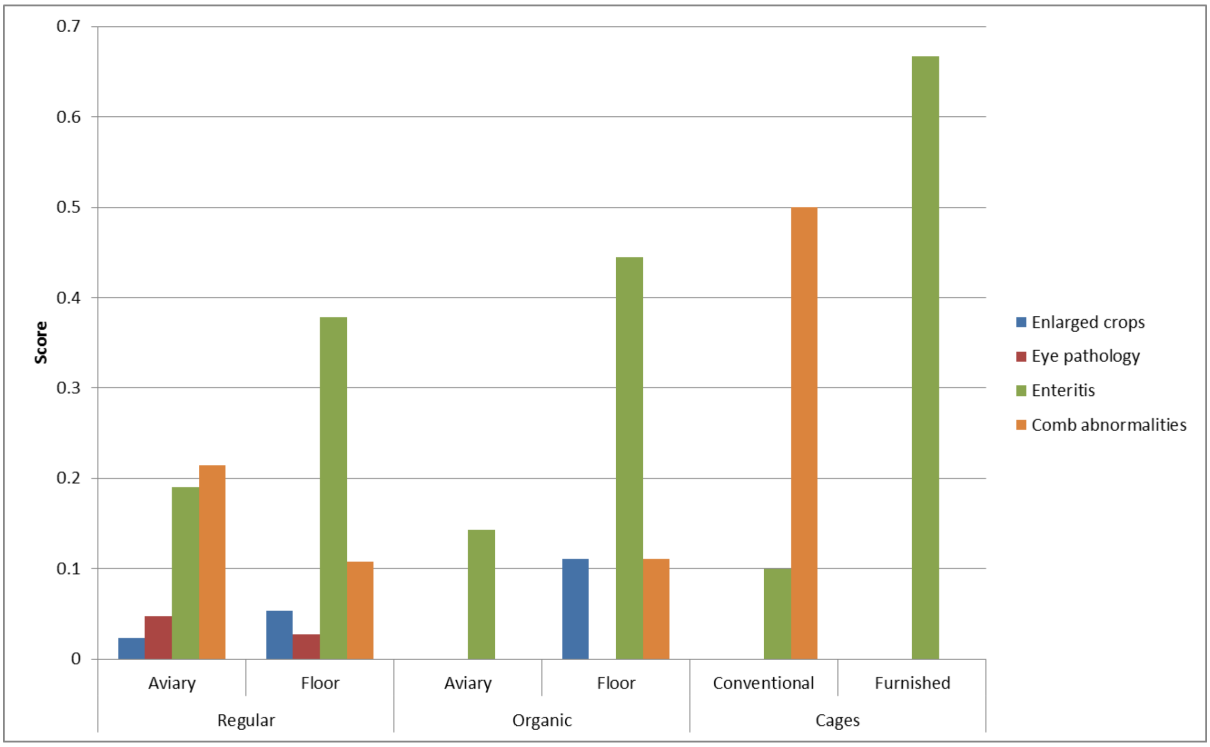


Figure 3.13: Average scores for enlarged crops, eye pathology, enteritis and comb abnormalities per housing system (scores for respiratory infections are not shown as none were found); These health issues are scored per flock on a 3-point scale: 0 = none, 1 = less than 3 birds and 2 = 3 or more birds.

Presence of flies and parasites were scored on a 2-point scale (0= not present, 1 = present). Flies themselves are not directly a health risk, but as flies can transmit diseases, they form a risk to bird health. No distinction was made between internal or external parasites, except for red mites that were scored separately (figure 3.14). The higher the number, the more flocks that were infected. As floor systems usually have a manure pit underneath the slatted floor, where parasites can thrive, it was to be expected that flies and parasites were seen there more often compared to aviaries with manure belts and thus frequent manure removal. Surprisingly a large number of flocks in conventional cages had high scores for flies and parasites .

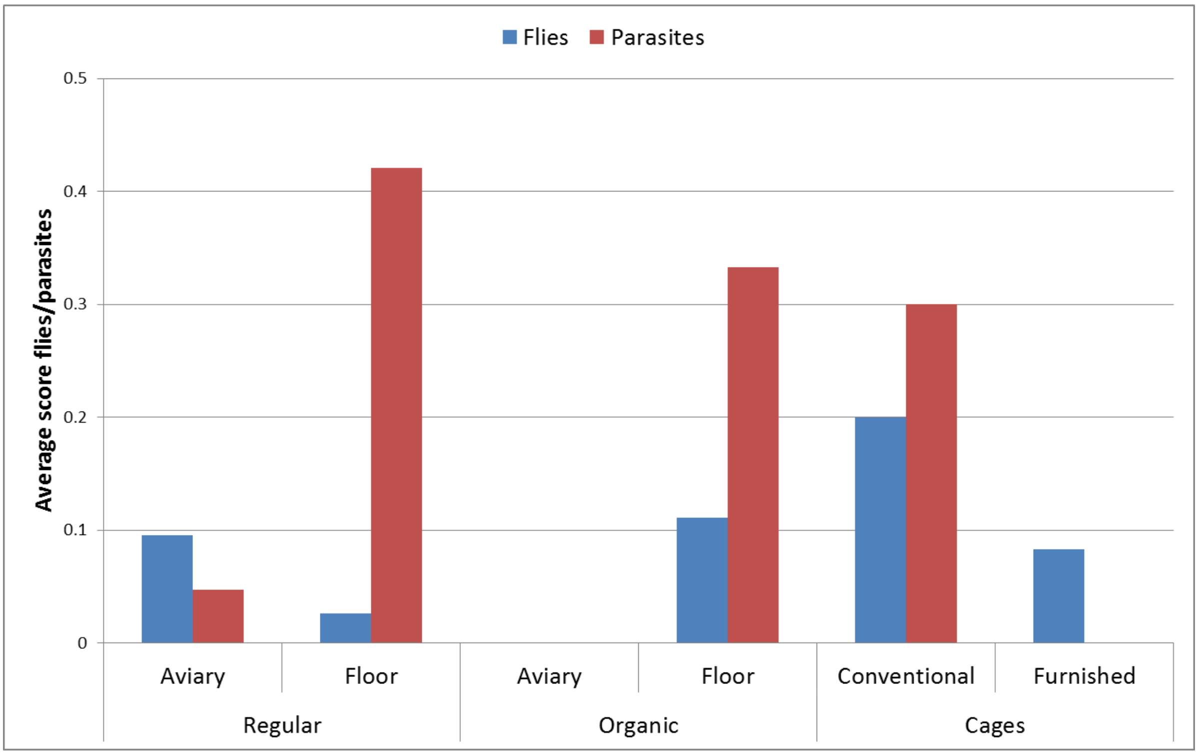


Figure 3.14 Average scores for presence of flies and parasites per housing system; This was scored on a 2-point scale: 0 = present, 1 = not present.

3.3.3 Absence of pain induced by management procedures

Beak trimming refers to severe treatments, whereas beak treatment refers to modern, more gentle ways to remove the beak tip (e.g. before 10 days of age, IR-method). Beak treatment was individually scored using 100 birds. Appraisal was performed based on the following 3-point scale: 0 = untreated and natural shape; 1 = moderate treatment and no abnormalities (i.e. lower beak not longer than upper beak); 2 = treated beaks with visible abnormalities.

Figure 3.15 provides a review of the results per housing system. Beak treatment is not allowed in Sweden and thus all Swedish flocks have low beak scores. This may have confounded the results to a certain extent, causing lower scores for regular flocks in aviaries and floor systems compared to conventional cages. Since conventional cages are banned in Sweden, all data concerning this system comes from the Netherlands. As expected beak scores in organic farming is almost zero, as beak treatments are not allowed. Some abnormalities such as crossed beaks or damaged beaks may be the reason why not all birds have scored 0 in these systems.

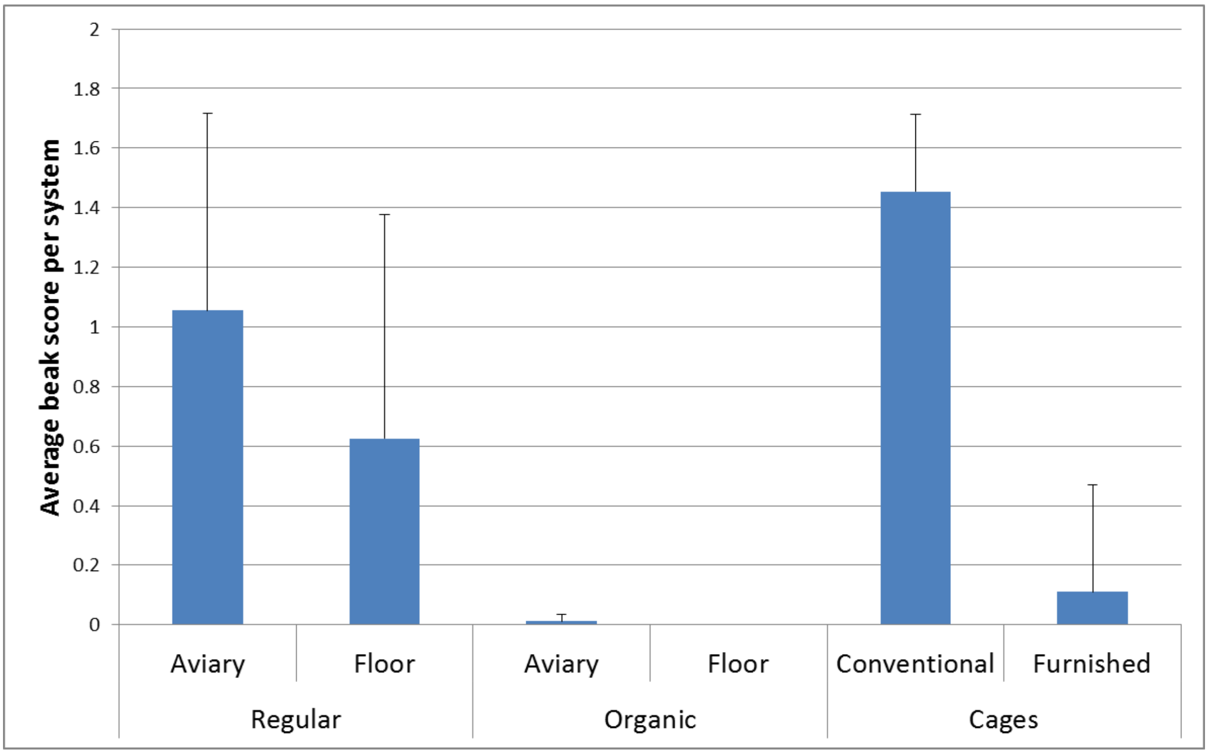


Figure 3.15: Average beak scores per housing system
(Per flock 100 birds have been evaluated using a 3-point scale: 0 = untreated and natural shape; 1 = moderate treatment and no abnormalities (i.e. lower beak not longer than upper beak); 2 = treated beaks with visible abnormalities)

3.3.4 Criteria and scores for principle 3

Scores for criterion 6 (Absence of injuries), criterion 7 (Absence of diseases), criterion 8 (Absence of pain induced by management procedures) and the integrated score for principle 3 (Good health) are presented per housing system in figure 3.16. Differences in principle scores are dictated mainly by the scores for criterion 8. This criterion reflects the beak scores, which in organic husbandry systems score highly due to the fact that beak treatment is prohibited. Because some birds are found with deformated (e.g. crossed beak) or damaged beaks, scores do not always add up to 100. The large difference in criterion 8 between conventional and furnished cages is caused by a country effect. Most furnished cage flocks are Swedish, where beak treatment is prohibited. All conventional cage flocks are Dutch and beak treated (see table 2.1 for distribution of systems over countries).This difference is also reflected in the scores for regular housing systems: most regular aviary systems are located in The Netherlands (beak treated), whereas regular floor systems are mainly located in Sweden (untreated).
As differences in between criterion 6 and 7 are not so very pronounced, the differences in scores for principle 3 are mostly predominantly reflecting the variation in scores for criterion 8, resulting in low scores for conventional cages.

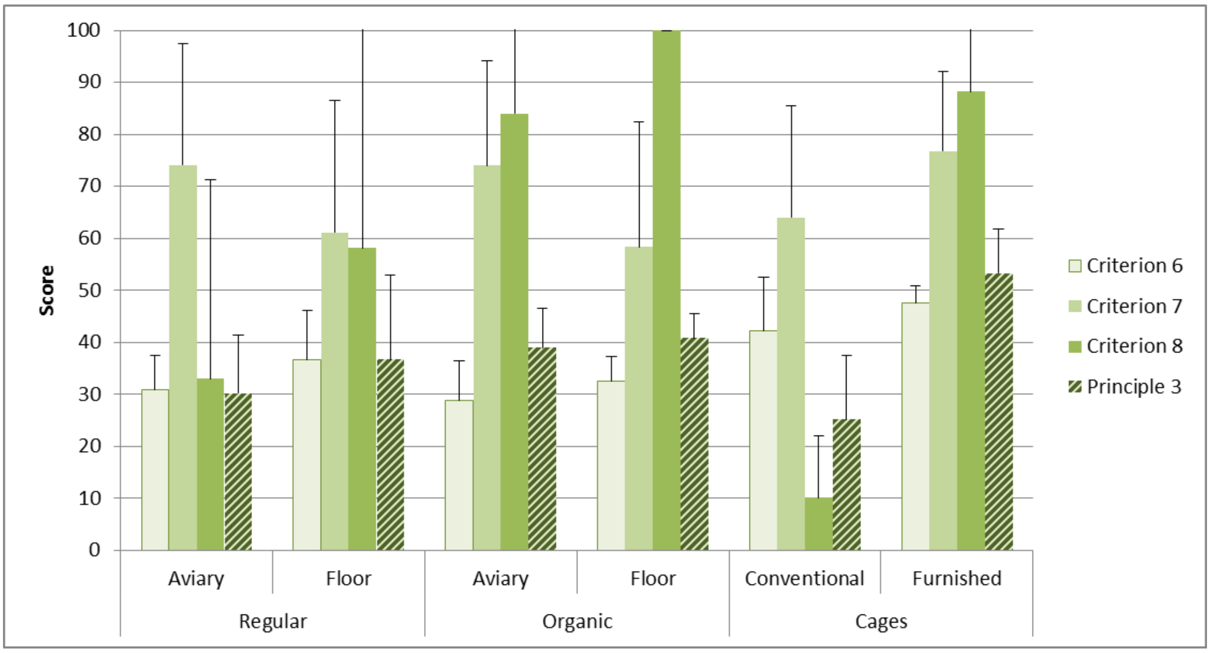


Figure 3.16: Scores for criterion 6 (Absence of injuries), criterion 7 (Absence of diseases), criterion 8 (Absence of pain induced by management procedures) and principle 3 (Good health) per housing system

3.4 Appropriate behaviour

According to welfare quality, the principle of appropriate behaviour is distinguished by four criteria: expression of social behaviours, expression of other behaviours, good human-animal relationship and a positive emotional state.

3.4.1 Expression of social behaviours

Determination of criterion 3 is based on the following measurements : absence or presence of aggression, plumage condition, comb pecks.

Aggression is scored once per flock at the end of the work in the hen house. A 2-point scale is used: 0 = no aggression observed; 1 = aggression seen. In figure 3.17 shows average scores per housing system for aggression. Aggression would appear to be lower in cage systems. However, is it difficult to determine whether there is actually less aggression or whether it is because it is more difficult observe behaviour in cages.

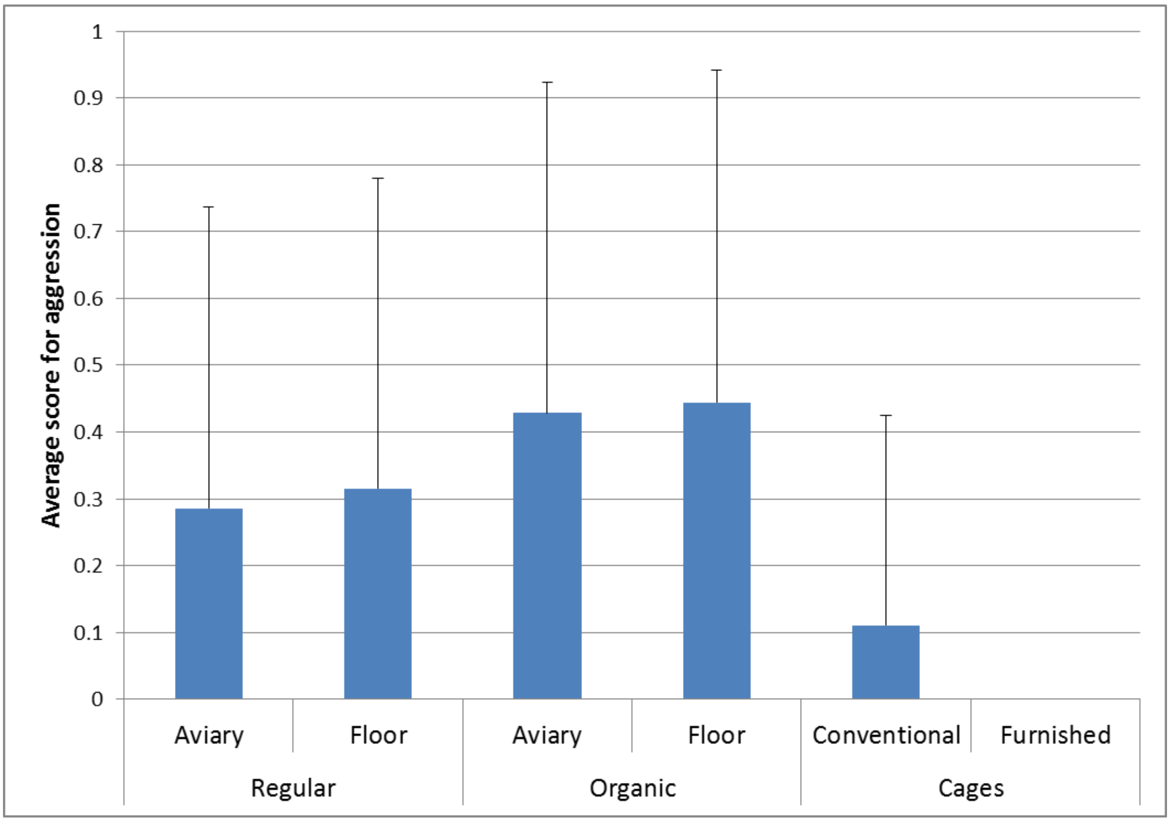


Figure 3.17: Average scores for aggression per housing system
(0 = no aggressive behaviour observed; 1 = aggressive behaviour observed)

Plumage condition was scored on 100 individually examined hens per flock on a 3-point scale: 0 = no or very slight wear, 1 = light wear; 2 = large featherless (bold) areas. Not all areas of the body are examined, only neck, back and the cloaca region, as these areas may be related to welfare(figure 3.18). Average plumage wear is 1.5. Considering all systems, 59% of the birds had an average score of 2, while 29% of the birds averaged a score of 1 and only 12% averaged a score of 0.

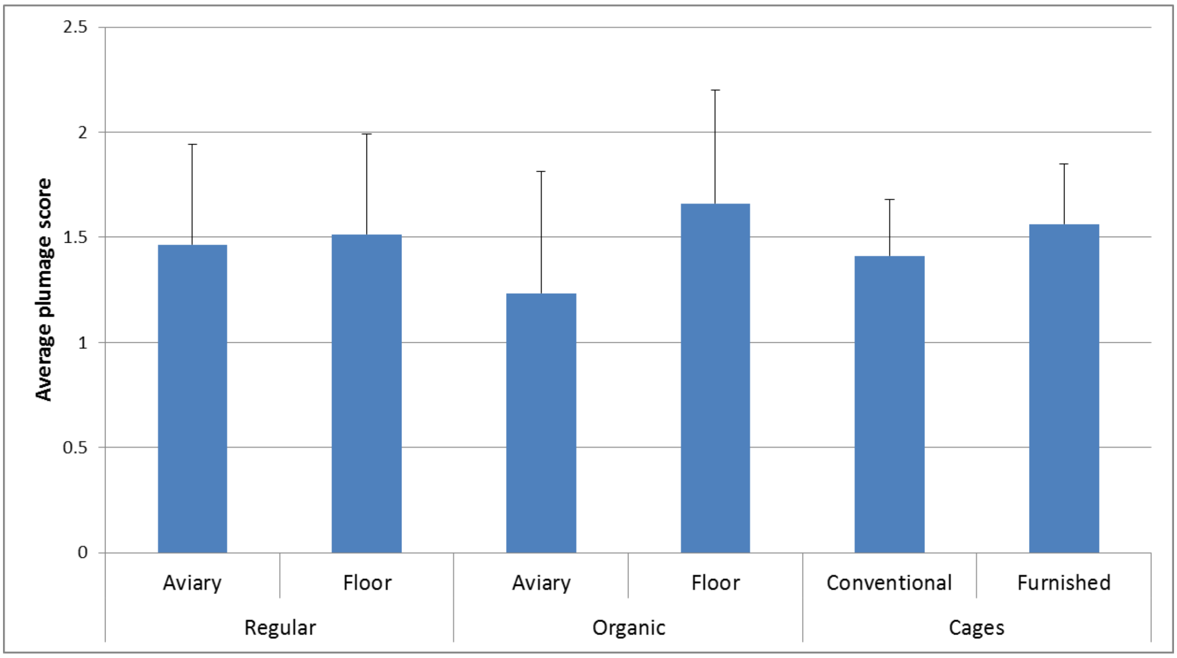


Figure 3.18: Average plumage scores per housing system
(0 = no or very slight wear, 1 = slight wear; 2 = large featherless areas)

Comb pecks were scored by 100 individually examined birds per flock. A 3-point scale is used: 0 = no comb pecks, 1 = less than 3 comb pecks; 2 = 3 or more comb pecks. Results per housing system are presented in table 3.19. Although there is quite some variation, organic flocks seem to score slightly higher than other flocks. This could be an effect of the sharp untrimmed beaks of the organic hens.

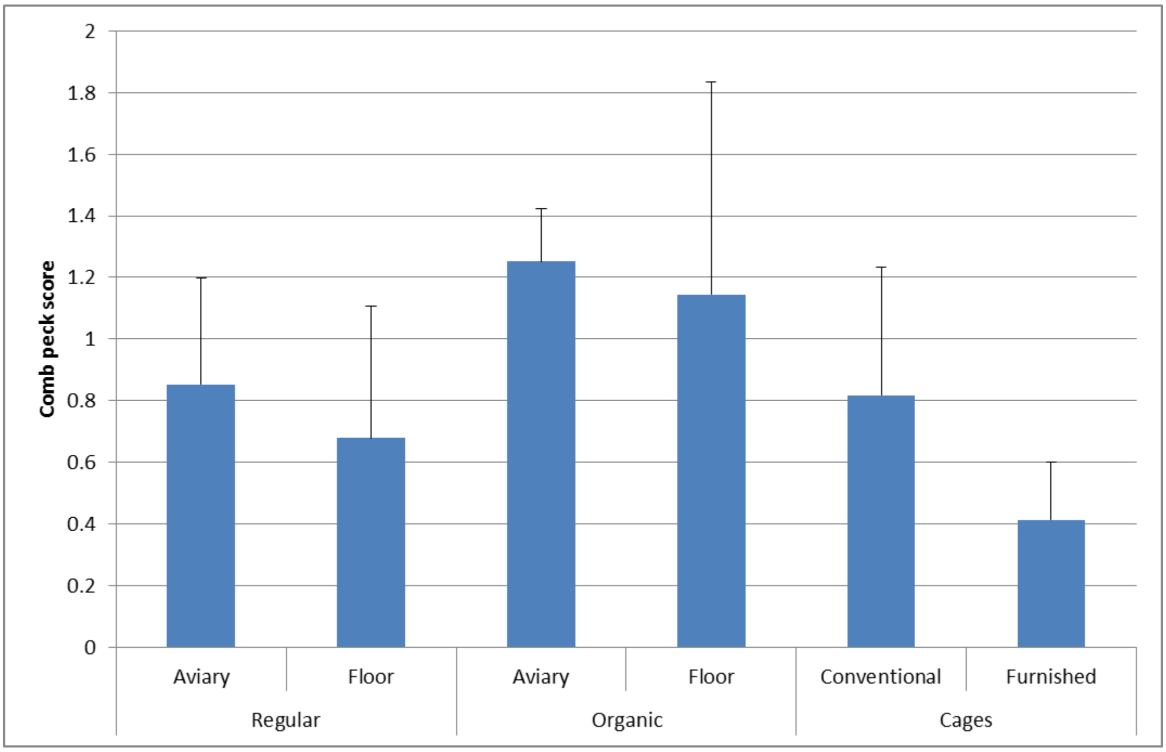


Figure 3.19: Average comb peck scores per housing system
(0 = no comb pecks, 1 = less than 3 comb pecks; 2 = 3 or more comb pecks)

3.4.2 Expression of other behaviours

Use of nest boxes is determined with 4 measurements:

- presence or absence of nest boxes
- distribution of nest boxes throughout the system (evenly spread, so hens don't have to travel far or negotiate obstacles to go to the nests)
- distribution of eggs over nest rows (evenly distributed or not)
- distribution of eggs within nest rows (evenly distributed or not)

Egg distribution is an expression of bird laying behaviour. Distribution of nest boxes was recorded while examining the hen house. Information on the distribution of eggs was provided by the farmer.

In addition to distribution and use of nest boxes the nesting space is also assessed by recording nest surface and number of nest boxes for group nests and number of nests for single nests. European legislation (EU, 1999) sets a maximum number of 7 hens per single nest and a minimum space of 1m^2 per 120 hens for group nests. Sizes for single nests were transformed into a figure for group nests as follows:

- average size for a single nest is 30*35 cm = 1050 cm
  with 7 hens per single nest this means 1050/7 = 150 cm^2/hen
- 120 hens per1m^2 means (100*100)/120 = 83.33 cm^2/hen
- space per hen in a single nest can be extrapolated to space in group nest by multiplying by 0.56 (=83.33/150).

This transformation provides an estimate for nest space pen hen in group nests based on available area per hen in single nests. Actually the space allowance for single nests is reduced to make it comparable to group nests, which is understandable since single nests provide more room that can't be used by hens (i.e. nest corners, space close to sidewalls of nest).

Figure 3.20 provides a view of average nest space per housing system. Conventional cages do not have nests and thus nest space is zero. Furnished cages provide small nests, while organic systems provide the most nest space per hen. This may be a side effect of lower stocking density.

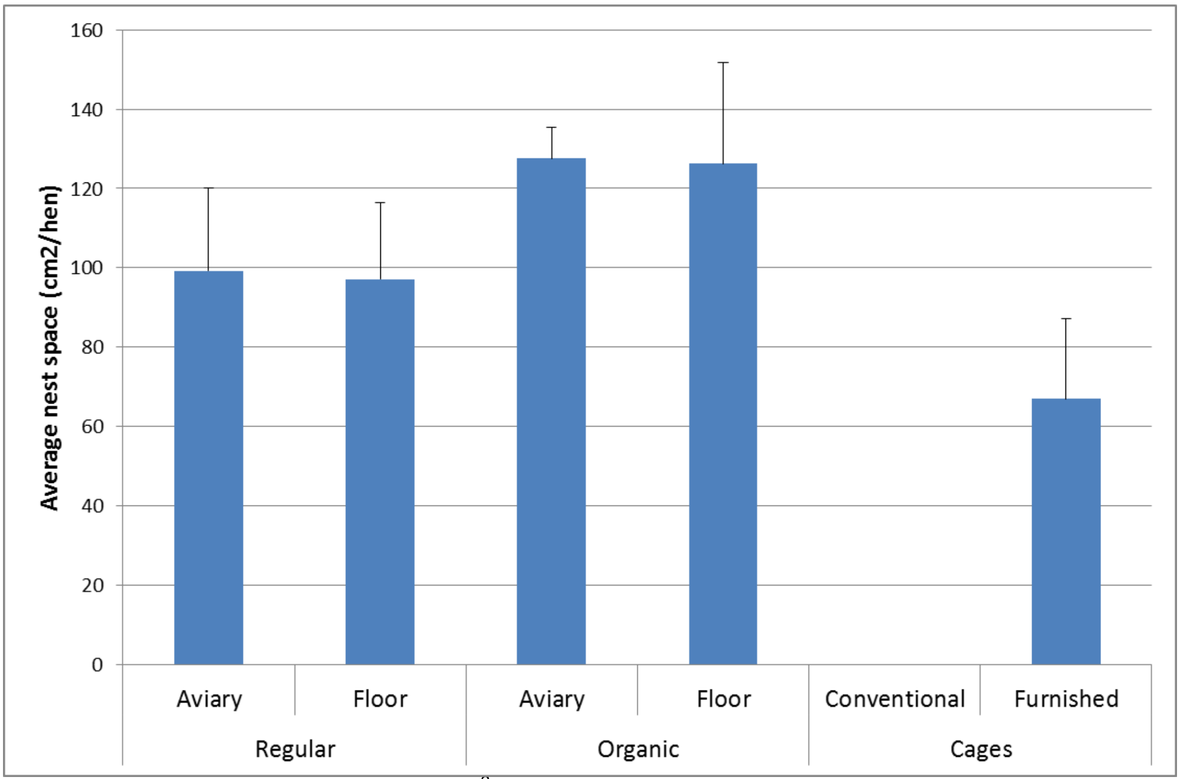


Figure 3.20: Average nest space (cm^2/hen) per housing system

Use of litter was scored after all work in the hen house was completed by filling in a diagram. The outcome is a score for litter usage on a 3-point scale: 0 = good use, 1 = moderate use, 2 = poor use or absence of litter.
Average litter use scores are presented in figure 3.21. Litter use scores for caged systems averaged 2, due to the fact that litter in cages was very limited or absent.. Other systems scored much better with slightly more variation in floor systems.

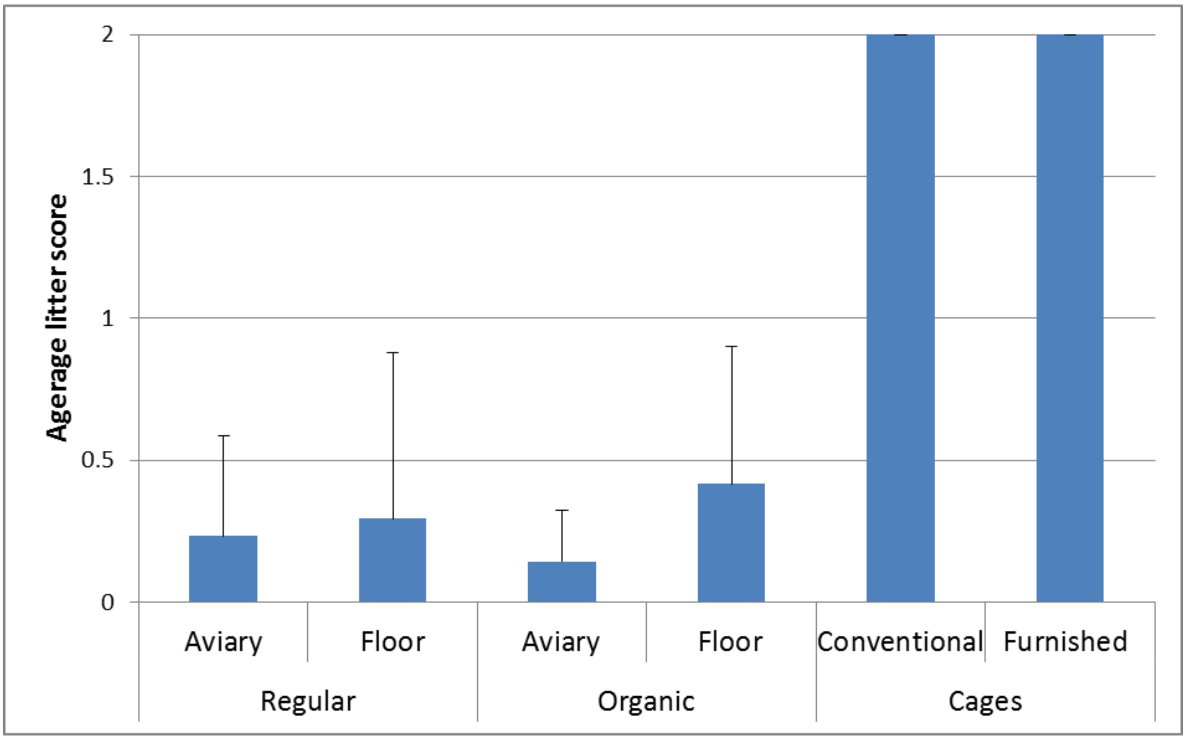


Figure 3.21: Average litter score per housing system (0 = good use, 1 = moderate use, 2 = poor use or absence of litter)

Scores for use of enrichment were determined after all work in the hen house had been performed. Assessment was based on a 3-point scale: 0 = good use, 1 = moderate use, 2 = poor use or absence of enrichment. Enrichments include bales of straw, scattering grain, etc. Specific enrichments scored were free range, cover on range (e.g. artificial roofs in the free range or natural cover such a trees) and use of a covered veranda (wintergarten).
Average scores for enrichment are presented in figure 3.22. Cages are devoid of enrichment, and thus they received a score 2. Other systems scored better, especially organic systems. The high scores for use of free range and use of cover on range for regular non-cage systems are partly due to limited use of the facilities, but also because free range is not always present.

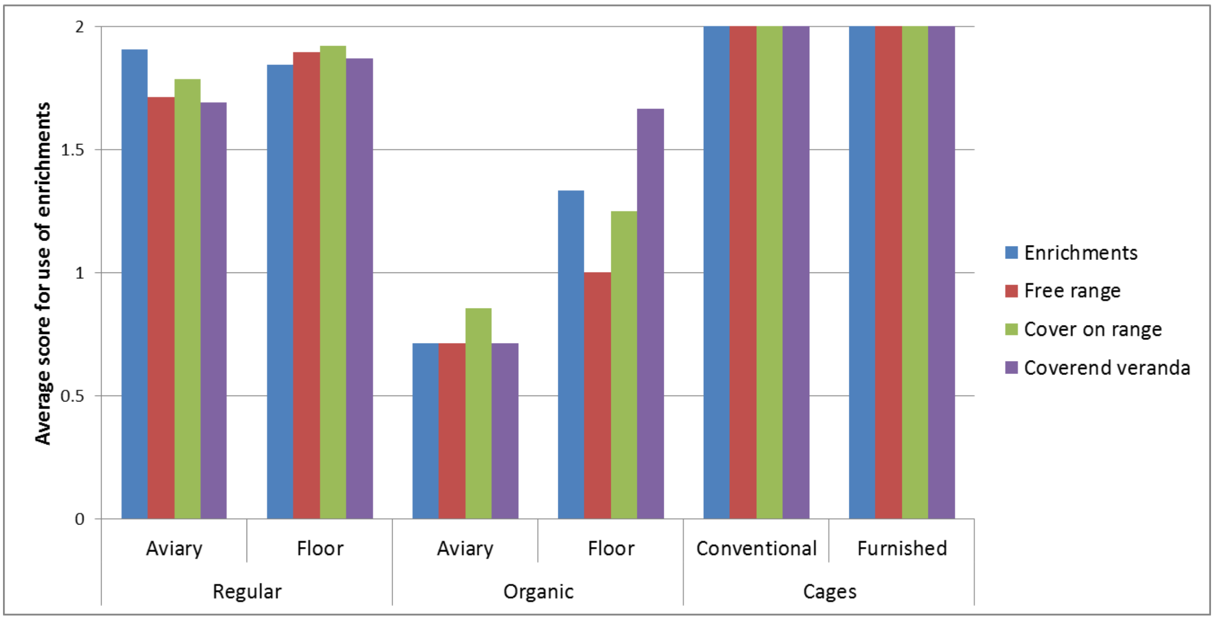


Figure 3.22: Average use of enrichements (0 = good use, 1 = moderate use, 2 = poor use or absence of enrichment)

3.4.3 Good human-animal relationship

Good human-animal relationship is measured with the Avoidance Distance Test (ADT), which is a measurement for fear of humans. This measurement was performed with 21 hens per flock. Larger distances (cm) reflect increased fearfulness of humans.
Results of ADT measurements per housing system are presented in figure 3.23 . No clear differences were found, although hens in organic floor systems seem to show slightly more fear.

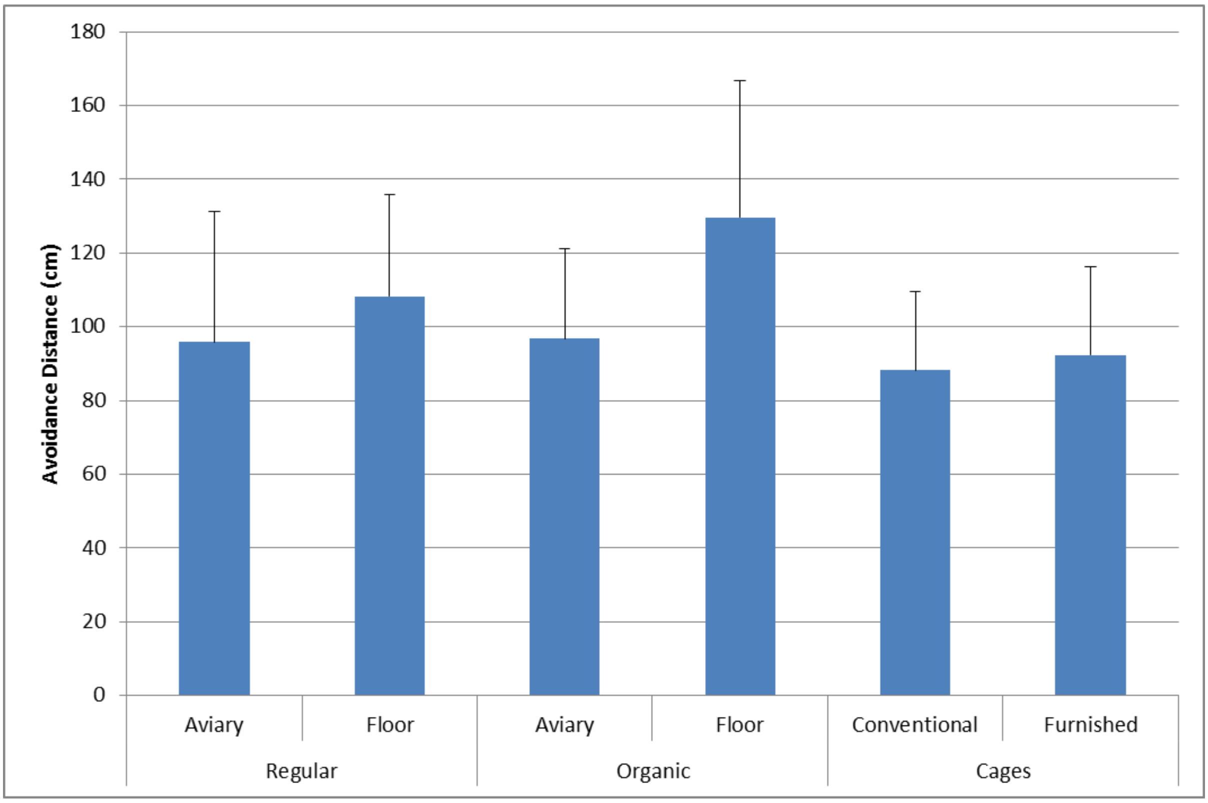


Figure 3.23: Average results of Avoidance Distance Test (ADT) per housing system. Numbers reflect the average closest distance in cm achieved between observer and obseved hens.

3.4.4 Positive emotional state

This comprises two measurements: the Novel Object Test (NOT) and the Qualitative Behaviour Assessment (QBA).

The Novel Object test reflects general fearfulness and is measured at 4 locations during 2 minutes per location. During these 2 minutes the numbers of hens close to the novel object are recorded every 10 seconds. High numbers indicate more hens close to the novel object and thus less general fearfulness.
The average outcome per housing system is given in figure 3.24. Variation in lowest in caged systems possibly because of the lack of free space and that the novel object is not easily visible to all hens in cages. Also, the number of hens in cages is limited, making it impossible to have large numbers near the object. Aviaries and regular floor systems show the largest numbers, indicating the least general fearfulness of the hens. Organic aviaries not only show large numbers, but also very high variation. There is no explanation for this.

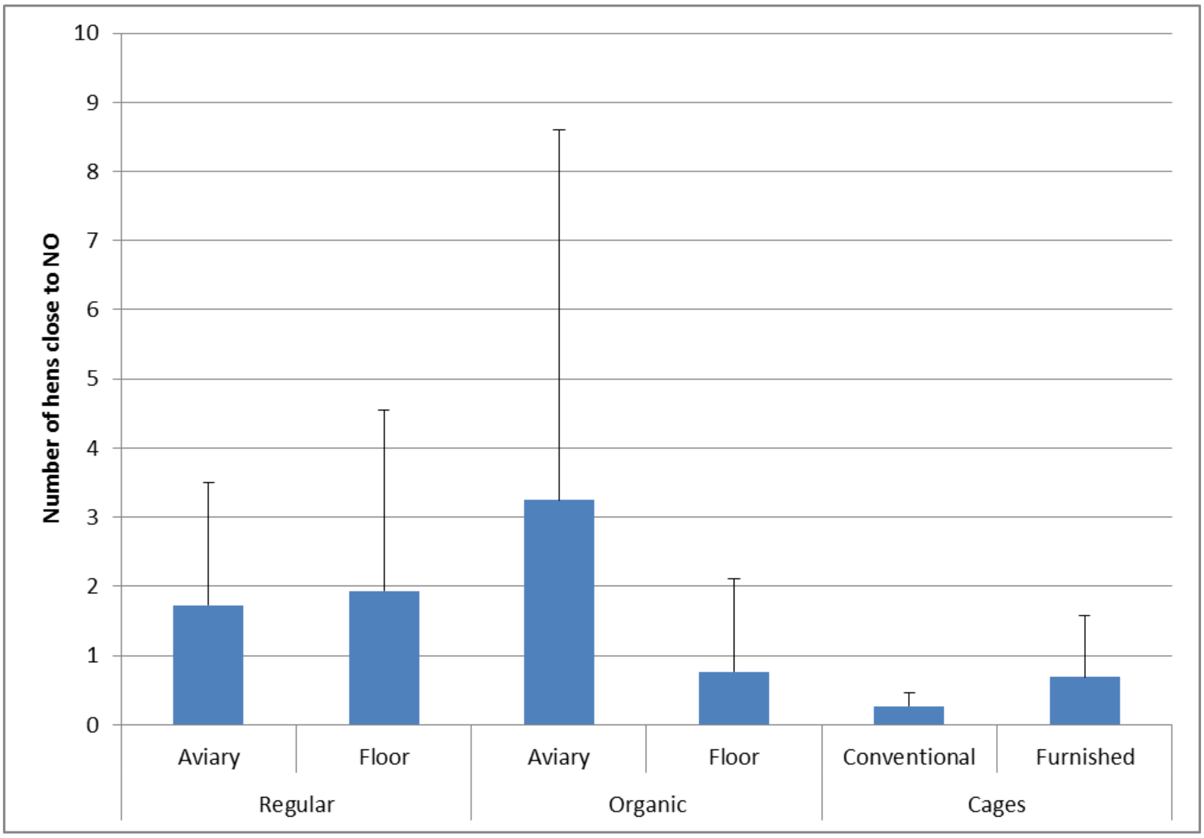


Figure 3.24: Average results of Novel Object Test (NOT) per housing system. The numbers indicate the average number of hens close to the novel object.

Qualitative Behavioural Assessment (QBA) is a measurement indicating the emotional state of the birds. The QBA is scored on, 10 positive and 10 negative emotions. The outcomes were combined to provide a single final score. A more negative outcome reflects a more negative emotional status of the birds and a more positive outcome reflects a more positive emotional status.

Figure 3.25 presents the average QBA scores per housing system. Considering the fact that QBA is based on human perception and human perception of cages usually is negative, low scores for cage systems, especially conventional cages are according to expectation. The same arguments however make it surprising that floor systems, especially regular floor systems, score low.

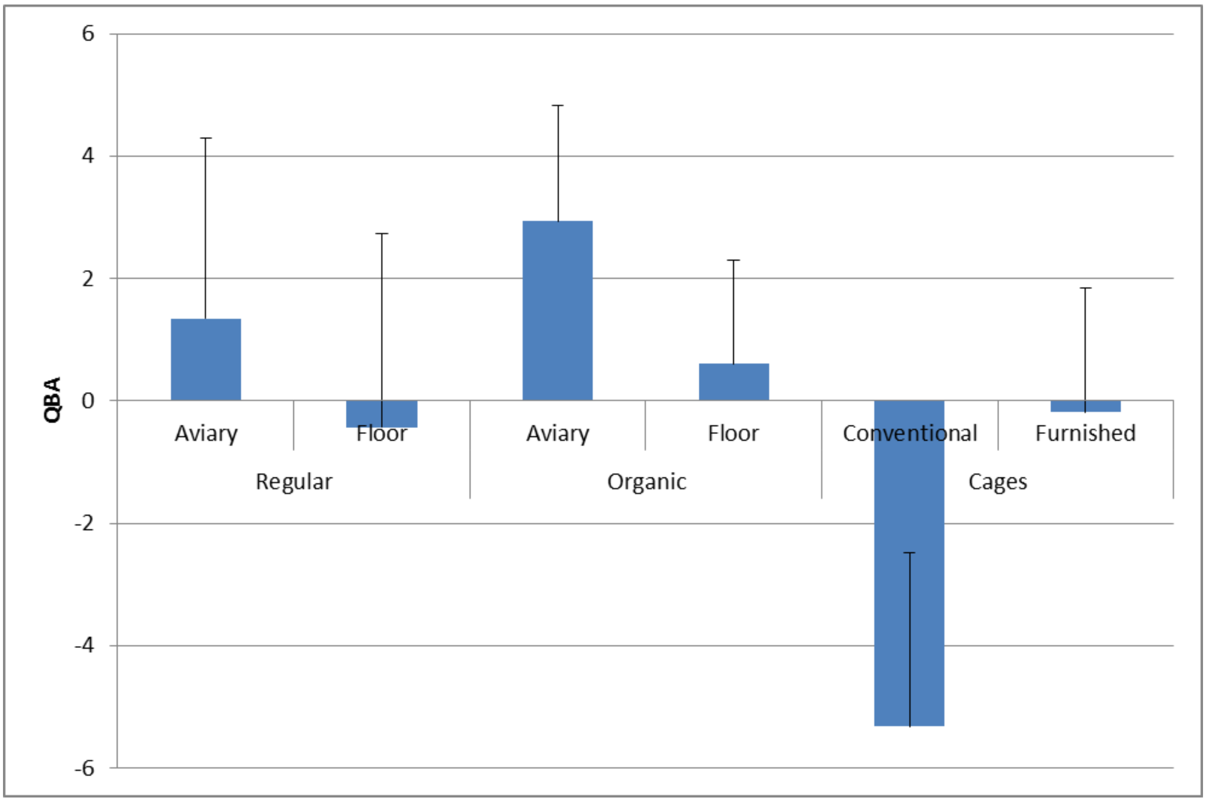


Figure 3.25: Average results of Qualitative Behavioural Assessment (QBA) per housing system. Positive scores indicate positive emotional status.

3.4.5 Criteria and scores for principle 4

Figure 3.26 presents scores for criterion 9 (Expression of social behaviour), criterion 10 (Expression of other behaviour), criterion 11 (Good human-animal relation), criterion 12 (Positive emotional state) and a score for principle 4 (Appropriate behaviour) per housing system.
In general, expression of social behaviour (criterion 9) was low for all housing systems. Expression of other behaviour is slightly higher in organic aviaries and zero in conventional cages. This is due to good use of facilities in organic aviaries (see figure 3.2.2) and the absence of facilities in conventional cages. Contrary to furnished cages, conventional cages have no litter and no nest space, resulting in a zero score for criterion 10
Criterion 11 reflects the outcome of the Avoidance Distance test. Conventional cage systems scored well in this test. It should be mentioned that this test may provide a slight bias in favour of cages, as there is always a wire door between the assessor and the birds, which is not the case in non-cage systems.
Criterion 12 reflects combined results of the QBA and NOT. Again there could be some bias due to housing system. Contrary to non-cage systems, the novel object was positioned on top of the feeder in cage systems. This feeder was always positioned outside the cage, which may influence both visibility and object appearance (in the open field or behind a wire fence). Therefore it remains unclear whether or not results from cages and non-cages are comparable. The QBA reflects a human interpretation of bird expression and can also be influenced by the impression a housing system makes on humans. This could be the cause of the very low score for conventional cages. As expression in laying hens is very difficult to observe and bird behaviour is often required for an indication of the birds emotional state, again cages may influence perception since the possibilities for bird behaviour in those systems is limited.
In general, all scores for principle 4 were fairly low. Conventional cages score lowest, mainly due to the lack of facilities to express behaviour and a very low QBA-score.

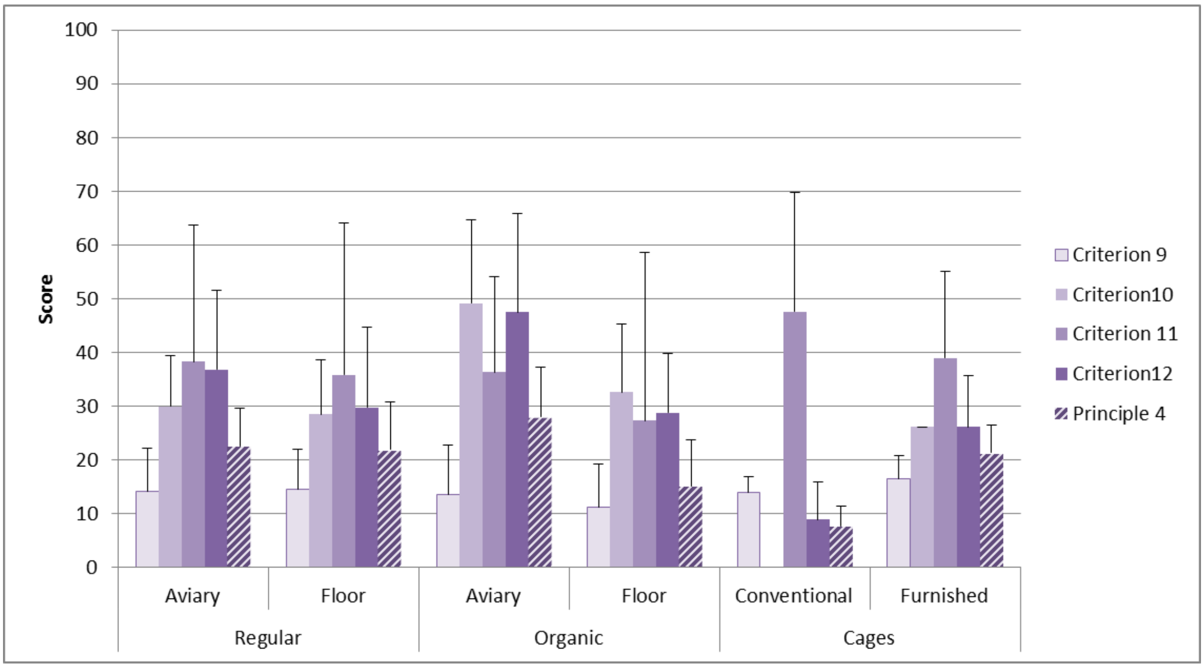


Figure 3.26: Scores for criterion 9 (Expression of social behaviour), criterion 10 (Expression of other behaviour), criterion 11 (Good human-animal relation), criterion 12 (Positive emotional state) and principle 4 (Appropriate behaviour) per housing system

3.5 Final assessment welfare categories

Figure 3.27 contains an overview of the scores for all 4 principles per housing system. Non cage systems and especially organic systems score high for principle 4 (good housing). Conventional cages score low for three of the four principles (principles 2, 3 and 4). Furnished cages score high for principle 1 and 3.

The scores for the four principles can be combined to provide an overall final assessment of welfare, based on the following four categories:

- Excellent: the welfare of the animals is of the highest level
- Enhanced: the welfare of the animals is good
- Acceptable: the welfare of the animals is above or meets minimal requirements
- Not classified: the welfare of the animals is low and considered unacceptable

In order to integrate the 4 principles into a final welfare assessment scores per principle are taken into account and limits are drawn at 10, 20, 55 and 80 points per criteria. The following system was used:

- Excellent: ≥ 55 all principles + 2 principles > 80
- Enhanced: ≥ 20 all principles + 2 principles > 55
- Acceptable: ≥10 all principles + 3 principles > 20
- Not classified: above is not achieved

Table 3.2 provides an overview of the total and individual classifications of the monitored flocks per housing system. Unfortunately, 51 flocks (43.2%) could not be given a classification due to missing information. The Welfare Quality® protocol is not very flexible on this point: a missing measurement makes it impossible to calculate one or more criteria and thus calculate scores for all principles. These flocks are indicated in the column 'Incomplete data' and should not be confused with the column Not Classified. The latter do have a classification, being ' Not classified' which means their scores were too low for the category acceptable (or higher).
From all flocks, 47 (39.8%) were classified as 'Acceptable', 15 (12.7%) were classified as 'Not classified' and 5 (4.2%) were classified as 'Enhanced'. No flocks were classified as 'Excellent'.

As number of flocks per housing system differs considerably the numbers in table 3.2 should be interpreted accordingly. Figure 3.28 contains the classifications from table 3.2 presented as percentage of the number of flocks monitored per housing system.

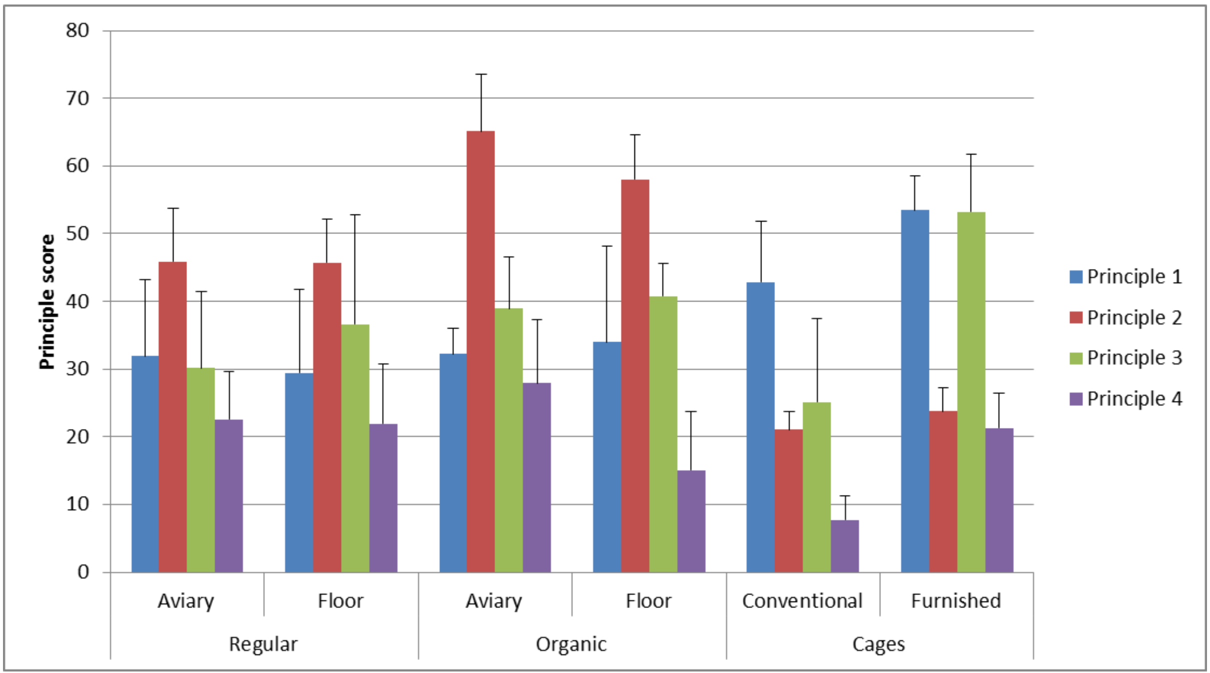


Figure 3.27: Scores for all 4 principles per housing system (Pr.1: Good feeding; Pr.2: Good housing; Pr.3: Good health; Pr.4: Appropriate behaviour)

Table 3.2: Classification of flocks per housing system according to the Welfare Quality® system

		Incomplete data	Enhanced	Acceptable	Not classified	Total
Regular	Aviary	13	1	25	4	43
	Floor	26		8	4	38
Organic	Aviary	2		4		6
	Floor	6		3		9
Cages	Conventional	4		1	5	10
	Furnished		4	6	2	12
Total		51	5	47	15	118

Surprisingly, 4 of the 5 enhanced flocks were housed in furnished cages and no flocks from organic systems could be classified as 'Enhanced'. On the other hand, no flocks in organic housing were classified as 'Not classified', which was not the case for the other systems and which indicates a higher base level than for the other systems.

Enhanced classification for furnished cage flocks is in accordance with findings of Sherwin et al. (2010) who also found the best overall welfare score for furnished cages compared to conventional cages, barn and free range. They used a different system to calculate overall welfare, but many measurements were comparable and showed the same tendency as the measurements reported in this report. Also findings reported in the EFSA report on the welfare of laying hens (EFSA, 2005) and the Laywel project (Nicol et al., 2007) are supporting the findings in this report.
Enhanced welfare classification for furnished cages can be explained by the fact that all four enhanced furnished cage flocks were Swedish flocks with good health and intact beaks. This gave them a high score for principle 3. They also had a fair amount of litter and nest boxes available, resulting in a score above 20 for principle 4. Although space is limited in furnished cages, the

presence of perches and a good climate resulted in a score for principle 2 above 20. Together with the ample feeder and drinker space, this resulted in the classification enhanced.

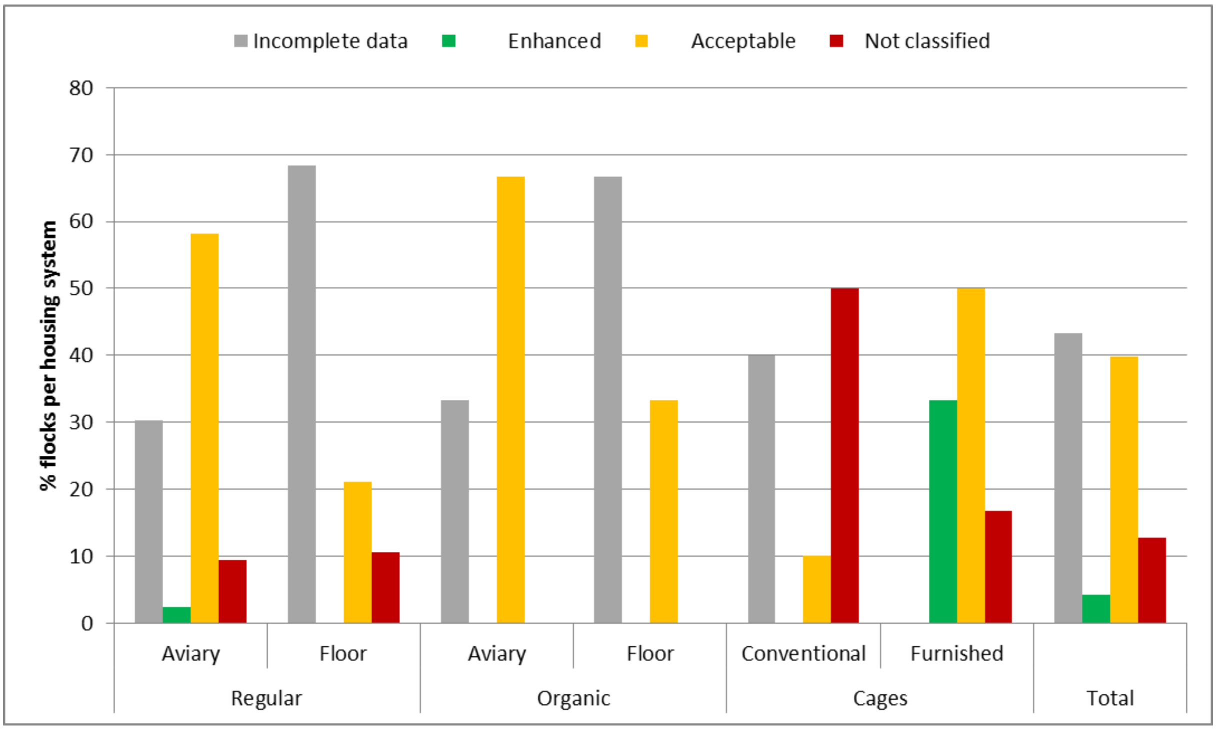


Figure 3.28: Classification of flocks per housing system according to the Welfare Quality® system, presented as percentage flocks per housing system

3.6 Modification of some measurements

Although the literature supports the finding of more furnished cage flocks classified as enhanced, one can question why only few non-cage flocks fall into the category 'enhanced'. To a certain extent this can be explained by some measurements that may need improvement. These measurements lack refinement under certain circumstances, resulting in an over estimation of a negative score for certain aspects. The following measurements may be responsible:

Principle 1:

- Feeder space: feeder space per bird does not necessarily reflect feeding status of the bird. Apart from available space also feeding management, more specifically, frequency and time of feeding, feed composition and feed structure have a large influence. A better measurement therefore would be an animal based measurement, for instance an indication of percentage of lean birds (e.g. Sherwin et al (2010) uses keel bone protrusion as an indicator for body condition).
- Drinker space: similar to feeder space the number of nipple drinkers does not necessarily reflect thirst of birds. To score thirst it would be better to record the percentage of birds with blue coloured combs (indicating dehydration of the birds).

Principle 3:

- Keel bone deformations: These are scored on a 2-point scale, indicating only whether or not there is a deformation. However, not all deformations are of similar severity. Some minor deformities may have been caused by decalcification during rearing or early lay, without any evidence of broken bones. One can question whether or not these deformities actually affect bird welfare. Schotz et al. (2008) carried out histological examinations of keel bone deformations in various severities. They stated that s-shaped deviations of keel bones were related to extended pressure loading while perching activities rather than short-duration trauma. These s-shapes ore slight flattenings of the keel bone therefore are not likely to affect welfare negatively. In the present system, these minor deformations are however scored in the worst category. It would be better to introduce an extra category for such minor deformations, improving refinement of scoring.

- Aggression: this is scored on a 2-point scale answering the question "have you seen aggressive behaviour". First of all, some aggression is part of the normal behavioural repertoire and usually does not have an impact on bird welfare. Only excessive aggression should be estimated as a threat to animal welfare and as undesirable. The 2-point scale however, only allows a yes or no answer. As birds in non-cage systems are more capable of expressing their specific behaviour, observation of aggressive behaviour is more likely than in cages. It would be better to refine the scoring of aggression by adding more categories, e.g. minor aggression (e.g. less than 5 encounters), frequent aggressive encounters (5 or more encounters). A further improvement would be to include additional information from the clinical scoring. Plumage condition is scored on 3 areas, one of them being the back of the head. This area reflects aggressive pecking. If this area of plumage scores is recorded separately from the other areas, it could provide additional information on aggressive behaviour.

Some other measurements requiring extra consideration:

- Toe damage: this can be wounds, but also missing toes. From observations in two particular flocks, where also clinical scoring during rearing was performed, strong indications have been obtained that the majority of toe damage originates during rearing. One cause of missing toes is small chicks being stuck in the feeding chain. To reflect welfare in the laying house, toe damage should therefore focus on fresh wounds.
- Use of litter: according to the NEN-protocol the schedule developed to monitor use of litter is only used once. However, there are several farms where ample use is made of litter in one area, but not in another area of the house. Also there are houses where little dust bathing is seen in the house, but a lot of it outside in the free range area. If this house is scored only once for the inside area, it would score low, whereas in fact the birds are expressing their behaviour in the outside area. In an attempt to avoid this and to give a more detailed score, we applied the schedule 4 times. It was left to the observer to determine the places to score. This could be four different areas inside the house, but it could also be 2 areas inside the house and 2 in the covered veranda. If a lot of dust bathing was seen in the free range area, but not in the house, one or two schedules could be used for the free range area.
- Novel Object Test: in non-cage systems the novel object is positioned in the litter area. In cages this is not possible as there is no litter or the litter area is too small. Also the novel object is too large to place in a cage. It therefore is positioned on the feed trough, outside the cage. First of all the number of hens in a cage is limited, making it impossible to have large numbers of birds near the object. Second, the sight of the object outside the cage, on top of a feeder may be completely different from a novel object lying in the litter. Also, it is quite likely that some of the hens in cages do not notice the object. One can therefore question whether the outcome in cages is comparable to that in non-cage systems. As yet, there is no acceptable solution to this problem.

4 Practical applications of the Welfare Quality® laying hen protocol

The Welfare Quality® protocols are often seen as classification systems for overall welfare of a particular flock and that is indeed the final outcome. This only becomes interesting when economic consequences are connected to such assessments, e.g. a higher price for eggs from highly qualified flocks. However, the Welfare Quality® protocol offers much more than just a general qualification of flock welfare and in fact can be used to good effect as an on-farm management tool to improve animal welfare .

For example, 5 regular aviary flocks were chosen from the database in an attempt to illustrate this. The principle scores for these 5 flocks are presented in figure 3.29 . In order to investigate further the basis for these scores the criteria scores are given in figure 3.30. Low criteria scores reduce the principle scores and thus improving one or more criteria scores improves the principle score. This not only provides a better overall welfare assessment, but more importantly, actually improves bird welfare.

Flock 'Acceptable 1' would progress to "enhanced" if one more principle scored above 55. According to figure 3.30 there are various possibilities for improvement:

- principle 1 by installing more feeder space
- principle 3 by reducing the criterion score for injuries. This score comprises keel bone deformations, skin lesions, foot pad lesions and toe damage. A closer examination of which measurements are responsible for the lower scores would provide an insight into where the opportunities lie for improvement of the criteria.
- r principle 4 by improving the score for criterion 9 (expression of social behaviour). This criterion comprises the measurement of aggressive behaviour, plumage condition and comb pecks. Measures taken to reduce pecking behaviour (e.g. providing roughage) could result in better feather quality, less aggression and ultimately in a better score for social behaviour.

For this flock this list of possibilities helps the farmer in making decisions of how to improve bird welfare. Information gained from such actions may not always be applicable directly but certainly for management subsequent flocks.

Flock 'Acceptable 2' would progress to "enhanced" if 2 principles scored above 55 points and principle 3 scored a minimum of 20 points. According to figure 3.30 there are various options for improvement:

- principle 1 by installing more feeder and water space.
- principle 2 by improving criterion 3 (comfort around resting) and 5 (ease of movement). Criterion 3 comprises measurements for perches, red mites and dust. Installing more perches or reducing red mite infestation would improve the score for criterion 3. Criterion 5 comprises stocking density and percentage area of perforated floor. Reducing stocking density would increase the score for criterion 5.
- principle 3 scores low specifically because the the score for criterion 8 (absence of pain induced by management procedures) is low. This implies that beak treatment in this flock is poor. Improvement is only possible for subsequent flocks.
- principle 4 scores low specifically because the score for criterion 9 (social behaviour) is low. This criterion comprises measurement of aggressive behaviour, plumage condition and comb pecks. Measures taken to reduce pecking behaviour (e.g. providing roughage) could improve feather quality, reduce aggressive behaviour and improve the score for social behaviour.

Flock 'Not classified 1' would progress to "acceptable" by improving the score for principle 3. According to figure 3.30 principle 3 scores low specifically because the score for criterion 8 (absence of pain induced by management procedures) is low. This indicates that beak treatment in this flock is poor. Improvement is only possible for subsequent flocks.

Flock 'Not classified 2' would progress to "acceptable" by improving the low score for principle 4. According to figure 3.30 options for improvement include:

- principle 4 scores low for all criteria, meaning that this flock has high levels of fear and low levels of social and other behaviour. Various measures taken to reduce fear and improve behaviour should be considered in order to improve principle 4.

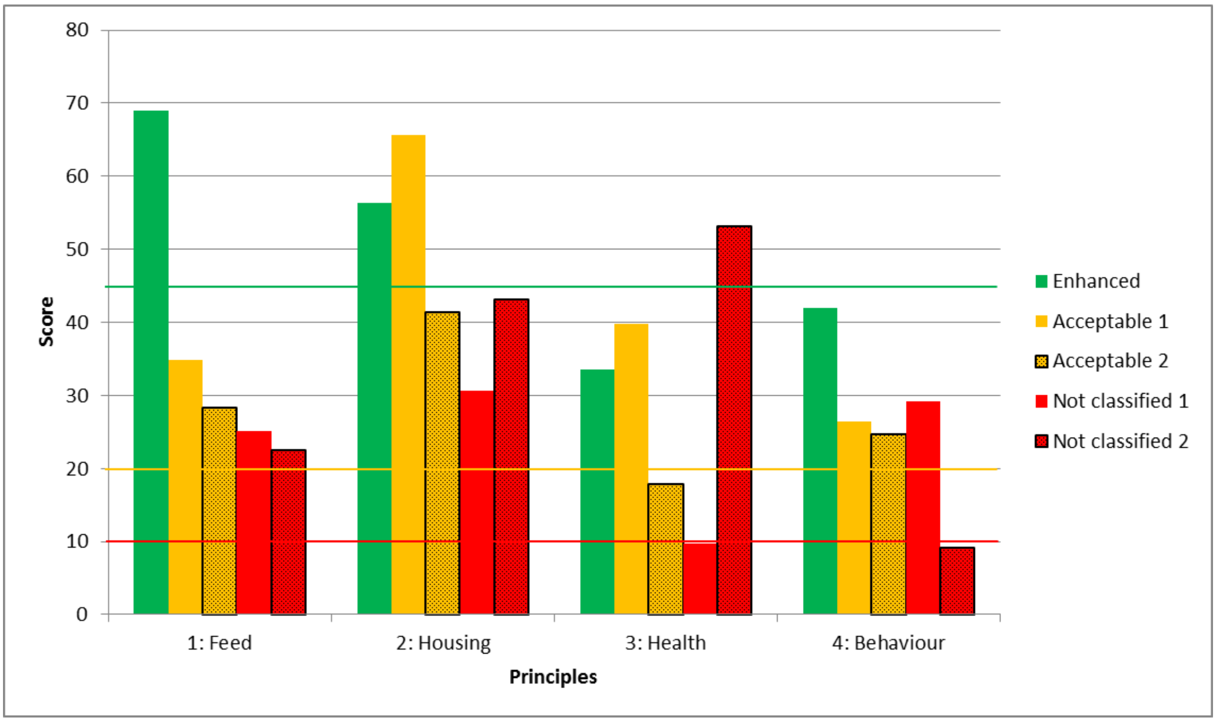


Figure 3.29: Principle scores and classification according to the Welfare Quality® system for 5 flocks from a regular aviary housing system

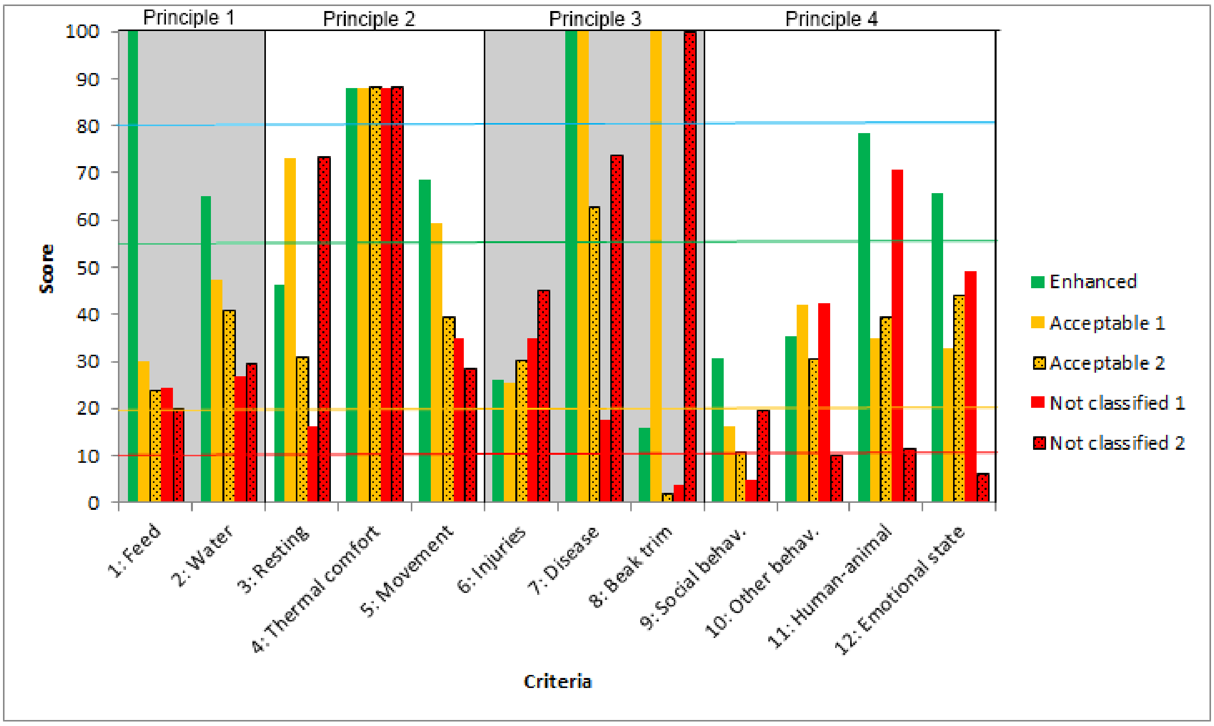


Figure 3.30: Criteria scores and classification according to the Welfare Quality® system for 5 flocks from a regular aviary housing system

The examples presented above indicate the potential of the Welfare Quality® protocol as a good management tool for:

- Identification of areas
- Identification of measures to improve animal welfare
- Comparison of overall on-farm welfare with individual welfare criteria and measurements
- Comparing flocks both on overall welfare as on various welfare criteria and measurements

By looking at and comparing outcomes of criteria farmers can identify where possibilities are to improve. When going back to measurement level, the protocol can be used as a management tool.

5 Possibilities for simplification of the laying hen protocol

5.1 Starting point

The search for possibilities for the simplification of the laying hen protocol was based on the assumption that meaningful correlations exist between various (animal-based) measurements recorded on laying hens. If this is the case, one measurement could be a predictor for another measurement. The Welfare Quality® framework would remain intact, and scores for all 12 Criteria and all 4 Principles could be calculated using a simplified version of the protocol.
Another simplification would be to reduce the sample size, i.e. the number of birds used for clinical examination, within the framework of the protocol.

5.2 Calculation

In order to examine the relationships of various measurements in the laying hen protocol a correlation analysis was performed alongside a principal component analysis . These analyses clearly showed that no meaningful correlations existed between parameters measured for the laying hen protocol (correlations were generally low, < 0.20, and not significantly different from 0: P > 0.20). Therefore, simplification strategies based on potential predictors were not considered further.
Reduction in sample size was analysed in a simulation study involving random sampling of half and 25% of a total sample size (100 birds), respectively, and this procedure was repeated 100 times.

5.3 Assessment of the quality of the simplification

The assessment of the quality of a simplification strategy involves three criteria:
- (a) % of agreement
- (b) Sensitivity
- (c) Specificity

These criteria can be explained using a 2x2 table.

Example of a 2x2 table for the calculation of criteria used for the assessment of the quality of the simplification of the Welfare Quality® protocol for laying hens

	Simplified version (reduced sample size)		
Reference	Positive	Negative	
Positive	A	B	A + B
Negative	C	D	C + D
		TOTAL	N

The outcome of a welfare assessment is expressed as either ‘Positive’ or ‘Negative’. The number (N) refers to the total number of farms where the protocol was applied. On each farm, the original ‘full’ protocol (reference) was applied along with a simplified version . The results of the on-farm assessments, can be allocated to a cell of the 2x2 matrix. Cell A contains results from farms assessed as ‘Positive’ according to both the reference and the simplified version, results in cell B are from farms assessed as ‘Positive’ according to the reference protocol but ‘Negative’ according to the simplified version. If there is complete agreement between reference and simplified version, all farms would be present in A and D, and cells B and C of the matrix would be empty.
Following the 2x2 matrix, the three criteria reflecting the quality of the simplification can be expressed as follows:
- (a) % of agreement = A+D/N x 100%
- (b) Sensitivity = A/A+B x 100%.
  100 – sensitivity = % false negatives = B/A+B x 100%
- (c) Specificity = D/C+D x 100%.
  100 – specificity = % false positives = C/C+D x 100%

The outcome of the Welfare Quality® protocol for laying hens was expressed in terms of two classes (i.e., 'Positive' versus 'Negative') as follows:

(i) An end qualification (Excellent, Enhanced, Acceptable or Not classified) where two classes were created by combining the worst two and the best two classes, i.e. Excellent + Enhanced versus Acceptable + Not classified.

(ii) At principle level, farms were divided across two classes based on Principle score, i.e. having a score higher or lower than a threshold value. Three thresholds were considered in this respect: 20, 55 en 80.

Below is an example of a 2x2 matrix based on the classification of farms according to the end qualification. NC = Not classified, A = acceptable, Enh = Enhanced en Exc = Excellent:

	Simplified version (reduced sample size)		
Reference	NC + A	Enh + Exc	
NC + A	A	B	A + B
Enh + Exc	C	D	C + D
		TOTAL	N

Similarly, a 2x2 matrix can be created based on the classification of farms according to the principle score:

	Simplified version (reduced sample size)		
Reference	Score < 20	Score ≥ 20	
Score < 20	A	B	A + B
Score ≥ 20	C	D	C + D
		TOTAL	N

In the above example, the threshold of 20 is used. Similar tables can be created using threshold values of 55 (i.e. score < 55 versus score ≥ 55), or 80 (i.e. score < 80 versus score ≥ 80).
For a simplified version to be acceptable in terms of performance relative to the reference, sensitivity and specificity should at least be higher than 90%.

As mentioned above, a simulation study involved 100 repetitions of random samples of 50 or 25% of the original sample size (100 birds). Thereafter, the following three steps were taken:

1) For each of the 100 repetitions, the % agreement, the sensitivity and the specificity were calculated.
2) The 100 simulated outcomes for % agreement, the sensitivity and the specificity were ordered from high to low.
3) Out of these 100 values, the 5th, 50th and 95th value were reported; these values are equal to the 5th, 50th, and 95th percentile points. The closer these three values are together, the higher the reliability of the simplification, i.e., the more acceptable the reduction of the sample size would be.

5.4 Results

For the end qualification, the outcome of the reduction of the sample size can be summarized as follows:

Reduction of sample size	Reference
%agreement	**50% (5%-95%)**
- n = 25	91.5 (87.3-95.8)
- n = 50	93.0 (89.4-97.2)
- n = 75	94.4 (90.1-97.2)
- n = 100	94.4 (90.8-97.2)

%specificity	
- n = 25	80.0 (40.0-100.)
- n = 50	80.0 (60.0-100.)
- n = 75	80.0 (60.0-100.)
- n = 100	90.0 (70.0-100.)

%sensitivity	
- n = 25	100. (98.5-100.)
- n = 50	100. (98.5-100.)
- n = 75	100. (98.5-100.)
- n = 100	100. (98.5-100.)

Reduction in sample size compromises specificity in particular.

The matrix below provides the results of the reduction in sample size for principle 3 scores.

Principle 3		20.0			55.0			80.0		
		50%	5%	95%	50%	5%	95%	50%	5%	95%
Reduction of sample size to 25 hens instead of 100	%agree	89.3	86.8	91.7	95.9	92.6	97.5	100.	100.	100.
	%se	72.0	68.0	82.0	97.7	95.4	99.1	100.	100.	100.
	%sp	92.7	91.7	94.8	80.8	61.5	92.3			

Sensitivity and specificity are too low in comparison with the reference situation (100 hens) when only 25 hens are observed.

5.5 Conclusions

Analyses clearly showed that no meaningful correlations exist between measurements of the laying hen protocol. Therefore, simplification strategies based on predictors were not an option.
Reduction in sample size, as analysed in a simulation study, clearly showed that reduction in sample size from 100 to 50 or 25 animals is not acceptable in terms of specificity and/or sensitivity.

6 Conclusions and recommendations

- Measurement of 122 flocks in 2 countries provided a lot of useful information not only for assessment of welfare quality in flocks , but also regarding implementation possibilities for the protocol as management tool. It also provided information concerning suggestions for improvements of certain measurements.
- Unfortunately 51 flocks (43.2%) could not be classified due to missing information. The Welfare Quality® protocol is not very flexible on this point: a missing measurement makes it impossible to calculate one or more criteria and thus eliminates calculation of principle scores. Ways should be found to overcome this by using a substitute score, e.g. an average score for that type of housing system, a score of a previous flock from the same house or an average score for that type of hen in that type of housing system.
- Variation in overall welfare assessment was highest in furnished cages, ranging from 'Enhanced' to 'Not classified'. Organic systems showed the least variation in overall welfare assessment. Results show that even with limited space such as in cages, there are possibilities to establish good animal welfare.
- Some of the measurements appear too simple and do not provide enough information for a detailed appraisal of the on-farm situation. A more refined classification would provide a more accurate indication of the welfare status in this respect. Also replacement of certain resource based measurements with animal based measurements would improve the Welfare Quality® assessment for laying hens.
- As, on average, it takes 7 hours to perform the full Welfare Quality® protocol for laying hens an investigation was made into possibilities for simplification , without compromising the general Welfare Quality® structure of the protocol. No correlations were found between measurements, indicating that there was no overlap between measurements. This also indicates that there are no possibilities to forecast measurements using predictor parameters. An alternative was to reduce the number of hens for the clinical observations. The results of the simulation study clearly indicate that reduction in sample size from 100 to 50 or 25 animals is not acceptable in terms of specificity and/or sensitivity.
- Although reduction of the time needed for the protocol appears impossible, this is not necessarily a large problem if the frequency of monitoring can be limited even more (normal frequency is already low, being once every flock and thus less than once per year). In order to reduce frequency, possibilities should be investigated to use certain measurements as indicators to warn when to perform a full protocol assessment. Examples of such indicators are possibly: mortality and culls. Sherwin et al. (2010) found second grade eggs (number or type) a good indicator for stress in flocks. Although second grade eggs are not included in the Welfare Quality® protocol for laying hens, the measurement can easily be carried out and thus could be considered as indicator when to perform the full Welfare Quality® protocol.

Literature

Bestman, M., P. Koene & J-P Wagenaar, 2009. Influence of farm factors on the occurrence of feather pecking in organic reared hens and their predictability for feather pecking in the laying period. Applied Animal Behaviour Science 121: 120–125

Bilcik, B. and L.J. Keeling,1999. Changes in feather condition in relation to feather pecking and aggressive behaviour in laying hens. British Poultry Science 40: 444-451.

Blokhuis, H.J., Veissier, I., Miele, M., Jones, B., 2010. The Welfare Quality® project and beyond: Safeguarding farm animal well-being. Acta Agric. Scand. Sect. A-Anim. Sci. 60, 129-140.

Brscic, M., Wemelsfelder, F., Tessitore, E., Gottardol, F., Cozzi, G., Van Reenen, C.G., 2009. Welfare assessment: correlations and integration between a Qualitative Behavioural Assessment and a clinical/health protocol applied in veal calves farms. Italian Journal of Animal Science 8, 601-603.

EFSA, 2005. Welfare aspects of various systems for keeping laying hens. European Food Safety Authority Journal, 197: 1-23

EU, 1999. COUNCIL DIRECTIVE 1999/74/EC of 19 July 1999 laying down minimum standards for the protection of laying hens. Official Journal of the European Communities (3.8.1999) L203:53-57.

Kappeli, S., Gebhardt-Henrich, S.G., Frohlich, E., Pfulg, A., Schaublin, H., Stoffel, M.H., 2011b. Effects of housing, perches, genetics, and 25-hydroxycholecalciferol on keel bone deformities in laying hens. Poult. Sci. 90, 1637-1644

Main, D.C.J., Kent, J.P., Wemelsfelder, F., Ofner, E., Tuyttens, F.A.M., 2003. Applications for methods of on-farm welfare assessment. Animal Welfare 12, 523-528.

Manten, A., De Jong, I.C., 2011. Protocollen voor het meten van dierenwelzijn. V-focus special 5: 16-17.

Nicol, C.J., Tauson, R., C.A. Weeks & H.A. Van de Weerd, 2007. The LayWel project: welfare implications of changes in production systems for laying hens. World's Poultry Science Journal, 63: 101-114

Pickel, T., Scholz, B., Schrader, L., 2010. Perch material and diameter affects particular perching behaviours in laying hens. Appl. Anim. Behav. Sci. 127, 37-42.

Scholz, B., Ronchen, S., Hamann, H., Surie, C., Neumann, U., Kamphues, J., Distl, O., 2008. Evaluation of bone strength, keel bone deformity and egg quality of laying hens housed in small group housing systems and furnished cages in comparison to an aviary housing system. Arch. Tierz.-Arch. Anim. Breed. 51, 179-186

Sherwin, C.M., Richards, G.J., Nicol, C.J., 2010. Comparison of the welfare of layer hens in 4 housing systems in the UK. Br. Poult. Sci. 51, 488-499.

Stockman, C.A., Collins, T., Barnes, A.L., Miller, D., Wickham, S.L., Beatty, D.T., Blache, D., Wemelsfelder, F., Fleming, P.A., 2011. Qualitative behavioural assessment and quantitative physiological measurement of cattle naive and habituated to road transport. Animal Production Science 51, 240-249.

Van Niekerk T.G.C.M., J.T.N.M. Thisse & C.G.van Reenen, 2012: Welfare Quality: description of the calculations to transform measurements into criterion for the welfare assessment protocol for laying hens. Report … (in preparation), Wageningen UR Livestock Research.

Welfare Quality®, 2009. Welfare Quality® assessment protocol for poultry (broilers, laying hens). Welfare Quality® Consortium, Lelystad, Netherlands.

Wemelsfelder, F., 2007. How animals communicate quality of life: the qualitative assessment of behaviour. Animal Welfare 16, 25-31.

Acknowledgements

Special thanks to the assessors (Jitske Westra, Gisabeth Binnendijk and Berry Reuvekamp) who made a tremendous effort with the farm visits. We are very grateful to them for their enthusiasm and efforts and admire their motivation, despite the long days.

We are also very grateful to the following people:

Geert André, for his assistance with the data analysis.

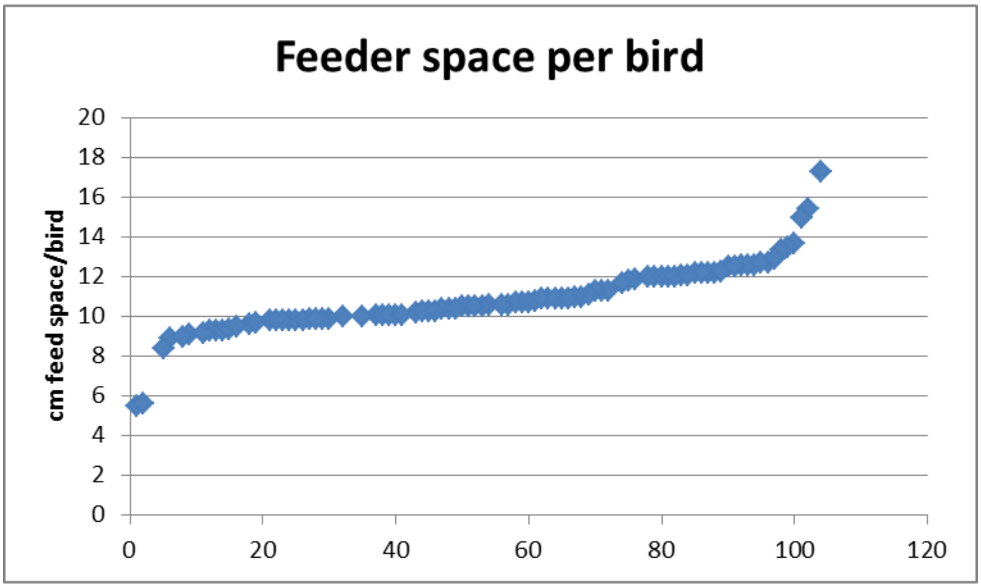
Prof. Dr. Linda Keeling and Anna Larson who very kindly provided us with information concerning the Swedish flocks included in the analysis.

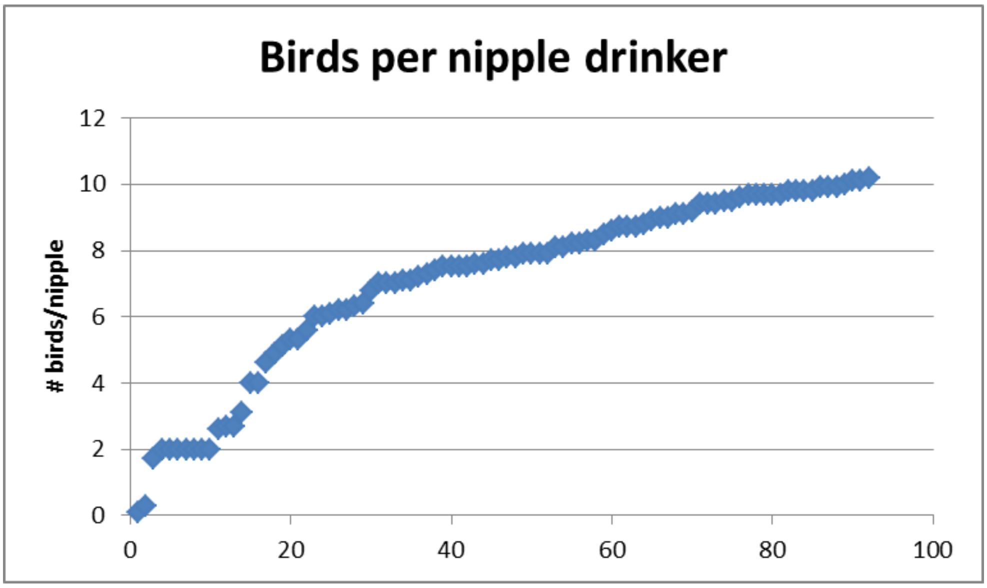
Hans van de Heuvel who provided us with software for PDA’s, enabling us to download data into the database, and who also provided us with a format for communication with the farmers.

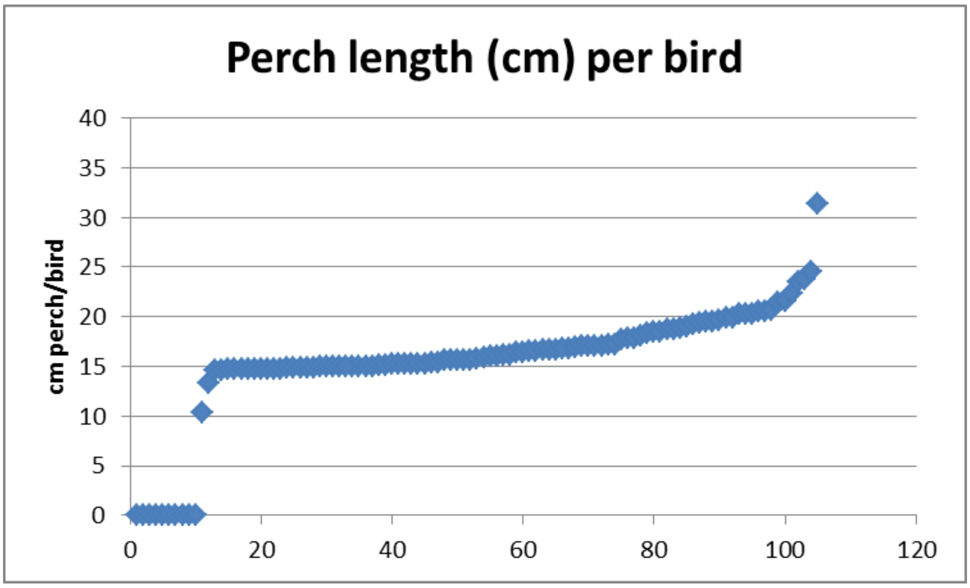
Experts Ingrid de Jong, Ute Knierim, Marion Staack and Claire Weeks for their expertise input in the multi criteria analysis.

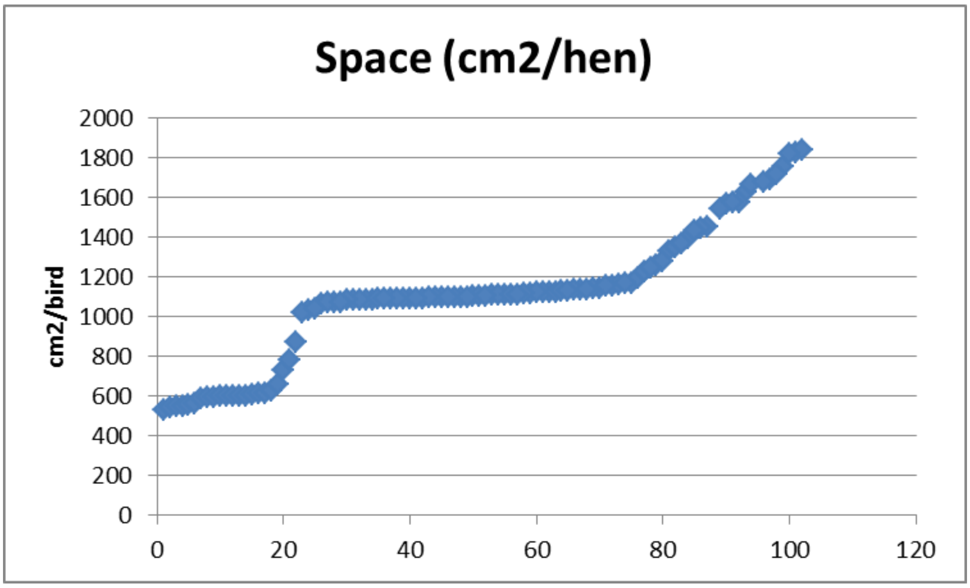
Biometris, especially Jac Thisse, who put much effort into the statistical analysis.

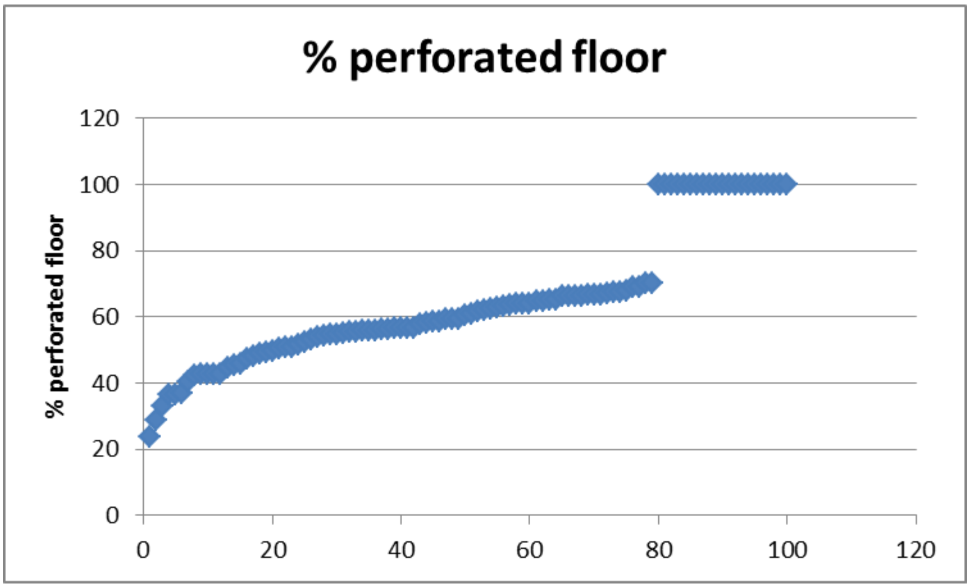
Appendix 1: Graphs showing variation between flocks for individual measurements.

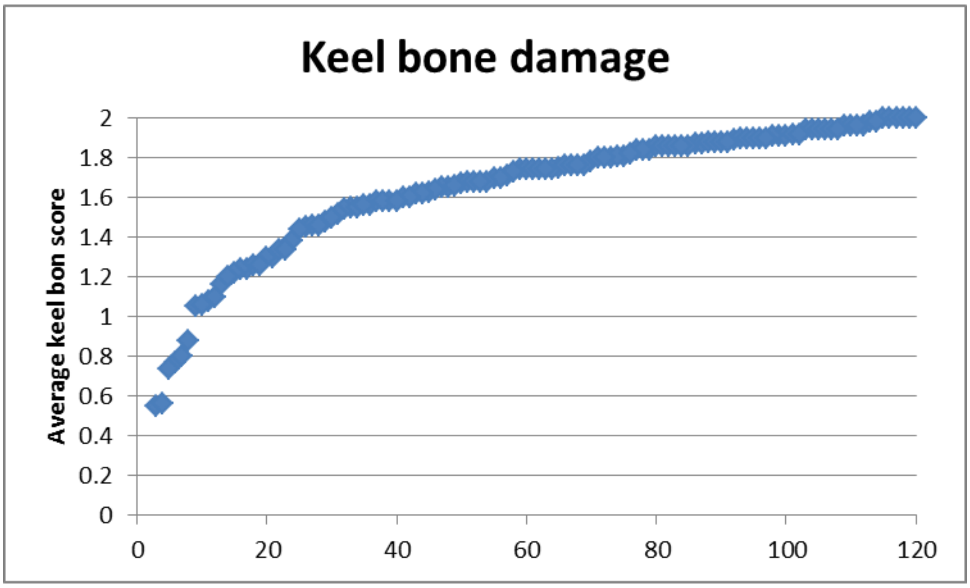


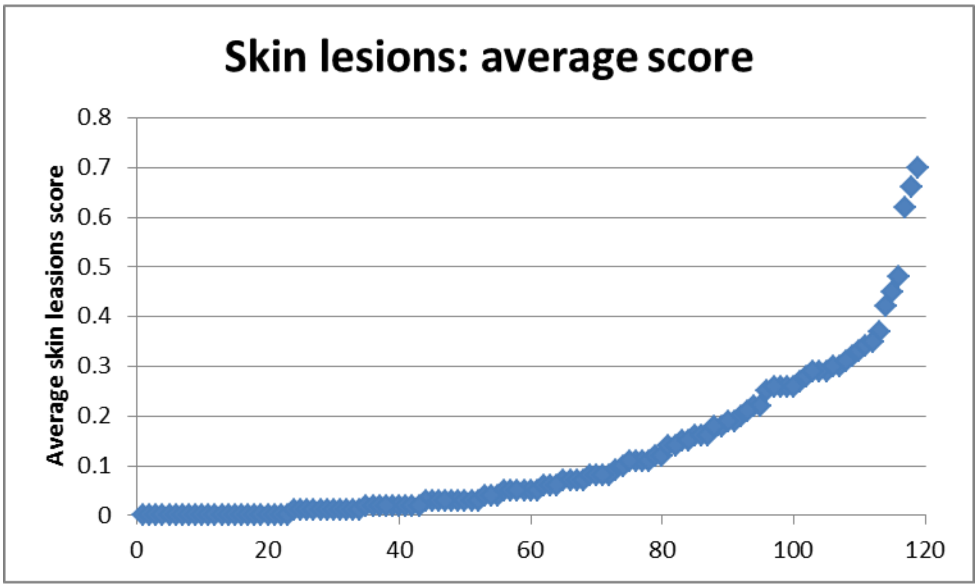


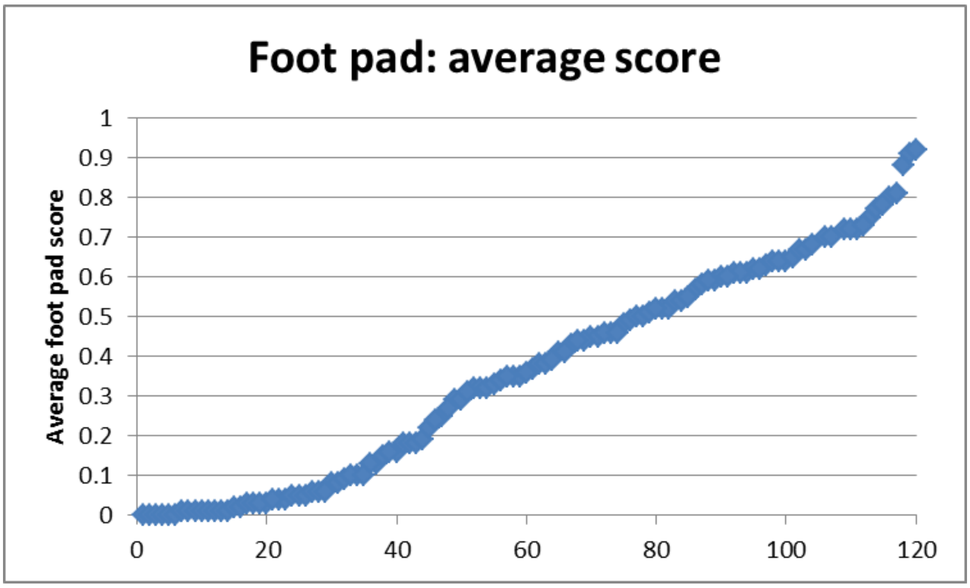

Perch length (cm) per bird
cm perch/bird
40
35
30
25
20
15
10
5
0
0
20
40
60
80
100
120

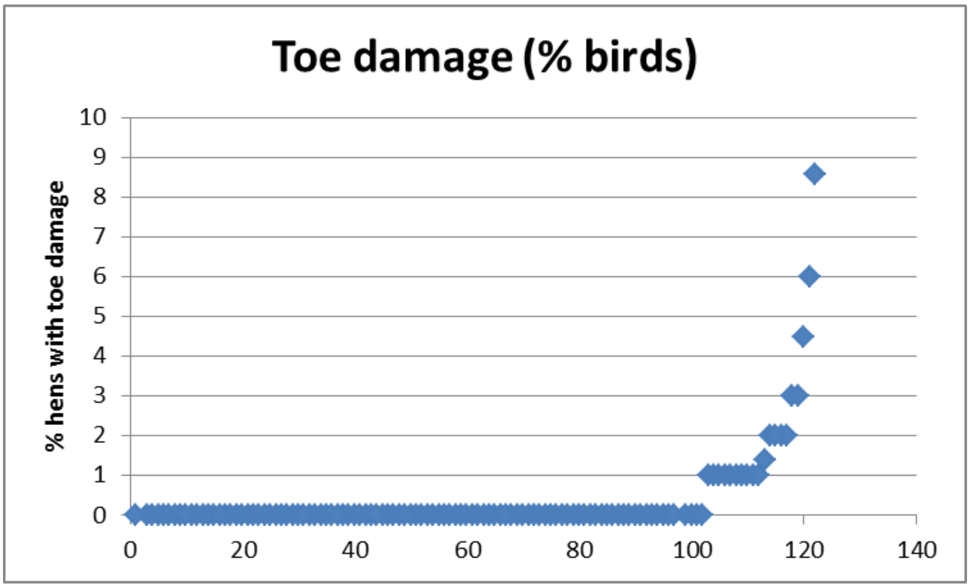

Space (cm2/hen)
cm2/bird
2000
1800
1600
1400
1200
1000
800
600
400
200
0
0
20
40
60
80
100
120

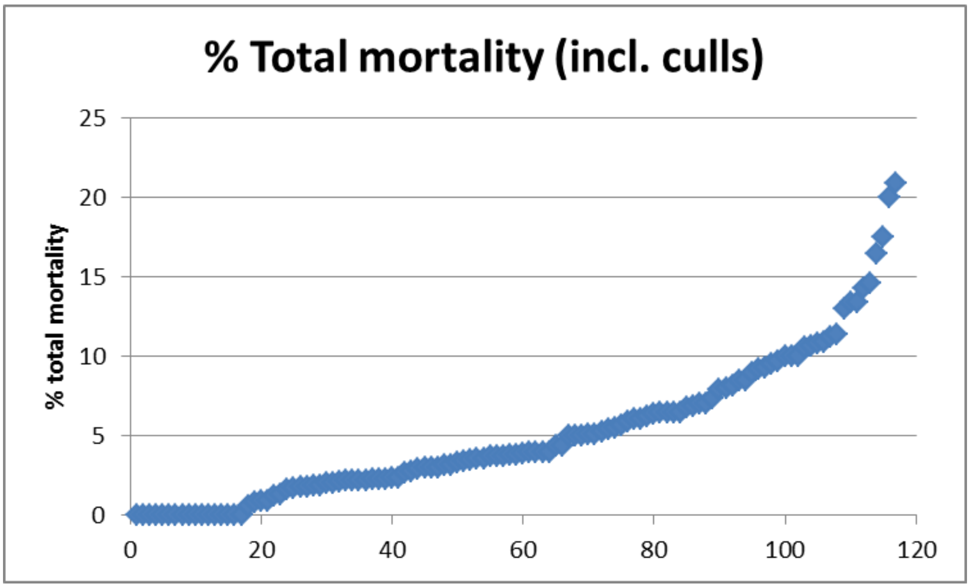

% perforated floor
% perforated floor
120
100
80
60
40
20
0
0
20
40
60
80
100
120

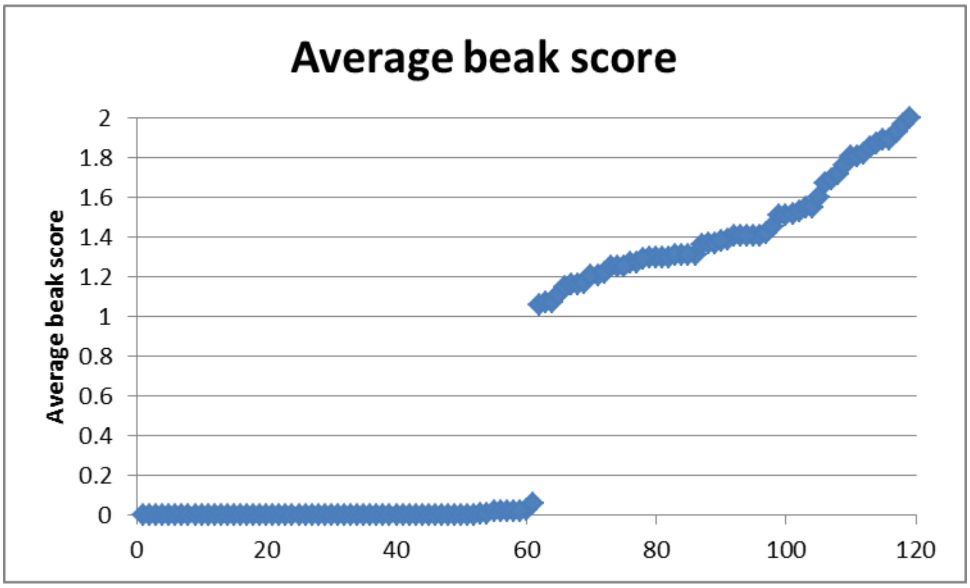

Keel bone damage
Average keel bon score
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
20
40
60
80
100
120

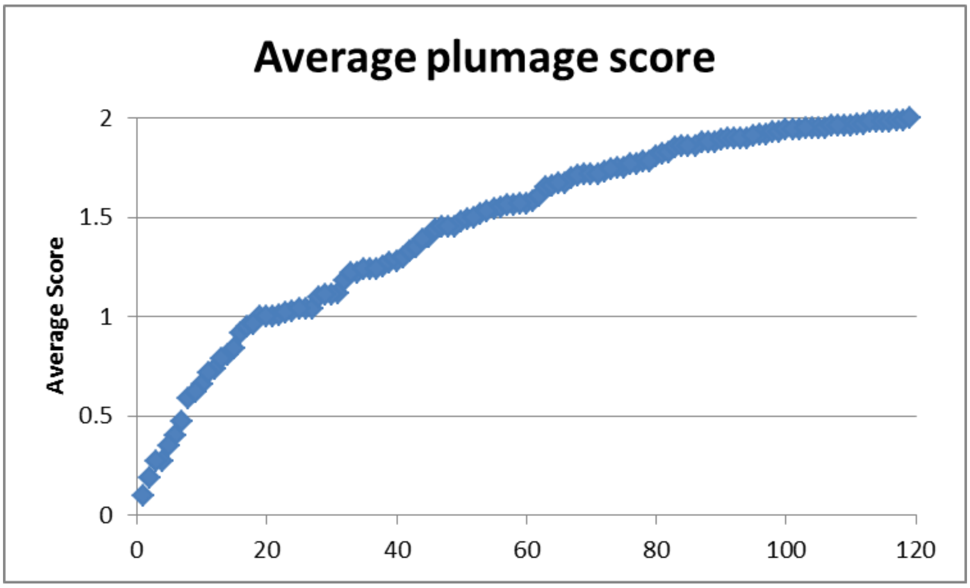

Skin lesions: average score
Average skin leasions score
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
20
40
60
80
100
120

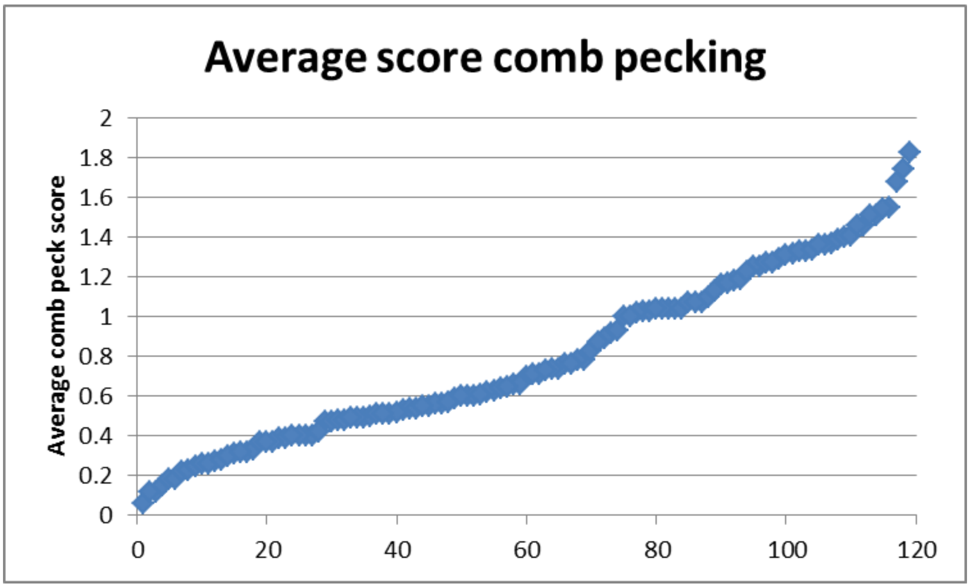

Foot pad: average score
Average foot pad score
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
20
40
60
80
100
120

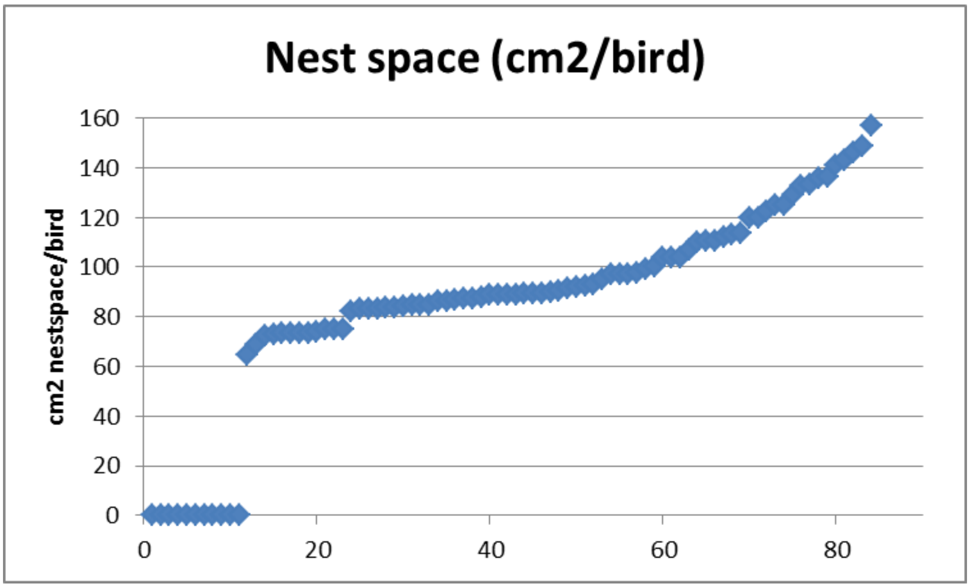

Toe damage (% birds)
% hens with toe damage
10
9
8
7
6
5
4
3
2
1
0
0
20
40
60
80
100
120
140

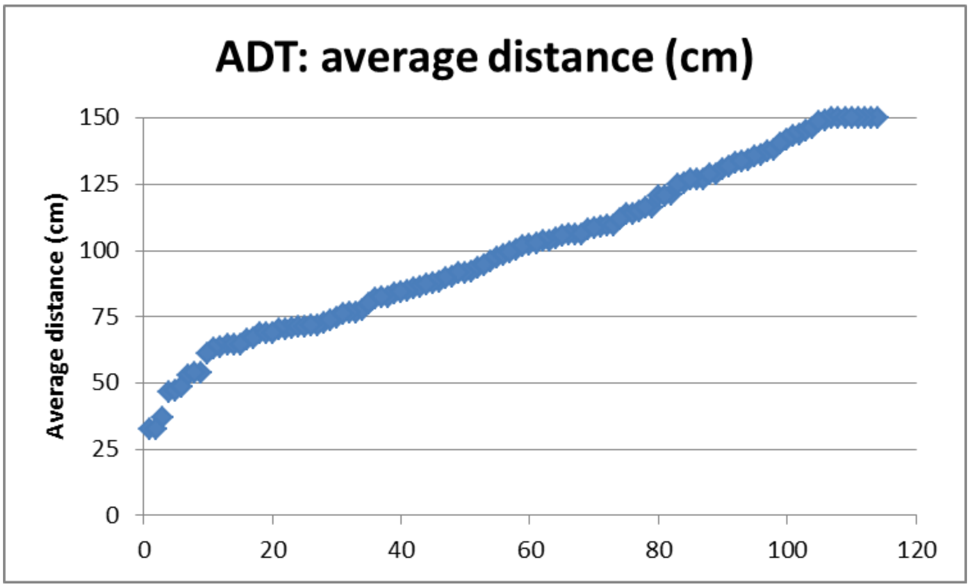

% Total mortality (incl. culls)
% total mortality
25
20
15
10
5
0
0
20
40
60
80
100
120

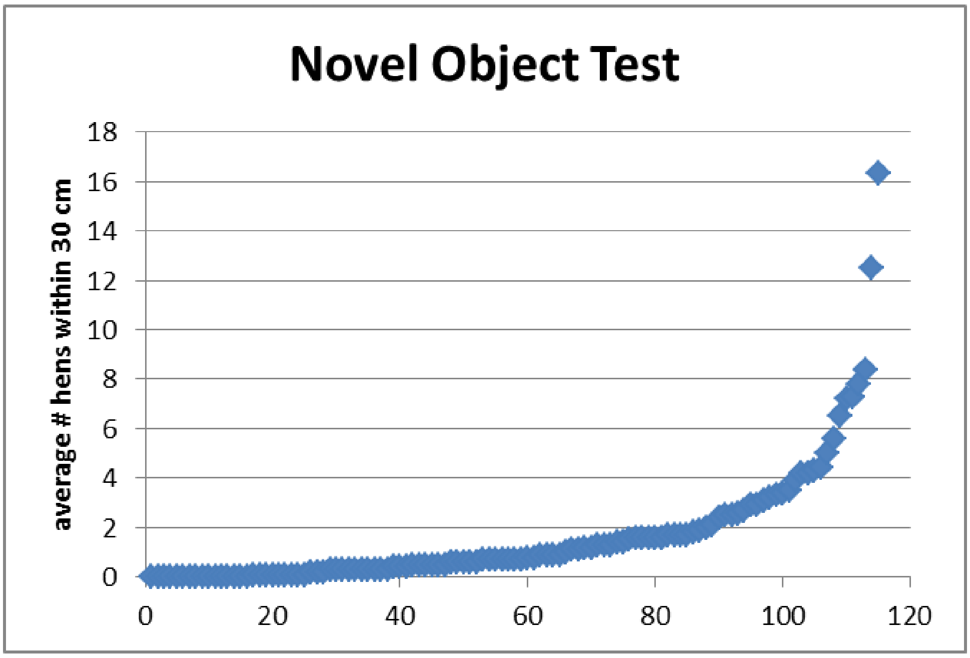

Average beak score
Average beak score
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
20
40
60
80
100
120

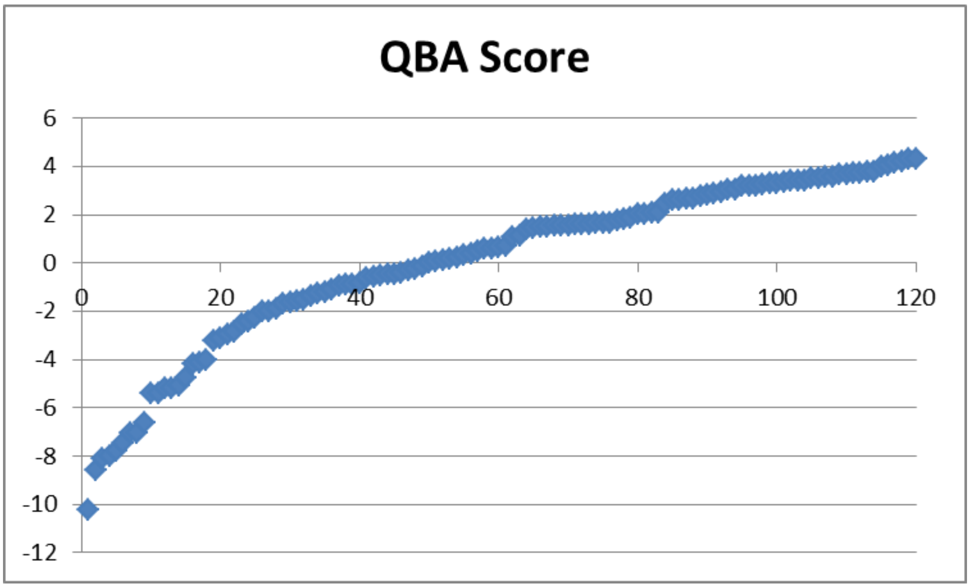

Average plumage score
Average Score
0
0.5
1
1.5
2
0
20
40
60
80
100
120


Average score comb pecking
Average comb peck score
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
0
20
40
60
80
100
120


Nest space (cm2/bird)
cm2 nestspace/bird
0
20
40
60
80
100
120
140
160
0
20
40
60
80


ADT: average distance (cm)
Average distance (cm)
150
125
100
75
50
25
0
0
20
40
60
80
100
120


Novel Object Test
average # hens within 30 cm
18
16
14
12
10
8
6
4
2
0
0
20
40
60
80
100
120


QBA Score
6
4
2
0
-2
-4
-6
-8
-10
-12
0
20
40
60
80
100
120



Wageningen UR Livestock Research

Edelhertweg 15, 8219 PH Lelystad T 0320 238238 F 0320 238050

E info.livestockresearch@wur.nl I www.livestockresearch.wur.nl