Horizon 2020 programme

Call: H2020-SFS-2018-2020 | Sustainable Food Security

Topic: SFS-08-2018-2019 | Improving animal welfare

Type of action: RIA

Sart date of the project: September 1st, 2019 Duration: 60 months

End date of the project: August 31st, 2024 Project ID: 816172

PPILOW

Poultry and Pig Low-input and Organic production systems' Welfare

D3.2 - Testing and evaluating animal welfare self-assessment & benchmarking tools

Deliverable number	D3.2
Deliverable title	Report on the efficiency of the welfare self-assessment tools
Due date of deliverable	31/3/2024 (originally 31/12/2023)
Actual submission date	29-03-2024
EC version	V1
Lead partner	EV ILVO
Author(s)	Evelien Graat, Charlotte Vanden Hole, Bas Rodenburg, Laura Warin, Frank Tuyttens
Dissemination level	CO

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°816172.

Table of contents

1. Summary

Objectives

This deliverable reports the results of PPILOW Task 3.2 (T3.2). The main aim of this task was to test the effect of regular animal welfare self-assessments by farmers with the welfare assessment apps PIGLOW and EBENE® developed in Task 3.1 (T3.1) on the welfare of pigs and poultry in commercial organic and low input production systems. Additional aims were to learn the opinions of farmers on the PIGLOW and EBENE® app and to compare animal welfare assessments performed by farmers and trained researchers using these apps.

Rationale

Free-range and organic farms come with specific animal welfare challenges. One way to give farmers more insight into the most important animal welfare problems and challenges on their farm and help them find solutions could be to have farmers perform welfare assessments of their own animals. In T3.1 of the PPILOW project, user friendly tools in the form of smartphone apps were developed that allow farmers to do this. These tools are the PIGLOW app for pigs and the EBENE® app for poultry. In developing these tools, care was taken to select welfare indicators that are: animal-based whenever possible, relevant for free-range and organic farms and suitable for on-farm assessment by farmers. Both apps contain benchmarking, in the hope that comparing their data with others will motivate farmers to improve aspects of welfare for which their scores are relatively low and make them feel good about aspects of welfare for which their scores are high. At the end of an assessment, the apps give automated feedback per welfare indicator in the form of risk factors and possible solutions that can help farmers identify the problem on their farm and find the right solution.

After the apps had been developed with input from stakeholders, the next step (T3.2) was to test whether regular use by farmers could lead to an improvement in animal welfare on-farm. This was tested with a 2-year longitudinal study in which free-range and organic fattening pig farmers and broiler chicken farmers participated. All participating farms were visited by a researcher (trained observer) at the beginning and at the end of the study. During these visits, the researcher and farmer performed the first and last animal welfare assessment with the PIGLOW or EBENE® app together to be able to compare scores of farmers and trained observers. Additionally, the researcher conducted a more detailed welfare assessment on each farm to be able to compare animal welfare at the beginning and at the end of the study. The farmers were asked to complete regular animal welfare assessments with the apps during the study. They were also asked to fill out a survey at the beginning and end of the study in which they gave their opinion on fattening pig or broiler chicken welfare. In the first survey the farmers additionally answered questions on their expectations of the app before using it and in the final survey they gave their opinion on the app after repeated use.

Results of the study did not support that the use of the PIGLOW and EBENE® app had a positive effect on the welfare of pigs and poultry on the participating farms. Farmers indicated that they found the apps easy to use and that they found most of the elements quite useful, but when asked about the specific influence of the app on their opinions on animal welfare and the welfare of their animals, they indicated that the apps only had small to medium-sized effects. It seems that the farmers mostly saw the potential of the apps for helping new farmers and other farmers are less knowledgeable about animal welfare. The results of the comparison of animal welfare assessments by farmers and researchers showed that there were some differences in how the 2 groups measured welfare that could be related to their different previous experiences. The level of agreement between farmers and researchers did not significantly change between the beginning and end of the study, thus a learning effect of farmers during the study was not supported by the data.

Teams involved

EV ILVO, ACTA (ITAVI) and UU were involved in the creation of protocols and execution of the farm visits, CRAW and BioForum were involved in the recruiting of participants for the study and INRAE was involved in general management of the task. EV-ILVO was leader of Task 3.2 and leader of WP3.

2. Introduction

Towards higher organic production in the EU

In 2021, an action plan on organic farming was presented as part of the European Union's (EU) 'Farm to Fork strategy' (European Green Deal). This plan outlines a set of actions to increase organic farming and have organic production reach at least 25% of the EU's agricultural land use by 2030 (European Commission, 2021). In 2019, organic production accounted for 8.5% of the EU's total 'utilized agricultural area' or 13.8 million hectares. In comparison, in 2009 organic farming only used 8.3 hectares, so over 10 years there was an increase of no less than 66%. However, the increase up to 25% land use will require some extra effort, more for some member states than others, as the extent of organic farming greatly differs between countries (ranging from 0.5% to more than 25%) (European Commission, 2021).

One of the axes on which the action plan is built is 'Organics leading by example: improving the contribution of organic farming to sustainability', in which action 21 on enhancing animal welfare is put forth. To quote: *"In the context of the Animal Welfare Platform, the Commission will: continue working with Member States and civil society to find concrete and operational ways to further improve animal welfare in organic production."* They explain that actions need to be taken to further improve animal welfare, mobilizing all instruments available to better respond to citizens' expectations and demands. They also highlight the importance of organic farming as a model in terms of animal welfare protection (European Commission, 2021).

Though the action plan mentions no specific targets per sector, one can assume that all are likely to increase (though not equally) in size and importance over the next few years. As such, we can safely assume that organic animal production, among which broiler chickens and pigs, will also increase. Organic animal production has already been experiencing a fast growth the last 2 decades, but in comparison to total animal production in the EU, it still remains rather small. In 2020, only 3.6% of poultry and only 1% of pigs was raised organically (European Commission, 2023). More specifically for broiler chickens in the EU, 90% are raised in intensive indoor systems, up to 5% in free-range systems and only 1% in organic systems (European Parliamentary Research Service, 2019). Free-range pig farms are not uncommon in the EU, although the percentages vary across member states, representing a median of 8% (interquartile range of 1.5–24%) of all pig farms in the 17 member states for which data are available (Nielsen et al., 2021).

Free-range vs. organic

Free-range rearing and organic rearing are not necessarily the same. Having an outdoor area is a prerequisite for both types of rearing, but in organic production, other compulsory measures apply, such as organic feed, no hormones, no excessive antibiotics or unnecessary mutilations, and increased space allowance. Farms that follow all regulations for organic production can have an organic production label, whereas farms that allow the animals to go outdoor without other measures are called free-range (Commission of the European Communities, 2008; Council of the European Union, 2007).

There are 3 main systems of free-range or organic **pig** farms in Europe: indoor, partly outdoor and outdoor (Früh et al., 2014). These names are slightly misleading as – according to the legislation on organic animal production – all animals in these systems have access to an outdoor area. The difference between these systems is in the type of outdoor area as well as the timing/duration of outdoor access. The 'indoor' system mainly constitutes indoor rearing, but the pigs have access to a limited outside run, often made of concrete or sometimes soil. In the 'outdoor' systems, on the

other hand, the animals are reared outside all year long, often on pasture, and have access to a shelter (e.g. huts, sometimes more permanent buildings). The 'partly outdoor' system is an intermediate form that combines indoor and outdoor access during different production stages or seasons. This can mean that at least one production stage (dry sows, lactating sows, group suckling, weaned piglets, or finishing pigs) lives outside, while the others are housed indoors, or that pigs spend certain seasons of the year outside and others inside.

In the case of **broiler chickens**, or even poultry in general, to the best of our knowledge, no clear categories have been described. However, similar to pigs, the outdoor area for poultry varies greatly, from a few square meters of concrete or dirt with no vegetation, to pasture with a lot of grass and/or other vegetation or even a forest-like environment. It is, however, important to note that even in the best-case scenario, the slower growing broiler breeds used in a free-range system only live for 2 to 3 months, and only have access to the outdoors for about half this time (Dawkins et al., 2003).

Animal welfare

Considering the large variation among and even within free-range rearing systems, it must be kept in mind that certain advantages or disadvantages of free-range/organic rearing (from here on out called 'outdoor production') might be more or less pronounced depending on the exact system (Früh et al., 2014; Leeb et al., 2019). The following paragraphs on animal welfare for both pigs and broiler chickens are a generalization and often represent the more extensive outdoor conditions. It must also be noted that variation with regard to e.g. breed, management, region or even national differences in the interpretation of regulations can affect animal welfare in any of the systems. We will discuss animal welfare in outdoor systems for pigs and broiler chickens separately. It is in no way our goal to provide an entire review of pig and broiler welfare in an outdoor setting, as this would be beyond the scope of this deliverable. For reviews on this topic, we refer to studies such as Åkerfeldt et al. (2021); Edwards et al. (2014); Kijlstra and Eijck (2006); Lund and Algers (2003) and Sutherland et al. (2013).

Pigs

A higher space allowance modifies the pig's behaviour in a positive way (Cornale et al., 2015). Especially pigs that have access to pasture increase their activity and show a wider range of behaviours, e.g. species-specific behaviour such as rooting or foraging (Høøk Presto et al., 2008; Jensen et al., 2010; Terlouw et al., 2009). Access to additional roughage (such as grass, clover, chicory, etc.) also has positive effects on pigs by leading to higher activity levels as well as more time spent foraging or rooting. Roughage also occupies the pigs for a longer time, resulting in less aggressive behaviour, fewer wounds and less tail biting (Holinger et al., 2018a; Høøk Presto et al., 2009; Jensen et al., 2010; Presto Åkerfeldt et al., 2019; Presto et al., 2013). Though tail biting may still be seen in outdoor production systems, it is less frequently observed in outdoor compared to indoor production systems. Aside from access to roughage, the lower stocking density, the ability to perform species-specific behaviour and general environmental enrichment (more diverse materials for exploration or investigation) are presumed to have a positive effect on tail biting (Leeb et al., 2019; Walker & Bilkei, 2006).

Similar to tail biting, lameness is a problem in all types of pig production systems and is an important welfare issue, as it is painful, restricts access to resources and often leads to premature culling in sows (Nalon et al., 2013). It is, however, less prevalent outdoors compared to conventional herds (Dippel et al., 2014; Knage-Rasmussen et al., 2014), with the best results being visible in systems with year-round or a lot of outdoor access (Leeb et al., 2019). Also in organic fattening pigs, fewer claw injuries and lower/less severe prevalence of bursitis were noted compared to conventional

fattening pigs (Gareis et al., 2016). Less lameness and fewer claw injuries can be explained by a softer underground, less exposure to manure and increased activity (Day et al., 2003; Knage-Rasmussen et al., 2014). However, with regard to leg health, outdoor rearing also poses some challenges, mainly an increased risk of arthritis (Alban et al., 2015). The likely reasons are higher infection risks, poorer hygiene leading to a higher general infection pressure, or even more mechanical stress and joint injuries predisposing the joint to arthritis. The latter seems somewhat contradictory, given the general positive effects of a softer underground on leg health. Indeed, Kongsted and Sørensen (2017) found that the increased risk of arthritis did not actually cause a higher level of lameness. The pigs may be clinically less affected as exercise may help strengthen the joint supportive tissues and lower pain levels (Etterlin et al., 2015). Osteochondrosis may also be a larger health issue in free-range pig production than previously assumed. As mentioned, exercise may help with strengthening the joints, but a large outdoor area may also be challenging. Given the complexity and multifactorial nature of leg health in general, prevention and monitoring are key (Etterlin et al., 2015; Etterlin et al., 2014; Wallenbeck et al., 2020).

In addition to leg problems, respiratory diseases and diarrhoea are common in any type of production system. However, a lower incidence of respiratory problems has been found in year-round outdoor systems compared to other free-range or conventional systems. This may be explained by higher levels of dust (caused by e.g. dry feed, straw bedding) in indoor conditions (Kijlstra & Eijck, 2006). In the organic system, pigs are weaned later (minimum of 40 days), which is a strategy to reduce the risk of post-weaning diarrhoea. However, diarrhoea and poor body condition of suckling, weaning and fattening pigs have been reported in outdoor systems (Früh et al., 2014; Leeb et al., 2014; Papatsiros, 2011; Sundrum et al., 2010). On the other hand, it has also been shown that in organic year-round outdoor systems, as compared with other outdoor systems, the frequency of diarrhoea is lower (Leeb et al., 2019), and the access to roughage, such as silage, has a positive effect on pigs' gastric health (Holinger et al., 2018b). However, Leeb et al. (2019) also state that the reliability of outdoor assessment of diarrhoea is possibly low, as it is easier to miss the signs in an outdoor setting (see also general remark on 'difficulty monitoring in outdoor systems').

Piglet mortality is also an important issue in outdoor systems, mainly due to the crushing of piglets and inadequate nursing of group-housed piglets (Lindgren et al., 2013; Westin et al., 2015). It has also been stated that the occurrence of stillbirths is higher outdoors, compared to conventional farms (Lindgren et al., 2013). For both stillbirths and piglet mortality, the identified risk factors are large litters, a higher parity number and the farrowing system (loose housing and group housing with free farrowing) (Grimberg-Henrici et al., 2019; Hales et al., 2014; Hales et al., 2015; Rangstrup-Christensen et al., 2017; Rangstrup-Christensen et al., 2018a; Rangstrup-Christensen et al., 2018b). Compared to conventional farms, organic sows show lower numbers in terms of replacement rate, though the number remains fairly high, with an annual replacement rate of around 30%. However, metritis, mastitis, agalactia (MMA) remain a problem, even in an outdoor setting (Früh et al., 2014; Sundrum et al., 2010).

Obviously, being outside also comes with some challenges, such as parasites, predators and adverse weather conditions (Delsart et al., 2020; Pietrosemoli & Tang, 2020). Both endo and ectoparasites are repeatedly reported to be a challenge for outdoor pig production (Alban et al., 2015; Baumgartner et al., 2003; Day et al., 2003; Früh et al., 2014; Katakam et al., 2016; Kongsted & Sørensen, 2017; Lindgren et al., 2014). In outdoor pig production, one of the most common endo-parasitic infections is *Ascaris suum*, both before and after weaning (Baumgartner et al., 2003; Nansen & Roepstorff, 1999). The exact risk of predation has rarely been quantified for outdoor pig production, but it is clear that piglets are most at risk (Fleming et al., 2016). The type(s) of predators are of course region dependent; some examples are foxes, wolves or birds of prey. Weather conditions that can cause stress for outdoor pigs include precipitation, temperature, solar radiation and wind. These can lead to changes in behaviour (e.g. restlessness or aggressiveness), but can also affect health (e.g. injuries, slower growth rate, lower reproductive performance). The impact of

weather conditions will depend on the exact system, but also the breed, age, weight, reproductive stage or general health of the pigs (Luković et al., 2017; Parrini et al., 2018; Pietrosemoli & Tang, 2020).

A risk that is very specific to outdoor pig production is African Swine Fever, due to the contact with wild boars. In outdoor production, both direct and indirect contact between wild boars and domestic pigs are difficult to control. A number of biosecurity measures have been identified for outdoor farms specifically to mitigate the risk of introducing or spreading African Swine Fever, though a certain risk always remains (especially in certain areas or seasons) (Nielsen et al., 2021).

Broilers

Similar to pigs, one of the main advantages of access to an outdoor area is more space for the birds, leading to a lower density. However, in broilers (or poultry in general) there generally is a lot of individual variation in the use of this outdoor range (Dawkins et al., 2003; European Commission, 2000; Weeks et al., 1994). Many of the birds tend to stay close to the house, and some of them do not even leave the house. This suggests that many outdoor systems are not providing a range that is suitable or attractive for poultry. Birds that do find the outdoor area attractive enough get more exercise, which can lead to better leg bone development and fewer gait problems (Aguado et al., 2015; Leterrier et al., 2008; Stadig et al., 2017).

The main health issues in organic broiler production are footpad dermatitis (FPD), as well as hock and breast lesions (Van De Weerd et al., 2009). Though FPD is a problem, it is still less common in organic broiler farms compared to conventional farms (Lund et al., 2017). In general, the prevalence and severity of lameness and other leg problems in slow-growing broilers in organic farms is lower than in broilers in conventional farms (Bergmann et al., 2016; Fanatico et al., 2005; Tahamtani et al., 2018; Tuyttens et al., 2008). The same trend was found in a study by Wilhelmsson et al. (2019), who also found mortality rate, contact dermatitis and plumage cleanliness to be better in organic production. According to Van De Weerd et al. (2009), most welfare problems among organic broiler chickens have a clear connection to the suitability of the breed, the banning of synthetic amino acids in organic production (which can lead to nutritional challenges and metabolic disorders), as well as the use of the outdoor area (see previous paragraph). Improved and well-balanced diets can be useful in avoiding nutrient deficiencies, as well as preventing injurious behaviours (Åkerfeldt et al., 2021). Especially fast-growing broiler breeds tend to do poorly in organic production, with higher incidences of mortality due to lameness and circulatory problems. This is linked to the fact that fast growing breeds are not suited for the longer rearing periods in organic production (Rezaei et al., 2018; Wallenbeck et al., 2016).

Similar to pigs, the outdoor environment poses some specific challenges such as parasites, predators and adverse weather conditions (Newberry, 2017). The most common parasites in broiler chickens are *Eimeria* spp. In organic production, paraphyletic treatments (coccidiostats) are not allowed, but vaccination as well as a decreased stocking density or a restricted group size can reduce the contamination risk (Åkerfeldt et al., 2021). Contrary to pigs, where mainly piglets might suffer predation, in broiler chickens this remains a risk during the entire production cycle (Sutherland et al., 2013). Though the type of predator will vary from region to region, some well-known examples are foxes, martens, and birds of prey (Stahl et al., 2002). Some weather conditions, such as rainfall, high solar radiation and wind strength have been related to fewer animals outside (Dawkins et al., 2003; Stadig et al., 2017).

A disease that has been more prevalent in the last decade(s) is Avian Influenza (AI). AI viruses are primarily introduced to domestic poultry through (in)direct contact with infected (wild) birds. Exposure can happen through contaminated organic material and movement of infected poultry, but even through movement or usage of contaminated equipment or vehicles. When the risk of AI

transmission is high (often from late autumn until early spring), strict biosecurity measures at farm level need to be implemented to prevent AI infections (Capua & Marangon, 2006). When the risk of transmission is too high, some countries will enforce a period of mandatory confinement for all poultry, which means the animals should be kept inside or under protective nets. Considering free-range broilers can roam outside, they are generally more at risk than those at a conventional farm. In addition, adapting to confinement will likely be more difficult or stressful for individual broilers that have already spent some weeks outdoors or breeds that are especially suitable for free-range production.

An important note when monitoring health and welfare in a free-range setting for any species is that there exists an added difficulty of correctly assessing the health status of animals if they have a lot of outdoor space. For example, Leeb et al. (2019) state that they observed lower treatment incidences of MMA in outdoor systems compared to indoors, or that the occurrence of diarrhoea was lower in outdoor systems. According to the authors, the lower treatment incidences might also be explained by farmers having difficulty observing MMA in an outdoor setting and therefore treating fewer animals. Similarly, a lower occurrence of diarrhoea can be explained by the fact that signs are easily missed in an outdoor environment (Leeb et al., 2019). Another example is the crushing of piglets, which is more difficult to monitor in a free-range setting and for which intervening is less safe when a sow can move freely (Kilbride et al., 2012).

All in all, it is clear that outdoor production systems have a lot of advantages with regard to e.g. positive welfare or naturalness (though there is still some room for optimization), but also comes with some challenges, particularly regarding animal health and monitoring (Åkerfeldt et al., 2021; Leeb et al., 2019).

Monitoring and improving animal welfare on-farm

Because of the remaining challenges for (the monitoring of) animal welfare in outdoor production systems, it is useful to look into the literature to see how people have tried to face those challenges in the past and which ways seem to be most effective for improving animal welfare. One approach that has been taken rather successfully is for researchers or other experts to monitor the welfare of animals on-farm and share their data with the farmers, sometimes with farm specific advice. Burke (2004) assessed several welfare parameters on 15 dairy farms in the UK and shared the results with the farmers and their veterinarians during a workshop. In response, 10 of the 15 farmers made animal welfare related changes on their farm. In a follow-up study in which qualitative interviews were held, the farmers indicated that they were very interested in hearing an outside opinion on animal welfare and that the possibility to compare results with other farms inspired discussion topics and motivated them to solve welfare issues on their farm (Burke, 2006). Atkinson et al. (2017) collected extensive data on animal-based welfare indicators on 18 dairy farms in Canada and subsequently shared their results with the farmers and their veterinarians during a face-to-face meeting. All farmers also received a benchmarking report comparing the data of their farm to the other farms. 83% of the farmers made at least 1 change on their farm because of this report and a second round of welfare observations by the researchers showed that the welfare outcomes related to those changes significantly improved. These improvements were not seen on the farms where no changes were made. Qualitative interviews with the farmers participating in this study were reported by Sumner et al. (2018) and showed that receiving the benchmarking reports motivated some of the farmers to seek out peers in their community so that they could regularly compare their farms going forward. Both studies thus seem to show a positive effect of benchmarking combined with contact with animal welfare researchers. However, Atkinson et al. (2017) also noted that in some cases it is difficult to know whether the benchmarking data or the presence of the researchers was the most influential factor. A study by van Dijk et al. (2018) which asked dairy cow and laying hen farmers for their opinions on different elements of welfare assurance schemes in the UK found similar results.

Several farmers indicated that they found it valuable and sometimes reassuring to be able to compare data of their farm to a national database. Additionally, following an exercise of joint welfare scoring by farmers and assessors of the welfare assurance schemes, many farmers found it useful to be able to discuss welfare problems and possible solutions with the assessors, even if the assessors were only allowed to give general rather than farm-specific advice. That farm-specific advice from experts can be beneficial was shown in a study on 60 dairy farms in The Netherlands (van Eerdenburg et al., 2021). Researchers conducted welfare assessments of animal-based indicators on 60 farms and subsequently gave personalised advice to all farms and organised sessions for the farmers and their veterinarians in which they could discuss the results and advice together. Many farmers made welfare related changes afterwards and additional welfare assessments showed that welfare improved more on farms that made more changes.

The studies described above show clear advantages of experts measuring animal welfare on-farm. However, it also comes with several disadvantages. One that immediately comes to mind – and is also mentioned by several authors of these studies – is that it requires a very large time investment of the experts to visit a considerable number of farms and perform extensive welfare assessments. Because of this, unfortunately only a relatively small group of farms compared to the total number of existing farms can be included in such initiatives. Another disadvantage is that not all farmers are eager to have researchers or other animal welfare professionals come to their farm to collect data. For instance, many farmers in the study by van Dijk et al. (2018) said that they were afraid that such data would be used against them and that if it was badly presented to the outside world, it could negatively impact the image of the sector. 44% of 525 surveyed dairy cow and laying hen farmers were against the establishment of a national dataset, partially for this reason. The same survey does perhaps suggest an alternative that is preferable to those farmers, as 86% of 536 farmers responded that they think they and their staff are the ones best equipped to monitor animal welfare on their own farm. Thus, for this large group of farmers it would be preferable if they were more actively engaged in the assessment of their animals' welfare.

Animal welfare monitoring by farmers

Many studies and initiatives have suggested or shown the possible beneficial effects of engaging farmers more directly in the monitoring of animal welfare and finding solutions to problems. A first step in involving farmers more directly could be the concept of "stable schools". In stable schools, groups of farmers are brought together to use farm-specific data to identify welfare issues, discuss possible solutions and set goals to improve animal welfare. While facilitators are sometimes involved in the discussions, there are also initiatives in which farmers work by themselves. A review of the effectiveness of stable schools for dairy farmers in 7 European countries showed that overall, the majority of farms involved showed improvements of welfare aspects discussed during the meetings (Ivemeyer et al., 2015). March et al. (2014) mention that the 19 German dairy farmers involved in their study on a modified stable school concept, in which farmers visited each other's farms to give advice, found the self-determined approach of this initiative very motivating, which was also apparent from the fact that farmers implemented over two thirds of the suggestions made by other farmers.

Besides stimulating active involvement of farmers in finding solutions to welfare problems, one can also go a step further and stimulate farmers to perform structured welfare assessments of their animals. For instance, since 2014, Article 11 Section 8 of the German animal welfare act states that farmers must collect information on animal-based welfare indicators on their own farm (German Animal Welfare Act (Tierschutzgesetz), 2006). The goal of this is to increase the farmers' awareness of weaknesses on their farm and stimulate them to improve in order to meet all requirements of Germany's animal welfare laws. Michaelis et al. (2022) also state that if farmers were to perform a systematic welfare assessment themselves, it could increase their awareness of the welfare of their animals and motivate them to find solutions to problems. In response to the change in the German

animal welfare act, experts from several institutes collaborated to design practical guides for welfare assessments of cattle, pigs and poultry that could be used be farmers (Zapf et al., 2015). Those guides were then tested on 44 cattle farms, 34 pig farms and 43 poultry farms and interviews were conducted to assess farmers' opinions (Schultheiß et al., 2023). The majority of interviewees said that their views on their animals had changed as a consequence of conducting the welfare assessments. Farmers rated the ease of integrating the welfare assessments into their normal work schedule as good, with many of the cattle and laying hen farmers saying that they would likely continue to perform the welfare assessments. For turkey and fattening pig farmers, this number was lower. Farmers indicated that target and alarm values for welfare indicators that were set by experts to serve as a means of benchmarking helped them identify weaknesses and motivated them. In the study by van Dijk et al. (2018), some of the farmers spoke positively about formal welfare self-assessments, saying that they kept them "on their toes" and made them go "a stage further". 23% of the farmers who had conducted self-assessments indicated that they had made changes on their farm as a consequence. On the other hand, some other farmers said that they viewed the suggestion that a structured welfare assessment would be necessary as an insult to what they already did on their farm. Thus, while these studies show that not all farmers are equally enthusiastic about conducting structured welfare self-assessments, they also indicate that for a large group of farmers this could be an effective method of improving animal welfare.

Very few studies have looked into whether the results of such welfare self-assessment by farmers are reliable and comparable with assessments by experts. One study with 146 German farmers of dairy cows, beef cattle, pigs and poultry found that interobserver reliability between farmers and trainers was quite good (Michaelis et al., 2022). In contrast, Katzenberger et al. (2020) found a very low interobserver reliability for experts and dairy farmers. However, in that study experts and farmers assessed the same animals at different moments with an average of 70 days between both assessments on the same farm, which allows for the possibility that the welfare status of the cows was actually different. Gibbons et al. (2012) did not include farmers in their study, but instead compared welfare scoring by experts and trainees without experience. One of their conclusions that could also be relevant for farmers is that agreement between trainees and experts was higher when less detailed and complex scoring methods were used. They argue that when the goal is to learn about animal welfare on farm level rather than individual level, the information that would be gained by a more detailed scoring method is not relevant and therefore, choosing higher accuracy rather than higher precision should be preferred. Thus, if our goal is to increase awareness of welfare issues and risk factors among farmers, a less detailed scoring method is likely more suitable for that purpose.

From the studies that are discussed in this chapter, we can also learn which elements related to structured welfare assessments seem to be valued the most by farmers and are most likely to motivate and help them to improve animal welfare. It is noticeable that almost all of these studies measured animal-based welfare indicators. van Dijk et al. (2018) showed that farmers also see the benefits of this, with 64% of 535 survey respondents agreeing that an animal-based assessment is more meaningful and robust than one concentrating on inputs and records. Something else that many of the studies discussed in this chapter show is that benchmarking was highly valued by farmers. In addition, in a study on animal welfare self-assessments in which benchmarking was not included, many farmers mentioned that they were unsure how to evaluate their results and know whether they were good or bad (Pfeifer et al., 2020), thus essentially asking for something similar to benchmarking. It seems that benchmarking is important for farmers to help identify problem areas. Farmers also indicated that they value getting feedback from experts. While live discussions with experts are not always possible because they are time consuming or because experts such as auditors visiting farms are not allowed to give farm-specific feedback (van Dijk et al., 2018), an alternative could be to provide farmers with automated feedback for specific welfare indicators. Although this might not be as useful as receiving farm-specific feedback, it would be much less time

consuming for experts and could help a large group of farmers identify risk factors and possible solutions for welfare problems. Finally, in studies on the German practical guide for animal welfare assessments, farmers also asked for a possibility to enter data digitally that would allow for automatic analyses (Pfeifer et al., 2020). The Excel application that was subsequently used by (Schultheiß et al., 2023) only received moderately positive scores. Thus, a more user-friendly digital tool to conduct animal welfare assessments would most likely be appreciated by farmers. Finally, we should move back to our specific goal of helping organic and free-range farmers. As was detailed in earlier parts of this introduction, certain welfare challenges or risk factors are different in/specific to outdoor production systems. Because of that, it would also be useful to have a tool of which the protocol and the included welfare indicators have been specifically tailored to outdoor production systems.

The PIGLOW and EBENE® apps for welfare self-assessment

With this knowledge in mind, the PIGLOW and EBENE® apps for animal welfare self-assessments were developed and/or updated in T3.1 of this project, with dedicated welfare assessments for outdoor pigs and poultry. Both apps can be downloaded in the Google Play Store or App Store. The PIGLOW app is available in 9 languages and the EBENE® app is available in 7 languages. Since their development, information about the apps has been disseminated through popular as well as scientific channels. The apps have also been used by PPILOW project partners to collect animal welfare data for other work packages. In this introduction, we will give a short description of both apps that will help the reader understand the results detailed later in this report. For a more extensive description, we refer to Deliverable 3.1, which details the development of both apps.

The PIGLOW app was newly developed within the PPILOW project and only contains welfare assessments for outdoor grower pigs, finishing pigs, pregnant sows and farrowing sows. The welfare assessments mainly contain questions about animal-based indicators that must be answered while observing the animals. After completing a welfare assessment, users receive an email with a link to a report with results that can be found on the PIGLOW website. In this report, the welfare scores for the different indicators are categorised under the 4 welfare principles from the Welfare Quality® protocol (Good Housing, Good Feeding, Good Health and Appropriate Behaviour) (Blokhuis et al., 2010). For each indicator, the app gives automated feedback in the form of risk factors for associated welfare problems. Benchmarking is available in the form of percentiles reflecting the percentage of other users with a worse score or the same score as the user. Thus, a percentile of 90 would indicate that the user scored among the best 10% of users. The report also includes a “welfare radar”, which is a radar-shaped graph in which the scores for 7 welfare indicators are highlighted (figure 1). These 7 indicators differ between the assessments for different types of pigs and reflect aspects of welfare that were thought to be especially important by the experts who developed the app. Finally, these same 7 indicators are shown in an “evolution graph”, which shows how the scores have changed compared to previous welfare assessments conducted by the same user (figure 2). Users can find the reports of all their previous welfare assessments in their account, which makes it possible to also compare their scores for the welfare indicators that are not shown in the evolution graph.

The EBENE® app already existed for conventional poultry and rabbit farms, but within the PPILOW project it was expanded with welfare assessments specifically for outdoor broiler chickens and laying hens. The welfare assessment for broiler chickens, which is most relevant for our study, contains behavioural and sanitary assessments that focus mainly on animal-based welfare indicators. The results of the assessment are shown inside the app itself and similar to PIGLOW, the scores for welfare indicators are divided among the 4 welfare principles from the Welfare Quality® protocol. For each welfare indicator, automated feedback in the form of risk factors for associated welfare problems is provided. In addition to average scores per welfare indicator, the EBENE® app also provides composite scores between 1 (worst) and 5 (best) for aspects of welfare in which several

indicators that relate to the same theme have been combined. These scores are also shown in the "spider map". This is a graph shaped like a spider web in which the user's own scores are shown as well as the comparison to scores of other users (benchmarking) (figure 3). The blue line represents the user's own scores, where the score is higher if the line is closer to the edge of the graph. The different colours of the backgrounds of the graph indicate the height of scores of other users.

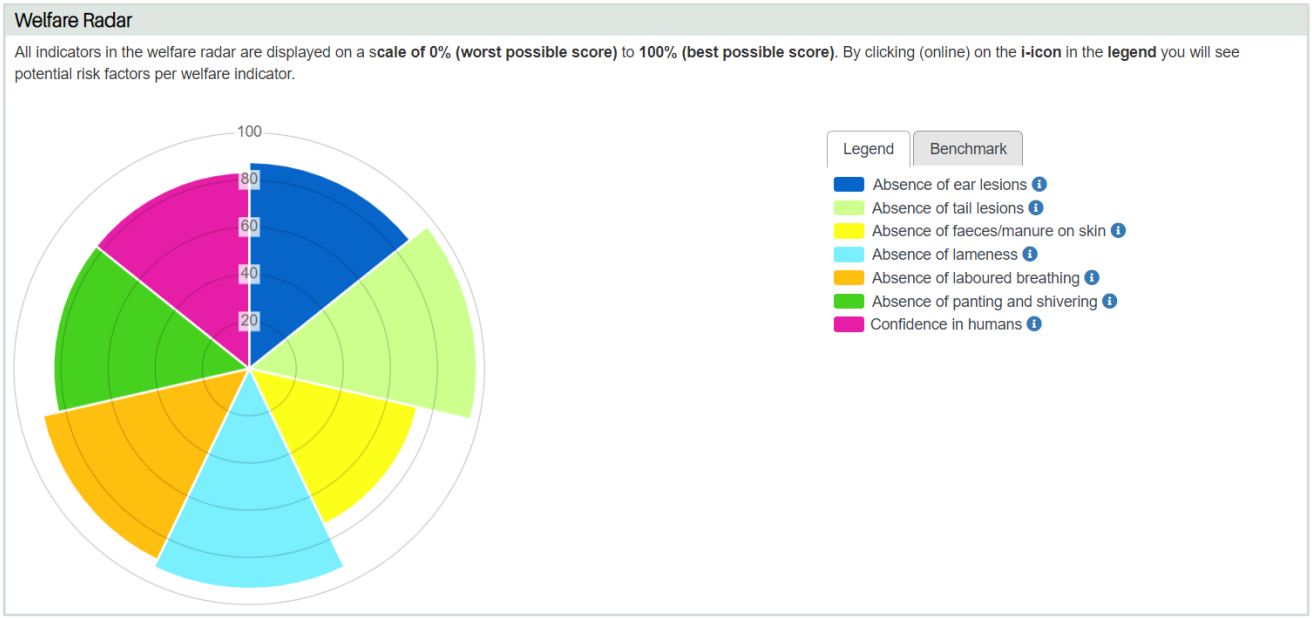

Figure 1: Welfare radar from the results for finishing pigs in the PIGLOW app, which shows the scores for 7 welfare indicators which were deemed to be very important by the developers are highlighted. The larger a segment, the better the score for the associated welfare indicator.

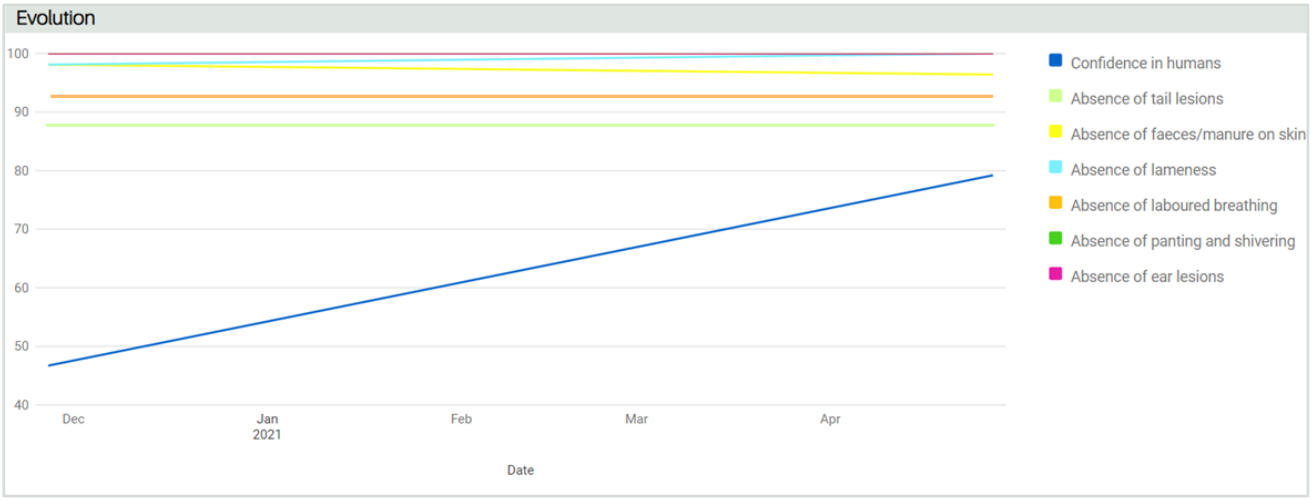

Figure 2: Evolution graph from the results for finishing pigs in the PIGLOW app, which shows the change of scores for 7 highlighted welfare indicators across time. A straight line is drawn for each indicator between the moments of each conducted welfare assessment.

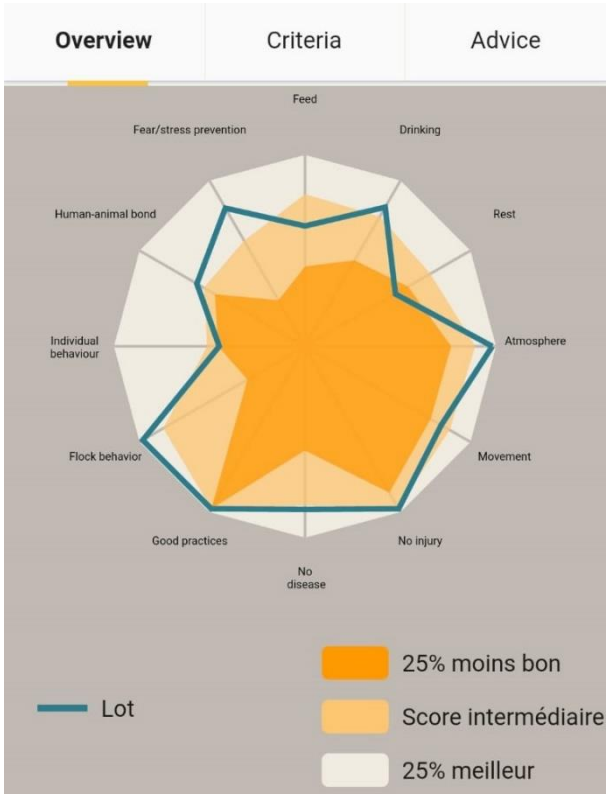

Figure 3: Spider map showing results for broiler chickens in the EBENE® app. The blue line represents the user's own scores, where the score is higher if the line is closer to the edge of the graph. The different colours of the backgrounds of the graph indicate the height of scores of other users.

3. Objectives

After the completion of the welfare self-assessment apps in T3.1, the objective of T3.2 was to test whether use of these apps by farmers could have an effect on animal welfare on outdoor farms. For this purpose, we set up a longitudinal study in which free-range and organic pig and poultry farmers were asked to perform regular assessments with the PIGLOW or EBENE® app during a 2-year period. Based on previous studies showing beneficial effects of welfare self-assessments, benchmarking and feedback from experts, we hypothesised that regular use of self-assessment apps (containing benchmarking and expert-feedback) would lead to an improvement of animal welfare on the farms. Because we were interested to know in what ways the apps would influence the farmers and which aspects of the apps would be most valuable to them, the farmers were also asked to fill out surveys on these subjects. In line with earlier studies, we expected farmers to particularly value the possibilities for benchmarking.

For this study, it was necessary for researchers to visit each of the participating farms at the beginning and end of the study to be able to gather our own welfare data with which we could evaluate changes in animal welfare during the study. This provided us with the opportunity to conduct the first and last welfare assessment with the PIGLOW or EBENE® app together with each farmer. This allowed us to also ask the question of how welfare assessments with the app by farmers and trained researchers compare to one another. This question was of interest, because the specific differences between farmers and researchers could give information about how farmers look at their animals and what factors might influence how they judge welfare problems. Additionally, it could provide useful insights into how the PIGLOW and EBENE® app could be improved. For instance, if the results suggest that farmers might misunderstand certain questions in the app, those questions could be altered to avoid such mistakes in the future. Because the farmers were not yet experienced with conducting structured welfare assessment but would gain experience during the study, we expected that a learning curve would be visible in the assessments by farmers. Because the researchers were already experienced, we expected the results of farmers and researchers to become more alike and therefore expected to see smaller differences between farmers and researchers at the end of the study than at the beginning.

4. Methods

For the legibility of the document, some recurring words are abbreviated in the methods and results sections in the following way:

- Welfare indicator(s): WI
- Welfare assessment(s): WAss
- Median at the beginning of the longitudinal study: median B
- Median at the end of the longitudinal study: median E

4.1. Participating farms

For this study, it was decided to only include farmers of finishing pigs and broiler chickens, because the cycle for fattening pigs and broiler chickens is shorter than that for sows and laying hens, which gives the farmers more opportunities to make changes on their farms based on the results of their WAss within the 2-year period of the study.

To recruit farmers to participate in the study, many different channels were used. As a first step, members of the National Practitioner Group (NPGs) and NPG facilitators in Belgium were approached to ask if they wanted to participate or knew other farmers who might be interested. These groups were made up of farmers and other professionals from the sector who were involved in the PPILOW project to give their opinion on the research being conducted. Unfortunately, the Dutch and French NPGs did not include pig or broiler chicken farmers. Afterwards, other colleagues from within and from outside PPILOW partner organisations were asked for their help to approach farmers they worked with, calls for participation were distributed through newsletters and other (social) media channels aimed at (organic) farmers, existing databases of organic farmers and farm stores were consulted and extensive Google searches were done to find farmers who fit the scope of the PPILOW project. Farmers were approached with phone calls and/or emails depending on which contact details were available.

Thirteen free-range and organic pig farmers from Belgium and The Netherlands and 15 free-range and organic poultry farmers from Belgium and France were initially recruited to participate in the longitudinal study. Originally, it was planned to recruit 20 to 30 farmers per species. Unfortunately, the Covid-19 pandemic as well as AI during the time of recruitment made it more difficult to find farmers willing to participate. Two poultry farmers withdrew from the study shortly after it had started and 1 poultry farmer withdrew approximately 1 year after the start for reasons related to AI. Two other poultry farmers stopped responding to attempts to contact them near the end of the study and a third had to withdraw from the study because their farm could not be visited at the end. One pig farmer was dropped from the study after approximately 1 year because - after repeated reminders - he had not yet completed any WAss with the PIGLOW app. This resulted in a final sample size of 12 pig farmers and 9 poultry farmers. General information on all participating farms can be found in table 1 (pigs) and table 2 (poultry). The study did not contain a control group with farmers who did not use the app. All farmers were asked to read and sign the information and consent form before the study (Annex A). This consent form contains the information that any collected research data may be used after a farmer has withdrawn from the study unless he or she specifically makes a request for the data to be destroyed. None of the farmers did this, meaning that all research data from their farms that had already been collected could be used for analysis.

Table 1: Overview of the pig farms participating in the longitudinal study on the effect of the use of the PIGLOW app.

Farm	Country	Average number of fattening pigs	Organic	Housing system	Date of first and last farm visit	Comment
1	Belgium	808	Yes	Stable with outdoor access	27-11-2020 13-01-2023	
2	Belgium	200	No	Huts	07-12-2020 16-01-2023	
3	Belgium	385	Yes	Stable with outdoor access	22-02-2021 27-02-2023	
4	Belgium	600	No	Huts	10-03-2021 03-03-2023	
5	Belgium	21	Yes	Mobile huts	15-04-2021 09-05-2023	
6	Belgium	250	No	Stable with outdoor access + mobile huts	08-03-2021 17-04-2023	
7	Belgium	150	Yes	Stable with outdoor access	12-02-2021	Was dropped from the study after approximately 1 year
8	The Netherlands	6000	Yes	Stable with outdoor access	14-01-2021 19-01-2023	
9	The Netherlands	988	Yes	Stable with outdoor access	18-01-2021 27-01-2023	
10	The Netherlands	350	No	Mobile huts	22-01-2021 23-01-2021	
11	The Netherlands	15000	No	Stable with outdoor access	17-03-2021 01-03-2023	
12	The Netherlands	1075	Yes	Stable with outdoor access	22-02-2021 21-02-2023	

13	The Netherlands	450	Yes	Stable with outdoor access	10-08-2021 14-07-2023	

Table 2: Overview of poultry farms participating in the longitudinal study on the effect of the use of the EBENE® app.

Farm	Country	Average number of broiler chickens	Organic	Date of first and last farm visit	Comment
1	Belgium	14550	Yes	20-07-2021 05-01-2024	
2	Belgium	13000	No	28-07-2021 27-10-2023	
3	Belgium	9100	Yes	26-07-2021 10-08-2023	
4	Belgium	9350	Yes	25-08-2021 06-11-2023	
5	Belgium	8800	Yes	23-11-2021 28-11-2023	
6	Belgium	6800	Yes	9-11-2021	Withdrew from the study (stopped responding) shortly before the end
7	Belgium	5000	Yes		Withdrew from the study shortly after the start
8	France	4000	Yes	04-06-2021 19-07-2023	
9	France	4000	Yes	18-06-2021 21-08-2023	
10	France	4400	No	19-08-2021 16-10-2023	
11	France	8800	No	20-08-2021 19-10-2023	
12	France	4000	Yes	5-10-2021	Withdrew from the study shortly before the end
13	France	12500	No	22-11-2021	Withdrew from the study (stopped responding) shortly before the end

14	France	4400	No	28-10-2021	Withdrew from the study after approximately 1 year
15	France	8000	Yes		Withdrew from the study shortly after the start

4.2. Farm visits

At the beginning and end of the 2-year study, each participating farm was visited by a researcher. All first visits were originally planned to take place between September and December of 2020, but due to Covid-19 as well as AI measures, many visits had to be postponed and eventually took place between November 2020 and November 2021. During this first visit, the farmer and researcher performed the first animal welfare assessment together, with the PIGLOW app for finishing pigs or with the EBENE® app for broiler chickens. On pig farms, where pigs of different ages were usually available, the farm visits could be planned at any time and pigs of different ages within the category of finishing pigs were included in the WAss to get a good overview of the animals on the farm. On broiler chicken farms all animals that were present were usually of the same age. Because many welfare problems are likely to increase with age, care was taken to plan the farm visits in the 2 weeks before the chickens were slaughtered so that data within and between farms would be as consistent as possible.

The WAss with the app was conducted simultaneously and on the same animals by both the researcher and the farmer, but each filled out their own WAss on their own mobile device and they did not discuss the answers until it was finished. This was done to ensure that the WAss would truly be independent, so that the results of farmers and researchers could reliably be compared. During the last visit, the farmer and researcher performed the last WAss of the study in the same way. This allowed for a comparison of the differences between WAss by farmers and researchers at the beginning and at the end of the 2-year study. An explanation of the 18 animal WI of the PIGLOW and EBENE® app that were used for the comparison of WAss by farmers and researchers can be found in table 3 and table 4, respectively.

Table 3: List of welfare indicators and corresponding scoring methods for finishing pigs in the PIGLOW app that were used to compare assessments by farmers and researchers.

Welfare indicator	Level	Positive or negative	Scoring method	Details of scoring method
Panting	Individual	Negative	Percentage of pigs during a group observation	The number of pigs that are panting is counted per group.
Shivering	Individual	Negative	Percentage of pigs during a group observation	The number of pigs that are shivering is counted per group.
Too small	Individual	Negative	Percentage of pigs during a group observation	The number of pigs that are at least 1/3 smaller than the group-average is counted per group.

Bad general state	Individual	Negative	Percentage of pigs during a group observation	The number of pigs that show signs of sickness or otherwise compromised health is counted per group. Examples: animals seem in pain, sick, needing further care to avoid complications, dull or apathic, isolated from the group, with dull/sunken eyes, blue/red ears or snout, pale skin colour, rapid respiration, or animals with a significant deformation or large hernia (bigger than the distance between the hernia and the floor).
Laboured breathing	Individual	Negative	Percentage of pigs during a group observation	The number of pigs that show laboured/heavy breathing (pumping) is counted per group.
Covered with faeces	Individual	Negative	Percentage of pigs during a group observation	The number of pigs covered with faeces on at least 50% of the skin surface (one side of the body) is counted per group. Only faeces is taken into account, not mud or sand.
Skin wounds	Individual	Negative	Percentage of pigs during a group observation	The number of pigs that have skin wounds larger than 5 cm on the flank or legs is counted per group.
Scratches	Individual	Negative	Percentage of pigs during a group observation	The number of pigs that have at least 15 scratches on one side of the body is counted per group.
Ear lesions	Individual	Negative	Percentage of pigs during a group observation	The number of pigs that have ear lesions is counted per group.
Tail lesions	Individual	Negative	Percentage of pigs during a group observation	The number of pigs that have tail lesions is counted per group.

Skin irritation	Individual	Negative	Percentage of pigs during a group observation	The number of pigs that show signs of skin irritation or parasites is counted per group.
Lameness	Individual	Negative	Percentage of pigs during a group observation	The number of pigs that are lame is counted per group. Lameness is defined as everything between 'visibly reduced weight bearing on one limb or 'limping') and being unable to walk.
Enrichment use	Individual	Positive	Percentage of pigs during a group observation	The number of pigs that are using any form of enrichment (toys, straw/roughage or soil) is counted per group.
Huddling	Group	Negative	Percentage of groups	Whether more than 50% of the pigs are huddling (lying close together and partially on top of each other) is observed for each group.
Lying on the flank	Group	Negative	Percentage of groups	Whether more than 50% of the pigs are spaced out across the pen and are lying on their flank is observed for each group.
Liquid faeces	Group	Negative	Percentage of groups	Whether there are any signs of liquid faeces inside the pen/enclosure is observed for each group.
Coughing / sneezing	Group	Negative	Percentage of groups	Whether there was any audible coughing or sneezing (at the time of observation) is observed for each group.
Water access	Group	Negative	Percentage of groups	Whether any animals might have difficulty in accessing good quality drinking water is observed for each group.

Table 4: List of welfare indicators and corresponding scoring methods for broiler chickens in the EBENE® app that were used to compare assessments by farmers and researchers.

Welfare indicator	Level	Positive or negative	Scoring method	Explanation / Details of scoring method
Still	Individual	Negative	Number of occurrences during 2 transect walks from one side of the building to the other	The number of birds that do not move (even when approached by an observer/surrounding animals) is counted per group.
Too small	Individual	Negative	Number of occurrences during 2 transect walks from one side of the building to the other	The number of birds that are at least ½ smaller than the average bird is counted per group.
Dirtiness	Group	Negative	Estimated percentage of birds after 2 transect walks from one side of the building to the other	The percentage of birds with dark marks easily visible on the back, wings or rear feathers is estimated.
Injuries	Individual	Negative	Number of occurrences during 2 transect walks from one side of the building to the other	The number of birds with visible wounds/lesions (healed or fresh) on the head, back or rear is counted.
Footpad dermatitis	Individual	Negative	Number of occurrences among 15 caught birds	The number of birds whose foot soles are affected by any degree of footpad dermatitis is counted.
Lameness	Individual	Negative	Number of occurrences during 2 transect walks from one side of the building to the other	The number of birds that have difficulty walking (the legs look stiff, the bird looks unbalanced, or the bird needs to stop walking every few steps) is counted.
Other anomalies	Individual	Negative	Number of occurrences during 2 transect	The number of birds with any other physical anomalies, such as

			walks from one side of the building to the other	feather damage or malformations, is counted.
Dead	Individual	Negative	Number of occurrences during 2 transect walks from one side of the building to the other	The number of birds that are dead is counted.
Litter quality	Group	Negative	Score from 0 (best) to 3 (worst)	The quality of the litter is scored after each transect walk. 0 = Dry and friable | 1 = Friable and slightly moist | 2 = Friable and crusty in some places | 3 = Totally crusty or moist
Dust bathing	Group	Positive	Frequency during 3x 5-minute behavioural group observations of a defined zone	The number of dust bathing incidences during the group observations is counted. Dust bathing is defined as: lying on the ground, the bird shakes the litter with its wings and claws to coat its feathers with dust.
Preening	Group	Positive	Frequency during 3x 5-minute behavioural group observations of a defined zone	The number of preening incidences during the group observations is counted. Preening is defined as: the bird cleans its own feathers with its beak.
Foraging	Group	Positive	Frequency during 3x 5-minute behavioural group observations of a defined zone	The number of foraging incidences during the group observations is counted. Foraging is defined as: the bird pecks the litter or another element – except the feed – and/or scrapes the litter with its claws.

Stretching / wing flapping	Group	Positive	Frequency during 3x 5-minute behavioural group observations of a defined zone	The number of stretching/wing flapping incidences during the group observations is counted. Stretching/wing flapping is defined as: the bird deploys one or both wings calmly - without locomotor activity- or extends one of its legs, flaps its wings.
Aggressive pecking	Group	Negative	Frequency during 3x 5-minute behavioural group observations of a defined zone	The number of aggressive pecking incidences during the group observation is counted. Aggressive pecking is defined as: pecking directed to one or several birds (the head or the neck are often targeted). Note: duels are not indicative of aggressive behaviours.
Positive interaction	Group	Positive	Frequency during 3 5-minute behavioural group observations of a defined zone	The number of incidences of positive interaction is counted. A positive interaction is defined as: mutual grooming, non-aggressive pecking, duelling.
Panting	Group	Negative	Percentage of birds in the group	The percentage of birds that are panting is estimated at the end of each behavioural observation. Panting is defined as: breathing quickly with the beak open.
Resting	Group	Positive	Percentage of birds in the group	The percentage of birds that are resting is estimated at the end of each behavioural observation. Resting is defined as: lying on the ground not showing any other activity, with the eyes open or closed. As an intermediate

				percentage of resting birds in a group is optimal, the formula $-0.0014X^2+0.1X+3$, where X is the percentage of resting animals, is used to calculate a score from 0 (worst) to 5 (best).
Distance from humans	Group	Negative	Score from 0 (best) to 3 (worst)	The distance of the chickens from the observer at the end of the behavioural observation is scored. 0 = Established contact | 1 = At your feet, no touching | 2 = Less than 1m away | 3 = More than 1m away

In addition to the WAss with the PIGLOW or EBENE® app, the researcher conducted a more detailed WAss of individual animals and groups during both farm visits to get a more precise view of animal welfare on each farm. These detailed WAss from the first and last farm visit were used as a baseline for animal welfare and as the final animal welfare status of the pigs and chickens on each farm in order to determine whether welfare on the farms had changed during the study. The WI of these detailed WAss were very similar to the WI in the PIGLOW and EBENE® app, so that the detailed WAss measured changes for aspects of welfare that could be expected to change due to use of the apps. For **finishing pigs**, this detailed welfare assessment included 7 group level WI and 13 individual level WI. Group level WI were either scored as a percentage of animals in a group that was "positive" for that WI or a score from 0-2 or 0-3 was assigned. Individual WI were all scored by answering a yes/no question for each pig or on a scale from 0 (no welfare problem) to 100 (worst possible welfare problem). If available on the farm, at least 5 pens/groups and at least 30 individual pigs were assessed. On farms that were big enough, care was taken to observe pigs from pens at different locations within the building or the area of the farm and to choose pigs of different ages that fell within the category of finishing pigs. As some farms were very small, the minimum number of observed groups was 1 and the minimum number of individual pigs that was observed was 15. In total, the WAss on the pig farms were completed by 3 different researchers. The list of assessed WI and an explanation of the scoring methods can be found in table 5.

For **broiler chickens**, 5x 5-minute group behavioural observations were conducted to evaluate 6 behaviours and 3 additional group level WI were assessed after each behavioural observation. In addition, 8 individual level WI were assessed, which were all scored on a scale from 0 (no welfare problem) to 100 (worst possible welfare problem). For the individual assessment, at least 50 chickens were caught and handled when possible. However, sometimes it was not possible to catch this number of chickens due to factors such as extreme temperatures or high perceived stress levels of the chickens. The minimum number of individual chickens and groups that were observed on the farms were 9 and 3, respectively. The WAss on poultry farms were completed by 3 different researchers. The list of assessed WI and an explanation of the scoring methods can be found in table 6.

Because this study did not contain a control group, it was not possible to make a comparison of the evolution of animal welfare on farms where the app was used and farms where the apps weren't used.

Table 5: Explanation of welfare indicators and corresponding scoring methods of the detailed welfare assessment by researchers for fattening pigs.

Welfare indicator	Level	Positive or negative	Scoring method	Explanation / Details of scoring method
Panting	Individual	Negative	Score on a continuous scale from 0 (best) to 100 (worst)	The following indications are used with the continuous scale: 1-33 = breathing is clearly more rapid and superficial than normal. 34-66 = breaths come in gasps and chest movements are clearly more rapid than normal. 67-100 = breaths come in short gasps and chest is moving rapidly.
Shivering	Individual	Negative	Yes | No	Shivering is defined as: body shakes with (very) small movements.
Too small	Individual	Negative	Yes | No	Too small pigs are defined as pigs that are at least 1/3 smaller than the average of the group.
Bad general state	Individual	Negative	Yes | No	Animals in a bad general state are defined as: animals that show general signs of sickness or otherwise compromised health. Examples: animals which are obviously in pain, sick, needing further care to avoid complications, dull or apathic, isolated from the group, with dull/sunken eyes, blue/red ears or snout, pale skin colour, rapid respiration.
Hernia	Individual	Negative	Score on a continuous scale from 0 (best) to 100 (worst)	The following indications are used with the continuous scale: 1-25 = small protrusion and no bleeding. 26-50 = small but bleeding protrusion or moderate protrusion, but not bigger than the distance between the hernia and the floor and not bleeding. 51-75 = moderate size and bleeding hernia or bigger than the distance between the hernia and the floor but not bleeding. 76-100 = (much) larger than the distance between

				the hernia and the floor up to touching the floor and bleeding.
Laboured breathing	Individual	Negative	Score on a continuous scale from 0 (best) to 100 (worst)	The following indications are used with the continuous scale: 1-25 = breathing is slightly more heavy than normal. 26-50 = breathing clearly sounds more heavy than normal, possibly a soft wheezing (high pitched sound) can be heard. 51-75 = more laboured breathing, possibly with more pronounced movements of the chest or a clear wheezing sound is present. 76-100 = very heavy breathing with loud wheezing or pumping and laboured movements of the chest with each breath.
Covered with faeces	Individual	Negative	Score on a continuous scale from 0 (best) to 100 (worst)	The score reflects the % of the skin surface that is covered with faeces (dirtiest side of the body, only faeces, no mud or sand).
Skin wounds	Individual	Negative	Score on a continuous scale from 0 (best) to 100 (worst)	Wounds on all sides of the body are assessed. Wounds on the ears or tail count only as ear and tail lesions. The following indications are used with the continuous scale: 1-25 = one or several small (<2 cm), shallow wounds that are not bleeding. 26-50 = one (26) or several (50) small wounds that are open/bleeding (towards 50), one (26) or several (50) medium size (2-5 cm) wounds that are almost healed. 51-75 = a large number of small/medium and open/bleeding wounds, one large open wound or a few large (>5 cm) but almost healed wounds. 76-100 = one or more deep, large wounds that are open and/or bleeding.
Scratches	Individual	Negative	Counting	A mark on the skin is considered to be a scratch when it is long, relatively narrow and shallower than a wound. The average number of scratches per side of the body (including flank, legs and head) is counted. If

				both sides of the body are visible, the total number of scratches is divided by two.
Ear lesions	Individual	Negative	Score on a continuous scale from 0 (best) to 100 (worst)	The following indications are used with the continuous scale: 1-25 = only very small scabs or lesions are visible. 26-50 = there are small crusts or healing wounds on the ears, but no blood. 51-75 = there are bigger crusts on the ears or smaller lesions with fresh blood. 76-100 = ears are severely damaged with big crusts and/or there are fresh wounds with blood. Scratches on the ears count as scratches, not ear lesions.
Tail lesions	Individual	Negative	Score on a continuous scale from 0 (best) to 100 (worst)	The following indications are used with the continuous scale: 1-25 = only small, minor lesions are visible. 26-50 = there are slightly bigger but healing lesions, some swelling or dried blood. 51-75 = there are open wounds, significant swelling or fresh blood. 76-100 = open wounds, significant swelling and fresh blood.
Skin irritation	Individual	Negative	Score on a continuous scale from 0 (best) to 100 (worst)	The most affected side of the body is scored. The following indications are used with the continuous scale: 1-25 = mild local skin inflammation or mild red spots (<10% of body surface on one side). 26-50 = larger area of mildly inflamed/spotted skin (>10%) or a small but clearly inflamed/spotted zone. 51-75 = a large area of the skin that is clearly inflamed/spotted. 76-100 = severely inflamed skin or dark spots over a large area of the skin or less severe inflammation/spots over a much larger area (>30%).
Lameness	Individual	Negative	Score on a continuous scale from 0 (best) to 100 (worst)	The following indications are used with the continuous scale: 1-25 = stiffness of one of the legs while walking. 26-50 = animal can walk, but weight bearing on the affected leg is clearly reduced. 51-75 =

				animal has clear difficulty walking. 76-100 = severe lameness that makes it (almost) impossible for the animal to walk.
Huddling	Group	Negative	Score from 0 (best) to 2 (worst)	The percentage of pigs in the group that are huddling (lying partially on top of each other)) is estimated/determined. 0 = No | 1 = More than 20% = | 2 = More than 50%
Enrichment use	Group / Positive	Negative	Counting / Percentage	A behavioural scan is performed, and the number of pigs in the group showing object play, playing in soil/mud or exploring their environment is counted.
Fear of humans	Group	Negative	Score from 0 (best) to 2 (worst)	After calmly entering the pen/enclosure, the percentage of pigs that respond in fear (try to get away from you, face away from you or huddle in a corner) is counted. 0 = <20% | 1 = 20-60% | 2 = >60%
Liquid faeces	Group	Negative	Score from 0 (best) to 3 (worst)	Whether any of the faeces in the pen (on floors or walls) are liquid is assessed 0 = None | 1 = Some | 2 = More than half | 3 = All visible faeces
Coughing	Group	Negative	Score from 0 (best) to 3 (worst)	Whether any animals were coughing during the observation of a certain group is noted 0 = No | 1 = Once | 2 = Up to 5 times | 3 = More than 5 times
Sneezing	Group	Negative	Score from 0 (best) to 3 (worst)	Whether any animals were sneezing during the observation of a certain group is noted. 0 = No | 1 = Once | 2 = Up to 5 times | 3 = More than 5 times

Table 6: Explanation of the welfare indicators and corresponding scoring methods of the detailed welfare assessment by researchers for broiler chickens.

Welfare indicator	Level	Positive or negative	Scoring method	Explanation / Details of scoring method
Still	Individual	Negative	Yes | No	A still animal is defined as: prostrated animal that does not move, even when the assessor

				approaches to catch it (less than 2 m) and the surrounding animals do move.
Too small	Individual	Negative	Yes | No	A small chicken is defined as: a chicken that is approximately 1/2 smaller than the average bird.
Dirtiness	Individual	Negative	Score on a continuous scale from 0 (best) to 100 (worst)	Dirtiness is defined as: bird with dark marks easily visible on the back, the wings or the rear feathers. The score reflects the % of the plumage that is dirty.
Wounds	Individual	Negative	Score on a continuous scale from 0 (best) to 100 (worst)	The following indications are used with the continuous scale: 1-25 = only 1 or a few small (<0.5 cm) wounds and no fresh blood on the head, back or rear. 26-50 = slightly larger wounds (0.5-1.5cm) and/or fresh blood. 51-75 = several larger (>1.50 cm) wounds or many small, bleeding wounds. 76-100 = a few large (>2.5 cm), bleeding wounds or many bleeding wounds >1.5 cm.
Footpad dermatitis	Individual	Negative	Score on a continuous scale from 0 (best) to 100 (worst)	The following indications are used with the continuous scale: 1-25 = only light discolouration of the footpad, very small superficial lesions or clear hyperkeratosis (tough thickening of the skin). 26-50 = clear discolouration of the footpad and/or slightly larger but still superficial lesions. 51-75 = clear discolouration and deeper lesions, dark papillae or scabs, possible swelling of the foot. 76-100 = deep lesions, large scabs or ulcers, signs of recent haemorrhage and/or a severely swollen foot. If needed, the birds' feet are washed/cleaned to remove dirt before scoring.
Hock burn	Individual	Negative	Score on a continuous scale from 0 (best) to 100 (worst)	The following indications are used with the continuous scale: 1-25 = only light discolouration of the skin (a brownish red) of the hock region up to very small superficial lesions.

				26-50 = clear discolouration of the skin (red and brown), slightly larger but still superficial lesions. 51-75 = dark coloured (dark brown or black), medium sized lesions or scabs in the hock region. 76-100 = dark coloured large lesions or scabs.
Lameness	Individual	Negative	Score on a continuous scale from 0 (best) to 100 (worst)	Movement is gently stimulated by one of the observers. The following indications are used with the continuous scale: 1-25 = legs look slightly stiffer when walking or weight baring on one leg is slightly reduced. 26-50 = legs look stiffer, weight baring on one leg is reduced (can be seen as irregular gait) and the animal might look a bit unbalanced. 51-75 = stiff legs, reduced weight baring on one leg, unbalanced and only walks slowly and short distances. 76-100 = has to stop after every few steps up to not being able to walk at all.
Other anomalies	Individual	Negative	Yes | No	Other anomalies are defined as: rare issues that are not often observed such as discolouration of the chest, feather damage, malformations, etc. The specific anomaly is noted.
Dust bathing	Group	Positive	Frequency during 5-minute behavioural group observation of a defined zone	The number of dust bathing incidences during the group observation is counted. Dust bathing is defined as: lying on the ground, the bird shakes the litter with its wings and claws to coat its feathers with dust.
Preening	Group	Positive	Frequency during 5-minute behavioural group observation of a defined zone	The number of preening incidences during the group observation is recorded. Preening is defined as: the bird cleans its own feathers with its beak.
Foraging	Group	Positive	Frequency during 5-minute behavioural group observation of a defined zone	The number of foraging incidences during the group observation is counted. Foraging is defined as: the bird pecks the litter or another

				element – except the feed – and/or scrapes the litter with its claws
Stretching / wing flapping	Group	Positive	Frequency during 5-minute behavioural group observation of a defined zone	The number of stretching/wing flapping incidences during the group observation is counted. Stretching/wing flapping is defined as: the bird deploys one or both wings calmly - without locomotor activity- or extends one of its legs, flaps its wings.
Aggressive pecking	Group	Negative	Frequency during 5-minute behavioural group observation of a defined zone	The number of aggressive pecking incidences during the group observation is counted. Aggressive pecking is defined as: pecking directed to one or several birds (the head or the neck are often targeted). Note: duels are not indicative of aggressive behaviours.
Positive interaction	Group	Positive	Frequency during 5-minute behavioural group observation of a defined zone	The number of incidences of positive interaction is counted. Positive interactions are defined as: mutual grooming, non-aggressive pecking, duelling.
Enrichment use	Group	Positive	Counting / Percentage	The birds in the zone of the behavioural observation using enrichment (such as perches, pecking objects, etc.) are counted to determine the percentage of animals in the group.
Panting	Group	Negative	Percentage	The percentage of birds in the zone of the behavioural observation panting are estimated. Panting is defined as: breathing quickly with the beak open.
Resting	Group	Positive	Percentage / Score from 0 (worst) to 5 (best)	The percentage of birds in the zone of the behavioural observation that are resting are estimated. Resting is defined as: lying on the ground not showing any other activity, with the eyes open or closed. As an intermediate percentage of resting birds in a group is optimal, the formula $-0.0014X^2+0.1X+3$, where X is the percentage of resting animals, is used to calculate a score from 0 (worst) to 5 (best).

Distance from humans	Group	Negative	Score from 0 (best) to 3 (worst)	The distance of the chickens from the observer at the end of the behavioural observation is scored. 0 = Established contact | 1 = At your feet, no touching | 2 = Less than 1 m away | 3 = More than 1 m away

4.3. Periodic animal welfare assessments with the apps

After the farm visit at the start of the study, farmers were asked to perform regular WAss with the app by themselves. Pig farmers were asked to perform 4 WAss with the PIGLOW app on their own in addition to the 2 WAss performed together with the researcher. Poultry farmers were asked to perform 3 WAss with the EBENE® app on their own in addition to the 2 WAss performed together with the researcher. The number of periodic WAss was lower for poultry than for pig farmers, because of practical reasons. In (organic) poultry farms all animals on the farm are usually of the same age and the WAss could only be completed in the 2 weeks before slaughter. This made it more difficult to conduct more frequent WAss. To ensure periodic and timely use of the apps, farmers received emails or phone calls to help remind them when it was time to perform their next WAss.

The results of these periodic WAss performed by the farmers were not used to assess the welfare status of the farms for the purposes of this study, since it cannot be said with certainty that the farmers performed the WAss correctly and honestly. The goal of these WAss was therefore only for the farmers to be engaged with the welfare of their animals and benefit from the functionalities of the apps.

4.4. Surveys

The farmers were also asked to fill out 2 surveys. The first survey was filled out at the start of the study before the farmers had used the app, while the second survey was filled out between 1 day and 4 weeks after the last WAss of the study had been completed. The main purpose of these surveys was to help answer the question whether the use of the apps could have a positive effect on animal welfare. Thus, the first survey was used to ask the farmers about their initial opinions on animal welfare in general, the welfare of their own animals and their expectations of an app for welfare self-assessments. In the final survey, the farmers were asked whether their opinions on animal welfare or the welfare of their animals had changed due to the use of the PIGLOW or EBENE® app. They were also asked to give their opinion on the user-friendliness and usefulness of different elements of the apps. The final survey was most important to help answer the question about the effect of the welfare self-assessment apps, while the answers to the first survey could be used to put the answers of the final survey into perspective. They were used to get a clearer idea of whether the thoughts of the farmers really changed during the study and whether their expectations had been met. In the questions about the farmers' opinions on animal welfare in general and animal welfare on their own farm, the farmers were specifically asked about welfare aspects that are also present in the PIGLOW and EBENE® app. All welfare aspects in the surveys were framed in a positive way so that the valence of the phrasing would not influence the opinion of the farmers. Thus, for example, the farmers were asked about their opinion on "Absence of wounds and lesions", rather than just "Wounds and lesions". Two types of questions were used for analysis, namely questions that required an answer on a scale from 1-7 or 1-10 and open questions. The questions of both surveys for pigs and poultry that were used for analysis can be found in table 7 and 8.

To design the content of the surveys, several PPILOW partners and other colleagues with experience in designing quantitative and qualitative surveys were consulted. The ethics committee of the University of Perugia was consulted on the necessity to submit these surveys for ethical approval. However, as the surveys do not ask the participants for any information that could be used to identify them, this was deemed unnecessary. The surveys were initially developed in English and subsequently translated into French and Dutch, the native languages of the farmers, by people with those same native languages. The surveys were designed with the LimeSurvey online survey system. All questions were mandatory, with the exception of questions asking farmers if they had any additional remarks to add to the answers of previous questions.

For each survey, participants received a link to the LimeSurvey website and a personal code via email to fill out the surveys online at the appropriate time. While this personal code pseudonymised the responses inside the database, the personal code of each farmer was known a small number of researchers so that it could be verified whether farmers had already responded to the surveys and so that, if necessary, survey answers of individual farmers could be linked to the welfare data from their farm.

Table 7: Overview of all questions of the first and final survey for pig farmers with the corresponding answer type and/or scoring method.

Question	Answer type / Scoring method	Survey
In your opinion, how important is each of the following elements for good animal welfare? • The pigs have enough space • The right type of floor • The pigs are not too cold or too hot (thermal comfort) • Proper hygiene • Having enough water available • Having enough food available • The right composition of the feed (structure) • The right composition of the feed (concentration of vitamins and minerals) • Disease control • The absence of wounds/lesions • The absence of lameness • The absence of diarrhoea • Possibilities to express positive behaviour • Possibilities to use enrichment • The absence of aggressive behaviour • The reaction to human presence	Rate on a scale of 1 (not important at all) to 7 (very important)	First
Please rate how well you think your own farm performs on these same elements of animal welfare	Rate on a scale from 1 (very badly) to 7 (very well)	First

How useful do you think an app could be to monitor the welfare of your animals?	Rate on a scale from 1 (not useful at all) to 7 (very useful)	First
Please rate how useful you think the following aspects of an app for welfare monitoring could be • Evaluating health/sanitary issues • Evaluating behaviour • Automated feedback on results (risk factors for possible welfare issues) • A comparison of your data with those of other farms (benchmarking) • A historical record of your own previous data	Rate on a scale from 1 (not useful at all) to 7 (very useful)	First
Overall, do you feel like the use of the PIGLOW app has changed how important certain aspects of animal welfare are to you?	Rate on scale from 1 (not at all) to 7 (very much)	Final
Please elaborate on your answer to the previous question. Which aspects of animal welfare do you find more or less important now than at the start of this study and why? Is this related to your using the PIGLOW app?	Free text	Final
Do you feel that the use of the PIGLOW app has changed how your farm performs on these welfare indicators? • The pigs have enough space • The right type of floor • The pigs are not too cold or too hot (thermal comfort) • Proper hygiene • Having enough water available • Having enough food available • The right composition of the feed (structure) • The right composition of the feed (concentration of vitamins and minerals) • Disease control • The absence of wounds/lesions • The absence of lameness • The absence of diarrhoea • Possibilities to express positive behaviour • Possibilities to use enrichment • The absence of aggressive behaviour	Rate on a scale from 1 severely deteriorated) to 7 (greatly improved)	Final

• The reaction to human presence		
Overall, do you think the use of the PIGLOW app has led to an improvement of the welfare of your animals?	Rate on a scale from 1 (not at all) to 7 (absolutely)	Final
Please add any additional remarks that you have concerning the previous question	Free text	Final
For each of the following aspects of the PIGLOW app, rate how easy you found them to understand • General use of the app • Choosing the right questionnaire • Content – Questions • Results - Opening the results on the website • Results - Own answers • Results – Benchmarking • Results - Welfare radar • Results - Automated feedback and risk factors • Results - Evolution graph	Rate on a scale from 1 (very difficult) to 7 (very easy)	Final
For each of these aspects of the PIGLOW app, rate how useful you found them • Content - Questions on housing • Content - Questions on health • Content - Questions on behaviour • Content - Additional information (i) for questions • Content - Photo material for questions • Results – Benchmarking • Results - Welfare radar • Results - Automated feedback and risk factors • Results - Keeping track of your previous data	Rate on a scale from 1 (not useful at all) to 7 (very useful)	Final
How would you rate the PIGLOW app?	Rate on a scale from 1 to 10	Final
Please add any additional remarks that you have concerning the previous question	Free text	Final
What changes would you like to see made to the PIGLOW app?	Free text	Final

Table 8: Overview of all questions of the first and final survey for poultry farmers with the corresponding answer type and/or scoring method.

Question	Answer type / Scoring method	Survey
In your opinion, how important is each of the following elements for good animal welfare? • The chickens have enough space • The quality of the litter • The absence of footpad dermatitis • The absence of dirty birds • Having enough water available • Having enough food available • The absence of immobility • The absence of wounds / lesions • The absence of lameness • The slaughter method on the farm • Possibilities to express positive behaviour • Possibilities to use enrichment • The absence of aggressive behaviour • The reaction to human presence • Foraging behaviour • Dustbathing	Rate on a scale of 1 (not important at all) to 7 (very important)	First
Please rate how well you think your own farm performs on these same elements of animal welfare	Rate on a scale from 1 (very badly) to 7 (very well)	First
How useful do you think an app could be to monitor the welfare of your animals?	Rate on a scale from 1 (not useful at all) to 7 (very useful)	First
Please rate how useful you think the following aspects of an app for welfare monitoring could be • Evaluating health/sanitary issues • Evaluating behaviour • Automated feedback on results (risk factors for possible welfare issues) • A comparison of your data with those of other farms (benchmarking) • A historical record of your own previous data	Rate on a scale from 1 (not useful at all) to 7 (very useful)	First
Overall, do you feel like the use of the EBENE® app has changed how important certain aspects of animal welfare are to you?	Rate on scale from 1 (not at all) to 7 (very much)	Final
Please elaborate on your answer to the previous question. Which aspects of animal welfare do you find more or less important	Free text	Final

now than at the start of this study and why? Is this related to your using the EBENE® app?		
Do you feel that the use of the EBENE® app has changed how your farm performs on these welfare indicators? • The chickens have enough space • The quality of the litter • The absence of footpad dermatitis • The absence of dirty birds • Having enough water available • Having enough food available • The absence of immobility • The absence of wounds / lesions • The absence of lameness • The slaughter method on the farm • Possibilities to express positive behaviour • Possibilities to use enrichment • The absence of aggressive behaviour • The reaction to human presence • Foraging behaviour • Dustbathing	Rate on a scale from 1 severely deteriorated) to 7 (greatly improved)	Final
Overall, do you think the use of the EBENE® app has led to an improvement of the welfare of your animals?	Rate on a scale from 1 (not at all) to 7 (absolutely)	Final
Please add any additional remarks that you have concerning the previous question	Free text	Final
For each of the following aspects of the EBENE® app, rate how easy you found them to understand • General use of the app • Choosing the right questionnaire • Content - Questions • Results - Spider map (own scores) • Results - Spider map (benchmarking) • Results – Criteria / detailed scores • Results - Automated feedback / advice • Results - Comparing current and previous assessments	Rate on a scale from 1 (very difficult) to 7 (very easy)	Final
For each of these aspects of the EBENE® app, rate how useful you found them • Content - Behavioural assessment • Content - Sanitary assessment	Rate on a scale from 1 (not useful at all) to 7 (very useful)	Final

• Content - Additional information (?) for questions • Content - Photo material for questions • Results - Spider map (own scores) • Results – Spider map (benchmarking) • Results – Criteria / detailed scores • Results - Automated feedback / advice • Results - Keeping track of your previous data		
How would you rate the EBENE® app?	Rate on a scale from 1 to 10	Final
Please add any additional remarks that you have concerning the previous question	Free text	Final
What changes would you like to see made to the EBENE® app?	Free text	Final

4.5. Data analysis

Welfare at the beginning and end

To statistically analyse whether welfare on the farms has improved during the study, the data for each WI that was evaluated in the detailed WAss by the researchers was modified to be comparable between farms and between observation moments (see below).

For **finishing pigs**, the values for the **individual level** WI “Shivering” and “Too small” constituted the percentage of observed pigs that were positive for that indicator. For all other individual level WI (Panting, Bad general state, Hernia, Laboured breathing, Covered with faeces, Skin wounds, Scratches, Ear lesions, Tail lesions, Skin irritation, Lameness) the values were the mean scores between 0 and 100 for all observed pigs. For the WI at **group level** “Huddling”, “Fear of humans”, “Liquid faeces”, “Coughing” and “Sneezing” the values were the scores between 0 and 2 and for “Enrichment use”, the value was the percentage of animals in each group that were using enrichment when a behavioural scan of the group was made. For all WI, the mean of each farm was used for the comparison between the beginning and end of the study. For this analysis, the data of the 12 pig farmers who completed the study were used.

For the **individual level** WI for **broiler chickens** “Still”, “Too small” and “Other anomalies”, the values comprised the percentage of observed chickens that were positive for that WI. For all other individual level WI “Dirtiness”, “Wounds”, “Footpad dermatitis”, “Hock burn” and “Lameness” the values were the mean scores between 0 and 100 for all observed chickens. For the 6 behaviours that were recorded at **group level** during the 5-minute behavioural observations (Dust bathing, Preening, Foraging, Stretching / wing flapping, Aggressive pecking, Positive interaction), all values were mean frequencies of the behaviour per chicken per observation. For the remaining group level WI “Enrichment use” and “Panting” the values were the percentage of chickens present at the end of each observation showing that behaviour. For “Resting”, the value was the score between 0 and 5 calculated based on the percentage of chickens present at the end of the observation showing that behaviour. Finally, for “Distance from humans” the value was the score between 0 and 3 for how close the closest chickens were at the end of each observation. For all WI, the mean of each farm was used for the comparison between the beginning and end. For the analysis of the group level WI,

the data of all 9 broiler chicken farmers that completed the study were used and for the analysis of the individual level WI, data of only 7 of those farms were available due to a loss of some of the data.

For the data on pig farms, linear mixed models could only be used for the indicator 'Scratches'. For all other WI, assumptions for linear mixed models were violated (these indicators were analysed by means of descriptive analysis, see further). In the model for 'Scratches', the response variable was the mean score across all farms; observation moment was included as a fixed effect and 'farm' as a random effect. An effect was considered significant if $p \leq 0.05$.

For the data on broiler chicken farms, sample sizes for both individual (7) and group level (9) WI were too small to use linear mixed models (power too low).

For all WI, descriptive statistics were used. Because the data were not normally distributed, medians for each WI at the beginning and end (medians B and E) were compared. In addition, it was quantified how many farms had scores higher than 0 for each WI and for how many farms the scores per WI improved or worsened from the beginning to the end of the study. Finally, boxplots were made in Rstudio to visually compare the data per WI for the beginning and end, but only for indicators for which at least 1 of the medians was higher than 0. Rstudio version 2023.12.1+402 was used for linear mixed models and boxplots, and Microsoft Excel 2016 was used for all descriptive statistics.

Surveys

Descriptive statistics were used to analyse the responses to the surveys from pig and poultry farmers. For all quantitative questions with numerical answers, the mean scores given by the farmers and the related standard deviations were compared. In visual representations of the data, all individual scores of the farmers were shown to provide a more complete overview of the data. For open questions, comments from farmers were selected to be part of the results. While for some questions not all comments are shown, care was always taken to make a fair selection of comments that reflected all of the farmers' opinions rather than a biased subset.

In order to relate the survey data (first and final) to the welfare data, only responses of farmers who completed the entire study were used for the analysis. For pig farmers, the sample size was 12 for the first survey and 11 for the final survey. The reason for the difference is that 1 farmer did complete the entire study but failed to fill out the final survey in spite of repeated reminders. Because the welfare data from that farmer were used in the analysis, the results of their first survey were linked to the welfare scores. For poultry farmers, the sample size for both surveys was 9.

Comparison farmers and researchers

To compare WAss with the PIGLOW or EBENE® app by farmers and researchers, descriptive statistics were used as well. It was not possible to use statistical models for these data, because each farmer/researcher couple as well as the farm on which each couple performed the WAss was different. The descriptive statistics that were used were the mean absolute difference between each farmer and researcher per WI, the number of farmer/researcher couples for which the score for each WI differed and the number of cases in which farmers and researchers scored higher than the other. For **finishing pigs**, the values that were compared for farmers and researchers for all **individual level** WI (Panting, Shivering, Too small, Bad general State, Laboured breathing, Covered with faeces, Skin wounds, Scratches, Ear lesions, Tail lesions, Skin Irritation, Lameness, Enrichment use) were the percentage of the total number of pigs observed in all groups that was scored as positive for each WI. For all **group level** WI (Huddling, Lying on the flank, Liquid faeces, Coughing /Sneezing, Water access) the percentage of groups that was positive for each WI was compared.

The data of 12 couples of pig farmers and researchers who completed the WAss with the PIGLOW app at the beginning and end of the study were used for this analysis.

For **broiler chickens**, the analysed values for all WI related to **individual level** physical anomalies (Still, Too small, Injured, Footpad dermatitis, Lameness, Other anomalies, Dead) were the total frequencies of the anomaly scored by farmers and researchers. For **group level WI** "Dirtiness" and "Panting" the percentages of the group of birds that were estimated to be positive for these indicators were compared between farmers and researchers. For "Litter quality" and "Distance from humans", the score between 0 (best) to 3 (worst) given by farmers and researchers was compared. For the behaviours observed during the 5-minute behavioural observations (Dust bathing, Preening, Foraging, Stretching / wing flapping, Aggressive pecking, Positive interaction) the total frequencies scored by farmers and researchers were compared. Finally, for "Resting" the estimated percentage of birds in the group that were resting was turned into a score between 0 (worst) and 5 (best) with the formula $-0.0014X^2 + 0.1X + 3$ where X was the percentage of resting birds. This was done because an intermediate percentage of resting birds is optimal and therefore directly comparing the percentages scored by farmers and researchers would not give information about whether one group gave better scores than the other. For all questions that were answered for multiple groups or animals, all answers were combined and the totals per WI were compared for farmers and researchers. At the beginning of the study, 13 couples of farmers and researchers completed the EBENE® WAss for broiler chickens. However, due to an error in the EBENE® app the data of 1 party of 2 of those couples was not synced to the database and was therefore lost, leaving the data of 11 couples to be analysed. The data of these 11 couples were used to analyse the difference between EBENE® WAss by farmers and researchers at the beginning of the study. At the end of the study, 9 couples of farmers and researcher completed the EBENE® WAss, but the data of 3 couples were lost due to the same error. Of the 6 remaining couples at the end, there were only 4 for which data from the beginning of the study were available as well. Therefore, for the comparison of the difference between farmers and researchers at the beginning and end of the study the data of only 4 couples could be used.

5. Results and discussion

The results of the longitudinal study are discussed separately for pigs and poultry, starting with those for pigs. For both species, the results on the effect of the use of the WAss app on animal welfare are discussed first, followed by the results of the surveys and finally by the comparison of WAss with the PIGLOW and EBENE® app by farmers and researchers. Because the majority of the results consist of descriptive statistics, it could be difficult for the reader to interpret the results without an immediate explanation of their meaning. For this reason, it was decided to combine each part of the results with a discussion of their meaning. This is followed by a general discussion in which possible explanatory factors for and the implications of all results will be discussed more in depth.

5.1. Pigs

5.1.1. Comparison of animal welfare at the beginning and end of the study

To determine whether animal welfare on the participating pig farms has changed during the study, the scores for each of the 19 WI from the detailed WAss by researchers were compared. The number of individual pigs and the number of groups that were included in those WAss on each of the farms are shown in table 9. The total number of WAss with the PIGLOW app that was performed by each farmer is also shown in the table. Farmers were asked to conduct 6 WAss.

Table 9: Number of individual pigs and groups observed on each of the pig farms at the beginning and end of the study and the number of welfare assessments for finishing pigs each farmer performed with the PIGLOW app. The order of the farms does not correspond with the order in earlier tables.

Farm	# Individual pigs observed		# Groups observed		# PIGLOW assessments by the farmer
	Beginning	End	Beginning	End	
1	30	35	6	5	6
2	16	19	1	3	6
3	47	38	14	5	6
4	43	37	19	5	6
5	17	22	2	3	5
6	30	40	5	6	6
7	54	42	19	5	5
8	30	32	1	2	6
9	30	32	5	5	5
10	59	33	27	6	5
11	15	19	1	2	5
12	30	36	5	2	6

The comparison of scores for all WI at the beginning and end of the study can be found in table 10. For 9 of the 19 WI that were scored (Panting, Shivering, Bad general state, Laboured breathing, Skin irritation, Huddling, Liquid faeces, Sneezing) the number of incidences of any welfare problem was so low at the beginning of the study that the median was 0. This means that the majority of the participating farms already had a perfect score for that WI at the beginning of the study and that there was little room for improvement (for that WI). For 8 of those WI, the median was still 0 at the end of the study, showing that those aspects of welfare did not change substantially during the study. For the 9th WI, "Hernia" (figure 4), the median was higher at the end of the study (0.41), suggesting that this WI deteriorated slightly. For the remaining 10 WI, there was more potential for improvement.

For 9 of the 19 WI the number of farms for which the score improved or worsened during the study was the same or differed by 1 digit only, showing that it differed per farm how these WI changed over time and that there was no clear pattern of those aspects of welfare improving or worsening during the study. Thus, even if the medians for one of these WI differed between the beginning and end, the evidence for a consistent change across farms for these aspects of welfare during the study was less compelling. For 6 of those 9 WI (Panting, Shivering, Laboured breathing, Skin irritation, Huddling, Liquid faeces) medians B and E were the same (0), confirming that there was no clear overall change for those aspects of welfare. For the previously discussed WI "Hernia", median E was slightly higher, but combined with the fact that the score improved and worsened on 5 farms each, the data did not support an overall change for this WI. For the WI "Too small" (median B = 1.67; median E = 3.08; figure 5) the increase of the median was larger, but the score improved and worsened on almost the same number of farms. Taken together, this suggests a slight deterioration of this aspect of welfare. For the last of the 9 WI, "Coughing" (median B = 0.23; median E = 0.20; figure 6), median E was lower. However, combined with the fact that the score for this WI improved and worsened for the same number of farms, the very small difference between medians did not support that this aspect of welfare improved.

There were 8 WI for which medians B and E differed and there was a pattern of scores on the farms predominantly changing in the same direction from the beginning to the end of the study, supporting that there was a general pattern of change. For one of those WI, "Fear of humans" (median B = 0.21; median E = 0.40; figure 7), median E was higher and the score worsened for the majority of the farms. This suggests that this WI deteriorated rather than improved. For 1 indicator, "Covered with faeces" (median B = 4.40; median E = 2.01; figure 8), median E was much lower, but still the scores of 4 farms were worse at the end than at the beginning, which is not convincing evidence of a general improvement for this WI. For 4 WI there was a clear pattern of improved scores at the end of the study on the majority of the farms, but the difference between the 2 medians was so small that the data only suggest a very small possible improvement during the study. These WI were "Skin wounds" (median B = 0.23; median E = 0.21; figure 9), "Ear lesions" (median B = 0.31; median E = 0.13; figure 10), Tail lesions (median B = 0.07; median E = 0; figure 11) and Enrichment use (median B = 20.15; median E = 19.18; figure 12). For the final 2 WI, Scratches (median B = 3.20; median E = 1.60; figure x) and "Lameness" (median B = 1.63; median E = 0. figure x) there was a clear pattern of improved scores on the majority of the farms at the end and the median was clearly lower. For "Scratches" and "Lameness" the data thus suggest that these aspects of welfare improved during the time of the study. This fits with the output of the linear mixed model for "Scratches", which showed a significant effect of observation moment on the welfare score ($t = -2.508$, $p = 0.03$).

Overall, the data of the detailed WAss for finishing pigs showed that there has been a consistent improvement for only 2 of the 19 measured WI during the longitudinal study. For the other 17 WI, the data supported no clear overall change or even suggested a slight deterioration. Thus, overall the data do not support that the use of the PIGLOW app by farmers has had a positive effect on the welfare of finishing pigs. The fact that the median score for many WI was already 0 at the start of the study, shows that the potential for improvement on these farms was already relatively low. Even if the use of the PIGLOW app could potentially have a positive effect on animal welfare, it would have

been difficult to detect on these particular farms. Whether any of the differences between the beginning and end of the study were related to the use of the PIGLOW app cannot be determined with certainty based on these data alone. The results of the surveys in which the farmers gave their own opinion on the app can give more information on this topic.

Table 10: Descriptive statistics for the differences in animal welfare on the pig farms at the beginning and end of the longitudinal study. The first and second column show the median of the mean scores per farm for each welfare indicator for the beginning and end, with the number of scores that were above 0 between brackets. The third and fourth column show the number of farms on which the score for each indicator decreased or increased, respectively, during the study. N = 12.

Welfare indicator	Median beginning (# Scores above 0)	Median end (# Scores above 0)	Farms with improved scores	Farms with worsened scores
Panting	0 (1)	0 (1)	1	1
Shivering	0 (0)	0 (0)	0	0
Too small	1.67 (6)	3.08 (8)	5	4
Bad general state	0 (4)	0 (1)	3	1
Hernia	0 (5)	0.41 (7)	3	4
Laboured Breathing	0 (5)	0 (4)	3	3
Covered with faeces	4.40 (9)	2.01 (8)	6	4
Skin wounds	0.23 (11)	0.21 (6)	8	3
Scratches	3.20 (10)	1.60 (10)	8	2
Ear lesions	0.31 (8)	0.13 (6)	7	4
Tail lesions	0.07 (8)	0 (2)	7	2
Skin irritation	0 (3)	0 (5)	3	4
Lameness	1.63 (9)	0 (4)	8	3
Huddling	0 (4)	0 (4)	4	4
Enrichment use	20.15 (11)	19.18 (12)	5	7
Fear of humans	0.21 (6)	0.40 (7)	1	6
Liquid faeces	0 (4)	0 (3)	4	3
Coughing	0.23 (7)	0.20 (7)	5	5
Sneezing	0 (5)	0 (2)	4	2

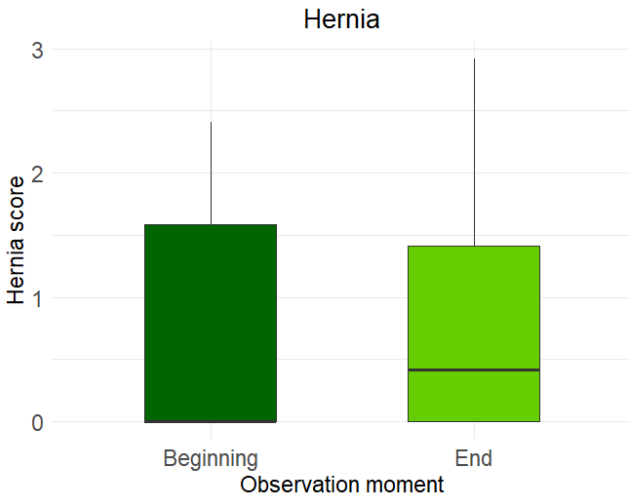

Figure 4: Boxplot showing the differences in the scores for the welfare indicator "Hernia" for finishing pigs at the beginning and end of the longitudinal study. The score for "Hernia" is the mean score from 0 (best) to 100 (worst) for all observed pigs. The line in each box represents the median.

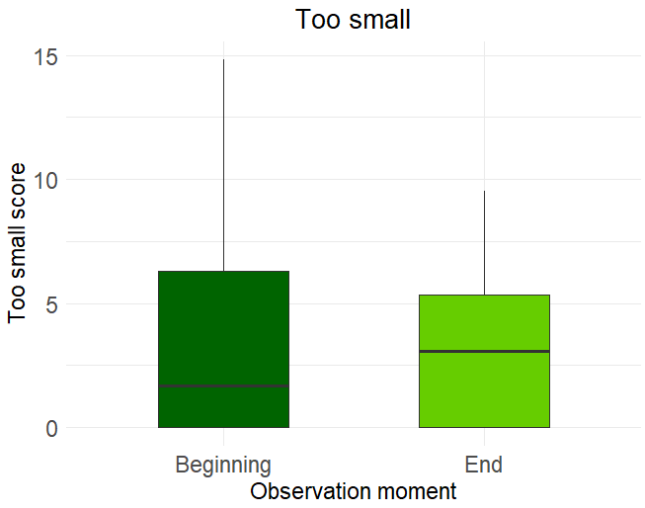

Figure 5: Boxplot showing the differences in the scores for the welfare indicator "Small" for finishing pigs at the beginning and end of the longitudinal study. The score for "Small" is the percentage of pigs that was scored as too small. The line in each box represents the median.

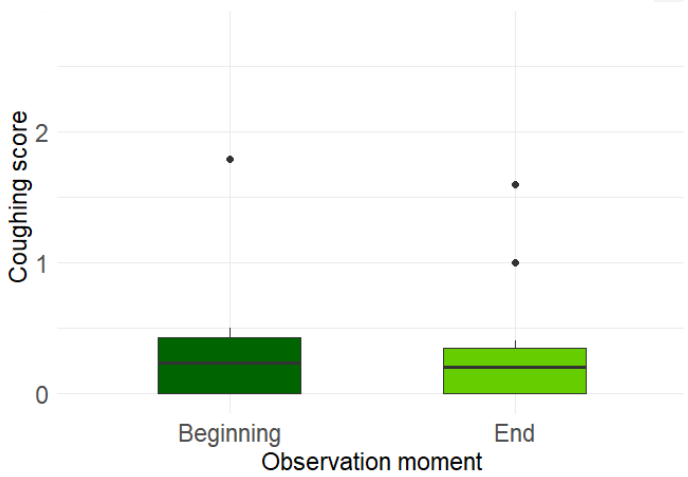

Figure 6: Boxplot showing the differences in the scores for the welfare indicator "Coughing" for finishing pigs at the beginning and end of the longitudinal study. The score for "Coughing" is the mean score between 0 (best) and 3 (worst) for coughing in a group. The line in each box represents the median.

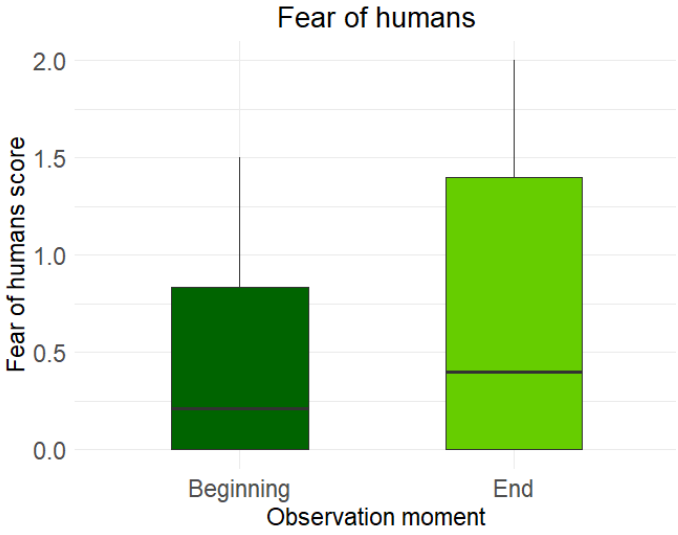

Figure 7: Boxplot showing the differences in the scores for the welfare indicator "Fear of humans" for finishing pigs at the beginning and end of the longitudinal study. The score for "Fear of humans" is the mean score between 0 (best) and 2 (worst) for the reaction of the pigs to humans entering an enclosure. The line in each box represents the median.

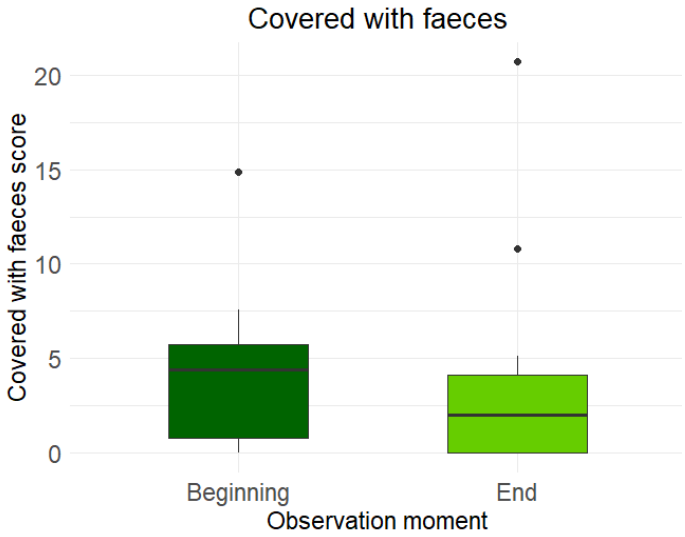

Figure 8: Boxplot showing the differences in the scores for the welfare indicator “Covered with faeces” for finishing pigs at the beginning and end of the longitudinal study. The score for “Covered with faeces” is the mean percentage of the skin surface on one side of the body that is covered with faeces for all observed pigs. The line in each box represents the median.

Figure 9: Boxplot showing the differences in the scores for the welfare indicator “Skin wounds” for finishing pigs at the beginning and end of the longitudinal study. The score for “Skin wounds” is the mean score from 0 (best) to 100 (worst) for all observed pigs. The line in each box represents the median.

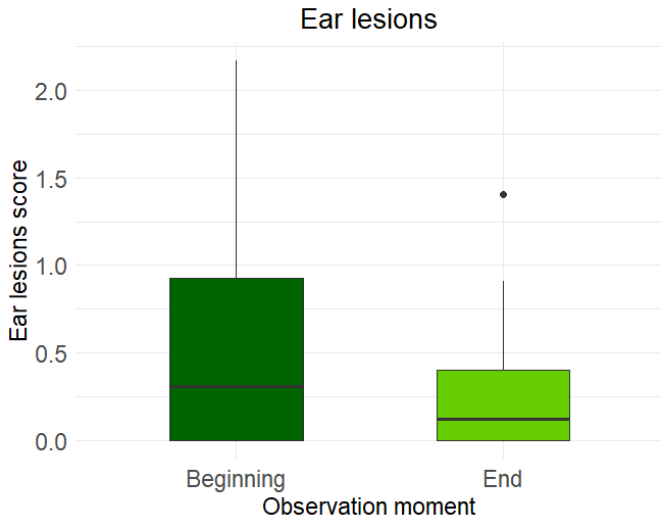

Figure 10: Boxplot showing the differences in the scores for the welfare indicator “Ear lesions” for finishing pigs at the beginning and end of the longitudinal study. The score for “Ear lesions” is the mean score from 0 (best) to 100 (worst) for all observed pigs. The line in each box represents the median.

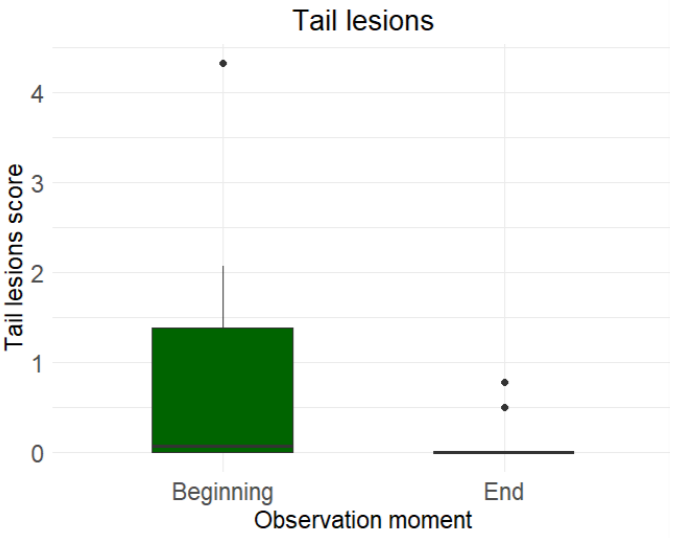

Figure 11: Boxplot showing the differences in the scores for the welfare indicator “Tail lesions” for finishing pigs at the beginning and end of the longitudinal study. The score for “Tail lesions” is the mean score from 0 (best) to 100 (worst) for all observed pigs. The line in each box represents the median.

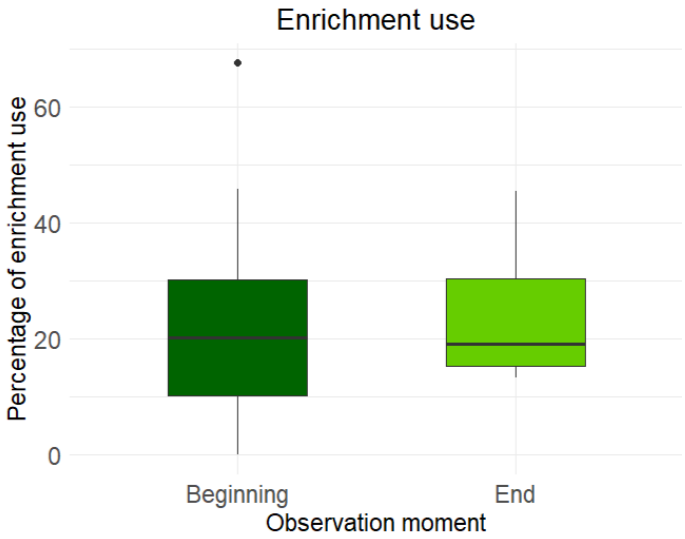

Figure 12: Boxplot showing the differences in the scores for the welfare indicator "Enrichment use" for finishing pigs at the beginning and end of the longitudinal study. The score for "Enrichment use" is the mean number of pigs per group using enrichment. The line in each box represents the median.

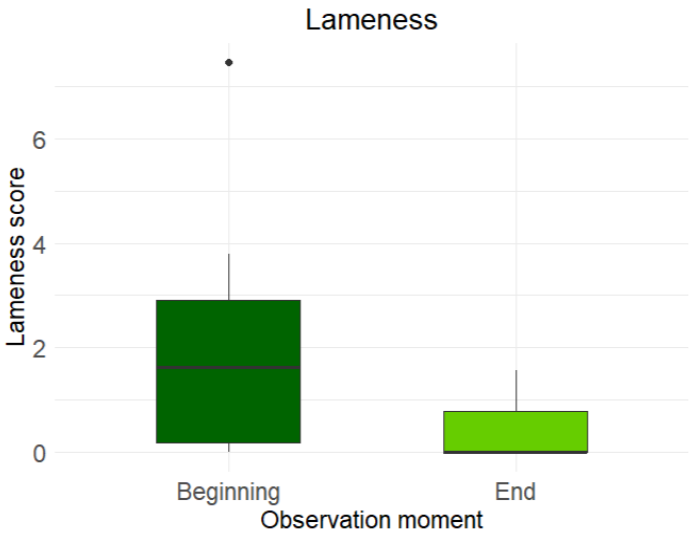

Figure 13: Boxplot showing the differences in the scores for the welfare indicator "Lameness" for finishing pigs at the beginning and end of the longitudinal study. The score for "Lameness" is the mean score from 0 (best) to 100 (worst) for all observed pigs. The line in each box represents the median.

Figure 14: Boxplot showing the differences in the scores for the welfare indicator "Scratches" for finishing pigs at the beginning and end of the longitudinal study. The score for "Scratches" is the mean number of scratches on one side of the body for all observed pigs. The line in each box represents the median.

5.1.2. Surveys

In the section on surveys, the terminology “welfare aspect” rather than “welfare indicator (WI)” is used. This is a deliberate choice, as the welfare aspects that are questioned in the surveys often encompass multiple WI (e.g. the welfare aspect ‘thermal comfort’ encompasses WI “panting” and “shivering”).

First survey

The results of how important the 12 pig farmers judged each of the 16 aspects of pig welfare to be on a scale from 1 (not important at all) to 7 (very important) at the beginning of the study can be found in figure 15. The mean scores for all welfare aspects ranged from 5.58 to 7.00, showing that all welfare aspects were of relatively high importance to the farmers. As all farmers who filled out this survey chose to volunteer for a study on animal welfare, this was perhaps not surprising. The highest mean scores were for “Water availability” (7.00 ± 0; mean ± SD), “Food availability” (6.92 ± 0.29), “Expressing positive behaviour” (6.58 ± 0.79), “Enough space” (6.42 ± 1.00) and “Absence of wounds / lesions” (6.17 ± 0.83). Water and food are the most basic necessities of all living animals, so it is to be expected that all farmers would find these welfare aspects very important. As mentioned in the introduction of this report, an important aspect of organic and most free-range pig farms is increased space allowance for the pigs, which provides them with increased opportunities to perform positive and species-specific behaviours (Høøk Presto et al., 2008; Jensen et al., 2010; Terlouw et al., 2009; Cornale et al., 2015). This could explain why the importance of these 2 aspects of welfare was also rated very highly by this group of free-range and organic farmers. “Absence of wounds / lesions” can include any types of lesions, such as skin wounds, ear or tail lesions and scratches. The scores of the detailed WAss completed by researchers on these farmers’ farms showed that the welfare indicator “Scratches” had one of the highest medians at both the beginning and end of the study. Thus, perhaps this aspect of welfare was important to these farmers because it is one of the problems they regularly experience on their farms.

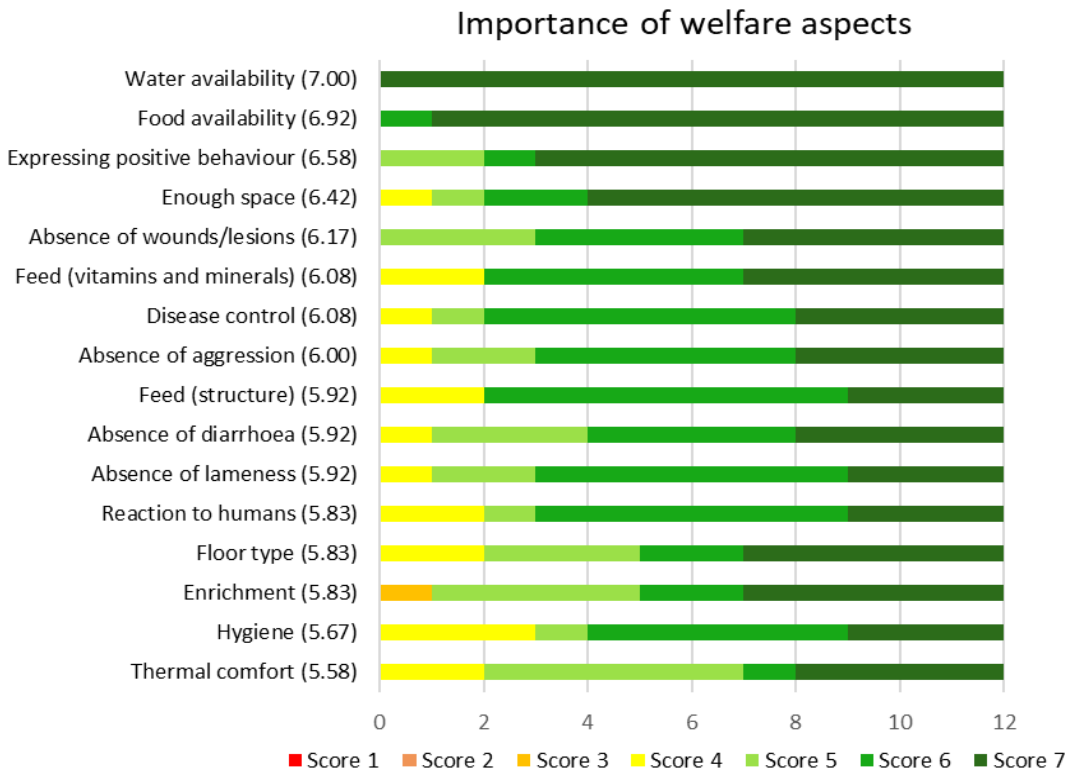

Figure 15: Importance of 16 aspects of pig welfare rated between 1 (not important at all) and 7 (very important) by pig farmers. The welfare aspects are ranked from the highest to the lowest mean rating, which is displayed between brackets behind the name of each welfare aspect, and all individual scores are shown. N = 12.

Figure 16 shows the mean scores given by farmers for the estimation of their own performance for 16 aspects of pig welfare on a scale from 1 (very badly) to 7 (very well). All mean scores were between 5.08 and 6.75, showing that in general the farmers thought quite highly of how they care for the welfare of their pigs. It is worth noting that 4 of the welfare aspects that ranked among the highest 5 for the question about how important farmers find these aspects of welfare were also at the top of the ranking for the question about the farmers' own performance. These were "Water availability" (6.75 ± 0.45), "Food availability" (6.67 ± 0.49), "Expressing positive behaviour" (6.25 ± 1.06) and "Enough space" (6.25 ± 0.97). This indicates that in general, farmers thought they performed well for the aspects of welfare that they found the most important. The exception is "Absence of wounds / lesions" (5.50 ± 1.17), which appeared 5th from the bottom in the ranking for the own performance. This enforces the idea that perhaps this aspect of welfare was important to the farmers precisely because they have experienced related welfare problems. The other aspects for which farmers indicated they performed the least well were "Absence of lameness" (5.08 ± 1.16), "Hygiene" (5.25 ± 1.48), "Thermal comfort" (5.42 ± 1.24) and "Floor type" (mean = 5.50 ± 1.31). Three of these 4 welfare aspects were also part of the bottom 5 of the ranking for the importance of the welfare aspects, showing again that in general, the farmers seemed to think they performed better for welfare aspects that they found more important than for the welfare aspects that they found less important. Once again there was one exception, namely "Absence of lameness" that appeared slightly higher in the ranking of the importance of the welfare aspects. At the beginning of the study, the median for the "Lameness" score from the detailed welfare assessments by researchers was also among the highest ones, which does seem to support the farmers' opinion that this was one of the welfare aspects for which they performed less well.

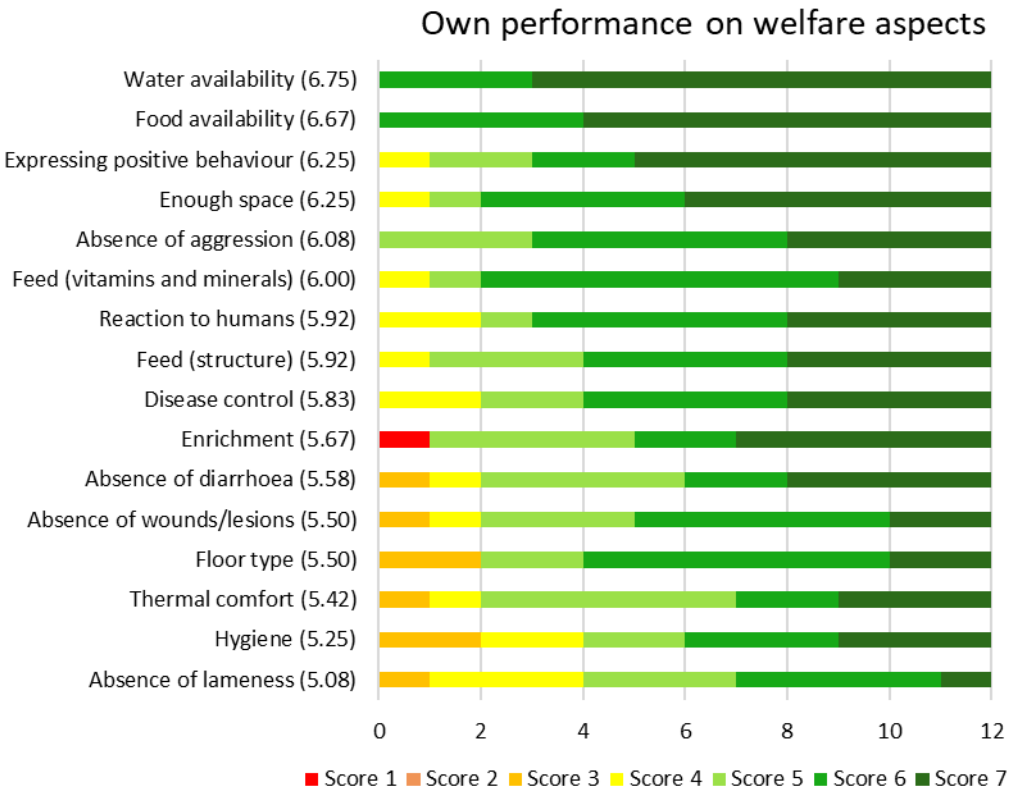

Figure 16: Farmers' estimates for their own performance on 16 aspects of pig welfare rated between 1 (very badly) and 7 (very well). The welfare aspects are ranked from the highest to the lowest mean rating, which is displayed between brackets behind the name of each welfare aspect, and all individual scores are shown. N = 12.

In figure 17, the mean rating for the farmers' own performance has been subtracted from the mean rating for the importance of the same welfare aspect and the differences have been ranked from highest to lowest. Thus, a high positive difference indicates that farmers rated the importance of that welfare aspect more highly than their own performance. It seems logical that the aspects with the largest positive difference would be the aspects of welfare for which the farmers would most wish to improve their performance. The differences between these 2 ratings were largest for "Absence of lameness" (0.83), "Absence of wounds/lesions" (0.67) and "Hygiene" (0.42). For the first 2, it was already established that based on the answers to survey questions and the results of the detailed WAss on these farms it seemed that farmers might experience problems with these aspects of welfare. Several WI that are related to hygiene were also observed in those WAss, such as "Bad general state", "Covered with faeces" and "Skin irritation". The median scores for the indicator "Covered with faeces" were the highest and second highest of all medians for negative WI at respectively the beginning and end of the study, indicating that perhaps farmers would want to improve on that aspect of hygiene, because they have experienced related welfare problems. It must be noted that, as mentioned in the discussion of the results of the detailed WAss, the incidences of most welfare problems were relatively low on these farms. Thus, even for the WI with a high median score relative to others, no serious welfare problems were observed on the majority of the farms. Even so, the welfare scores do support that relative to other welfare aspects, these farmers might have experienced more issues related to the WI "Covered with faeces". The same is true for the previously discussed "Lameness" and "Scratches".

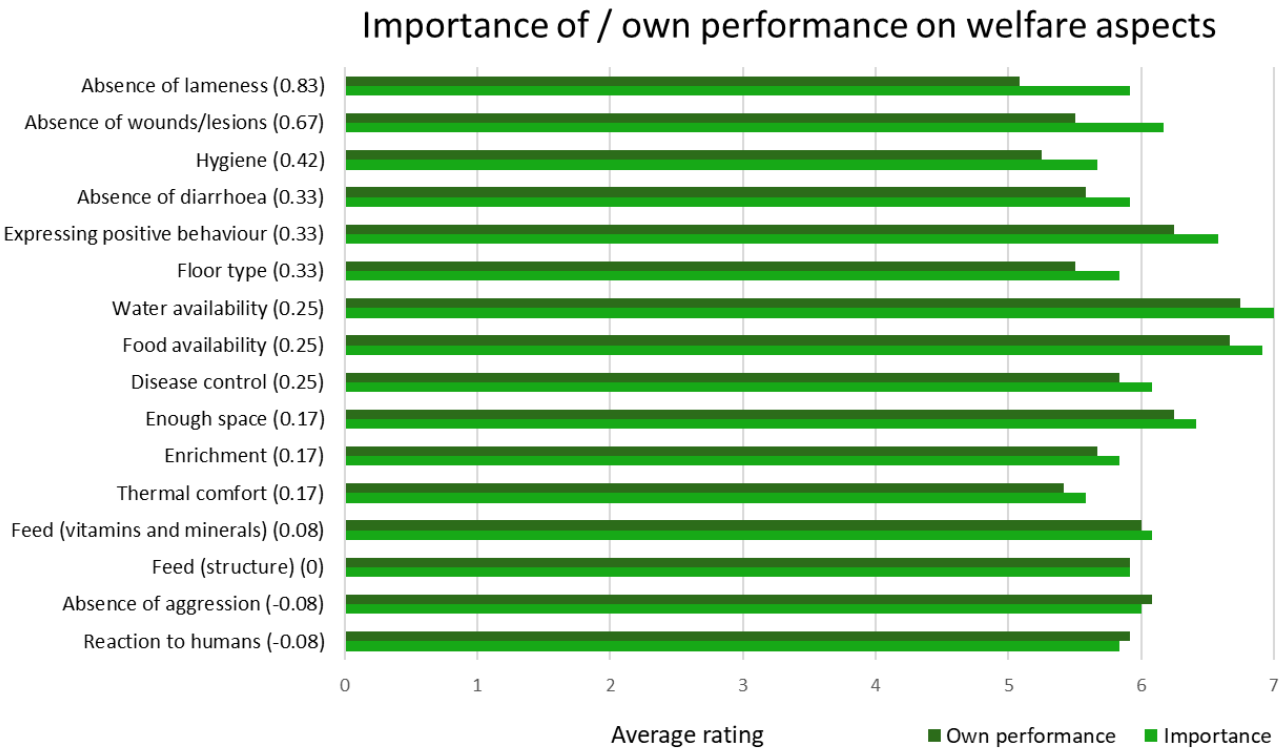

Figure 17: Mean ratings out of 7 for the importance of and farmers' own performance on 16 aspects of pig welfare. The aspects are ranked from the highest positive to highest negative difference when the mean rating for farmers' performance is subtracted from the mean rating for the importance. This difference is shown between brackets behind the name of each indicator. N = 12.

The expected usefulness of the PIGLOW app and its most important elements as indicated by farmers before using the app for the first time are shown in figure 18. The farmers rated the usefulness between 1 (not useful at all) and 7 (very useful). The farmers' mean rating of the usefulness of an animal welfare self-assessment app like the PIGLOW app was 5.25 (SD = 1.48), indicating that they expected to find it quite useful. As this question was only answered by farmers who had already agreed to participate in a study about the app, this score might be higher than it would have been for a random group of free-range and organic pig farmers. The rather large standard deviation does show that even among participants of the study, farmers differed in their opinion on the expected usefulness of the app. The highest mean ratings for specific elements of the PIGLOW app that the farmers were asked about were for "Historical record of data" (5.92 ± 1.08) and "Evaluating health issues" (5.92 ± 1.08), followed by "Benchmarking" (5.58 ± 1.51) and "Automated feedback" (5.50 ± 1.24). The range of all these ratings was quite small, indicating that on average the farmers expected to find each element almost equally useful. Farmers were slightly more interested in using the app to evaluate health issues compared to behaviour. Keeping in mind that "Expressing positive behaviour" was rated as the third most important welfare aspect in an earlier question, this is perhaps surprising. It is possible that farmers expected the assessment of health issues with an app to be easier than the assessment of behaviour. The farmers also seem to have been more interested in keeping a record of their own historical data than in benchmarking and receiving automated feedback.

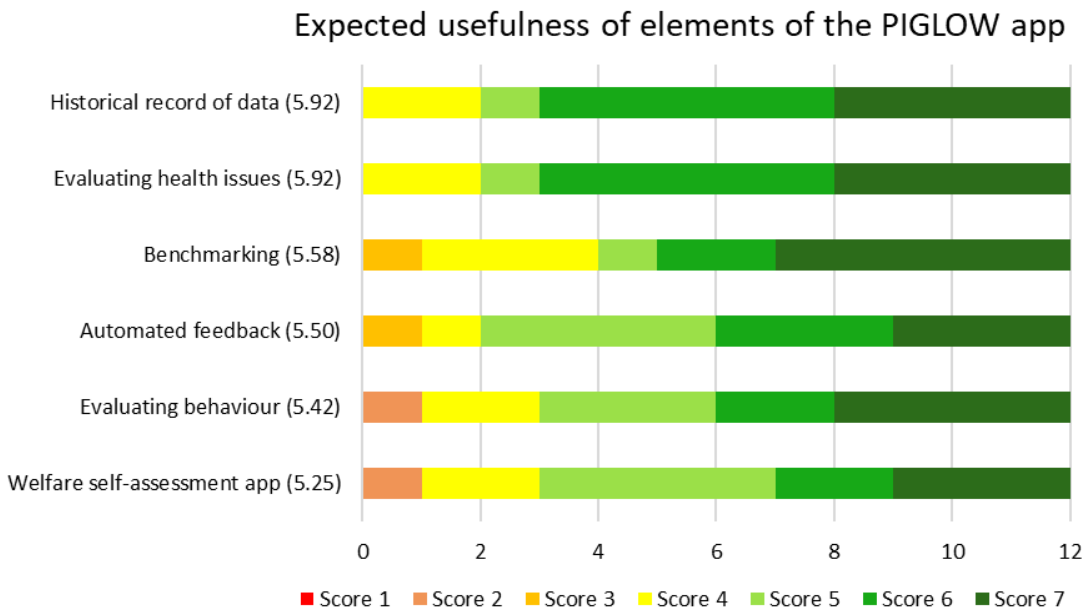

Figure 18: Mean ratings for the expected usefulness from 1 (not useful at all) to 7 (very useful) of the PIGLOW app and the most important elements that the app contains. All individual scores are shown. N = 12.

Final survey

Eleven farmers responded to the question on how much the use of the PIGLOW app had influenced their opinion on the importance of the 16 welfare aspects assessed in the first survey on a scale from 1 (not at all) to 7 (very much). Thus, any score higher than 1 indicated that the farmers noticed an influence of the PIGLOW app. The mean rating given by all farmers was 4.00 (SD = 2.00), which shows that the farmers have noticed a medium-sized effect of the use of the PIGLOW app on their opinions on these aspects of welfare. However, the large standard deviation indicates that there was a wide range of answers. Farmers were asked to elaborate on their answers if possible and their comments could help determine what the mean rating for this question meant. Quotes of farmers are shown together with the rating they gave as an answer to this question. Answers of the farmers to this question included:

1. “If you assess the pigs with the app, you do look more closely.” (6)
2. “The assessment of lesions, scratches, etc. on the bodies of pigs has evolved positively in my opinion. I pay more attention to it.” (5).
3. “Because we muck out the enclosures every day, we have a good image of the welfare of the animals. I don’t think we treat the animals differently.” (4)
4. “Our whole farm already revolves around obtaining the best possible animal welfare, it’s our main goal.” (1)

Motivating the farmers to take a closer look at their animals than they are used to is one of the main goals of the PIGLOW app. Thus, quote 1 showed that for that particular farmer, who gave a higher score than the mean, that goal was achieved. Quote 2 also showed that this farmer has started to pay more attention to at least certain aspects of welfare that are important to them. Interestingly, this farmer mentioned scratches, which was one of the welfare aspects for which the median score had improved at the end of the study. Quote 3 and 4 came from farmers who have given an average score and a very low score and both indicated that they were already very focused on the welfare of their animals and therefore did not feel that they benefited much from the use of the PIGLOW app. These answers illustrate a possible issue with our study, which is that by looking for farmers who

were willing to volunteer for a study on animal welfare, our sample was likely biased towards farmers who are very motivated to provide good welfare for their animals. It is likely that the farmers who are the most motivated to do this are not the farmers who have to most to learn on the topic of animal welfare and therefore are not the farmers who can benefit the most from a tool like the PIGLOW app. This could explain why the mean rating for the question on the effect of the PIGLOW app on the farmers' opinions on the importance of animal welfare was only 4.00.

The farmers' opinions on the effects of the use of the PIGLOW app on their own performance for each of the 16 welfare aspects are shown in figure 19, in order from the largest to smallest perceived effect. Farmers were asked to rate the impact they think the PIGLOW app has had on a range from 1 (severely deteriorated) to 7 (greatly improved), where 4 meant that nothing had changed. The mean ratings for all welfare aspects were between 3.91 and 4.91, indicating that overall farmers have noticed only a very small positive effect of the use of the PIGLOW app on the welfare of their animals. "Food availability" (3.91 ± 1.04), "Water availability" (4.00 ± 0.77), "Feed (structure)" (4.00 ± 1.18) and "Feed (vitamins and minerals)" (4.00 ± 0.18) were at the bottom of the ranking. For food and water this was not surprising given that the farmers rated their own performance on these welfare aspect very highly at the beginning of the study, meaning that there was almost no room for improvement. Feed is a topic that was not directly assessed in the PIGLOW app through a question, but only mentioned in the automated feedback as a risk factor to look at if problems with other aspects of welfare are encountered. Thus, farmers were perhaps less likely to make changes to feed due to the PIGLOW app than they were to make changes to aspects of welfare they observed directly during a WAss. At the top of the ranking were "Reaction to humans" (4.91 ± 1.14), "Expressing positive behaviour" (4.91 ± 1.14), "Enrichment" (4.82 ± 1.08) and "Absence of wounds / lesions" (4.82 ± 0.98). It is noticeable that 3 of these welfare aspects were related to animal behaviour. Although farmers indicated ahead of time that they were more interested in using the PIGLOW app to evaluate health, these results suggest that they learned the most about behaviour. Unfortunately, none of the data from the detailed WAss by researchers showed that scores for these aspects of welfare had indeed improved on the farms, but as the welfare scores were generally already quite good at the beginning of the study it is possible that the differences noticed by the farmers were too subtle to be seen in the welfare scores. The other aspect in the top 4 of the ranking, "Absence of wounds / lesions" was one for which the farmers previously rated their own performance relatively low compared to how important they considered it to be. This suggests that this was one of the aspects of welfare for which the farmers would most like to improve their performance. Although the mean score of 4.82 was still relatively low, it seems that compared to the other welfare aspects, farmers do indeed feel that they have improved their performance more for this one. This fits with the fact that the median score for "Scratches" from the detailed WAss has improved during the study.

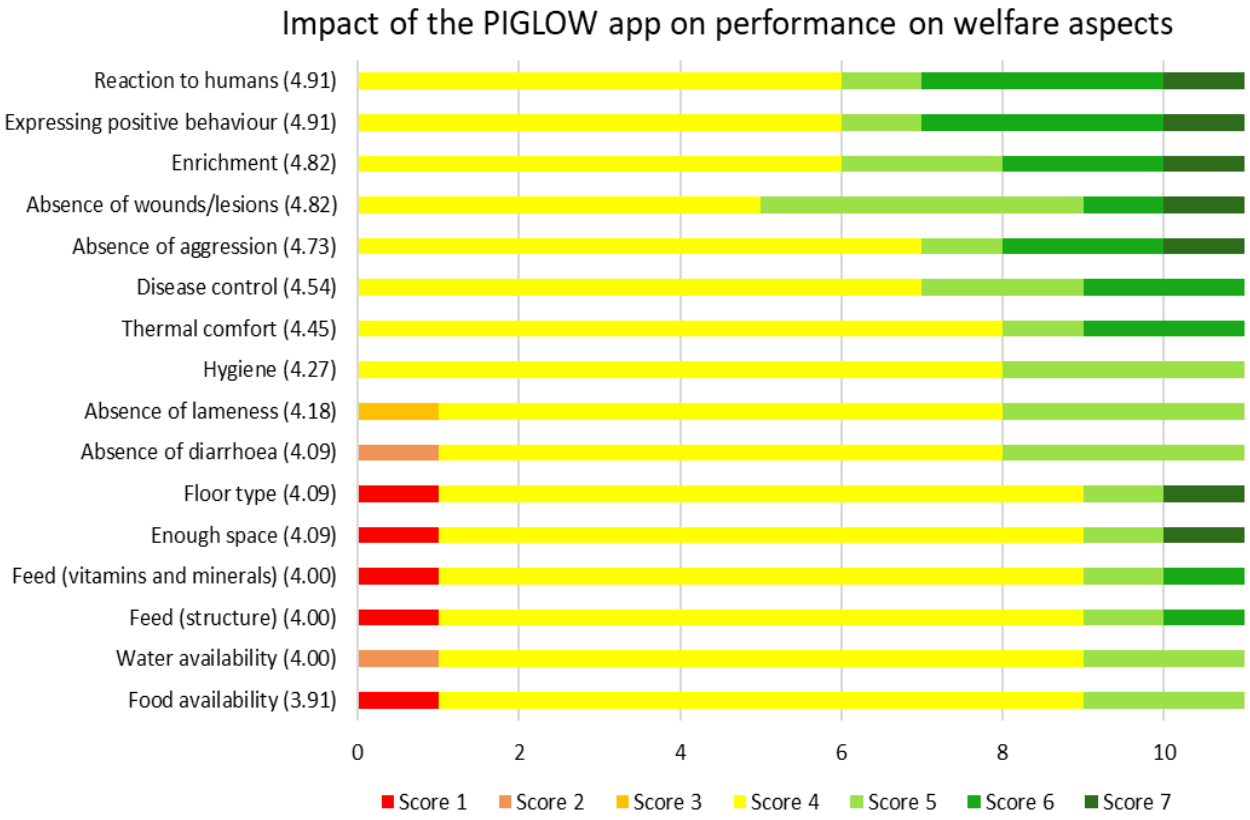

Figure 19: Farmers' opinion on how the use of the PIGLOW app influenced their own performance on 16 aspects of pig welfare rated between 1 (severely deteriorated) and 7 (greatly improved). The welfare aspects are ranked from the highest to the lowest mean rating, which is displayed between brackets behind the name of each welfare aspect, and all individual scores are shown. N = 11.

Farmers were also asked to give an overall score for whether they thought the use of the PIGLOW app had a positive effect on the welfare of their animals, rated from 1 (not at all) to 7 (absolutely). Therefore, any score higher than 1 indicated that the farmers noticed an effect. The average score of 3.82 (SD = 1.60) shows that farmers perceived a small to medium-sized effect of the use of the PIGLOW app on the overall welfare of their pigs. As indicated by the large standard deviation, the answers differed largely between farmers. Several farmers included a comment to explain their answer:

1. *"More reflection on pig behaviour" (5)*
2. *"It has made me look at the animals slightly differently, but we were already very focused on" (2)*
3. *"We regularly have young people walking around here, and I find that the app has added value for them. Now they know what they should look at." (4)*

The first farmer, who gave a relatively high score, mentioned that it made him focus more on the behaviour of his pigs, which fits with the relatively high mean scores for the effect that the PIGLOW app had on several behavioural welfare aspects. Another farmer mentioned that the effect of the PIGLOW app was relatively small, because animal welfare was already important to them. This was similar to some of the comments for the previous question that showed that perhaps the sample of this study was biased towards farmers who already have a lot of knowledge on animal welfare. The third quote actually confirmed that for young people, the app can be a useful tool to teach them what you should look at when assessing the welfare of pigs. Personal conversations with this farmer revealed that these people are mostly students of animal/veterinary sciences. This seems to confirm

that the PIGLOW app can provide benefits for people who are less knowledgeable about animal welfare.

Farmers were also asked what they thought about the PIGLOW app in general. Figure 20 shows the ratings of the farmers for how easy they found it to use different elements of the PIGLOW app on a scale from 1 (very difficult) to 7 (very easy). For this question, farmers were allowed not to answer if they did not think they used an aspect of the app well enough to form an opinion. Therefore, the number of answers that was given indicated something about the extent to which certain parts of the app were used. With a mean rating of 6.73 (SD = 0.47) for “General use of the app”, the farmers clearly found the PIGLOW app very user-friendly. The mean ratings for individual elements ranged from 6.13 to 6.55, indicating that the farmers found all elements user-friendly. However, it is noticeable that none of the elements of the results were scored by everyone, indicating that some farmers did not use some of the elements. The data showed that all elements of the results of the WAss were slightly less easy to understand than the content of the WAss themselves. However, as all mean ratings were very high, this does not seem cause for concern.

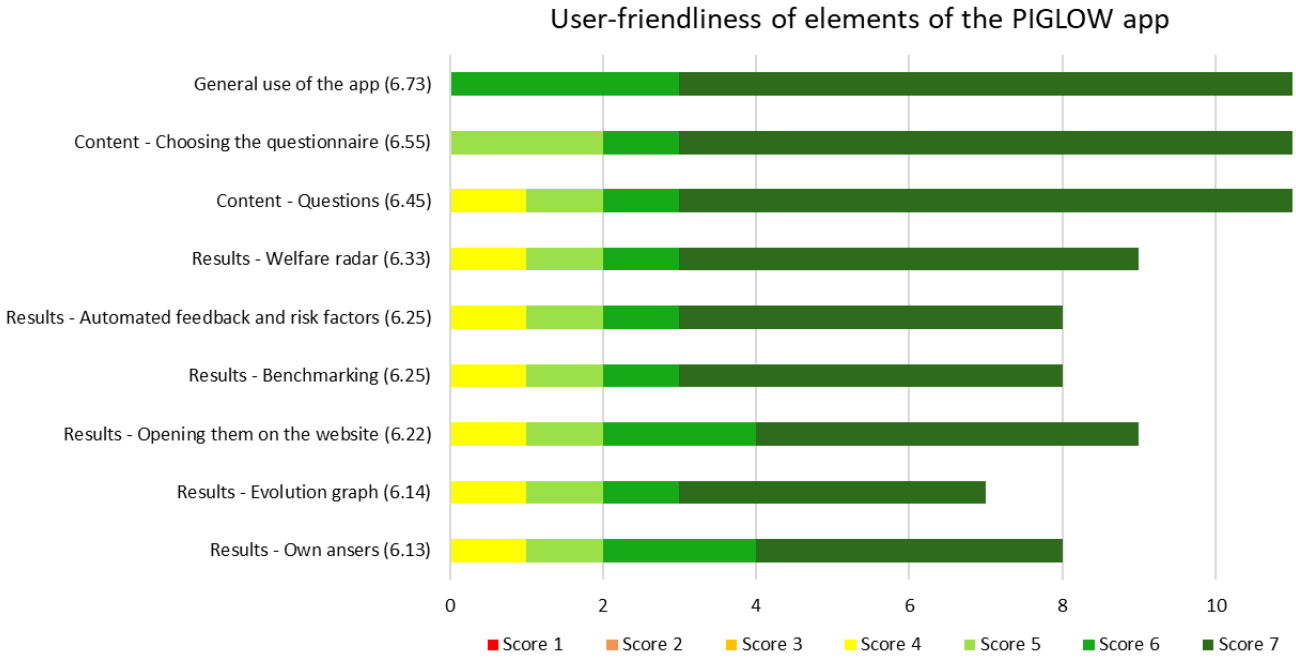

Figure 20: Mean ratings for how easy it was to use different elements of the PIGLOW app on a scale from 1 (very difficult) to 7 (very easy). The elements are ranked from the highest to the lowest mean rating, which is shown in brackets behind the name of each element, and all individual scores are shown. N = 11.

The mean ratings for the usefulness of the different elements of the PIGLOW app are shown in figure 21. The mean ratings were all between 5.63 and 6.70, indicating that the farmers found all aspects of the PIGLOW app very useful. Keeping in mind that the mean ratings from the same farmers for the effect of the PIGLOW app on their opinions on animal welfare and on the welfare of their animals were 4.00 and 3.82 out of 7, it does seem as if the farmers were thinking differently when they answered this question about the usefulness of different elements of the app. It is possible that the farmers were not just thinking about how much they benefited from each part of the app, but also about how useful they thought it could be for farmers in general and that this is why their answers were more positive. This would indicate that they saw the potential of the PIGLOW app for others. It is noticeable that for some elements of the results, more farmers answered the question about the

user-friendliness than about the usefulness. For “Benchmarking” there were 8 and 6 answers and for “Welfare radar”, there were 9 and 8 answers. This could mean that some farmers looked at those elements once or twice, but did not find them useful enough to keep using them and did not feel like they could provide their opinion. Farmers found “Additional information (i) for questions” (6.70 ± 0.48) the most useful and “Photo material for questions” (5.62 ± 1.51) the least useful. The additional information for questions in text form in the app are sometimes combined with photos of examples of welfare indicators. That farmers found the first one very useful and the second one less so could indicate that farmers did not need the photos, since they were familiar enough with the welfare problems. Other possible explanations are that the photo material could be improved (by adding more photos or increasing the quality) or that farmers simply did not like to look at photos on a small smartphone screen. Comments from the farmers did not provide extra information about this, so it could be useful to follow up on this with some farmers to see if the PIGLOW app could be improved or not. In line with previous answers showing that farmers thought certain behavioural welfare aspects have improved the most on their farm, the average rating for the usefulness of “Questions on behaviour” (6.64 ± 0.81) was higher than those for “Questions on health” (6.55 ± 0.82) and “Questions on housing” (6.55 ± 0.82), which were the other large categories of questions. Among the elements of the results, “Keeping track of your previous data” (6.43 ± 0.53) was rated highest. This suggests that farmers were more interested in the data of their animals that they gather themselves than they were in getting “help” from others through the means of automated feedback or benchmarking.

The answers to both questions about elements of the PIGLOW app show that not everyone used all elements of the app. This could be because farmers were not interested in those elements, but it is also possible that they did not realise what the possibilities of the app were. This is indicated by a comment that was added by 1 farmer: “*I didn't take the time to look for it and didn’t know at this stage that it was possible to compare ourselves to other farms. That might be the most interesting aspect”.* Although the results of the welfare assessment were explained by the researcher during the first farm visit and the farmers also received a document with an explanation of each aspect of the results, it seems that some farmers did not pay attention well enough to remember everything. Even though this farmer mentioned afterwards that benchmarking could have been very interesting, he was apparently not motivated enough to take a closer look at everything himself. This means that perhaps for farmers to fully benefit from the app, more interaction with people who are experienced in using the app would be useful.

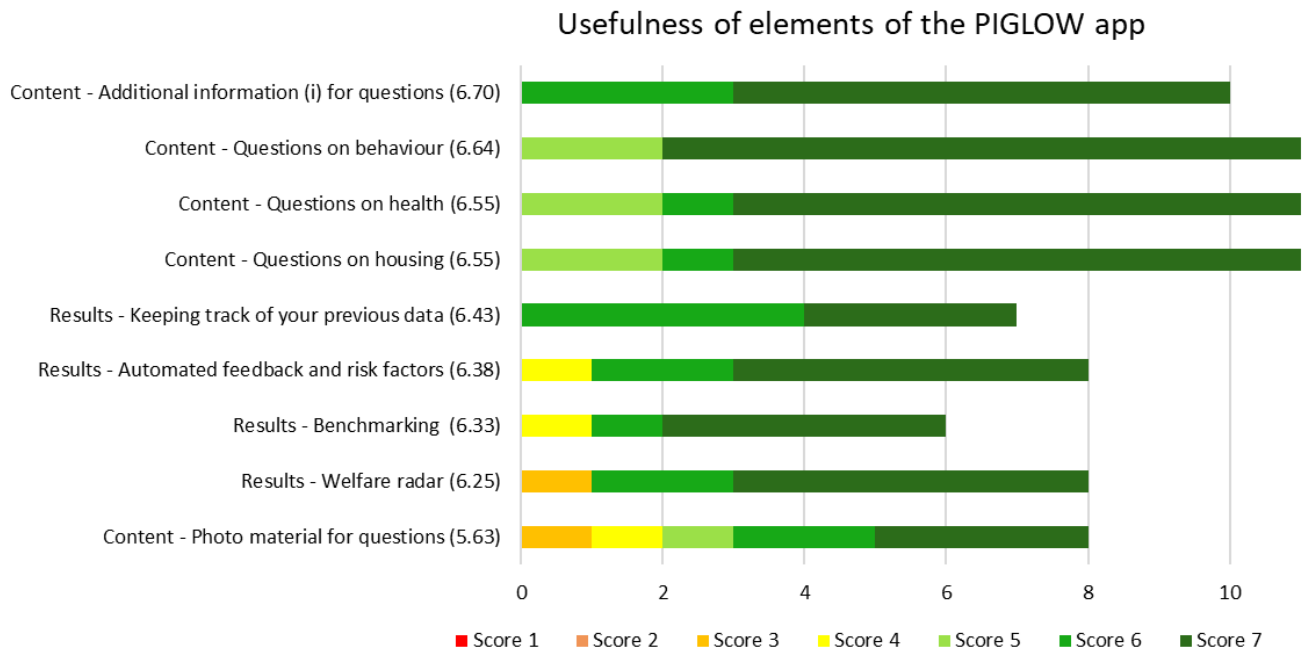

Figure 21: Mean ratings for how useful the farmers found each element of the PIGLOW app on a scale from 1 (not useful at all) to 7 (very useful). The elements are ranked from the highest to the lowest mean rating, which is shown in brackets behind the name of each element, and all individual scores are shown. N = 11.

Farmers were also asked how they would rate the PIGLOW app on a scale from 1 (worst) to 10 (best). The mean rating of an 8.09 (SD = 1.38) means that the farmers thought of PIGLOW as a good app, which fits with the previous results showing that they found all aspects of the app easy to use and quite useful. For some of the farmers, who gave above average scores for the effect of the app on their opinions on animal welfare and the welfare of their animals, this score likely reflects that they experienced benefits of the use of the app themselves. However, on average the farmers only saw a small to medium-sized effect of the PIGLOW app on how they rated the importance of animal welfare aspects and on the welfare of their own animals. Therefore, it does seem likely that at least some of the farmers gave this high score partially because they saw the potential of the app for others and not necessarily because they experienced large benefits from using the app themselves. This was also evident from the following comment from a farmer: *"I found the app very easy and user friendly. I think it is very suitable for a farmer who wants to improve welfare" (9).* This quote, from a farmer who gave a score of 9/10 for the PIGLOW app in general, was given by the same farmer who gave a score of 2/7 for the impact of the PIGLOW app on the importance of different aspects of welfare and for the impact on the welfare of his animals. Previously discussed comments of this farmer and others seemed to indicate that among the participants of this study, many of the farmers were already so focused on animal welfare beforehand that the PIGLOW app did not have much left to teach them. This is corroborated by the relatively good welfare scores for all aspects of welfare included in the detailed WAss, showing that there were no large welfare problems on the majority of these farms. The fact that these farmers already had a lot of knowledge on animal welfare and provided good welfare for their animals possibly hindered the goal of our study to find out whether the PIGLOW app can help farmers improve animal welfare. A comment from another farmer in response to this final question offered a suggestion for a more specific target audience: *"I think this app has added value for new comers in the sector. They can learn how to look at an animal, what you should pay attention to, and how to see whether an animal feels comfortable" (5).* While this farmer had experience with students of animal sciences in particular, the same could be true for new farmers. To put the potential that the farmers from this study saw in the PIGLOW app to use

and help improve the welfare of organic and free-range pigs, it could thus be useful to target farmers and possibly other professionals or students that are new to the sector.

Finally, farmers were asked whether they had any suggestions for improvements of the PIGLOW app. As for previous questions, the answers differed among farmers:

1. *"None, it's very good like this" (9)*
2. *"I would maybe give some more practical tips in the apps, or examples from real farms" (9)*
3. *"More depth and better feedback on the results" (7)*
4. *"Sometimes a little bit too precise" (9)*

As became clear from the answers to some previous survey questions, some farmers were very happy with the PIGLOW app as it is, e.g. the farmer from quote 1. However, quote 2 and 3 from farmers who previously indicated that they were already focussed on animal welfare before the start of the study showed that for farmers who are already very knowledgeable on the subject, the feedback offered in the app would have to contain more depth to be truly helpful. It is worth exploring the options for these adaptations to the PIGLOW app in the future in the hope of helping more farmers improve the welfare of their animals. However, quote 4 indicates that for some farmers, apparently the app was already too precise. A risk of adding more depth to certain elements of the app would be that it would ask for a larger time investment of the farmers, which could also decrease motivation for some to use the app in the first place. Therefore, this compromise must be kept in mind when thinking about making changes to the PIGLOW app.

5.1.3. Comparison of PIGLOW assessments by farmers and researchers

The PIGLOW assessments of farmers and researchers at the beginning of the study were compared first. Subsequently, the differences between farmers and researchers at the beginning and end of the study were compared to one another. Table 11 shows the number of pigs and the number of groups that were observed with the PIGLOW app by each farmer and researcher couple.

Table 11: Number of pigs and groups that were observed with the PIGLOW app by each farmer/researcher couple at the beginning and end of the study. The order of the farms does not correspond with the order in earlier tables.

Farm	# Pigs observed		# Groups observed	
	Beginning	End	Beginning	End
1	54	47	3	2
2	56	59	2	2
3	52	46	4	4
4	48	56	4	5
5	26	19	2	3
6	67	82	2	3
7	60	58	4	3
8	35	63	1	2
9	45	61	4	4
10	107	115	4	3

11	13	12	2	1
12	82	74	2	2

In table 12, the differences between farmers and researchers for all 18 WI for which a comparison was possible are ranked from the highest to the lowest mean absolute difference. In general, WI at group level appeared relatively high in the ranking. This was related to the fact that fewer groups than individual pigs were observed and therefore a different answer for 1 group would have had a larger impact on the overall difference between the farmer and researcher than a different answer for 1 individual pig. Because of this difference, WI should be compared within their own category. Among group level WI, the differences between farmers and researchers were largest for “Water access” (mean abs. diff. = 37.5; 5 different scores), “Lying on flank” (mean abs. diff. = 35.42; 7 different scores) and “Coughing / sneezing” (mean abs. diff. = 14.58; 5 different scores). For the WI “Water access”, 5 out of 12 farmers and 0 researchers gave a better score than the other party in their couple. This was likely due to a difference in judgement by farmers and researchers of what constitutes “difficulty accessing water”. Researchers tended to judge access to water as difficult if the only drinking spots were located in a tight corner or in a busy area where pigs were sometimes pushed aside while trying to drink, whereas farmers might have only based their answers on the number of drinking spaces per group of pigs. For the WI “Lying on flank” there were 7 researchers who gave a better score than farmers. This was most likely related to a misunderstanding of the welfare indicator by farmers. The question clearly stated that one should only score this WI as positive if more than 50% of the pigs in a group were lying on their flank as well as widely spread out through the enclosure, as an indicator for thermal comfort. However, it seems likely based on the experiences of the researchers that several farmers scored this WI as positive if the pigs were lying on their flank, but close to one another. This indicates that the farmers might not have taken the time to read the questions clearly. For the WI “Coughing / sneezing”, 5 farmers and 0 researchers gave a better score. This could be related to the fact that for this WI, the PIGLOW user was asked at the end of a group observation to remember whether they heard any coughing or sneezing at any time during the observation. Perhaps the farmers, who were likely less experienced than the researchers with structured animal welfare assessments, did not remember what they had previously heard.

Among individual level WI, the differences were largest for “Enrichment use” (mean abs. diff. = 21.22; 11 different scores), “Scratches” (mean abs. diff. = 6.48; 8 different scores), and “Too small” (mean abs. diff. = 2.44; 6 different scores). For the WI “Enrichment use”, 7 farmers and 4 researchers gave a better score, while only 1 farmer/researcher couple gave the same score. A possible explanation for the fact that the scores of most couples differed, is that “Enrichment use” was measured through a behavioural scan taken of an entire group of pigs during which every pig using enrichment must be counted. As it can change from one second to another whether a pig is using enrichment or not, this could easily lead to two observers counting a different number of pigs. However, the data did show that overall, farmers were more likely than researchers to give the better score. This could indicate that farmers are more likely than researchers to still count pigs that have just stopped using enrichment or to count some pigs twice. In general, the fact that the scores of almost all couples differed is also an indication that this method is not suitable to get a precise result on enrichment use. However, the only way to increase precision would be to observe pigs for a longer time. That would not be suitable for the PIGLOW app, as asking for a larger time investment of the farmers could reduce their motivation to use the app at all. Because the main goal of the use of the PIGLOW app is to sensitise farmers to early signs of welfare problems, the fact that the score for enrichment use is not very precise should not stand in the way of reaching this goal. For the WI “Scratches”, 7 farmers and 1 researcher gave a better score, indicating that researchers scored more strictly in general. This is possibly related to the fact that seeing scratches, which are narrow and shallow, on the pigs’ skin requires close observation. On average, researchers are likely more experienced with

structured WAss than farmers, which could explain why they counted more scratches than farmers. However, as we do not know which scores are correct, another possible explanation is that researchers mistook dirt for scratches more often than farmers and therefore counted too many scratches. For the WI "Too small", 2 farmers and 4 researchers gave the better score, indicating that in general farmers were slightly stricter when scoring this. This might be related to the fact that the size of pigs is not only an important WI, but also an important economic parameter, which is more relevant for farmers than for researchers. Perhaps for this reason farmers are better than researchers at spotting pigs that are too small. However, it is also possible that because size is an important economic aspect, farmers scored certain pigs as small even though they did not meet the criterion that was stated in the PIGLOW app, which is that a pig should be at least 1/3 smaller than the average for the group.

For the remaining WI, the mean absolute differences were quite small and the number of different scores combined with the number of cases in which either the farmer or researcher gave the better score showed no clear patterns of one group scoring differently than the other. This could be partially related to the fact that in general, the occurrence of welfare problems for most indicators was quite low. It seems likely that the larger the number of pigs with different welfare problems, the more potential for disagreement between observers there would be. Overall, the data support that the judgement and observation methods of farmers and researchers were likely to be influenced by their different experiences and that the measurement of certain WI perhaps requires more experience than others. It also became clear that some WI, such as "Enrichment use" were perhaps more difficult to measure precisely within a short time window.

Table 12: Difference between farmers and researchers in the assessment of welfare indicators for finishing pigs with the PIGLOW app at the beginning of the longitudinal study. The median score given by farmers and researchers, the mean absolute difference between all farmer/researcher couples, the number of farmer/researcher couples that gave different scores for each indicator and the number of times that each party gave the better score are shown. The indicators are shown in order from the highest to lowest mean absolute difference. N=12.

Welfare indicator	Median	Mean absolute difference	# Different scores	Better score
Water access	0	37.5	5	5F
Lying on flank	0	35.42	7	7R
Enrichment use	31.37	21.22	11	7F, 4R
Coughing / sneezing	0	14.58	5	5F
Scratches	0	6.48	8	7F, 1R
Liquid Faeces	0	6.25	2	1F, 1R
Too small	0.75	2.44	6	2F, 4R
Covered with faeces	0	2.25	5	2F, 3R
Skin wounds	0	1.68	5	2F, 3R
Lameness	0	1.30	3	2F, 1R
Tail lesions	0	1.26	4	3F, 1R
Laboured breathing	0	0.93	3	2F, 1R
Ear lesions	0	0.54	3	2F, 1R

Bad general state	0	0.43	2	2F
Panting	0	0.40	2	2F
Skin irritation	0	0.20	1	1F
Shivering	0	0.16	1	1R
Huddling	0	0	0	-

The comparison of differences for PIGLOW assessments by farmers and researchers at the beginning and end of the study are shown in table 13. WI for which the change between the beginning and the end was largest were “Water access”, “Enrichment use”, “Scratches”, Liquid faeces” and “Huddling”. For “Water access”, “Enrichment use” and “Scratches” the difference between farmers and researchers was smaller at the end of the study than at the beginning. This indicates that the two groups grew closer in their judgement of these indicators, perhaps because one or both groups gained experience with the assessment. However, because it is not known which welfare scores are correct, it cannot be said whether a smaller difference between groups means that either group has scored more accurately. For “Liquid faeces” and “Huddling”, the difference between farmers and researchers was larger at the end of the study. However, for both of these indicators the difference is caused by only 3 out of 12 couples. Because both indicators are measured at group level, very few different answers to individual questions from those 3 couples were necessary to cause a large difference in the total mean absolute difference. Thus, the increase in the difference between farmers and researchers at the end of the study likely does not indicate a meaningful change in the assessment of those aspects of welfare.

The mean number of couples with different scores for each welfare indicator was lower at the end of the study (3.82) than at the beginning of the study (4.54). However, while for 9 out of 19 WI there were fewer couples with a different score at the beginning of the study, there were also 7 welfare indicators for which there were more couples with a different score at the end of the study. Thus, although the mean number of different scores has decreased, overall, there did not seem to be a clear pattern of improved agreement between farmers and researchers at the end of the study. This conclusion is supported by the fact that the means of the mean absolute difference for all welfare indicators at the beginning (8.08) and end (8.87) were very similar. Furthermore, the mean absolute difference was largest at the beginning for 9 indicators, but largest at the end for the other 9 indicators.

Table 13: Difference between farmers and researchers in the assessment of welfare indicators for finishing pigs with the PIGLOW app, compared for the beginning and end of the longitudinal study. N = 12.

Welfare indicator	Median		Mean absolute difference		# Different scores		Better score	
	Beginning	End	Beginning	End	Beginning	End	Beginning	End
Water access	0	0	37.50	9.17	5	2	5R	2F
Lying on flank	0	16.67	35.42	36.11	7	5	7F	5R
Enrichment use	31.37	0	21.22	13.91	11	12	7F, 4R	8F, 4R
Coughing / sneezing	0	0	14.58	15.97	5	4	5R	1F, 3R

Scratches	0	23.68	6.48	1.17	8	4	1F, 7R	4F
Liquid Faeces	0	0	6.25	14.58	2	3	1F, 1R	3R
Too small	0.75	1.73	2.44	2.56	6	8	4F, 2R	1F, 7R
Covered with faeces	0	0	2.25	0.83	5	4	3F, 2R	2F, 2R
Skin wounds	0	0	1.68	1.63	5	5	3F, 2R	4F, 1R
Lameness	0	0	1.30	0.28	3	1	1F, 2R	1R
Tail lesions	0	0	1.26	0.07	4	1	1F, 3R	1R
Laboured breathing	0	0	0.93	2.35	3	4	1F, 2R	3F, 1R
Ear lesions	0	0	0.54	0.76	3	3	1F, 2R	1F, 2R
Bad general state	0	0	0.43	0.47	2	3	2R	2F, 1R
Panting	0	0	0.40	0	2	0	2R	-
Skin irritation	0	0	0.20	0.58	1	2	1R	1F, 1R
Shivering	0	0	0.16	0	1	0	1F	-
Huddling	0	0	0	13.19	0	3	-	3R
Mean			**8.08**	**8.87**	**4.54**	**3.82**		

5.2. Poultry

5.2.1. Comparison of animal welfare at the beginning and end of the study

To determine whether animal welfare on the participating poultry farms has changed during the study, the scores for each of the 18 WI assessed in the detailed WAss by researchers were compared. The number of individual chickens and the number of groups included in those WAss on each of the farms are shown in table 13. The total number of WAss with the EBENE® app that were performed by each farmer is also shown in the table. Farmers were asked to conduct 5 WAss.

Table 13: Number of individual chickens and groups observed on each of the poultry farms at the beginning and end of the study and the total number of welfare assessments broiler chickens each farmer performed with the EBENE® app. The order of the farms does not correspond with the order in earlier tables.

Farm	# Individual chickens observed		# Groups observed		# EBENE® assessments by the farmer
	Beginning	End	Beginning	End	
1	19	9	5	5	3
2	40	40	5	5	4
3	51	50	5	4	5
4	50	50	4	5	4
5	51	50	5	4	2
6	55	50	5	5	4
7	50	50	5	5	2
8	0	30	3	5	2
9	0	30	3	5	2

The comparison of scores for WI at the beginning and end of the study can be found in table 14. For 5 of the 11 assessed WI that are negative, meaning a low score is better, the number of incidences of any welfare problem was so low at the beginning of the study that the median was 0 (Still, Dirtiness, Other anomalies, Lameness, Panting). This means that the majority of the participating farms already had a perfect score for those WI at the beginning of the study and that there was little room for improvement for that WI. For 4 of those WI, the median was still 0 at the end of the study, showing that those aspects of welfare did not change substantially. For the 5th indicator, “Lameness” (median B = 0; median E = 2.40; figure 22), the score has worsened for the majority of farms and the median has increased, showing that this WI has deteriorated. There were no positive WI for which median B was equal to the perfect score, meaning that for the remaining 6 negative and all 7 positive WI, there was some potential for improvement. However, for 1 of those positive WI, “Dust bathing’, median B and E were both 0, indicating that no clear change for this WI took place during the study.

For 6 of the 18 WI the number of farms for which the score has improved or worsened was the same or only differed by 1 digit, showing that how these indicators changed over time was different per farm and that there was no clear pattern of those aspects of welfare improving or worsening during the study. Thus, even if the medians for one of these WI differed between the beginning and end, the evidence for a consistent change across farms for these aspects of welfare during the study was less compelling. For 2 of those 6 indicators (Still, Other anomalies) medians B and E were both 0, confirming that there was no clear overall change for those WI. For the WI “Footpad dermatitis” (median B = 23.42; median E = 36.20; figure 23), median E was much higher, suggesting that even though the scores have improved or worsened on almost the same number of farms, this aspect of welfare has deteriorated overall. For 2 of the remaining WI ”Preening” (median B = 0.22; median E = 0.24; figure 24) and “Positive interaction” (median B = 0.05; median E = 0.04; figure 25), there was a very small improvement of the median. However, combined with the fact that scores improved and worsened for almost the same number of farms, the change of these medians is too small to

support that there was any consistent change for these WI. For the last of the 6 WI, “Foraging” (median B = 0.38; median E = 0.54; figure 26), there was a larger increase in the median, which suggests a small improvement of this WI during the study. Combined with the knowledge that the score improved or worsened for almost the same number of farms, the data only partially supported that this WI changed substantially.

There are 9 WI for which medians B and E differed and there was a pattern of scores on the farms predominantly changing in the same direction, supporting a general pattern of change. However, for 4 of those WI the direction of change of the median and of the scores of individual farms suggested that these aspects of welfare worsened rather than improved during the study. These indicators were “Hock burn” (median B = 4.14; median E = 11.90; figure 27), the previously mentioned “Lameness” (median B = 0; median E = 2.4), “Resting” (median B = 4.06; median E = 3.00; figure 28) and “Distance from humans” (median B = 2.25; median E = 3.00; figure 29) with worsened scores on the majority of farms. For the WI “Aggressive pecking” (median B = 0.008; median E = 0.007; figure 30) there were more improved than worsened scores, but the difference in the medians was too small to indicate a clear change for this WI. For the final 4 WI both the direction of change of the median and of the scores of individual farms suggested that these aspects of welfare improved. These were the WI “Too small” (median B = 5.88; median E = 2.00; figure 31), “Wounds” (median B = 0.88; median E = 0; figure 32), “Stretching / wing flapping” (median B = 0.26; median E = 0.49; figure 33) and “Enrichment use” (median B = 0; median E = 2.44; figure 34), all with improved scores for the majority of farms.

Overall, the data from the detailed WAss for broiler chickens showed that there has been a clear improvement for 4 of the 18 measured WI during the longitudinal study, but that there are also 4 other WI that have clearly deteriorated. In addition, there was one more WI for which the data partially supported an improvement and one more WI for which the data partially supported a deterioration. Thus, the data did not provide convincing evidence that the use of the EBENE® app by farmers had a positive effect on the overall welfare of broiler chickens. It must be said that, as table 13 shows, many of the farmers did not complete as many welfare assessments with the EBENE® app as they were asked to. Even though they were reminded to perform welfare assessments, it was difficult for many farmers to find time for this in the 2 weeks pre slaughter. This was partially due to issues caused by AI in the regions where these farmers were located. Additionally, due to avian influenza some farmers had no chickens on their farm for longer periods of time than usual. The fact that some farmers did not use the EBENE® app as often as was asked made it more difficult to see if the use of the app can have a positive effect on animal welfare if it is used as intended. Looking at the specific WI for which the scores were worse at the end of the study, it is noticeable that all indicators related to leg health (Footpad dermatitis, Hock burn, and Lameness) were among them. This could be related to the weather and climate, as the year in which the final farm visits took place was very wet (Torfs, 2023). This could lead to more damp litter, which is an important cause of leg problems (Mayne, 2005). At the beginning of the study, 8 out of 9 farm visits took place between June and August, while at the end of the study, 6 out of 9 visits took place between October and January. This could unfortunately not be avoided due to problems on some of the farms that prevented us from visiting during the summer. However, partially because of this the birds were confined indoors on a much larger number of farms at the end of the study, compared to the beginning (8 instead of 2). Outdoor range use has been shown to have a positive effect on leg health (Aguado et al., 2015; Stadig et al., 2017) and thus confinement (and its effect on litter quality as well as opportunities for exercise) could have been a factor in the deterioration of leg health. For the welfare aspects that have improved during the study, no clear pattern was visible. Some of the aspects related to physical welfare problems, whereas others related to animal behaviour. Whether any of the differences

between the beginning and end of the study were related to the use of the EBENE® app cannot be concluded based on this data alone. The results of the surveys in which the farmers give their own opinion on the app can give more information on this topic.

Table 14: Descriptive statistics for the differences in animal welfare on the poultry farms at the beginning and end of the longitudinal study. The first and second column show the median of the average scores per farm for each welfare indicator for the beginning and end, with the number of scores that were above 0 between brackets. The third and fourth column show the number of farms on which the score for each indicator decreased or increased, respectively, during the study. N = 9.

Welfare indicator	Median beginning (# Scores above 0)	Median end (# Scores above 0)	Farms with improved scores	Farms with worsened scores
Still	0 (1)	0 (1)	1	1
Too small	5.88 (5)	2 (4)	4	1
Dirtiness	0 (3)	0 (0)	3	0
Wounds	0.88 (5)	0 (1)	5	0
Footpad dermatitis	23.42 (7)	36.20 (7)	3	4
Hock burn	4.14 (7)	11.90 (7)	2	5
Lameness	0 (0)	2.40 (5)	0	5
Other anomalies	0 (2)	0 (3)	2	1
Dust bathing	0 (1)	0 (4)	3	1
Preening	0.22 (9)	0.24 (9)	5	4
Foraging	0.38 (9)	0.54 (9)	5	4
Stretching / wing flapping	0.26 (9)	0.49 (9)	7	2
Aggressive pecking	0.01 (6)	0.01 (6)	6	3
Positive interaction	0.05 (8)	0.04 (9)	4	5
Enrichment use	0 (3)	2.44 (5)	5	1
Panting	0 (4)	0 (1)	7	2
Resting	4.06 (9)	3.00 (7)	2	7
Distance from humans	2.25 (9)	3.00 (9)	1	4

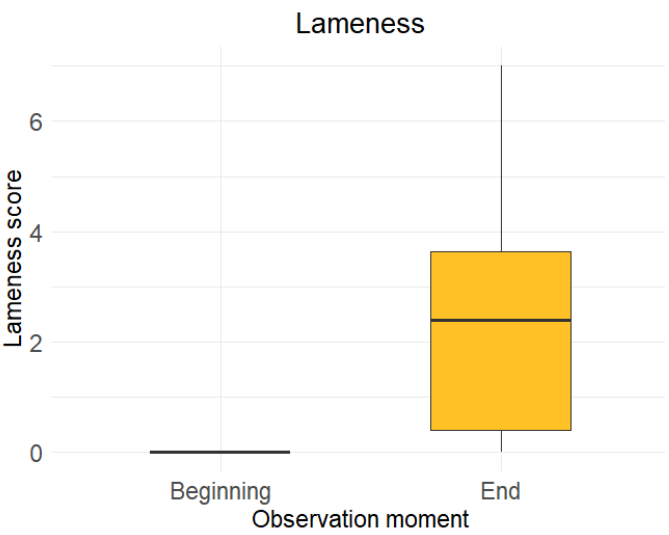

Figure 22: Boxplot showing the differences in the scores for the welfare indicator "Lameness" for broiler chickens at the beginning and end of the longitudinal study. The score for "Lameness" is the mean score from 0 (best) to 100 (worst) for all observed chickens. The line in each box represents the median.

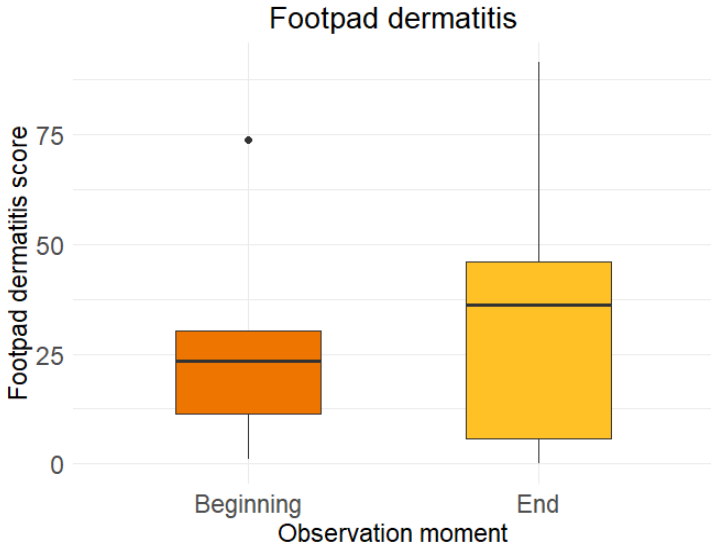

Figure 23: Boxplot showing the differences in the scores for the welfare indicator "Footpad dermatitis" for broiler chickens at the beginning and end of the longitudinal study. The score for "Footpad dermatitis" is the mean score from 0 (best) to 100 (worst) for all observed chickens. The line in each box represents the median.

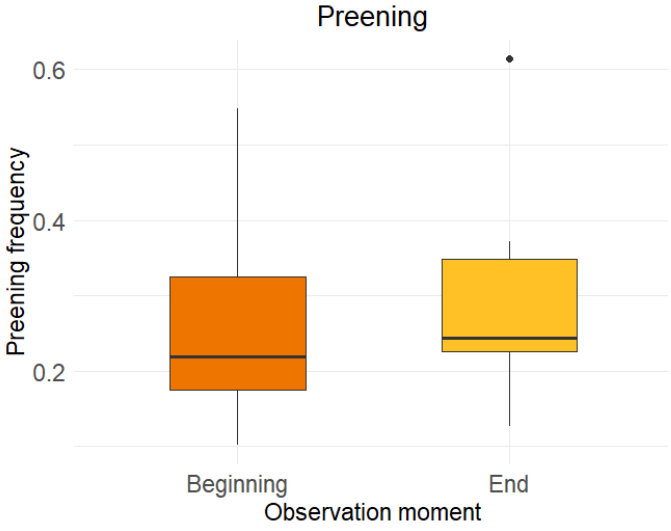

Figure 24: Boxplot showing the differences in the scores for the welfare indicator "Preening" for broiler chickens at the beginning and end of the longitudinal study. The score for "Preening" is the mean frequency of this behaviour per chicken during a 5-minute behavioural observation. The line in each box represents the median.

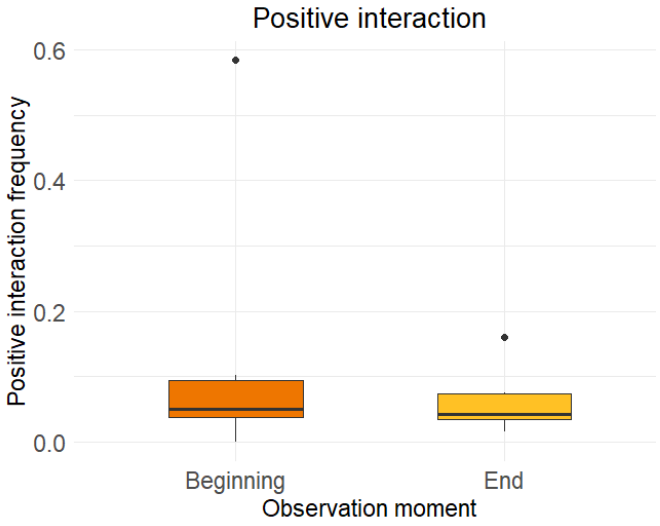

Figure 25: Boxplot showing the differences in the scores for the welfare indicator "Positive interaction" for broiler chickens at the beginning and end of the longitudinal study. The score for "Positive interaction" is the mean frequency of this behaviour per chicken during a 5-minute behavioural observation. The line in each box represents the median.

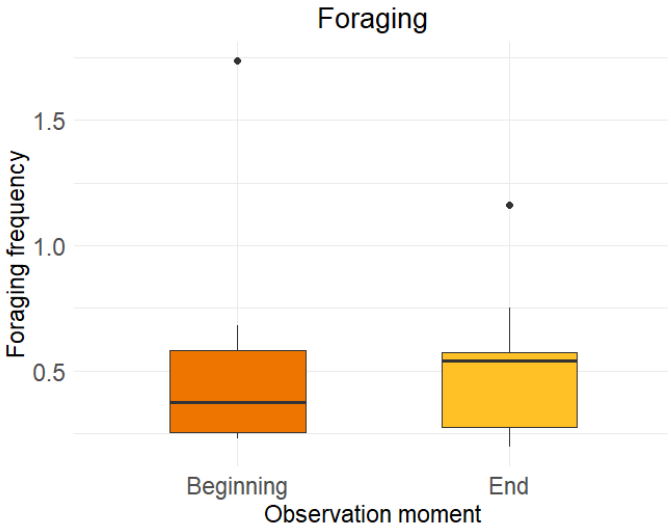

Figure 26: Boxplot showing the differences in the scores for the welfare indicator “Foraging” for broiler chickens at the beginning and end of the longitudinal study. The score for “Foraging” is the mean frequency of this behaviour per chicken during a 5-minute behavioural observation. The line in each box represents the median.

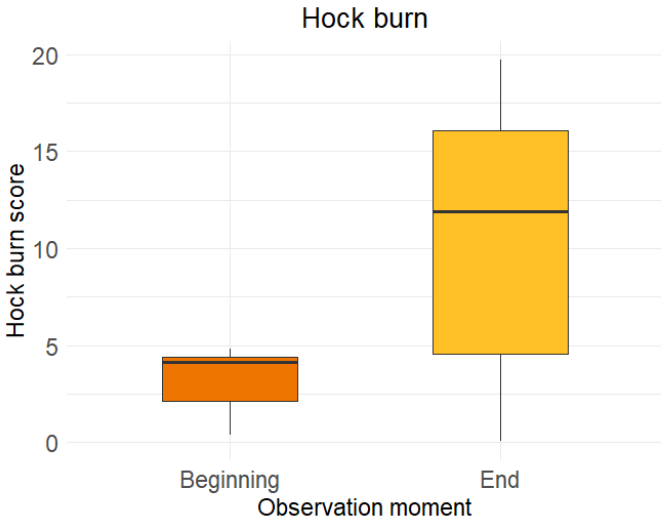

Figure 27: Boxplot showing the differences in the scores for the welfare indicator “Hock burn” for broiler chickens at the beginning and end of the longitudinal study. The score for “Hock burn” is the mean score from 0 (best) to 100 (worst) for all observed chickens. The line in each box represents the median.

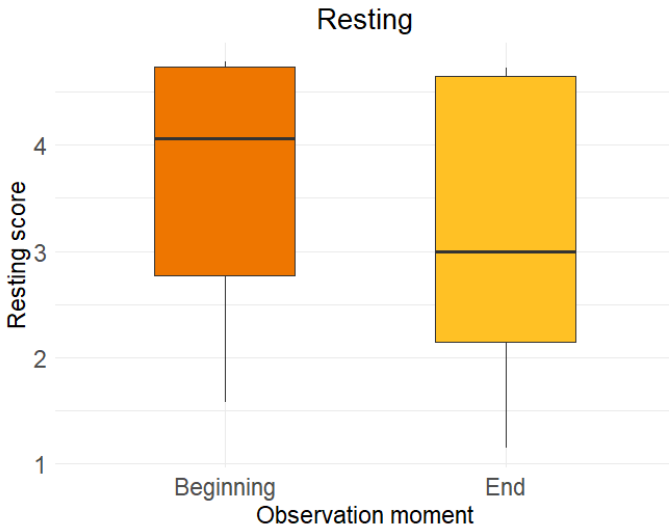

Figure 28: Boxplot showing the differences in the scores for the welfare indicator “Resting” for broiler chickens at the beginning and end of the longitudinal study. The score for “Resting” is a score from 0 (worst) to 5 (best) for the percentage of birds that is resting. The line in each box represents the median.

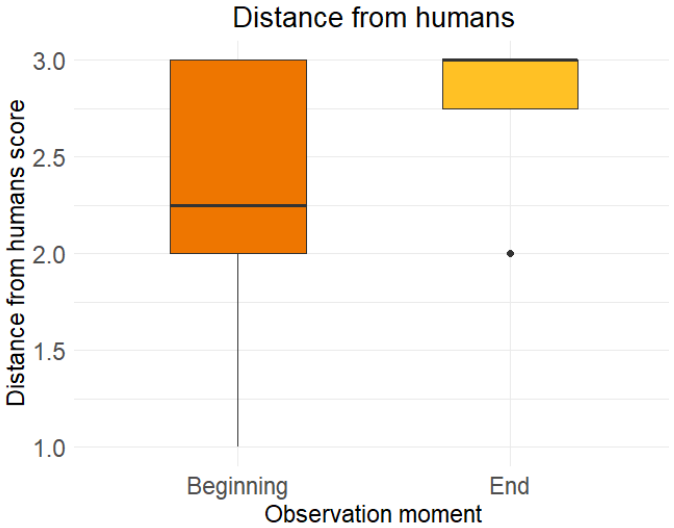

Figure 29: Boxplot showing the differences in the scores for the welfare indicator “Distance from humans” for broiler chickens at the beginning and end of the longitudinal study. The score for “Distance from humans” is a score from 0 (best) to 3 (worst) for the distance from humans at the end of a behavioural observation. The line in each box represents the median.

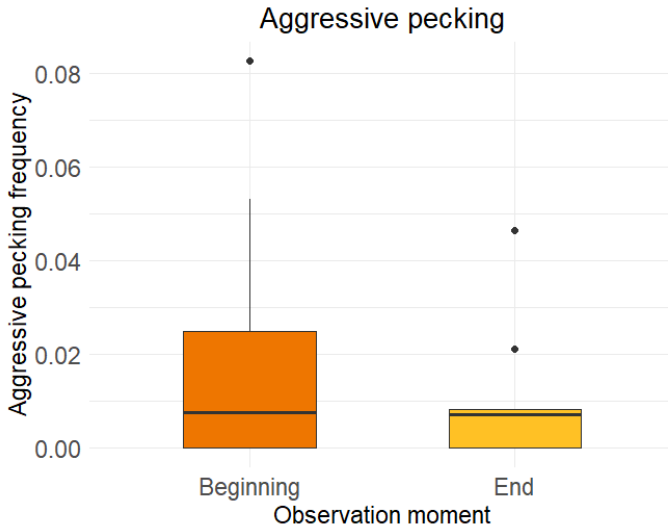

Figure 30: Boxplot showing the differences in the scores for the welfare indicator "Aggressive pecking" for broiler chickens at the beginning and end of the longitudinal study. The score for "Aggressive pecking" is the mean frequency of this behaviour per chicken during a 5-minute behavioural observation. The line in each box represents the median.

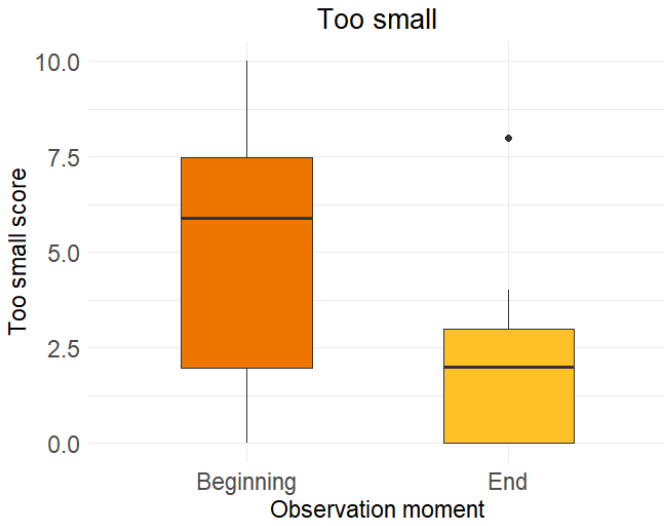

Figure 31: Boxplot showing the differences in the scores for the welfare indicator "Small" for broiler chickens at the beginning and end of the longitudinal study. The score for "Small" is the percentage of pigs that was scored as too small. The line in each box represents the median.

Figure 32: Boxplot showing the differences in the scores for the welfare indicator "Wounds" for broiler chickens at the beginning and end of the longitudinal study. The score for "Wounds" is the mean score from 0 (best) to 100 (worst) for all observed chickens. The line in each box represents the median.

Figure 33: Boxplot showing the differences in the scores for the welfare indicator "Stretching / wing flapping" for broiler chickens at the beginning and end of the longitudinal study. The score for "Stretching / wing flapping" is the mean frequency of this behaviour per chicken during a 5-minute behavioural observation.

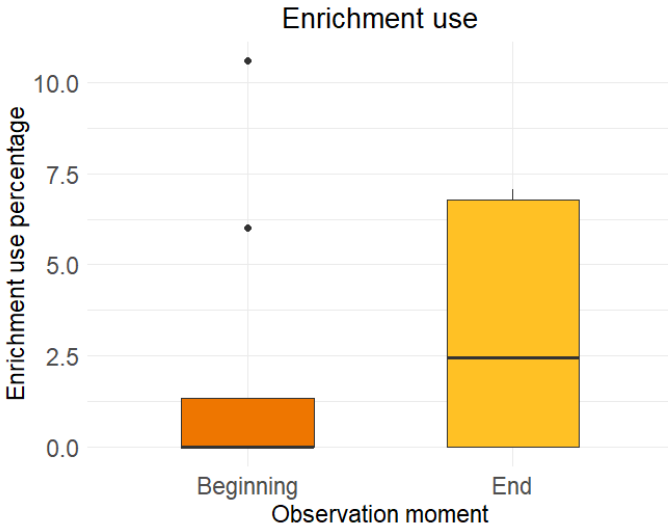

Figure 34: Boxplot showing the differences in the scores for the welfare indicator "Enrichment use" for broiler chickens at the beginning and end of the longitudinal study. The score for "Enrichment use" is the percentage of birds in the group using enrichment. The line in each box represents the median.

5.2.2. Surveys

In the section on surveys, the terminology "welfare aspect" rather than "welfare indicator (WI)" is used. This is a deliberate choice, as the welfare aspects that are questioned in the surveys often encompass multiple WI.

First survey

The results of how important the 9 poultry farmers judged each of the 15 aspects of pig welfare to be on a scale from 1 (not important at all) to 7 (very important) at the beginning of the study can be found in figure 35. The mean scores for all welfare aspects range from 5.44 to 6.78, indicating that all welfare aspects were relatively important to the farmers. As this concerns a group of farmers who volunteered to participate in a study on animal welfare, this was not surprising. The highest mean ratings were for the welfare aspects "Water availability" (6.78 ± 0.67; mean ± SD), "Food availability" (6.56 ± 0.73), "Enough space" (6.56 ± 0.73) and "Absence of dirty birds" (6.44 ± 0.88). Because water and food are basic necessities, it is not surprising that all farmers found these aspects very important. As mentioned previously for pigs, increased space allowance is an important characteristic of organic and other outdoor farms. Thus, it is logical that a group of farmers who chose an organic or free-range farm would value this aspect of welfare. For the high rating for the importance of "Absence of dirty birds", there was no immediate obvious explanation. The scores from the detailed WAss by researchers on these farms did not show that the farmers had a particular problem with this aspect of welfare. At the bottom of the ranking were the welfare aspects "Dustbathing" (5.44 ± 1.81), "Slaughter method on-farm" (5.78 ± 1.30), "Enrichment" (5.78 ± 1.39) and "Reaction to humans" (5.89 ± 1.27). It is noticeable that all these aspects were related to animal and/or human behaviour, suggesting that the farmers found behaviour slightly less important than more resource-based and health related aspects of animal welfare. However, it must be noted that "Foraging behaviour" (6.33 ± 0.71) and "Absence of aggression" (6.33 ± 0.87) were rated a bit higher and that even the lowest mean ratings were still relatively high on the scale from 1 to 7.

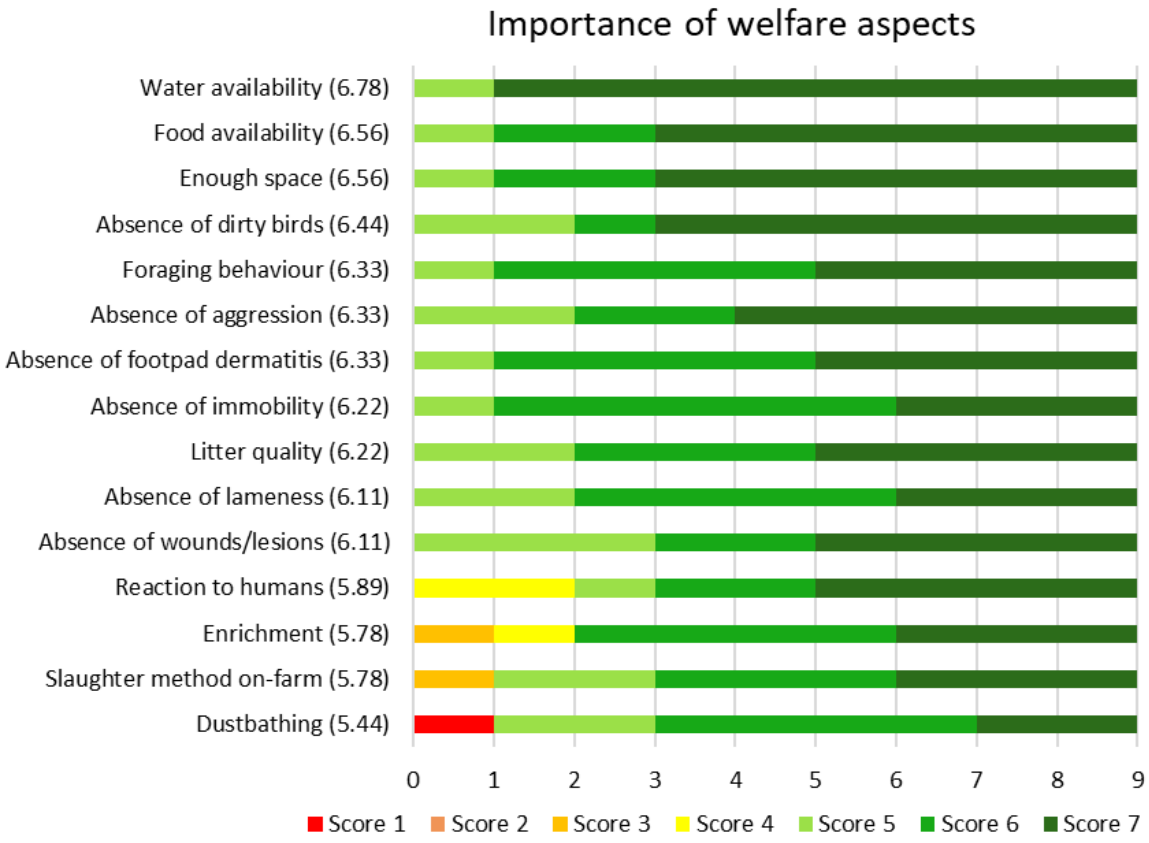

Figure 35: Importance of 15 aspects of poultry welfare rated between 1 (not important at all) and 7 (very important) by poultry farmers. The welfare aspects are ranked from the highest to the lowest average rating, which is displayed between brackets behind the name of each welfare aspect, and all individual scores are shown. N = 9.

Figure 36 shows the mean scores given by farmers for the estimation of their own performance for 16 aspects of poultry welfare on a scale from 1 (very badly) to 7 (very well). All mean scores were between 5.22 and 6.67, indicating that farmers estimated their own performance on all of these welfare aspects to be good to very good. It is noticeable that both the top and bottom of the ranking for the farmers' own performance was very similar to the ranking of the importance of the welfare aspects, meaning that in general, the farmers thought they performed better for aspects of welfare that they thought were more important and vice versa. In the top 4 of the ranking we found 3 of the same welfare aspects, namely "Food availability" (6.67 ± 0.71), "Water availability" (6.56 ± 1.01) and "Enough space" (6.56 ± 0.73). The 4th welfare aspect in the top 4 was "Foraging behaviour" (6.44 ± 0.53), of which the importance was also ranked quite highly. In the bottom 4 were the same 3 welfare aspects "Dustbathing" (5.22 ± 1.72), "Reaction to humans" (5.44 ± 1.33) and "Slaughter method on-farm" (5.56 ± 0.88), with the addition of "Absence of lameness" (5.56 ± 1.01). Interestingly, the scores for "Lameness" from the detailed WAss by researchers at the beginning of the study were quite good, but a higher median E showed that this aspect of welfare deteriorated during the study. Thus, it is possible that farmers thought that they performed less well on this aspect of welfare because they have indeed faced problems related to lameness on their farm.

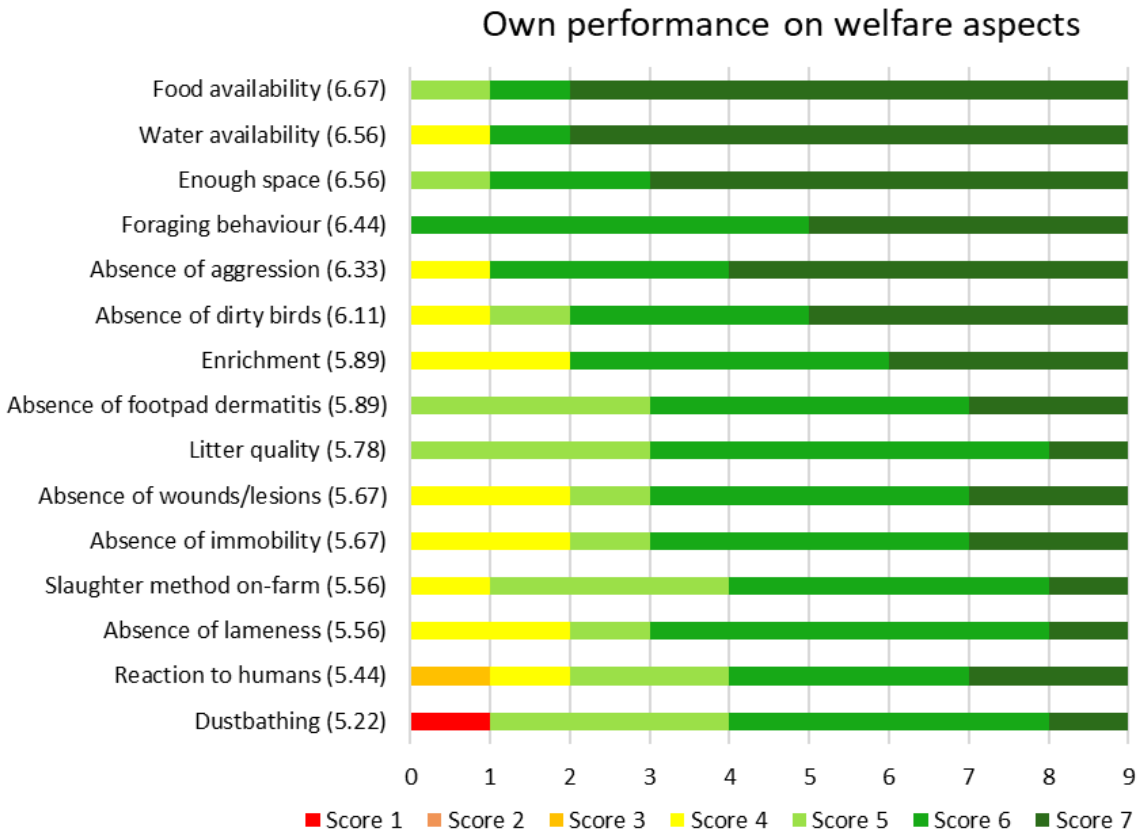

Figure 36: Farmers' estimates for their own performance on 15 aspects of poultry welfare rated between 1 (very badly) and 7 (very well). The welfare aspects are ranked from the highest to the lowest average rating, which is displayed between brackets behind the name of each welfare aspect, and all individual scores are shown. N = 9.

In figure 37, the mean rating for the farmers' own performance has been subtracted from the mean rating for the importance of the same welfare aspect and the differences have been ranked from highest to lowest. Thus, a high positive difference indicates that farmers rated the importance of that welfare aspect higher than their own performance regarding this aspect. It seems logical that the aspects for which this difference was highest would be the aspects of welfare for which the farmers would most wish to improve their performance. The differences between these two ratings were largest for "Absence of lameness" (0.56) and "Absence of immobility" (0.56). As already established, based on other survey answers and the scores from the detailed WAss, it seems possible that the farmers have experienced problems related to lameness, which would explain why they would want to improve on this aspect. "Absence of immobility" refers to birds who do not move when approached. Welfare scores of the detailed WAss did not indicate that farmers had specific problems with this, as the occurrence of still birds was very low at both the beginning and end of the study. Thus, our data did not offer an explanation for why farmers would rate their own performance so much lower than the importance of this welfare aspect. Next in the ranking came 4 welfare aspects with the same difference between the 2 previous answers, namely "Reaction to humans" (0.44), "Litter quality" (0.44), "Absence of wounds and lesions" (0.44) and "Absence of footpad dermatitis" (0.44). For "Footpad dermatitis" and "Reaction to humans" the idea that farmers want to improve on those aspects was supported by the detailed welfare assessments, in which the reaction to humans is measured through "Distance from humans". The median scores for these 2 welfare indicators were relatively high compared to the median scores for others.

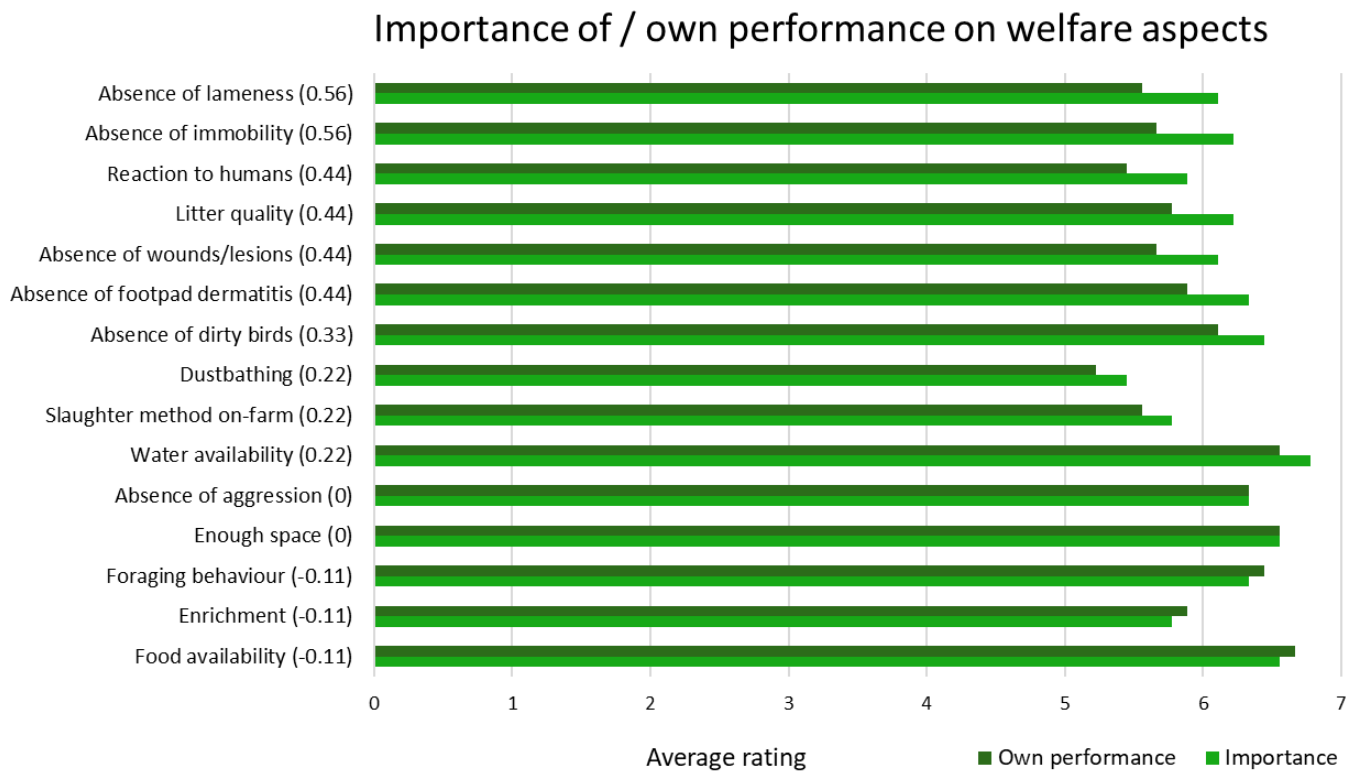

Figure 37: Mean ratings out of 7 for the importance of and farmers' own performance on 15 aspects of poultry welfare. The aspects are ranked from the highest positive to highest negative difference when the mean rating for farmers' performance is subtracted from mean rating for the importance. This difference is shown between brackets behind the name of each indicator. N = 9.

The expected usefulness of the EBENE® app and its most important elements as indicated by farmers before using the app for the first time is shown in figure 38. The farmers rated the usefulness between 1 (not useful at all) and 7 (very useful). Farmers rated the usefulness of an animal welfare self-assessment app like the EBENE® app at 5.11 (SD = 1.36). The large standard deviation showed that farmers differed in their opinion, but overall it seemed that the farmers expected an app for animal welfare self-assessments to be quite useful. Keeping in mind that this question was only answered by farmers who had already agreed to participate in a study about the app, this score was most likely higher than it would have been for a random group of free-range and organic pig farmers. Looking at specific elements of the app for which farmers were asked to rate the expected usefulness, farmers seemed the most interested in keeping a "Historical record of data" (6.00 ± 0.87) and expected "Benchmarking" (5.67 ± 1.00) and "Automated feedback" (5.67 ± 1.12) to be equally useful. They expected "Evaluating behaviour" (5.67 ± 1.12) to be more useful than "Evaluating health issues" (4.56 ± 1.74). This was somewhat surprising, given that several behavioural aspects were at the bottom of the ranking for the importance of welfare aspects.

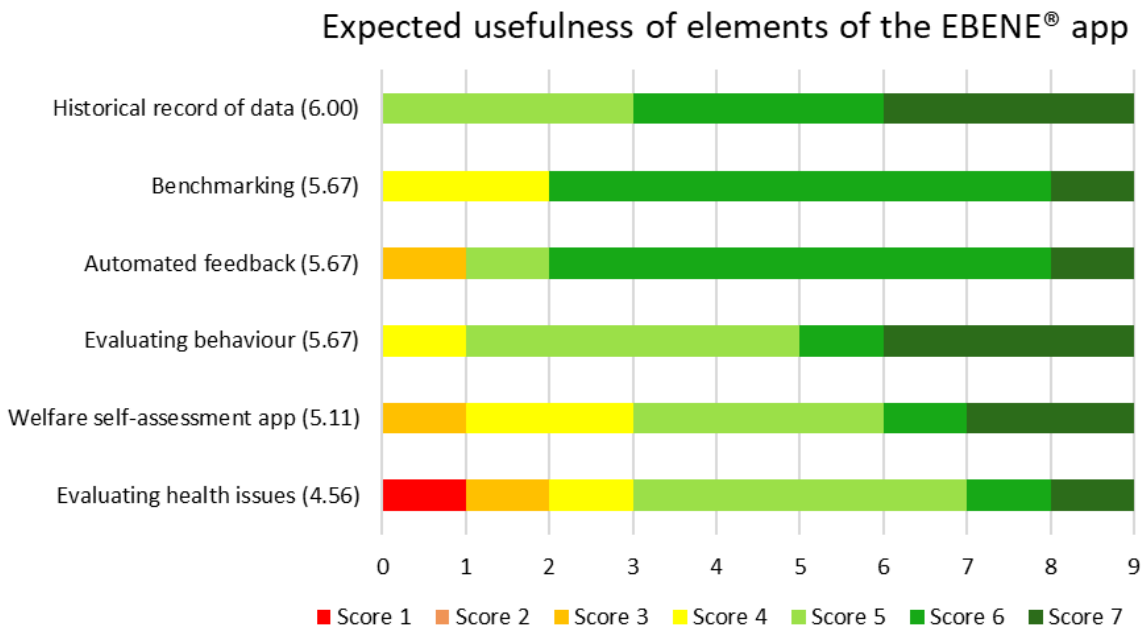

Figure 38: Mean ratings for the expected usefulness from 1 (not useful at all) to 7 (very useful) of the EBENE® app and the most important elements that the app contains. All individual scores are shown. N = 9.

Final survey

Nine farmers responded to the question on how much the use of the EBENE® app influenced their opinion on the importance of the 15 welfare aspects assessed in the first survey on a scale from 1 (not at all) to 7 (very much). Thus, any score higher than 1 indicated that the farmers noticed an influence of the EBENE® app. The mean rating given by all farmers was 5.11 (SD = 0.78), indicating that the farmers noticed a medium to large effect of the use of the EBENE® app. Farmers were asked to give an explanation for their answer if possible and their comments provided interesting information about the specific effect of the EBENE® app on their opinions on animal welfare:

1. “The app makes us observe the animals for a longer time in the zones. Because of that we can see their behaviour more easily and judge their welfare.” (5)
2. “Analysis of exploration” (6)
3. “Allows you to quantify animal behaviour” (5)
4. “General behaviour” (4)
5. “I was already sensitive to animal welfare before use, the fact of applying it nevertheless encouraged me to be more concerned about it” (5)
6. “I have always been aware of animal welfare because I consider that an animal, to express its full potential, must be in the best possible conditions. A bit like a sportsman!” (4)

The first 4 quotes from farmers all seemed to express that the use of the EBENE® app has changed how they looked at (certain aspects of) chicken behaviour. With 4 out of 9 farmers specifically mentioning this, it seems that the EBENE® app was particularly useful for farmers to evaluate the behaviour of their animals. One farmer mentioned that the app encouraged him to be more involved with the welfare of his chickens. This shows that at least for this farmer, the goal of our study to sensitize farmers towards animal welfare was reached. What this quote as well as quote 6 also illustrated is that these farmers were already sensitive to the importance of animal welfare before they participated in this study. As was mentioned previously for the pig farmers participating in this study, it is very possible that the fact that all of our participants volunteered to take part in a study on animals welfare has led to a biased sample. These farmers might have already been very motivated

to improve the welfare of their animals and already have had above average knowledge on the subject of animal welfare. Although an average rating of 5.11 is not low, it is possible that there are free-range and organic poultry farmers who are less knowledgeable who could have learned more from the use of the EBENE® app.

The farmers' opinions on the effects of the use of the EBENE® app on their own performance for each of the 15 welfare aspects are shown in figure 39, in order from the largest to smallest perceived effect. Farmers were asked to rate the impact they thought the EBENE® app had had on a range from 1 (severely deteriorated) to 7 (greatly improved), where 4 meant that nothing had changed. The answer of one farmer for the question about "Absence of wounds / lesions" was unfortunately erased due to an error in the survey system, explaining why only 8 individual scores for this question are displayed in the graph. The mean ratings for all welfare aspects were between 3.63 and 4.89, which would indicate a small deterioration for some welfare aspects and a small improvement for others. However, the comments added by farmers to explain their answers did not indicate that any of them have noticed decreased welfare of their animals related to the EBENE® app, so it seems possible that some farmers misunderstood the scoring method for the question and gave answers below 4 when they did not notice any change. However, as we cannot be certain about this, we will avoid going too deeply into individual scores and primarily discuss the ranking of the welfare aspects. Even if some farmers misunderstood the scoring method, the positions of the welfare aspects relative to one another should correctly reflect their opinion. "Absence of wounds / lesions" (3.63 ± 1.69), "Enough space" (3.67 ± 1.41), "Water availability" (3.78 ± 1.86) and "Absence of aggression" (3.78 ± 1.72) were at the bottom of the ranking. The last 3 of those 4 aspects were near the top of the ranking of the farmers' own performance at the beginning of the study. Thus, it is likely that there was relatively little room for improvement on those aspects and therefore the EBENE® app could not help the farmers very much regarding those aspects. "Absence of wounds / lesions" appeared lower in the ranking of the farmers' own performance. Thus, even though the farmers saw room for improvement, it seems the EBENE® app was not able to help them improve in this area. Farmers saw the largest positive impact of the EBENE® app on the welfare aspects "Foraging behaviour" (4.89 ± 0.78), "Dustbathing" (4.78 ± 0.79) and "Litter quality" (4.67 ± 0.50). The fact that 2 aspects of animal behaviour were at the top of this ranking was in line with the quotes from farmers from the previous question showing that the app helped them look at the behaviour of their chickens more closely. For "Foraging behaviour", the data from the detailed WAss by researchers supported that these farms were indeed performing better on this subject, as the median E was higher than median B. "Litter quality" was among the welfare aspects for which the data suggested that farmers would want to improve their performance, so it is positive that they feel as if they did that. While welfare scores for litter quality did not clearly change during the study, litter quality is also related to the occurrence of footpad dermatitis, for which the median B and E welfare scores from the detailed WAss were relatively high, indicating a possible welfare problem on many of the farms. It is possible that being made attentive to the problem of footpad dermatitis made farmers pay more attention to litter quality as well. However, "Footpad dermatitis" ranked low for this question, indicating that farmers did not feel the EBENE® app has helped them resolve these problems.

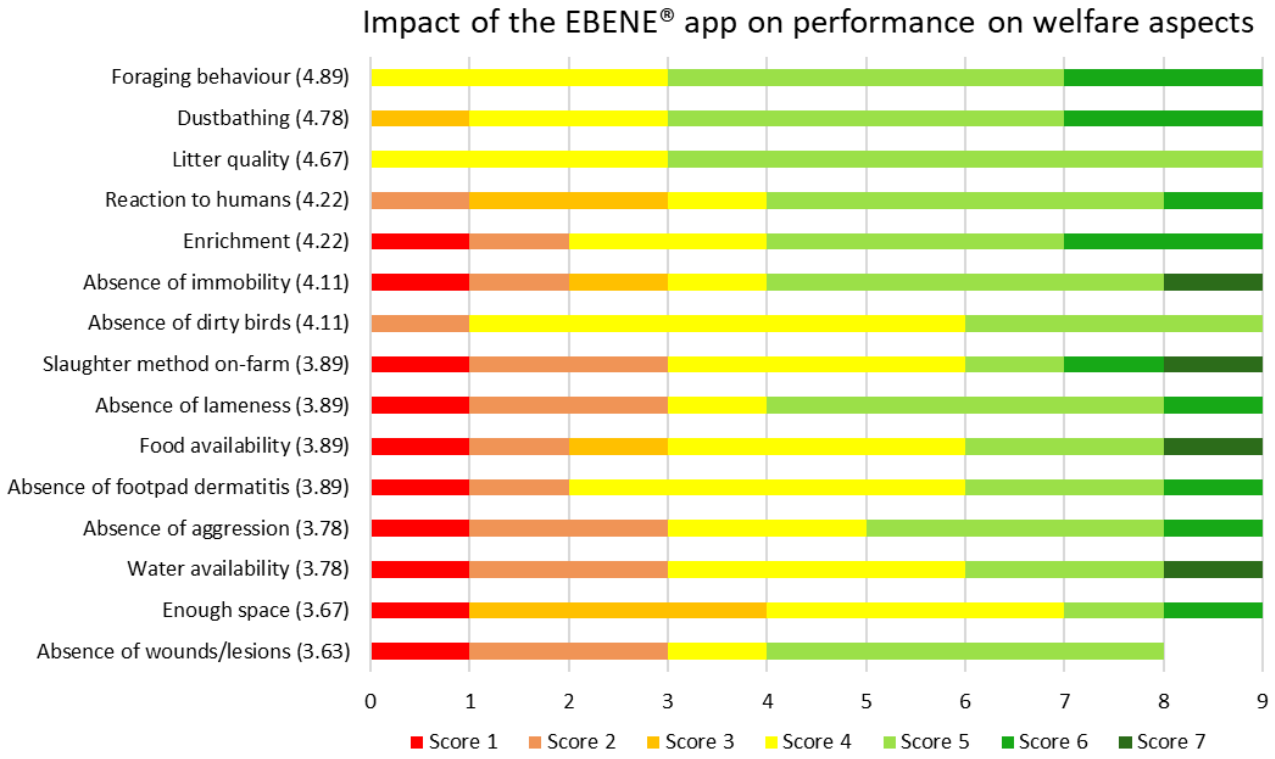

Figure 39: Farmers' opinion on how the use of the EBENE® app influenced their own performance on 15 aspects of poultry welfare rated between 1 (severely deteriorated) and 7 (greatly improved). The welfare aspects are ranked from the highest to the lowest mean rating, which is displayed between brackets behind the name of each welfare aspect, and all individual scores are shown. N = 9.

Farmers were also asked to give an overall score for whether they thought the use of the EBENE® app had had a positive effect on the welfare of their animals, rated from 1 (not at all) to 7 (absolutely). Thus, any score higher than 1 indicated that the farmers noticed an effect. The average score of 4.22 (SD = 0.97) indicates that farmers noticed a medium-sized effect of the use of the EBENE® app on the welfare of their chickens. Several farmers added a comment to complement their answer:

1. *"Allows for a different look at my farm" (4)*
2. *"Learning to observe" (5)*
3. *"Observation confirmed to me that the greater the density, the less animal well-being there is." (4)*

The first 2 quotes showed that the EBENE® app helped these farmers look at their animals in a different way than they were used to, but did not express that any aspects of welfare on the farm changed specifically. The third quote implied that while it helped this farmer confirm his beliefs, which can be useful, it did not teach him anything new. This probably explains why the mean score only supports a medium-sized effect of the EBENE® app on animal welfare.

Farmers were also asked what they thought about the EBENE® app in general. Figure 40 shows the ratings of the farmers for how easy they found it to use different elements of the EBENE® app on a scale from 1 (very difficult) to 7 (very easy). For this question, farmers were allowed not to answer if they did not think they used an element of the app enough to form an opinion. Therefore, the number of answers that was given says something about the extent to which certain parts of the app were used. With a mean rating of 6.11 for the general use of the app (SD = 0.78), farmers clearly found the app easy to use. However, the mean ratings for individual elements of the app ranged between 4.57 and 5.75, indicating that at least for some there was a little room for improvement. It is noticeable

that the farmers found the content of the welfare assessment itself easier to understand than the different elements of the results. Among elements of the results, farmers found “Comparing current and previous results” (5.13 ± 0.99) the easiest and found “Spider map (benchmarking” (4.57 ± 1.90) the most difficult to use. The question about benchmarking was also the only one for which more than one farmer did not answer, indicating perhaps that some of the farmers did not understand how to use this element.

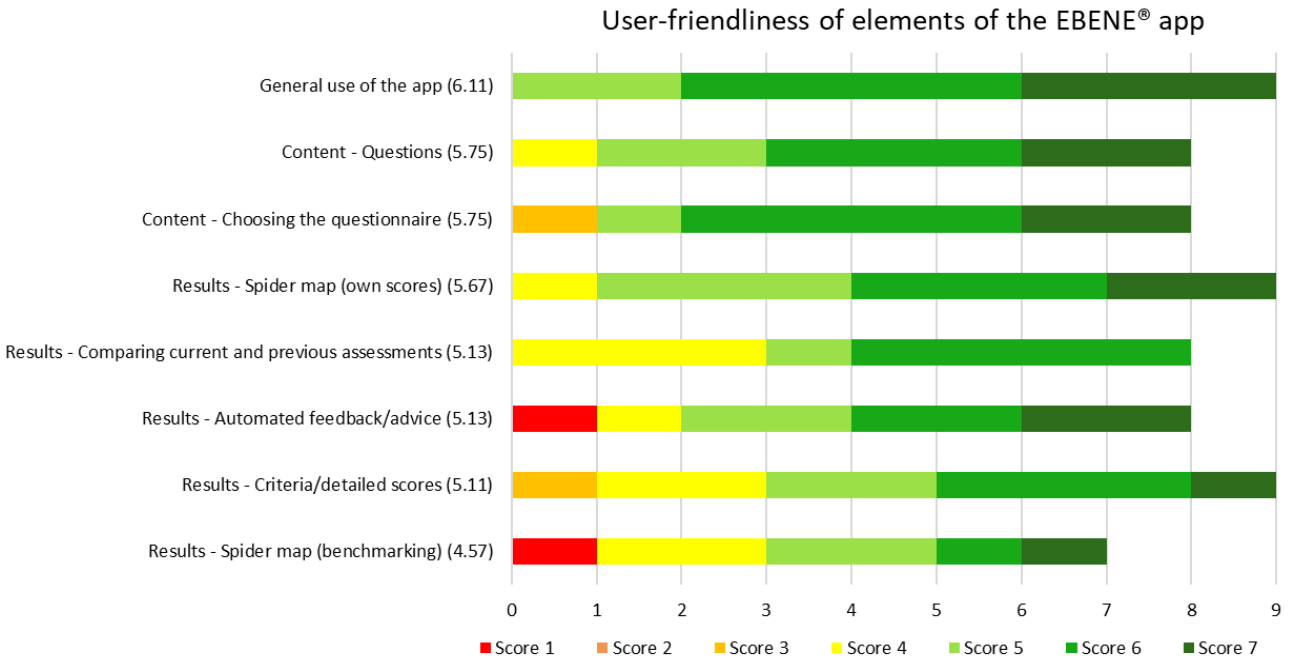

Figure 40: Mean ratings for how easy it was to use different elements of the EBENE® app on a scale from 1 (very difficult) to 7 (very easy). The elements are ranked from the highest to the lowest average rating, which is shown in brackets behind the name of each element, and all individual scores are shown. N = 9.

The mean ratings for the usefulness of different elements of the EBENE® app are shown in figure 41. The average ratings varied between 4.50 and 6.22, indicating a clear difference in how useful the farmers found different elements. These ratings did seem relatively high compared to the mean ratings of the farmers for the effect of the EBENE® app on their opinions on animal welfare and on the welfare of their animals, which were 5.11 and 4.22 out of 7, respectively. Thus, it is possible that while rating the usefulness of these elements, farmers were not only thinking of how much they benefited from the app themselves, but also took into account how useful they thought each element of the app could be to farmers in general. Farmers found “Behavioural assessment” (6.22 ± 0.83) the most useful part of the welfare assessment itself, which aligned with several earlier comments of farmers stating that the app helped them look at animal behaviour differently. Farmers rated the usefulness of “Photo materials for questions” (5.83 ± 1.17) more highly than the usefulness of “Additional information (?) for questions” (5.00 ± 1.07), suggesting that they preferred an explanation of a welfare indicator with a visual representation over an explanation in text form. However, fewer farmers have answered the question about photo material than the question about additional information, so perhaps not all farmers preferred the use of photos. It is noticeable that for some elements of the results, more farmers answered the question about the user-friendliness than about the usefulness. A possible explanation for this is that some farmers noticed those elements often enough to give their opinion on the user-friendliness, but were not motivated to use them properly and therefore felt like they could not give an opinion on their usefulness. Among the elements of the results, farmers found “Keeping track of your previous data” (5.86 ± 1.07) the most useful. This

element also ranked highest when the farmers were asked about the expected usefulness of different elements before using the EBENE® app. It suggests that farmers were more interested in the data of their animals that they gathered themselves than they are in getting “help” from others through the means of automated feedback or comparing their results with other farmers. “Spider map (benchmarking)” ranked lowest of all aspects of the results, with a relatively large distance from the others. It was also one of the questions to which the smallest number of farmers responded, indicating that it is likely that several farmers did not really use the benchmarking function. As this element also ranked lowest for user-friendliness, it is possible that not all farmers understood how to interpret the benchmarking data. An extra explanation inside the EBENE® app of how to interpret the benchmarking data could be useful to help farmers take full advantage of this function.

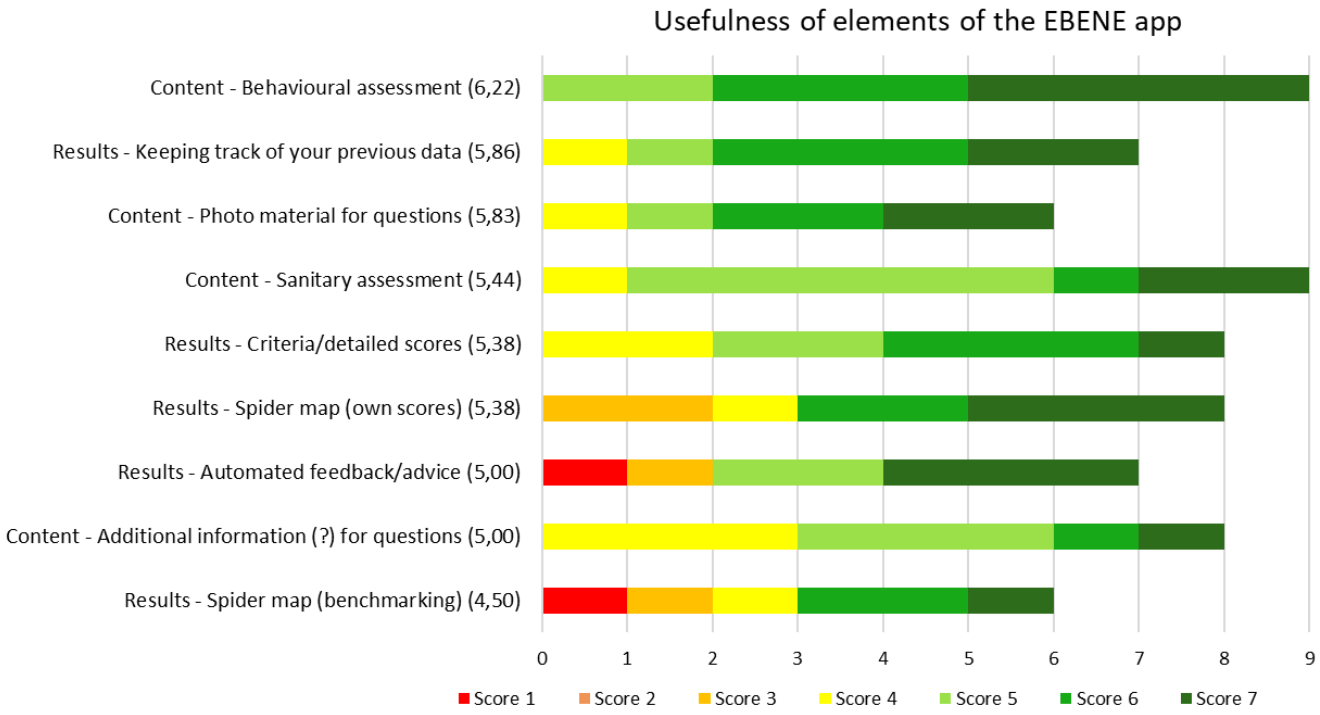

Figure 41: Mean ratings for how useful the farmers found each element of the EBENE® app on a scale from 1 (not useful at all) to 7 (very useful). The elements are ranked from the highest to the lowest average rating, which is shown in brackets behind the name of each element, and all individual scores are shown. N = 9.

Farmers were also asked how they would rate the EBENE® app on a scale from 1 (worst) to 10 (best). The average rating of a 7.78 (SD = 0.97) indicates that the farmers thought of EBENE® as a good app, with perhaps some room for improvement. This was generally supported by the medium to high ratings for the user-friendliness and the usefulness of most elements of the app. Looking at the answers of the previous 2 questions, the key to improving the app could be to increase the user-friendliness of several elements of the results of the welfare assessment, with special attention to the benchmarking data. Keeping in mind that farmers only noticed a medium-sized effect of the use of the EBENE® app on their opinions on animal welfare and on the welfare of their animals, it is possible that some of the farmers partially rated the app for the potential they see in it for other farmers, rather than for the benefits that they experienced themselves. Although it is not mentioned by any of the poultry farmers, just as for the PIGLOW app it could be useful to specifically target new farmers who have the most left to learn.

Finally, farmers were asked if they had any suggestions for improvements of the EBENE® app. The answers to this question offered some useful suggestions:

1. *"More photos for explanations" (6)*
2. *"Being able to go back. It's not always possible to go back and so sometimes you have to start all over again." (9)*
3. *"Recording of breeding parameters once encoded for the first time. Very restrictive to have to provide building information each time (number of pipettes, size of plates, etc.). Difficult to understand the results. Solutions to problems are not targeted." (7)*

As shown by the answer to the previous question, farmers found the photo material in the app quite useful already. Thus, the suggestion from quote 1 to add more photos could perhaps help farmers to better understand questions in the app that do not have photo material yet. The second and third quote were related to the user-friendliness of the app. The first farmer referred to some issues they had while completing the welfare assessment with going back to a previous screen in the app to change the answer to a question. Sometimes when trying to do this, their app would accidentally close instead and they were not able to re-open the ongoing assessment. The farmer from the third quote was referring to the general characteristics describing the building that must be filled in in the app at the beginning of a welfare assessment. While the app should fill those in automatically when farmers complete a second assessment for the same building, unfortunately this did not always happen and it was tedious for the farmers to have to spend time on this again. Thus, together these quotes showed that farmers experienced several technical issues. While the general user-friendliness of the app was already rated quite highly, fixing these types of issues could help motivate the farmers to use the EBENE® app more frequently and therefore benefit more from it. The third farmer also commented that they found it difficult to understand the results. The mentioned difficulty with understanding the results confirmed what was suggested by some of the results of the previous survey questions. As mentioned before, adding additional explanations for some elements of the results to the EBENE® app could help farmers benefit more from what the app has to offer. Finally, this farmer mentioned that more targeted solutions to specific welfare problems would be useful. As was discussed for the PIGLOW app, adding targeted solutions without being present on the farms and while still making sure that the offered advice is relevant to a range of different farms is challenging. However, it is worth exploring in order to make the EBENE® app even more useful to a larger group of farmers.

5.2.3. Comparison of EBENE® assessment by farmers and researchers

The EBENE® WAss of farmers and researchers at the beginning of the study were compared first and subsequently, the differences between farmers and researchers at the beginning of the study were compared to the differences at the end of the study. Table 14 shows the number of groups that were observed and the number of chickens that were observed in groups with the EBENE® app by each farmer/researcher couple. The number of individual chickens included in the assessment of individual level WI that were evaluated during the transect walks from one side of the building to the other was not known exactly. This number is estimated by the EBENE® app based on the total number of chickens in the building and the width and length of each transect walk. However, as these groups were always very large and only absolute frequencies for these indicators are compared within each couple, the exact number of chickens should not have an effect on the data.

Table 14: This table shows the combined area of the 2 transect walks during which individual level welfare indicator are assessed, the number of groups that were observed and the number of chickens that were observed in groups with the EBENE® app by each farmer/researcher couple at the beginning and end of the study. The order of the farms does not correspond with the order in earlier tables. N = 11 for the beginning and N = 4 for the end.

Farm	# Groups observed		# Chickens observed in groups	
	Beginning	End	Beginning	End
1	3	3	100	45
2	3	3	63	75
3	3	3	62	65
4	3	3	75	120
5	3	-	125	-
6	3	-	115	-
7	3	-	80	-
8	3	-	65	-
9	3	-	125	-
10	3	-	85	-
11	3	-	80	-

In table 15, the differences between farmers and researchers for all 18 WI for which a comparison was possible are ranked from the highest to the lowest mean absolute difference. The medians showed that in general, indicators with a higher number of occurrences appeared higher in the ranking. This might be explained by the fact that the potential for disagreement is higher when there are more occurrences, since in general people are probably more likely to miss the presence of a welfare problem or sign of positive welfare than to record the presence of something that is not there. The number of occurrences was generally higher for behaviours than for physical anomalies. In addition, different types of scoring methods also influenced the maximum possible difference between farmers and researchers. For instance, for WI for which a score out of 3 or 5 is given, the mean absolute difference could not be as large as for indicators for which the total frequency was scored. Therefore, all of WI should be compared within their own category.

For the behavioural indicators “Foraging” (mean abs. diff. = 17.55; 11 different scores), “Preening” (mean abs. diff. = 8.36; 11 different scores) and “Stretching / wing flapping” (mean abs. diff. = 5.91; 11 different scores) at the top of the ranking, a similar pattern can be seen with substantially more researchers than farmers giving a better score, indicating that researchers quite consistently scored more of these behaviours during the 3x 5-minute behavioural observations than farmers. A possible explanation for this is that performing behavioural observations was new to the majority of the farmers, whereas the researchers had previous experience with it. As seen from the medians, these behaviours occurred quite often during the behavioural observations and therefore it could be quite challenging to see all occurrences. It is likely that researchers missed fewer occurrences than farmers because of their previous experience with behavioural observations. The other 3 behaviours

that were observed during the 5-minute behavioural observations (Positive interaction, Dust bathing and Aggressive pecking) were rarer, as shown by their medians. For all 3 of these behaviours, there were more farmers than researchers who scored more occurrences of the behaviours. While the difference was relatively small for “Dust bathing” and “Aggressive pecking”, it was larger for “Positive interaction”. These differences between farmers and researchers might have to do with the fact that these behaviours are relatively complex and farmers, who had less experience observing them, were more likely to count a similar looking behaviour that did not quite meet the description of the behaviour given in the EBENE® app. It is also possible that farmers thought they were missing occurrences of the rare behaviours because they had not scored any yet at a certain point during an observation. This could have made them unsure of themselves, which might have motivated them to try to spot occurrences of those behaviours that were not really there. However, none of these possible explanations for the differences in recorded behaviours can be proven, as it is not known what the “true” frequencies of the behaviours were. For the 3 remaining behaviours that were observed in the EBENE® app, different scoring methods were used. For the WI “Panting”, the percentage of birds in the group showing this behaviour was recorded. Therefore, the mean absolute difference of 1.73 was quite small and did not suggest a clear difference between farmers and researchers in scoring this WI. For the WI “Resting” a score between 0 and 5 was calculated based on the percentage of birds in the group that was resting. The mean absolute difference of 0.68 was therefore relatively large and the fact that only 1 farmer and 10 researchers gave the better score seemed to show a clear pattern. However, this is complicated by the fact that an intermediate percentage of resting birds leads to the best score out of 5. While 10 researchers gave a better score than the farmer in their couple, approximately half of those researchers recorded a lower percentage of resting chickens, and the other half recorded a higher percentage of resting chickens. Thus, there was no clear pattern of one group recording more or fewer resting birds. The large mean absolute difference for this WI therefore suggested that it might be difficult for farmers as well as researchers to accurately estimate the percentage of resting chickens in a large group. For the last behavioural indicator “Distance from humans”, a score between 0 and 3 was always awarded. Relatively speaking, the mean absolute difference of 0.23 was therefore of moderate size. Four farmers gave a better score for this indicator than the researchers in their couple, which could indicate that farmers judge the distance between them and the chickens more positively. While it is of course possible that the chickens did approach the farmer more closely than the researchers, this should not explain the difference in answers. The farmer and researcher were always standing close to one another and answered this question for the chickens that were closest to either person.

For 6 WI regarding physical anomalies (Too small, Lameness, Still, Injuries, Other anomalies and Dead) the absolute frequency was recorded during 2 transect walks through the building. Of all those WI, there was only a moderately clear difference between farmers and researchers for "Too small" (mean abs. diff. = 0.81; 4 different scores), with 3 farmers and 1 researcher giving a better score. This might suggest that researchers judged the size of chickens more severely than farmers. The final WI assessed during the transect walk was “Dirtiness”, for which the percentage of dirty birds during the walk was scored. The small mean absolute difference of 0.32 and the fact that the majority of the couples gave the same score shows that there was no clear difference in how this indicator was scored by farmers and researchers. One other physical anomaly, “Footpad dermatitis” (mean abs. diff. = 0.64; 3 different scores), was scored for 15 chickens that were caught randomly in the building. While the mean absolute difference was of moderate size, the majority of the couples gave the same score. Thus, there was no pattern of one group scoring differently than the other.

The mean absolute difference of “Litter quality”, for which a score between 0 and 3 was given, was of moderate size (0.27). However, as the majority of couples gave the same score there was no clear pattern of one group scoring differently than the other.

Overall, the data seem to support that experience with behavioural observations made researchers less likely to miss occurrences of behaviours that were very frequent and less likely to overestimate

the occurrence of rare behaviours. However, as the true frequencies of the behaviours were not known, this cannot be said for certain. The indicator “Resting” was probably most difficult to be observed accurately with or without experience. For WI related to physical anomalies, the differences between groups were generally much smaller and did not support clear patterns of differences between farmers and researchers. However, it is possible that differences between farmers and researchers in measuring physical anomalies would have been larger if the physical anomalies had occurred more often. The low occurrence of these welfare problems likely limited the potential for disagreement between different observers.

Table 15: The difference between farmers and researchers in the assessment of welfare indicators for broilers chickens with the EBENE® app at the beginning of the longitudinal study. The median score given by farmers and researchers, the mean absolute difference between all farmer/researcher couples, the number of farmer/researcher couples that gave different scores for each indicator and the number of times that each party gave the better score are shown. The indicators are shown in order from the highest to lowest mean absolute difference. N=11.

Welfare indicator	Median	Mean absolute difference	# Different scores	Better score
Foraging	34	17.55	11	2F, 9R
Preening	23.5	8.36	11	3F, 8R
Stretching / wing flapping	19.5	5.91	11	3F, 8R
Positive interaction	8.5	3.09	9	6F, 3R
Panting	0	1.73	4	2F, 2R
Too small	1	0.81	4	3F, 1R
Dustbathing	0	0.73	2	2F
Resting	4.28	0.68	11	1F, 10R
Footpad dermatitis	3.5	0.64	3	2F, 1R
Lameness	1.5	0.55	2	2R
Aggressive pecking	0	0.45	5	2F, 3R
Dirtiness	0	0.32	2	1F, 1R
Litter quality	0	0.27	2	3F
Distance from humans	3	0.23	4	4F
Still	0	0.18	2	2F
Injuries	0	0.09	1	1R
Other anomalies	0	0.09	1	1R
Dead	0.5	0	0	-

The comparison of differences for EBENE® assessments by farmers and researchers at the beginning and end of the study are shown in table 16. For this comparison, only the data for the 4 farmer/researcher couples for which data of both moments were available was used. This small

sample size made it difficult to draw proper conclusions about the changes for individual WI, as many differences between groups are determined by only 1 couple. For the majority of WI, the pattern of the difference between farmers and researchers has remained the same. For more details, the reader is referred to table 16.

The mean of the mean absolute difference (1.61 and 1.89) and the mean number of different scores (1.61 and 1.89) for all WI were both slightly larger at the end of the study than at the beginning, which shows that the level of agreement between farmers and researchers has not increased during the study. The differences were not large enough to conclude that the level of agreement between the groups has decreased either. Thus, it seems that on average for all indicators, there has been no clear change in the level of agreement.

Table 16: The difference between farmers and researchers in the assessment of welfare indicators for broiler chickens with the EBENE® app, compared for the beginning and end of the longitudinal study. The median score given by farmers and researchers, the mean absolute difference between all farmer/researcher couples, the number of farmer/researcher couples that gave different scores for each indicator and the number of times that each party gave the better score are shown. The indicators are shown in order from the highest to lowest mean absolute difference at the beginning. For the mean absolute difference and the number of different scores, the averages of all welfare indicators are also shown. N=4.

Welfare indicator	Median		Mean absolute difference		# Different scores		Better score	
	Beginning	End	Beginning	End	Beginning	End	Beginning	End
Foraging	31.5	32.5	15.5	24.5	4	4	1F, 3R	2F, 2R
Preening	12.5	18	9.00	4.50	4	4	2F, 2R	3F, 1R
Stretching / wing flapping	26.5	28	7.25	7.25	4	4	4R	4R
Panting	2.67	4.06	4.67	0.10	3	2	1F, 2R	2R
Positive interaction	9	6	2.50	8.50	2	3	2F	3F
Resting	4.09	0	1.41	2.08	4	1	1F, 3R	1R
Footpad dermatitis	3	5	1.00	2.00	1	2	1F	2F
Too small	1	1	0.75	2.00	1	2	1F	2R
Litter quality	0	0	0.63	0	2	0	2F	0
Injuries	0	1	0.25	0.25	1	1	1R	1F
Lameness	1	1	0.25	0.75	1	3	1R	3F
Other anomalies	0	0	0.25	1.25	1	1	1R	1F
Still	0	0	0	0	0	0		0
Dirtiness	0	0	0	0.75	1	1	1F	1R
Dead	0.5	0	0	0.25	0	1	-	1R
Dustbathing	0	0	0	2.25	0	1	-	1R

Aggressive pecking	0.5	1	0	2.50	0	3	-	3R
Distance from humans	2.75	3	0.00	0.13	0	1	-	1F
Mean			**1.61**	**1.89**	**1.61**	**1.89**		

6. Discussion

Welfare at the beginning and end of the study

The main aim of this study was to find out whether the use of the PIGLOW and EBENE® welfare self-assessment apps by farmers could have a positive effect on the welfare of their animals. The results from the detailed welfare assessments conducted by researchers at the beginning and end of the longitudinal study on pig as well as poultry farms did not show a clear effect of the use of the PIGLOW and EBENE® app on animal welfare on the farms. While the welfare data from the farms showed that scores for a few welfare indicators improved during the time of the study, no improvement was observed for the majority of welfare indicators. Results from the surveys in which the farmers were asked for their own opinion on the effect of the welfare self-assessment apps showed some positive effects of the use of the apps, but otherwise largely supported the result of the detailed welfare assessments. The pig farmers indicated that on average, they noticed small to medium-sized effects of the use of the PIGLOW app on how important they considered different aspects of welfare to be and on the actual welfare of their own pigs. Scores differed among farmers, indicating that while for some farmers the PIGLOW app had almost no effect, some other farmers did see large effects of the use of the app. The poultry farmers saw a medium to large effect of the use of the EBENE® app on how they considered the importance of different welfare aspects and a medium-sized effect on the actual welfare of their chickens. Furthermore, the survey answers from poultry farmers specifically showed that for many of them, the EBENE® app taught them about the importance of animal behaviour. Pig and poultry farmers also gave relatively high scores for the easiness of use as well as the usefulness of different elements of the apps, with especially the PIGLOW app scoring very well for user-friendliness of all its elements. When asked to give a general score out of 10 for the app, both apps also scored quite well with and 8.09 for PIGLOW and a 7.78 for the EBENE® app. However, when asked to rate the effect of the use of the apps on the evolution of specific welfare aspects on their farms, the average scores for all welfare aspects for pigs and poultry indicated that farmers had only noticed a very small (positive) effect of the use of the apps. In addition, comments that were added by farmers to explain their answers showed that while some farmers started to pay more attention to certain specific elements of welfare thanks to the apps, most farmers indicated that the use of the app did not truly change how they looked at the welfare of their animals. Especially because the farmers did give high scores for the apps in general, it is worth exploring possible reasons for why the use of the apps did not have a substantial impact on animal welfare on the farms in our study.

Recruitment of participants

A limitation of the study that must be mentioned is that unfortunately, the sample size was smaller than foreseen. Originally, it was planned to include 20 to 30 farms per species. While we used all available channels and spared no effort to recruit as many farmers as possible to participate in the study, particularly the Covid-19 pandemic during the time of recruitment and periods of avian influenza made it more difficult to find farmers who were willing to participate and/or follow the instructions for periodic assessment correctly. Some farmers did not feel comfortable allowing unnecessary visitors to their farms while preventative measures for Covid-19 were in place, even outside of lock down periods, which made it impossible to include them in the study. Avian influenza similarly made poultry farmers more hesitant to allow visitors, either because they were afraid for the health of their birds or because birds had been confined indoors and farmers were worried that visitors would cause too much stress. To find as many participants as possible and be able to plan as many farm visits as possible around Covid-19 and avian influenza, we extended the time period for recruitment and the first round of farm visits as far as possible. This is why the date for this

deliverable was postponed. An additional difficulty for finding farmers who were willing to participate is that animal welfare is a sensitive topic and that not all farmers are comfortable with the idea of researchers visiting their farm to assess the welfare of their animals. During several phone calls, farmers or representatives of farmer organisations indicated that they were afraid that welfare data of their farms would be made public and would get them into trouble, sometimes because they had had previous experience with that. This is something that was also experienced by other researchers in similar situations (van Dijk et al., 2018). If the sample of the study had been larger, it might have been possible to compare characteristics of farms on which welfare scores had and had not improved. This could have helped us determine if perhaps there are specific types of farms or farmers for which the apps are more helpful than for others.

The fact that some farmers might be uncomfortable to have animal welfare research conducted on their farm could also be a part of the explanation for why the results of our study did not show an overall improvement of animal welfare on the participating farms, despite the fact that both apps received high scores in the surveys. Scores from the detailed welfare assessments by researchers showed that the prevalence of most welfare problems was already quite low on the majority of the farms at the beginning of the study and that animal welfare was already quite good, especially for the pig farms. Because animal welfare is a sensitive issue and scientists and farmers do not always share the same opinions on this topic, it may be that farmers who are proud of the welfare of their animals are more likely to welcome visitors to their farm. It is possible that at least on average, the welfare of those farmers' animals is already better than animal welfare on farms that are reluctant to participate. Additionally, the fact that these particular farmers were motivated enough to want to be part of our study on improving animal welfare might reflect a general high motivation of these farmers to give their animals a good life. This is also supported by some survey answers from farmers, in which these farmers state that they were already very focused on animal welfare. More pig farmers than poultry farmers specifically mentioned this, which could explain why on average the poultry farmers noticed a larger impact of the app on how important they found certain aspects of welfare than the pig farmers. This is in line with the results of a study on German pig farmers' attitudes towards participation in animal welfare programmes. The study found correlations between a high willingness to participate in such programmes and a strong believe that animals must be able to show natural behaviour, less strong objections against public discussions on animal welfare, and less negative attitudes towards animal welfare related demands made by politicians and consumers (von Hardenberg & Heise, 2018). While this does not explicitly show that these farmers have already worked harder to improve the welfare of their animals than others, it does suggest that the farmers who are willing to participate in animal welfare programmes have a generally more positive attitude towards providing good welfare for their animals. The fact that the welfare on their farms was already good, limited the potential for improvements of animal welfare during the study. This would also make it more difficult for the study to show a positive effect of the use of the apps on animal welfare.

Motivating farmers

Especially from the study with the EBENE® app it became apparent that it is difficult to motivate farmers to use the app on a regular basis. Pig farmers were asked to perform at least 6 welfare assessments with the PIGLOW app and the majority of them did indeed do this, with a few farmers only performing 5 assessments. In contrast, poultry farmers were asked to perform at least 5 welfare assessment with the EBENE® app, but only one of the 9 farmers did this. The other only performed 2, 3 or 4 assessments. The survey results showed that both the PIGLOW and EBENE® app received relatively high scores in general and for aspects such as user-friendliness. And yet, we noticed that it was necessary to contact farmers to let them know they should perform another welfare assessment with the app. For some farmers it was necessary to contact them a second time, either because they had forgotten to do the assessment or, in the case of poultry, were not able to find the

time right before slaughter of the animals and were then forced to wait until their next round of animals had reached the right age. This seemed to be a larger issue for poultry than for pig farms, because poultry farms often only have birds from the same age at any one time, meaning that once a lot of chickens has gone to the slaughterhouse, it will be several months before an organic farm has a new lot of chickens at slaughter age. Pig farms, on the other hand, usually have pigs of many different ages at the same time and will therefore almost always have pigs available that can be assessed. The fact that regular reminders by researchers were necessary in this case does beg the question whether it will be easy to motivate farmers outside of this study to regularly use the app. Others working with farmers on the topic of welfare self-assessments have experienced similar problems. A study in which welfare scores by farmers and experts on 188 Italian dairy farms were compared experienced similar problems with farmers who kept delaying the welfare assessment they were asked to do (Katzenberger et al., 2020). The researchers responsible for the Dierenwelzijnscan (Thys et al., 2019), a mobile app for welfare assessments on conventional pig, poultry and cattle farms, have noticed that when they stopped actively promoting the app among farmers, the number of users started to drop. The number of farmers using the Dierenwelzijnscan app to evaluate pig welfare has increased due to a collaboration with the Belgian Pork group, who made it mandatory for their farmers to complete one welfare assessment with the app each year. However, although all of these farmers do indeed conduct the welfare assessment with the app, it is difficult to say how much effort they actually put into it and how reliable the results are. It is possible that some farmers fill out the assessment sitting behind their desk instead of inside the stables or purposely pick the best animals to observe to obtain good welfare scores (personal communication, 26th of March 2024). That making it obligatory for farmers to perform welfare assessments would likely not be the best solution also became clear in a study in the UK by van Dijk et al. (2018), who suggested that forcing farmers to perform welfare assessments when they do not see the value of it themselves is unlikely to stimulate them to make changes related to animal welfare (van Dijk et al., 2018). Similarly, a study in Denmark showed that mandatory participation in “stable schools” where farmers learned to identify and find solutions to welfare problems, did not motivate farmers to take action (Vaarst & Fisker, 2013). A solution to the problem of motivating farmers could be to promote the app among veterinarians, farm technicians and consultants who already work together with many farmers. If they believe in the benefits of the PIGLOW and EBENE® apps, they could motivate the farmers they work with and see fairly regularly to use the apps and discuss the results with them.

Besides being an issue in and of itself, the fact that not all poultry farmers used the EBENE® as often as they were asked to makes it more difficult to draw proper conclusions about whether the app has the potential to help farmers improve the welfare of their animals when it is used regularly. The reason that farmers were asked to perform regular welfare assessments during a longer period of time instead of just one assessment is because sensitisation towards possible welfare issues and related behavioural changes in farmers are more likely to happen when farmers go through the material of the app multiple times. Thus, the fact that some farmers did not do this as often as planned might have made them less likely to make any changes related to animal welfare.

Possible improvements of the apps

All of the reasons above could explain why the use of the apps in this study has not led to a clear improvement of animal welfare even if the apps did have the potential to help. However, survey results also indicated that some aspects of the apps can perhaps be improved. In the survey, farmers were allowed not to answer questions about the easiness to use or the usefulness of a certain element of the apps if they did not use it. The missing answers showed that some farmers did indeed not use some of the elements of the app, or at least not enough to feel that they could rate those elements. For both pigs and poultry, several farmers did not use the photo material attached to the questions to show visual examples of welfare indicators. For both apps, photos are not instantly

shown on the screen. Instead, users need to click somewhere on their screen to open them. It is therefore possible that it was not always clear to some of the farmers that photos were available. What could help resolve this issue is to add a short sentence behind each question for which photo material is available to let users know this. Alternatively, photos could be shown immediately on the screen. However, this could make the screen look too full and would often require users to scroll down to be able to answer the question, which could decrease user-friendliness. It must be said that it is also possible that farmers did see where they had to click to find the photos, but already understood all questions well enough to feel like they did not need to see them. It is worth following up on this to learn from the farmers which explanation is the right one before making any changes to the apps.

Besides photo material, the elements of the apps for which fewer farmers answered the questions were elements of the results. For both pigs and poultry, there were no elements of the results for which all farmers have answered the questions. While it is possible that farmers did not answer the questions because they did not understand how to use certain elements of the results, the scores that were given by farmers for the user-friendliness for most elements of the results were rather high, indicating that this is unlikely. In addition, all farmers were given an explanation of how to interpret the results of the welfare assessments during the first farm visit of the study (and a researcher was available via email or telephone for the entire duration of the longitudinal study to answer possible questions). Thus, at least for elements of the results for which the largest difficulty would be to know where to find them, that explanation should have been more than sufficient. Therefore, it is more likely that some farmers did not make the effort to look at the results of their welfare assessments properly. Similar to the issue of farmers who did not use the app as often as they were asked to, this was most likely a question of motivation. Thus, to allow farmers to benefit more from what the apps have to offer, we must find a way to motivate them to invest the time to properly evaluate the results of their welfare assessments. A change to the apps themselves that could possibly help accomplish this would be to integrate a kind of virtual reward system. For instance, the apps could keep track of the users “streak”, which they could hold on to by completing at least one welfare assessment every few months. Similarly, users could earn virtual badges for completing a certain number of welfare assessments or for improving certain scores compared to their previous assessments. There are many possibilities for such virtual rewards that could perhaps motivate people to keep using the apps. However, as the costs of making such changes would be high, this would at the very least not be a short term solution. As for previous issues, another solution for this could be to promote the apps among veterinarians, technicians or other advisors so that they can stimulate the farmers they work with.

One element for which a lack of motivation might not explain the low number of answers to the questions is benchmarking. Previous studies have found that farmers often found benchmarking useful and that it stimulated many farmers to make animal welfare related changes on their farms (Atkinson et al., 2017; Burke, 2004; Schultheiß et al., 2023; van Dijk et al., 2018). However, for both pig and poultry farmers in our study, this was the element of the results for which the smallest number of farmers answered the question, namely 6 out of 9 and 6 out of 11. For PIGLOW, benchmarking was rated among the least useful elements of the app and for EBENE®, it was the only element for which the average score for the easiness of use as well as the usefulness was lower than 5. This suggests that the reason that farmers found benchmarking less useful and did not use the function as much might be that they found it more difficult to understand. One pig farmer even commented in the survey that he did not realise until the end of the study what the possibilities for benchmarking were. It is logical that benchmarking would be the most difficult element to understand, because putting one’s own scores into perspective next to the scores of others is less straightforward than simply reading a summary of one’s own scores or reading the automated feedback. While the PIGLOW app represents the comparison with data of other app users in a numerical way and the EBENE® app uses a graph for a visual comparison of the data, both require an explanation to

understand. The explanation given by the researchers during the farm visit and the explanations currently present in the apps might not have been enough. A solution to this could be to add a clear explanation in text form of how to interpret the benchmarking data to the results pages of both apps. For PIGLOW, such an explanation is already available, but for the moment the users need to click on an 'i' icon next to the column with benchmarking data to see this explanation. It is possible that more users would read the explanation if it was included in the main text of the report with results. In the EBENE® app, only a legend for the graph showing the visual comparison of data of other farms is currently included in the app, which might be the reason that the easiness of use for benchmarking in the EBENE® app received a lower average score. Another reason for the fact that the farmers found benchmarking less useful than other elements of the results could also be the small sample size of the study. In the EBENE® app, benchmarking was only done among participants of the longitudinal study to ensure that only farms of a similar type were included in the comparison. Because the EBENE® app can also be used for conventional poultry farms, using the entire database for benchmarking might not have provided a relevant comparison of this group of free-range and organic farmers. For PIGLOW, which is exclusively for free-range and organic farms, all validated results were used for benchmarking. However, as the app is still very new the total number of users is still quite small. Thus, for both apps, the data that could be used for benchmarking was relatively limited, which most likely decreased its usefulness to the farmers. Finally, it is also possible that some of these farmers simply do not value a comparison with others as much as they value their own data. For both pig and poultry farmers, the highest average score for the usefulness of any element of the results was for keeping track of their own previous data, suggesting that what these farmers value most is a record of the welfare of their animals. As mentioned previously, it seems likely from the results as well as from our recruitment process that the farmers participating in our study already had above average knowledge of animal welfare and were already focused on improving the welfare of their animals. Therefore, it is possible that they feel they also already know how their farm performs compared to others in the same sector, or they already know what their aims concerning good animal welfare are and are not interested in a comparison with other farms. In the study by van Dijk et al. (2018) the majority of farmers was positive about the use of benchmarking, but there was also at least one farmer who said that benchmarking was only useful if you thought you had welfare problems or that otherwise it was not worth it to invest time in it. Because the farmers in our study had relatively good welfare scores, it is possible that they felt the same way and did not feel that they needed to be reassured by comparing their scores to others.

Farmers were also asked directly which changes they would like to see made to the apps. For the EBENE® app, several farmers commented on some technical issues that they experienced while using the app that decreased their motivation to use the app again. Additionally, we experienced some issues ourselves with saving assessments in the central database which led to a loss of some research data. These issues should be fixed to improve the user-friendliness for farmers and help avoid a future loss of data. For both apps, but especially for PIGLOW, farmers indicated that they would like more depth in the app and more practical tips and targeted solutions as part of the feedback on results. For some farmers, it was not clear if they are just asking for increased depth of the results or also for the content of the welfare assessment. For the content of the welfare assessment, it was a very conscious choice not to increase the depth by asking more questions or more complicated questions. We made this choice based on surveys that were filled out by NPG members of the PPILOW project, of which the results can be found in Deliverable 3.1. In this survey, NPG members were asked what they thought would be a realistic time investment for a welfare assessment with the app. The majority of both pig and poultry farmers answered that they would not like to spend more than 1 hour on such an assessment, which was taken into account when designing the welfare assessments. If questions about additional welfare topics were added to the apps or scoring methods for questions were made more complicated, the necessary time investment for the apps would increase and this might cause farmers to lose the motivation to use the apps. In

addition, other studies on welfare scoring by people who are not yet experts have suggested that when the goal of the assessment is to gain information about animal welfare on farm level, less detailed scoring systems can actually be more suitable. This is the case because a less detailed score leads to a higher scoring accuracy and the loss of precision from giving less detailed scores does not hinder the goal (Gibbons et al., 2012; Katzenberger et al., 2020; Michaelis, 2022). An addition of more (automated) feedback would not necessarily ask for a larger time investment, as farmers would only have to read the feedback for the welfare indicators that they are interested in. Additionally, they can choose how much time they are willing to spend on this after having already completed the welfare assessment. Thus, it is worth exploring possibilities to improve the automated feedback of both apps, perhaps by including more practical examples of occurrences of certain welfare problems and how other farmers have solved them. However, the difficulty for improving the feedback lies in the fact that it has to be automated for each possible app user without specific knowledge of the farms' characteristics. Researchers working on a similar study also stated that as problems on farms can be related to multiple risk factors and farm specific conditions, it is very difficult to provide feedback that is tailored to a specific farm (van Dijk et al., 2018). Providing targeted solutions for a specific farm will therefore unfortunately not be possible through an app like PIGLOW or EBENE®. We can only try to increase the number of examples of possible causes for welfare problems in the hope that farmers can recognise which cause might be relevant on their farm. In addition, while changing the automated feedback care must be taken that the feedback is still relevant for the wide range of farms that exists within the categories of organic and free-range farms.

The question for more depth also illustrates that perhaps the participants of our study are not the best possible target audience for these apps. Some answers to survey questions by pig farmers showed that the farmers think the PIGLOW app could be more useful for people who are new to the sector to learn about animal welfare. Thus, while we can change the apps to fit better with the needs of the farmers who were a part of our study, we can also try to specifically target farmers, and perhaps others in the field, who are new to the sector or who do not have as much knowledge on animal welfare yet for other reasons. We have already made a start on targeting this group. For instance, in several countries such as France and Italy, new farmers have to follow mandatory training sessions of several days when they start their own farm. We have started to make connections with some of the teachers of those training sessions in the hope that if they believe in the benefits of the welfare self-assessment apps, they will present the apps to the farmers they train. This could help the apps reach many new users.

Differences between farmers and researchers

The second main question that was asked in this study was how animal welfare assessments with the PIGLOW and EBENE® app by farmers and researchers compare to one another and what insights the differences between the two can give us about the welfare assessments in the app. The results showed that the level of agreement differed greatly per indicator, which makes it difficult to draw conclusions about the overall level of agreement between the groups. In previous studies on this topic, the opposite is true and researchers have almost exclusively drawn conclusions about overall levels of agreement rather than about individual welfare indicators. The level of agreement between farmers and researchers was relatively high in a previous study on dairy cows, beef cattle, pigs and poultry (Michaelis, 2022), but low in the study by Katzenberger et al. (2020) on dairy cows. However, in the last study farmer and researcher did not observe simultaneously, which could explain the low level of agreement. A third study in which pig, poultry and dairy cow farmers and researchers scored animal welfare only stated that farmers for whom the inter observer reliability was low were excluded from the study (Schultheiß et al., 2023), but did not report anything on the remaining farmers. Unfortunately, none of the studies on relevant species discuss any differences

in levels of agreement for different indicators, making it impossible to make comparisons with our study.

What must be kept in mind when drawing conclusions about the differences between farmers and researchers is that we do not know what the true value for each of the observed welfare indicators on each of the farms was. Although we might be inclined to assume that researchers will measure more accurately than farmers because they have more experience performing structured welfare assessments, this does not have to be the case. Farmers could also see more than researchers in some cases, because they see their animals every day and might notice differences with previous days that researchers who see the animals for the first time cannot notice. Farmers might also already know before starting the welfare assessment that the animals have a certain welfare problem that the researcher does not know about and could therefore be more alert to detect animals with that welfare problem. Therefore, we have to be very careful when drawing conclusions about the differences in assessments by farmers and researchers. This also means that we cannot use the data of these welfare assessments to definitively say whether farmers measure accurately or not and whether their welfare assessments are reliable. We can only discuss the most probable explanations for differences between farmers and researchers and potentially use these explanations to make changes to the apps that will make it easier for the farmers to measure welfare as well as possible. What the results also show is that differences between farmers and researchers are generally larger for welfare indicators with a higher occurrence. As many welfare problems had low occurrences, it is possible that there would have been more differences between farmers and researchers in the way of scoring those indicators if there had been more occurrences on which they could disagree. This must be kept in mind when comparing the differences between farmers and researchers for different welfare indicators.

This being said, results of the comparison of welfare assessments by farmers and researchers do give us some interesting information on general differences between farmers and researchers and on small changes that could be made to the apps to increase the understanding of users. Results for both pigs and poultry seem to suggest that there are some welfare indicators that are easier to measure by people who have experience with conducting structured welfare assessments, such as counting scratches on pigs or listening for coughing and sneezing while observing other indicators, and observing behaviour during 5-minute behavioural observations for poultry. For some of those indicators, small changes to the app could perhaps help farmers to improve quicker. For instance, in the PIGLOW app a notification could be added at the beginning of the assessment to remember to listen whether you hear any coughing or sneezing during the assessment in order to be able to answer a question about that at the end of the assessment. For the behavioural observations for chickens, a description of the 6 behaviours that are observed is available in the app, but needs to opened by clicking somewhere on the screen. Because the results suggest that farmers are too quick to count occurrences of rare behaviours that do not actually fit the definition of that behaviour, it could be helpful to state specifically that users should read the description of the behaviours before starting in order to avoid scoring the wrong behaviours. For pigs, the fact that farmers scored pigs as too small more often than researchers suggested that perhaps some farmers were inclined to judge too severely for welfare indicators that also have economical relevance. While there is a description of what we define as “Too small”, currently app users need to click somewhere on the screen for this description to be shown. As with some other information in the apps, it would perhaps be helpful if this description was visible directly with the question and farmers do not need to specifically choose to read the extra information. For poultry farmers, this effect was not seen and farmers did not score more chickens as small than the researchers did.

Finally, for both pigs and poultry there was one welfare indicator for which almost none of the farmers and researchers agreed on a score, which were enrichment use and resting, respectively. For those indicators, the data seemed to suggest that it might have been difficult to score accurately independent of experience. That the scores for these indicators are likely somewhat less accurate

is not an immediate reason for concern. As previously mentioned, the main goal of the use of the welfare self-assessment apps is to help farmers look at their animals in a different way than they are used to and make them aware of possible welfare problems and early signs pointing to those problems. This is something that can still be accomplished by measuring these 2 welfare indicators. However, the fact that these scores might not be completely accurate is something that should be kept in mind for benchmarking, because it does make a fair comparison between farms more difficult. In addition, for enrichment use, it could help to add a notification to the app that someone should observe this indicator for a longer time if they wanted to have a more representative view of enrichment use over time.

The differences in the level of agreement between farmers and researchers at the beginning and end of the study was very small for both pigs and poultry. Other studies that have looked at the level of agreement between farmers and researchers did not make comparisons over time, thus it is not possible to relate this result to previous studies. It was assumed that farmers were less experienced with structured welfare assessments at the beginning of the study and would gain experience during the study. If experience was an important factor in the correct use of the PIGLOW and EBENE® app, that learning effect should have led to a higher level of agreement between farmers and researchers at the end of the study. However, this was not supported by the data. This could be the case because the farmers had not yet used the app often enough for this learning effect to be visible or because overall, experience with structured welfare assessments does not have a large influence on average scores given for each welfare indicator.

7. Conclusions

The results of this study did not support that the use of the PIGLOW and EBENE® app had an overall effect on the welfare of pigs and poultry on the participating farms. Farmers indicated that they found the apps easy to use and that they found most of their elements quite useful, but when asked about the specific influence of the app on their opinions on animal welfare and the welfare of their animals, they indicated that the apps only had small to medium-sized effects. It seems that the majority of the farmers did not personally experience large benefits from the use of the apps, but that they did see the potential of the apps for helping other farmers. Responses show that these apps could likely be very helpful to new farmers and other newcomers in the sector, who still have more to learn about animal welfare. In addition, the apps could be improved by resolving some small technical issues, adding more detailed and practical automated feedback to the results, and making small changes to questions that will increase user-friendliness.

A comparison of animal welfare assessments by farmers and researchers showed that there were some differences in how the two groups measured welfare that could be related to their different previous experiences. The specific differences have offered some useful insights into small changes that could be made to the app to avoid misunderstandings on how certain welfare indicators should be measured. The level of agreement between farmers and researchers did not significantly change between the beginning and end of the study, thus a learning effect of farmers during the study is not supported.

8. References

Aguado, E., Pascaretti-Grizon, F., Goyenvalle, E., Audran, M., & Chappard, D. (2015). Bone mass and bone quality are altered by hypoactivity in the chicken. *PloS one*, *10*(1), e0116763.

Åkerfeldt, M. P., Gunnarsson, S., Bernes, G., & Blanco-Penedo, I. (2021). Health and welfare in organic livestock production systems—a systematic mapping of current knowledge. *Organic Agriculture*, *11*(1), 105-132.

Alban, L., Petersen, J. V., & Busch, M. E. (2015). A comparison between lesions found during meat inspection of finishing pigs raised under organic/free-range conditions and conventional, indoor conditions. *Porcine Health Management*, *1*, 1-11.

Atkinson, D. J., von Keyserlingk, M. A. G., & Weary, D. M. (2017). Benchmarking passive transfer of immunity and growth in dairy calves. *Journal of Dairy Science*, *100*(5), 3773-3782. https://doi.org/https://doi.org/10.3168/jds.2016-11800

Baumgartner, J., Leeb, T., Gruber, T., & Tiefenbacher, R. (2003). Husbandry and animal health on organic pig farms in Austria. *Animal Welfare*, *12*(4), 631-635.

Bergmann, S., Louton, H., Westermaier, C., Wilutzky, K., Bender, A., Bachmeier, J., Erhard, M., & Rauch, E. (2016). Field trial on animal-based measures for animal welfare in slow growing broilers reared under an alternative concept suitable for the German market. *Berl. Münch. Tierärztl. Wochenschr*, *129*, 453-461.

Blokhuis, H. J., Veissier, I., Miele, M., & Jones, B. (2010). The Welfare Quality® project and beyond: Safeguarding farm animal well-being. *Acta Agriculturae Scandinavica, Section A — Animal Science*, *60*(3), 129-140. https://doi.org/10.1080/09064702.2010.523480

Burke, J. R., S.; Huxley, J. N.; Whay, H. R.; Main, D. C. J. . (2004, 20-22 April, 2004). *Health and welfare benchmarking as a tool for the development of dairy herd health plans on organic farms.* Organic farming: science and practice for profitable livestock and cropping, Shropshire.

Burke, J. (2006). Welfare benchmarking and herd health plans on organic dairy farms.

Capua, I., & Marangon, S. (2006). Control of avian influenza in poultry. *Emerging Infectious Diseases*, *12*(9), 1319.

Commission of the European Communities. (2008). EC No 889/2008 laying down detailed rules for the implementation of Council Regulation (EC) No 834/2007 on organic production and labelling of organic products with regard to organic production, labelling and control. *Official Journal of the European Union*.

Cornale, P., Macchi, E., Miretti, S., Renna, M., Lussiana, C., Perona, G., & Mimosi, A. (2015). Effects of stocking density and environmental enrichment on behavior and fecal corticosteroid levels of pigs under commercial farm conditions. *Journal of veterinary behavior*, *10*(6), 569-576.

Council of the European Union. (2007). EC No 834/2007 on organic production and labelling of organic products and repealing Regulation (EEC) No 2092/91. *Official Journal of the European Union*.

Dawkins, M. S., Cook, P. A., Whittingham, M. J., Mansell, K. A., & Harper, A. E. (2003). What makes free-range broiler chickens range? In situ measurement of habitat preference. *Animal Behaviour*, *66*(1), 151-160.

Day, J., Kelly, H., Martins, A., & Edwards, S. (2003). Towards a baseline assessment of organic pig welfare. *Animal Welfare*, *12*(4), 637-641.

Delsart, M., Pol, F., Dufour, B., Rose, N., & Fablet, C. (2020). Pig farming in alternative systems: strengths and challenges in terms of animal welfare, biosecurity, animal health and pork safety. *Agriculture*, *10*(7), 261.

Dippel, S., Leeb, C., Bochicchio, D., Bonde, M., Dietze, K., Gunnarsson, S., Lindgren, K., Sundrum, A., Wiberg, S., & Winckler, C. (2014). Health and welfare of organic pigs in Europe assessed with animal-based parameters. *Organic Agriculture*, *4*, 149-161.

Edwards, S., Prunier, A., Bonde, M., & Stockdale, E. (2014). organic pig production in Europe—animal health, welfare and production challenges. *Organic Agriculture*, *4*, 79-81.

Etterlin, P. E., Morrison, D. A., Österberg, J., Ytrehus, B., Heldmer, E., & Ekman, S. (2015). Osteochondrosis, but not lameness, is more frequent among free-range pigs than confined herd-mates. *Acta Veterinaria Scandinavica*, *57*, 1-10.

Etterlin, P. E., Ytrehus, B., Lundeheim, N., Heldmer, E., Österberg, J., & Ekman, S. (2014). Effects of free-range and confined housing on joint health in a herd of fattening pigs. *BMC veterinary research*, *10*, 1-14.

The Welfare of Chickens kept for Meat Production (Broilers). Report of the Scientific Committee on Animal Health and Animal Welfare., (2000). http://europa.eu.int/comm/dg24/health/sc/scah/index–en.html.

COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS ON AN ACTION PLAN FOR THE DEVELOPMENT OF ORGANIC PRODUCTION, (2021). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021DC0141R%2801%29

European Commission. (2023). Organic farming in the EU. A decade of organic growth. https://agriculture.ec.europa.eu/system/files/2023-04/agri-market-brief-20-organic-farming-eu_en.pdf

European Parliamentary Research Service. (2019). The EU poultry meat and egg sectors. Main features, challenges and prospects. . https://www.europarl.europa.eu/RegData/etudes/IDAN/2019/644195/EPRS_IDA(2019)644195_EN.pdf

Fanatico, A., Pillai, P., Cavitt, L., Owens, C., & Emmert, J. (2005). Evaluation of slower-growing broiler genotypes grown with and without outdoor access: Growth performance and carcass yield. *Poultry Science*, *84*(8), 1321-1327.

Fleming, P. A., Dundas, S. J., Lau, Y. Y., & Pluske, J. R. (2016). Predation by red foxes (Vulpes vulpes) at an outdoor piggery. *Animals*, *6*(10), 60. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5082306/pdf/animals-06-00060.pdf

Früh, B., Bochicchio, D., Edwards, S., Hegelund, L., Leeb, C., Sundrum, A., Werne, S., Wiberg, S., & Prunier, A. (2014). Description of organic pig production in Europe. *Organic Agriculture*, *4*, 83-92.

Gareis, M., Oberlaender, S., Zipplies, J., Reese, S., Schade, B., Boehm, B., & Schwaiger, K. (2016). Prevalence of auxiliary bursae and injuries of claws in fattening pigs at time of slaughter-results of a study at four slaughterhouses. *Berliner und Münchener Tierärztliche Wochenschrift*, *129*(9/10), 428-436.

German Animal Welfare Act (Tierschutzgesetz). (2006). Retrieved 25 March 2024 from https://www.gesetze-im-internet.de/tierschg/BJNR012770972.html

Gibbons, J., Vasseur, E., Rushen, J., & de Passillé, A. M. (2012). A training programme to ensure high repeatability of injury scoring of dairy cows. *Animal Welfare*, *21*(3), 379-388. https://doi.org/10.7120/09627286.21.3.379

Grimberg-Henrici, C., Büttner, K., Lohmeier, R., Burfeind, O., & Krieter, J. (2019). The effect of group-housing with free-farrowing pens on reproductive traits and the behaviour of low-risk and high-risk crushing sows. *Applied Animal Behaviour Science*, *211*, 33-40.

Hales, J., Moustsen, V., Nielsen, M., & Hansen, C. (2014). Higher preweaning mortality in free farrowing pens compared with farrowing crates in three commercial pig farms. Animal, 8: 113-120. In.

Hales, J., Moustsen, V., Nielsen, M., & Hansen, C. (2015). Temporary confinement of loose-housed hyperprolific sows reduces piglet mortality. *Journal of animal science*, *93*(8), 4079-4088.

Holinger, M., Früh, B., Stoll, P., Graage, R., Wirth, S., Bruckmaier, R., Prunier, A., Kreuzer, M., & Hillmann, E. (2018a). Chronic intermittent stress exposure and access to grass silage interact differently in their effect on behaviour, gastric health and stress physiology of entire or castrated male growing-finishing pigs. *Physiology & Behavior*, *195*, 58-68.

Holinger, M., Früh, B., Stoll, P., Kreuzer, M., & Hillmann, E. (2018b). Grass silage for growing-finishing pigs in addition to straw bedding: Effects on behaviour and gastric health. *Livestock Science*, *218*, 50-57.

Høøk Presto, M., Algers, B., Persson, E., & Andersson, H. (2009). Different roughages to organic growing/finishing pigs—Influence on activity behaviour and social interactions. *Livestock Science*, *123*(1), 55-62.

Høøk Presto, M., Andersson, H. K., Folestam, S., & Lindberg, J. E. (2008). Activity behaviour and social interactions of pigs raised outdoors and indoors. *Archives Animal Breeding*, *51*(4), 338-350.

Ivemeyer, S., Bell, N. J., Brinkmann, J., Cimer, K., Gratzer, E., Leeb, C., March, S., Mejdell, C., Roderick, S., Smolders, G., Walkenhorst, M., Winckler, C., & Vaarst, M. (2015). Farmers taking responsibility for herd health development—stable schools in research and advisory activities as a tool for dairy health and welfare planning in Europe. *Organic Agriculture*, *5*(2), 135-141. https://doi.org/10.1007/s13165-015-0101-y

Jensen, M. B., Studnitz, M., & Pedersen, L. J. (2010). The effect of type of rooting material and space allowance on exploration and abnormal behaviour in growing pigs. *Applied Animal Behaviour Science*, *123*(3-4), 87-92.

Katakam, K. K., Thamsborg, S. M., Dalsgaard, A., Kyvsgaard, N. C., & Mejer, H. (2016). Environmental contamination and transmission of Ascaris suum in Danish organic pig farms. *Parasites & Vectors*, *9*, 1-12.

Katzenberger, K., Rauch, E., Erhard, M., Reese, S., & Gauly, M. (2020). Inter-rater reliability of welfare outcome assessment by an expert and farmers of South Tyrolean dairy farming. *Italian Journal of Animal Science*, *19*(1), 1079-1090. https://doi.org/10.1080/1828051X.2020.1816509

Kijlstra, A., & Eijck, I. (2006). Animal health in organic livestock production systems: a review. *NJAS-Wageningen Journal of Life Sciences*, *54*(1), 77-94.

Kilbride, A. L., Mendl, M., Statham, P., Held, S., Harris, M., Cooper, S., & Green, L. E. (2012). A cohort study of preweaning piglet mortality and farrowing accommodation on 112 commercial pig farms in England. *Preventive veterinary medicine*, *104*(3-4), 281-291.

Knage-Rasmussen, K. M., Houe, H., Rousing, T., & Sørensen, J. T. (2014). Herd-and sow-related risk factors for lameness in organic and conventional sow herds. *Animal*, *8*(1), 121-127.

Kongsted, H., & Sørensen, J. T. (2017). Lesions found at routine meat inspection on finishing pigs are associated with production system. *The Veterinary Journal*, *223*, 21-26.

Leeb, C., Hegelund, L., Edwards, S., Mejer, H., Roepstorff, A., Rousing, T., Sundrum, A., & Bonde, M. (2014). Animal health, welfare and production problems in organic weaner pigs. *Organic Agriculture*, *4*, 123-133.

Leeb, C., Rudolph, G., Bochicchio, D., Edwards, S., Früh, B., Holinger, M., Holmes, D., Illmann, G., Knop, D., & Prunier, A. (2019). Effects of three husbandry systems on health, welfare and productivity of organic pigs. *Animal*, *13*(9), 2025-2033.

Leterrier, C., Vallée, C., Constantin, P., Chagneau, A.-M., Lessire, M., Lescoat, P., Berri, C., Baéza, E., Bizeray, D., & Bouvarel, I. (2008). Sequential feeding with variations in energy and protein levels improves gait score in meat-type chickens. *Animal*, *2*(11), 1658-1665.

Lindgren, K., Bochicchio, D., Hegelund, L., Leeb, C., Mejer, H., Roepstorff, A., & Sundrum, A. (2014). Animal health and welfare in production systems for organic fattening pigs. *Organic Agriculture*, *4*, 135-147.

Lindgren, Y., Lundeheim, N., Boqvist, S., & Magnusson, U. (2013). Reproductive performance in pigs reared under organic conditions compared with conventionally reared pigs. *Acta Veterinaria Scandinavica*, *55*, 1-4.

Luković, Z., Škorput, D., & Karolyi, D. (2017). Pig welfare at different production systems. Proceedings of the 11th International Symposium,

Lund, V., & Algers, B. (2003). Research on animal health and welfare in organic farming—a literature review. *Livestock Production Science*, *80*(1-2), 55-68.

Lund, V., Nielsen, L., Oliveira, A., & Christensen, J. (2017). Evaluation of the Danish footpad lesion surveillance in conventional and organic broilers: Misclassification of scoring. *Poultry Science*, *96*(7), 2018-2028.

March, S., Brinkmann, J., & Winckler, C. (2014). Improvement of animal health in organic dairy farms through 'stable schools': selected results of a pilot study in Germany. *Organic Agriculture*, *4*(4), 319-323. https://doi.org/10.1007/s13165-014-0071-5

Michaelis, Schubbert, Gieseke, Cimer, Zapf, Lühken, March, Brinkmann, Schultheiß, & Knierim. (2022). A comparison of online and live training of livestock farmers for an on-farm self-assessment of animal welfare. *Front. Anim. Sci.*, *Volume 3 - 2022*.

Nalon, E., Conte, S., Maes, D., Tuyttens, F., & Devillers, N. (2013). Assessment of lameness and claw lesions in sows. *Livestock Science*, *156*(1-3), 10-23.

Nansen, P., & Roepstorff, A. (1999). Parasitic helminths of the pig: factors influencing transmission and infection levels. *International journal for parasitology*, *29*(6), 877-891.

Newberry, R. C. (2017). Commercial free-range egg production practices. In *Egg innovations and strategies for improvements* (pp. 89-102). Elsevier.

Nielsen, S. S., Alvarez, J., Bicout, D. J., Calistri, P., Canali, E., Drewe, J. A., Garin-Bastuji, B., Gonzales Rojas, J. L., & Herskin, M. (2021). EFSA Panel on Animal Health

Welfare - African swine fever and outdoor farming of pigs. *EFSA Journal*, *19*(6), e06639.

Papatsiros, V. (2011). Impact of animal health management on organic pig farming in Greece. *Biotechnology in Animal Husbandry*, *27*(1), 115-125.

Parrini, S., Acciaioli, A., Franci, O., Pugliese, C., & Bozzi, R. (2018). Grazing behaviour of Cinta senese and its crossbreed pigs. *Italian Journal of Animal Science*.

Pfeifer, M., Koch, A., Lensches, C., Schmitt, A. O., & Hessel, E. F. (2020). Acceptance and feasibility of a guideline for the animal welfare assessment of fattening pigs from farmers' point of view. *Animals*, *10*(4), 711.

Pietrosemoli, S., & Tang, C. (2020). Animal welfare and production challenges associated with pasture pig systems: A review. *Agriculture*, *10*(6), 223.

Presto Åkerfeldt, M., Nihlstrand, J., Neil, M., Lundeheim, N., Andersson, H., & Wallenbeck, A. (2019). Chicory and red clover silage in diets to finishing pigs—influence on performance, time budgets and social interactions. *Organic Agriculture*, *9*, 127-138.

Presto, M., Rundgren, M., & Wallenbeck, A. (2013). Inclusion of grass/clover silage in the diet of growing/finishing pigs–Influence on pig time budgets and social behaviour. *Acta Agriculturae Scandinavica, Section A–Animal Science*, *63*(2), 84-92.

Rangstrup-Christensen, L., Krogh, M., Pedersen, L., & Sorensen, J. (2017). Sow-level risk factors for stillbirth of piglets in organic sow herds. Animal 11 (6): 1078–1083. In.

Rangstrup-Christensen, L., Krogh, M. A., Pedersen, L. J., & Sørensen, J. T. (2018a). Sow level risk factors for early piglet mortality and crushing in organic outdoor production. *Animal*, *12*(4), 810-818.

Rangstrup-Christensen, L., Schild, S.-L. A., Pedersen, L. J., & Sørensen, J. T. (2018b). Causes of preweaning mortality in organic outdoor sow herds. *Research in veterinary science*, *118*, 171-180.

Rezaei, M., Yngvesson, J., Gunnarsson, S., Jönsson, L., & Wallenbeck, A. (2018). Feed efficiency, growth performance, and carcass characteristics of a fast-and a slower-growing broiler hybrid fed low-or high-protein organic diets. *Organic Agriculture*, *8*, 121-128.

Schultheiß, U., Zapf, R., Brinkmann, J., Cimer, K., March, S., Schrader, L., Schubbert, A., Lühken, S., Gieseke, D., & Michaelis, S. (2023). Tools for on-farm self-assessment of animal welfare. *LANDTECHNIK*, *78*(3), 114-123.

Stadig, L. M., Rodenburg, T. B., Ampe, B., Reubens, B., & Tuyttens, F. A. (2017). Effect of free-range access, shelter type and weather conditions on free-range use and welfare of slow-growing broiler chickens. *Applied Animal Behaviour Science*, *192*, 15-23.

Stahl, P., Ruette, S., & Gros, L. (2002). Predation on free-ranging poultry by mammalian and avian predators: field loss estimates in a French rural area. *Mammal review*, *32*(3), 227-234.

Sumner, C. L., von Keyserlingk, M. A. G., & Weary, D. M. (2018). How benchmarking motivates farmers to improve dairy calf management. *Journal of Dairy Science*, *101*(4), 3323-3333. https://doi.org/https://doi.org/10.3168/jds.2017-13596

Sundrum, A., Goebel, A., Bochicchio, D., Bonde, M., Bourgoin, A., Cartaud, G., Dietze, K., Dippel, S., Gunnarsson, S., & Hegelund, L. (2010). Health status in organic pig herds in Europe. Proc. of the 21st Int. Pig Veterinary Society (IPVS) Congress,

Sutherland, M. A., Webster, J., & Sutherland, I. (2013). Animal health and welfare issues facing organic production systems. *Animals*, *3*(4), 1021-1035.

Tahamtani, F. M., Hinrichsen, L. K., & Riber, A. B. (2018). Welfare assessment of conventional and organic broilers in Denmark, with emphasis on leg health. *Veterinary Record*, *183*(6), 192-192.

Terlouw, C., Berne, A., & Astruc, T. (2009). Effect of rearing and slaughter conditions on behaviour, physiology and meat quality of Large White and Duroc-sired pigs. *Livestock Science*, *122*(2-3), 199-213.

Thys, M., Watteyn, A., Vaes, R., Vandepoel , G., Sonck, B., & Tuyttens, F. (2019). Cattle welfare self-assessment and benchmarking tool for farmers via mobile application. 70th Annual Meeting of the European Federation of Animal Science, Ghent, Belgium.

Tuyttens, F., Heyndrickx, M., De Boeck, M., Moreels, A., Van Nuffel, A., Van Poucke, E., Van Coillie, E., Van Dongen, S., & Lens, L. (2008). Broiler chicken health, welfare and fluctuating asymmetry in organic versus conventional production systems. *Livestock Science*, *113*(2-3), 123-132.

Vaarst, M., & Fisker, I. (2013). Potential contradictions connected to the inclusion of stable schools in the legislation for Danish organic dairy farms. *The Open Agriculture Journal*, *Vol. 7*, 118-124.

Van De Weerd, H., Keatinge, R., & Roderick, S. (2009). A review of key health-related welfare issues in organic poultry production. *World's Poultry Science Journal*, *65*(4), 649-684.

van Dijk, L., Elwes, S., Main, D. C. J., Mullan, S. M., & Jamieson, J. (2018). Farmer perspectives on welfare outcome assessment: learnings from four farm assurance scheme consultation exercises. *Animal Welfare*, *27*(1), 1-11. https://doi.org/10.7120/09627286.27.1.001

van Eerdenburg, F. J. C. M., Di Giacinto, A. M., Hulsen, J., Snel, B., & Stegeman, J. A. (2021). A New, Practical Animal Welfare Assessment for Dairy Farmers. *Animals*, *11*(3).

von Hardenberg, L., & Heise, H. (2018). German pig farmers' attitudes towards animal welfare programs and their willingness to participate in these programs: An empirical study. *International Journal on Food System Dynamics*, *9*(3), 289-301.

Walker, P., & Bilkei, G. (2006). Tail-biting in outdoor pig production. *The Veterinary Journal*, *171*(2), 367-369.

Wallenbeck, A., Eliasson, C., Lundeheim, N., & Nilsson, K. (2020). Leg health, growth and carcass characteristics in growing-finishing pigs of two different genotypes reared on Swedish organic farms. *Organic Agriculture*, *10*, 97-103.

Wallenbeck, A., Wilhelmsson, S., Jönsson, L., Gunnarsson, S., & Yngvesson, J. (2016). Behaviour in one fast-growing and one slower-growing broiler (Gallus gallus domesticus) hybrid fed a high-or low-protein diet during a 10-week rearing period. *Acta Agriculturae Scandinavica, Section A—Animal Science*, *66*(3), 168-176.

Weeks, C., Nicol, C., Sherwin, C., & Kestin, S. (1994). Comparison of the behaviour of broiler chickens in indoor and free-range environments. *Animal Welfare*, *3*(3), 179-192.

Westin, R., Holmgren, N., Hultgren, J., Ortman, K., Linder, A., & Algers, B. (2015). Post-mortem findings and piglet mortality in relation to strategic use of straw at farrowing. *Preventive veterinary medicine*, *119*(3-4), 141-152.

Wilhelmsson, S., Yngvesson, J., Jönsson, L., Gunnarsson, S., & Wallenbeck, A. (2019). Welfare Quality® assessment of a fast-growing and a slower-growing broiler hybrid, reared until 10 weeks and fed a low-protein, high-protein or mussel-meal diet. *Livestock Science*, *219*, 71-79.

Zapf, R., Schultheiß, U., Achilles, W., Schrader, L., Knierim, U., Herrmann, H.-J., Brinkmann, J., & Winckler, C. (2015). Indicators for on-farm self-assessment of animal welfare—Example: Dairy cows. *LANDTECHNIK Agric. Eng*, *70*, 221-230.

9. Annex

A. Consent forms for participants of the longitudinal study - pigs

PARTICIPANT INFORMATION FORM

Study: Testing and evaluating animal welfare self-assessment and benchmarking tools

Dear participant,

We kindly invite you to participate in our study 'Testing and evaluating animal welfare self-assessment and benchmarking tools' of project PPILOW (www.ppilow.eu). Before consenting to participate, it is however important that you read the following information with clear attention. If anything is unclear, please do not hesitate to ask further information.

Project: PPILOW (Poultry and PIg Low-input and Organic production systems' Welfare)

Funding: The PPILOW project aiming at improving the welfare of pigs and poultry in low-input outdoor and organic production systems has received funding from the European Union Horizon 2020 'Research and Innovation' Programme under Grant Agreement No. 816172. This project gathers 23 contractually-engaged partners who signed a Consortium Agreement version 2019-04-25). The information provided here reflects the views of its authors. The Executive Agency for Research of the European Commission cannot be held responsible for the use of the information provided.

Introduction

XXX (organism) offers to participate in a research multi-actor project, based on National Practitioner Groups networking for proposing and testing innovations for improving Pig and Poultry welfare in low input outdoor and/organic production systems. The legal basis for this project is the performance of a public interest task requiring your consent in accordance with the general European Data Protection Regulation nº 2016/67.

Before deciding to take part in this research project, please take the time to read the following information. You are free to agree or refuse to take part in this research. If you do agree you may also choose not to answer all the questions put to you, or at any time stop contributing without having to explain why.

Context of the PPILOW research project

The aim of the European PPILOW project is to propose and study innovations related to animal welfare in low-input outdoor and organic poultry and pig production systems. In order to choose and test the best possible means of action and to facilitate change, we need to closely interact with practitioners throughout Europe for knowledge transfer and co-building of strategies. To do this, we

would like to invite you to participate in this study. If you choose to participate, we thank you for agreeing to contribute to the 'PPILOW' project.

Objective/description of the research

Welfare self-assessment apps have been developed as part of the PPILOW project co-created with National Practitioner Groups (NPG). One of the main goals of the PPILOW project is to test how effective animal welfare self-assessment via a mobile application by the farmer combined with automated feedback (including anonymous benchmarking) is in improving the welfare of animals in commercial organic and low input broiler chicken and pig production systems.

In addition, we will be testing agreement in scoring of animal welfare measures between farmers and trained researchers, and we will be collecting data for the central data base that will be useful for researchers and the farming sector to highlight positive aspects of animal welfare as well as document animal welfare issues in low-input farming systems and to identify differences in time or between systems.

The study will be conducted by an international team of researchers from different partner institutes (Frank Tuyttens and Evelien Graat from ILVO, Sophie Herremans from CRA-W, Laura van Vooren and An Jamart from BioForum, Laura Warin, Mathilde Stomp, Marion Pertusa and Claire Bonnefous from ACTA - ITAVI, Bas Rodenburg and Mona Giersberg from Utrecht University, and Anne Collin from INRAE) on farms in Belgium, The Netherlands and France.

Process description

We aim to include 20 to 30 broiler chicken farms and 20 to 30 pig farms from three (BE, NL, FR) and two (BE, NL) countries, respectively. Each farm will be visited by (a) researcher(s) at the onset of the study. During this visit, the researcher will explain how to use the animal welfare self-assessment tool (smartphone app). The farmer and researcher then carry out the animal welfare assessment (using the app) at the same time but independently of one another. These data will be used for testing the agreement between the scores of the farmer with those of the trained researcher and as baseline data in the central database. During this visit, the researcher will also conduct additional welfare assessments that are similar to the self-assessment app, but more extensive and refined. The farmer will be required to periodically (at least three times a year) carry out the self-assessment scan after which he receives automated feedback (including anonymous benchmarking with other comparable farms). This feedback details which animal welfare scores have improved or worsened since the previous assessment(s) and how these scores compare to the scores of other anonymous farms as derived from the central reference database. The feedback will direct the farmer's attention to aspects of, for example, management and housing that are related to animal welfare and, if relevant, encourage the farmer to discuss the situation with a veterinarian or other expert and subsequently take actions to improve certain welfare problems. The farmer will be asked to complete the questionnaire "loading" in the PIGLOW app at least once during the study. This is a welfare assessment that should be completed during depopulation when the animals are being loaded into the transport vehicle. The farmer will also be asked to record and inform the researcher of any remedying actions that are taken or considered and of any relevant changes to housing or management. A form designed for this purpose will be provided. At the end of the study the researcher will visit each of the farms again and carry out the last animal welfare assessment at the same time as, but independently of, the farmer as well as conduct additional animal welfare assessments.

At the beginning as well as at the end of the two-year study, the farmer will be asked to fill out a survey containing questions about his/her ideas about animal welfare, his/her opinion on (the effect of) the self-assessment tool (app), and the One Welfare concept (including social, economic and

environmental aspects). Answering the questions about animal welfare and the self-assessment tool is a required part of the study, but the farmer is free to answer any or none of the additional questions related to the One Welfare concept. This will be an important and useful evaluation measure for our welfare self-assessment tool.

Participant selection

We received your contact details from the PPILOW National Practitioners Group based on expressed interest in volunteering to participate in this study or you were contacted based on the expressed interest of your company and yourself in the use of a welfare assessment tool or based on the information available on your low input outdoor and/or organic farm

be a farmer rearing either pigs or broiler chickens organically or with access to an outdoor range either in the Netherlands, Belgium or France.

Your rights regarding confidentiality and privacy:

For the use of the pig welfare self-assessment app, only an e-mail address will be collected. This is required for automated e-mailing of the output (report of the self-assessment) to the farmer. After uploading the data (collected by the farmer) via the pig welfare app, the e-mail address of the farmer will automatically be replaced by a unique artificial identifier (or pseudonym) prior to data processing and analysis. Thus, your e-mail address will not be stored together with your data.

All additional data that will be collected by the researcher (including the results of completed surveys and other information regarding animal and human welfare indicators and opinion of the farmers about the self-scan app) will be pseudonymized as well prior to processing and analysis. Hence, only pseudonymized data will be stored and analysed. All data that will be reported / made public will be fully anonymised. However, please be aware that publications may include quotes mentioning your general position in the supply chain (e.g. farmer, adviser) and your country of origin. Publication types may include reports to the European Commission and scientific papers. Relevant general outcomes may also be used to promote organic and/or free range poultry productions in specific websites. The private data will not be transmitted to any other recipient, nor used in any context other than that described above.

The information will be kept, under the best conditions of security and confidentiality, for the entire duration of the PPILOW project, i.e. 5 years, and the following 5 years (until September 2029) and will possibly be reused for contacting you for a subsequent project related to the topic, if you consent. At the end of this period, the information collected is intended to be archived, in accordance with the law, in a pseudonymized form, i.e. without any possibility of access to your identity.

Your rights regarding your questions:

You can ask questions about the research project at any time (before, during and after your participation) by contacting *partner* by email at *xxx@xxx (or by phone at xxx)*. In the event of unavailability, you can contact Frank Tuyttens from ILVO at the following address*:* Frank.Tuyttens@ilvo.vlaanderen.*be (or by telephone at +32(0)92722605)* for questions about this study and Anne Collin from INRAE via Anne.Collin@inrae.fr *(or by telephone at +33(0)247427929)* for general questions about the PPILOW project.

Voluntary participation

Participating in this study is completely voluntary, meaning that you have the right to refuse participation. You also have the right to end participation at any time, upon simple request to the researcher, even after signing the informed consent form (see the section below).

If you agree to participate in this research, we kindly ask you to keep this information sheet, and provide us with your informed consent, by completing and signing the attached consent form, in twofold. One copy will remain with you (the participant), and the second copy will be securely stored by the researcher.

Your right to withdraw from the study at any time

In accordance with the European Regulation on the protection of personal data and the National Data Protection Act, you have the right to access, rectify, oppose and delete information concerning you. Unless you object, the personal data collected during this research project may be the subject of a subsequent research project with a similar research purpose by the project partner *xxx* and under the same conditions of confidentiality and security. If you oppose further use of your data, all of your data (personal and knowledge provided to the involved organisms) will be deleted at the end of the project.

You can terminate your participation in the study at any given time and without delay, after informing the researcher. You can deactivate your self-scan user account online. You may also withdraw your consent for the use of your research data (meaning any non-personal information collected in the context of this project) for future projects and request the deletion of your personal information at any time. Consent for the use of your research data for this study can only be withdrawn until the time when analysis of that data has started. After that time, you can only request that your personal information is deleted, making the research data fully anonymous, and that your research data is deleted at the end of the project.

If you wish to exercise these rights and/or obtain information about yourself, please contact *XXX (xxx@xxx) or YYY (yyy@yyy).* Your decision to participate, refuse to participate, or cease participation will not affect future relationships with *involved organism(s).* In case you cease your activity within the PPILOW project before the intended end-date of this study you will not be allowed to claim for the financial compensation of €300 and the confidentiality clause will remain until 5 years after the project.

If, after contacting us, you consider that your rights and freedoms are not respected, you have the possibility to file a complaint with the *XXX (National organism for IT and freedom)* by mail: *address* or online at http://www...../

Associated risks

As appropriate privacy and data protection measures will be taken, there are no known risks associated with participating in this study related to this topic.

The researchers collecting the data for this study will visit all participating farms within several months. This could pose a risk for hygiene and the spread of pathogens between farms. The researchers are aware of this and will take all reasonable measures to ensure proper hygiene and reduce this risk as much as possible.

Associated benefits

The use of the welfare self-assessment app can help you gain a better understanding of the welfare of your animals through the means of feedback based on scientific research as well as benchmarking. During the visit of the researcher at the start of the study, you will have the opportunity to ask any question about the content and the use of the welfare self-assessment app. A good understanding of the app ensures that you can benefit maximally from its use. At the end of the study, the involved PPILOW partners will provide you with a personalized report on the results within your own farm as well as the overall study.

You get to play an active part in an innovative study that could increase the social license for the animal production sector, making you a pioneer for your peers.

Your participation in the study can also be used as a marketing tool. You will be able to communicate to the production chain and the public that you are concerned about animal welfare and that you are using a method designed by scientists to monitor the welfare of your animals.

Finally, a financial compensation of €300 will be given to you if you participate for the entire duration of the study.

Dissemination of the results

The effect of using the self-assessment tool (smartphone app) and being given automated feedback on the animal welfare status on farm will be tested by comparing differences in animal welfare outcomes before and after using the tools. The complete set of data will be useful for researchers and the farming sector to highlight positive aspects of animal welfare as well as document animal welfare issues in low-input farming systems and to identify differences in time or between systems.

This research will be disseminated in conferences, meetings with practitioners, videos, e-learning, project website www.ppilow.eu and published in conference proceedings and academic journal articles.

XXX (researcher) is accompanied by the Personal Data Protection Officer (*DPO, if any)* or the person who is responsible for data security and storage of its supervisory institution. His contact details are *Address; Tel: xxxx ; E-mail: xxx@yyy.*

In case you have any further questions on the research in the future, you can contact me.

Sincerely,

Name Researcher

Consent Form

To guarantee your privacy rights, we ask you to give your explicit consent (tick the corresponding boxes):

	Yes	No
1- I hereby certify that I have read the information on the PPILOW research project mentioned above and that I have obtained the answers to my questions	□	□
2- I have had the necessary time to reflect on my involvement in this study and I am aware that my participation is entirely voluntary	□	□
3- I agree to take part in this study and periodically carry out welfare self-assessments with the PIGLOW app	□	□
4- I agree to allow researchers to conduct animal welfare assessments of the pigs or broiler chickens on my farm	□	□
5- I agree to be photographed during the on-farm trial and that my image may be published on the project website or other communication or dissemination means. The pictures will be stored by *XXX (organism)* until 5 years after the end of the project (until August 2029).	□	□
6- I agree to record and inform the researcher of any remedying actions taken/considered and any relevant changes in housing or management	□	□
7- I accept that so-called sensitive information concerning the welfare of my animals, my own welfare and economic or environmental indicators concerning my farm may be collected in dedicated surveys	□	□
8- I agree that all the information collected in the context of this study may be published in publications as pseudonymized quotes (without my surname and first name being mentioned).	□	□
9- I agree that my personal data collected through this project may be the subject of a subsequent project to refine or consolidate the research outcomes resulting from this project (excluding any exploitation for commercial purposes), under the same conditions of confidentiality and security.	□	□

I have noted that I may withdraw my consent at any time.

Made in two original copies, one of which must be given to the volunteer by hand.

Date:

Name, first name of the researcher(s): Name, first name of the volunteer:

Mailing address or e-mail address: Mailing address or e-mail address:

Signature: Signature:

B. Consent forms for participants of the longitudinal study - Poultry

PARTICIPANT INFORMATION FORM

Study: Testing and evaluating animal welfare self-assessment and benchmarking tools

Dear participant,

We kindly invite you to participate in our study 'Testing and evaluating animal welfare self-assessment and benchmarking tools' of project PPILOW (www.ppilow.eu). Before consenting to participate, it is however important that you read the following information with clear attention. If anything is unclear, please do not hesitate to ask further information.

Project: PPILOW (Poultry and PIg Low-input and Organic production systems' Welfare)

Funding: The PPILOW project aiming at improving the welfare of pigs and poultry in low-input outdoor and organic production systems has received funding from the European Union Horizon 2020 'Research and Innovation' Programme under Grant Agreement No. 816172. This project gathers 23 contractually-engaged partners who signed a Consortium Agreement version 2019-04-25). The information provided here reflects the views of its authors. The Executive Agency for Research of the European Commission cannot be held responsible for the use of the information provided.

Introduction

XXX (organism) offers to participate in a research multi-actor project, based on National Practitioner Groups networking for proposing and testing innovations for improving Pig and Poultry welfare in low input outdoor and/organic production systems. The legal basis for this project is the performance of a public interest task requiring your consent in accordance with the general European Data Protection Regulation nº 2016/67

Before deciding to take part in this research project, please take the time to read the following information. You are free to agree or refuse to take part in this research. If you do agree you may also choose not to answer all the questions put to you, or at any time stop contributing without having to explain why.

Context of the PPILOW research project

The aim of the European PPILOW project is to propose and study innovations related to animal welfare in low-input outdoor and organic poultry and pig production systems. In order to choose and test the best possible means of action and to facilitate change, we need to closely interact with practitioners throughout Europe for knowledge transfer and co-building of strategies. To do this, we would like to invite you to participate in this study. If you choose to participate, we thank you for agreeing to contribute to the 'PPILOW' project.

Objective/description of the research

Welfare self-assessment apps have been developed as part of the PPILOW project co-created with National Practitioner Groups (NPG). One of the main goals of the PPILOW project is to test how effective animal welfare self-assessment via a mobile application by the farmer combined with automated feedback (including anonymous benchmarking) is in improving the welfare of animals in commercial organic and low input broiler chicken and pig production systems.

In addition, we will be testing agreement in scoring of animal welfare measures between farmers and trained researchers, and we will be collecting data for the central data base that will be useful for researchers and the farming sector to highlight positive aspects of animal welfare as well as document animal welfare issues in low-input farming systems and to identify differences in time or between systems.

The study will be conducted by an international team of researchers from different partner institutes (Frank Tuyttens and Evelien Graat from ILVO, Sophie Herremans from CRA-W, Laura van Vooren and An Jamart from BioForum, Laura Warin, Mathilde Stomp, Marion Pertusa and Claire Bonnefous from ACTA - ITAVI, Bas Rodenburg and Mona Giersberg from Utrecht University, and Anne Collin from INRAE) on farms in Belgium, The Netherlands and France.

Process description

We aim to include 20 to 30 broiler chicken farms and 20 to 30 pig farms from three (BE, NL, FR) and two (BE, NL) countries, respectively. The study for pigs will take two years. For broiler chickens, it will take either two years or one year, depending on the impact of avian influenza on the timeline of the study. Participating broiler farmers will be informed of the duration of the study before the end of the first year. Each farm will be visited by (a) researcher(s) at the onset of the study. During this visit, the researcher will explain how to use the animal welfare self-assessment tool (smartphone app). The farmer and researcher then carry out the animal welfare assessment (using the app) at the same time but independently of one another. These data will be used for testing the agreement between the scores of the farmer with those of the trained researcher and as baseline data in the central database. During this visit, the researcher will also conduct additional welfare assessments that are similar to the self-assessment app, but more extensive and refined. The farmer will be asked to complete three assessments with the app in the first year, including the initial one together with the researcher. In case the study lasts for two years, the farmer will be asked to perform two additional welfare assessments, amounting to a total of five assessments in two years. After each scan he will receive automated feedback (including anonymous benchmarking with other comparable farms). This feedback details which animal welfare scores have improved or worsened since the previous assessment(s) and how these scores compare to the scores of other anonymous farms as derived from the central reference database. The feedback will direct the farmer's attention to aspects of, for example, management and housing that are related to animal welfare and, if relevant, encourage the farmer to discuss the situation with a veterinarian or other expert and subsequently take actions to improve certain welfare problems. Following the depopulation of the chickens from the first welfare assessment, the farmer will be asked to fill out an online questionnaire on the depopulation practices they use on their farm. The farmer will also be asked to record and inform the researcher of any remedying actions that are taken or considered and of any relevant changes to housing or management. A form designed for this purpose will be provided. At the end of the study the researcher will visit each of the farms again and carry out the last animal welfare assessment at the same time as, but independently of, the farmer as well as conduct additional animal welfare assessments.

At the beginning as well as at the end of the study, the farmer will be asked to fill out a survey containing questions about his/her ideas about animal welfare, his/her opinion on (the effect of) the

self-assessment tool (app), and the One Welfare concept (including social, economic and environmental aspects). Answering the questions about animal welfare and the self-assessment tool is a required part of the study, but the farmer is free to answer any or none of the additional questions related to the One Welfare concept. This will be an important and useful evaluation measure for our welfare self-assessment tool.

Participant selection

We received your contact details from the PPILOW National Practitioners Group based on expressed interest in volunteering to participate in this study or you were contacted based on the expressed interest of your company and yourself in the use of a welfare assessment tool or based on the information available on your low input outdoor and/or organic farm

be a farmer rearing either pigs or broiler chickens organically or with access to an outdoor range either in the Netherlands, Belgium or France.

Your rights regarding confidentiality and privacy:

For the use of the poultry welfare self-assessment app, e-mail address, name, surname and farm ID will be collected. This is required for the account creation. After uploading the data (collected by the farmer) via the poultry welfare app, your identity will be manually replaced by a unique artificial identifier (or pseudonym) prior to data processing and analysis. Thus, your personal information will not be stored together with your data.

All additional data that will be collected by the researcher (including the results of completed surveys and other information regarding animal and human welfare indicators and opinion of the farmers about the self-scan app) will be pseudonymized as well prior to processing and analysis. Hence, only pseudonymized data will be stored and analysed. All data that will be reported / made public will be fully anonymised. However, please be aware that publications may include quotes mentioning your general position in the supply chain (e.g. farmer, adviser) and your country of origin. Publication types may include reports to the European Commission and scientific papers. Relevant general outcomes may also be used to promote organic and/or free range poultry productions in specific websites. The private data will not be transmitted to any other recipient, nor used in any context other than that described above.

The information will be kept, under the best conditions of security and confidentiality, for the entire duration of the PPILOW project, i.e. 5 years, and the following 5 years (until September 2029) and will possibly be reused for contacting you for a subsequent project related to the topic, if you consent. At the end of this period, the information collected is intended to be archived, in accordance with the law, in a pseudonymized form, i.e. without any possibility of access to your identity.

Your rights regarding your questions:

You can ask questions about the research project at any time (before, during and after your participation) by contacting *partner* by email at *xxx@xxx (or by phone at xxx)*. In the event of unavailability, you can contact Frank Tuyttens from ILVO at the following address*:* Frank.Tuyttens@ilvo.vlaanderen.*be (or by telephone +32(0)92722605)* for questions about this study and Anne Collin from INRAE via Anne.Collin@inrae.fr *(or by telephone at +33(0)247427929)* for general questions about the PPILOW project*.*

Voluntary participation

Participating in this study is completely voluntary, meaning that you have the right to refuse participation. You also have the right to end participation at any time, upon simple request to the researcher, even after signing the informed consent form (see the section below).

If you agree to participate in this research, we kindly ask you to keep this information sheet, and provide us with your informed consent, by completing and signing the attached consent form, in twofold. One copy will remain with you (the participant), and the second copy will be securely stored by the researcher.

Your right to withdraw from the study at any time

In accordance with the European Regulation on the protection of personal data and the National Data Protection Act, you have the right to access, rectify, oppose and delete information concerning you. Unless you object, the personal data collected during this research project may be the subject of a subsequent research project with a similar research purpose by the project partner *xxx* and under the same conditions of confidentiality and security. If you oppose further use of your data, all of your data (personal and knowledge provided to the involved organisms) will be deleted at the end of the project.

You can terminate your participation in the study at any given time and without delay, after informing the researcher. You may also withdraw your consent for the use of your research data (meaning any non-personal information collected in the context of this project) for future projects and request the deletion of your personal information at any time. Consent for the use of your research data for this study can only be withdrawn until the time when analysis of that data has started. After that time, you can only request that your personal information is deleted, making the research data fully anonymous, and that your research data is deleted at the end of the project.

If you wish to exercise these rights and/or obtain information about yourself, please contact *XXX (xxx@xxx) or YYY (yyy@yyy).* Your decision to participate, refuse to participate, or cease participation will not affect future relationships with *involved organism(s).* In case you cease your activity within the PPILOW project before the intended end-date of this study you will not be allowed to claim for the specific budget allocation for innovation implementation and the confidentiality clause will remain until 5 years after the project.

If, after contacting us, you consider that your rights and freedoms are not respected, you have the possibility to file a complaint with the *XXX (National organism for IT and freedom)* by mail: *address* or online at http://www...../

Associated risks

As appropriate privacy and data protection measures will be taken, there are no known risks associated with participating in this study related to this topic.

The researchers collecting the data for this study will visit all participating farms within several months. This could pose a risk for hygiene and the spread of pathogens between farms. The researchers are aware of this and will take all reasonable measures to ensure proper hygiene and reduce this risk as much as possible.

Associated benefits

The use of the welfare self-assessment app can help you gain a better understanding of the welfare of your animals through the means of feedback based on scientific research as well as benchmarking. During the visit of the researcher at the start of the study, you will have the opportunity to ask any question about the content and the use of the welfare self-assessment app. A good understanding of the app ensures that you can benefit maximally from its use. At the end of the study, the involved PPILOW partners will provide you with a personalized report on the results within your own farm as well as the overall study.

You get to play an active part in an innovative study that could increase the social license for the animal production sector, making you a pioneer for your peers.

Your participation in the study can also be used as a marketing tool. You will be able to communicate to the production chain and the public that you are concerned about animal welfare and that you are using a method designed by scientists to monitor the welfare of your animals.

Finally, a financial compensation of €300 will be given to you if you participate for the entire duration of the study.

Dissemination of the results

The effect of using the self-assessment tool (smartphone app) and being given automated feedback on the animal welfare status on farm will be tested by comparing differences in animal welfare outcomes before and after using the tools. The complete set of data will be useful for researchers and the farming sector to highlight positive aspects of animal welfare as well as document animal welfare issues in low-input farming systems and to identify differences in time or between systems.

This research will be disseminated in conferences, meetings with practitioners, videos, e-learning, project website www.ppilow.eu and published in conference proceedings and academic journal articles.

XXX (researcher) is accompanied by the Personal Data Protection Officer (*DPO, if any)* or the person who is responsible for data security and storage of its supervisory institution. His contact details are *Address; Tel: xxxx ; E-mail: xxx@yyy.*

In case you have any further questions on the research in the future, you can contact me.

Sincerely,

Name Researcher

Consent Form

To guarantee your privacy rights, we ask you to give your explicit consent (tick the corresponding boxes):

	Yes	No
1- I hereby certify that I have read the information on the PPILOW research project mentioned above and that I have obtained the answers to my questions	□	□
2- I have had the necessary time to reflect on my involvement in this study and I am aware that my participation is entirely voluntary	□	□
3- I agree to take part in this study and periodically carry out welfare self-assessments with the EBENE app	□	□
4- I agree to allow researchers to conduct animal welfare assessments of the pigs or broiler chickens on my farm	□	□
5- I agree to be photographed during the on-farm trial and that my image may be published on the project website or other communication or dissemination means. The pictures will be stored by *XXX (organism)* until 5 years after the end of the project (until August 2029).	□	□
6- I agree to record and inform the researcher of any remedying actions taken/considered and any relevant changes in housing or management	□	□
7- I accept that so-called sensitive information concerning the welfare of my animals, my own welfare and economic or environmental indicators concerning my farm may be collected in dedicated surveys	□	□
8- I agree that all the information collected in the context of this study may be published in publications as pseudonymized quotes (without my surname and first name being mentioned).	□	□
9- I agree that my personal data collected through this project may be the subject of a subsequent project to refine or consolidate the research outcomes resulting from this project (excluding any exploitation for commercial purposes), under the same conditions of confidentiality and security.	□	□

I have noted that I may withdraw my consent at any time.

Made in two original copies, one of which must be given to the volunteer by hand.

Date:

Name, first name of the researcher(s): Name, first name of the volunteer:

Mailing address or e-mail address: Mailing address or e-mail address

Signature: Signature: