Development and Evaluation of an Online Training-tool for the Assessment of Animal-based Welfare Parameters in Cattle

Josef SCHENKENFELDER (✉)
Christoph WINCKLER

Summary

Self-evaluation of animal welfare by farmers has recently been encouraged, e.g. by the Austrian organic farming association BIO AUSTRIA. However, rather little is known how a resource-efficient training of a large number of farmers could take place and which level of agreement might be achieved. For this purpose, in the present study, an online training-tool for the assessment of 10 animal-based parameters of dairy cattle welfare was established. This tool included online quizzes containing pictures or video clips of selected animal-based parameters which had to be assessed by the test persons. IOR as compared to a gold standard (calculated as Cohen's Kappa κ) was investigated. Furthermore, it was of interest whether practice in terms of repeated trials leads to improvement.

In total, 938 κ values from 111 users were obtained from the 10 different quizzes. The average agreement per quiz in round 1 reached values of $\kappa \geq 0.40$ (n = 58–100 users). For the parameters cleanliness and diarrhoea, κ exceeded 0.40 for all test persons in round 1. Agreement was lowest for the parameters body condition, hairless patches and lameness. Retaking the quizzes (round 2, n = 14–24) led to significant improvement of agreement for all parameters, except for hairless patches and lameness.

In conclusion, the results of this study are promising as regards the intended use of the training-tool. However, its potential to improve reliability of live on-farm assessments needs to be further investigated, e.g. with regard to transferability to live observations.

Key words

animal-based welfare parameters, cattle, online-training, inter-observer reliability

Division of Livestock Sciences, Department of Sustainable Agricultural Systems, University of Natural Resources and Life Sciences (BOKU), Gregor-Mendel-straße 33, A-1180 Vienna, Austria
✉ e-mail: josef.schenkenfelder@boku.ac.at

Received: May 15, 2017 | Accepted: August 8, 2017

ACKNOWLEDGEMENTS
The authors would like to thank the Austrian organic farmer's association 'BIO AUSTRIA' for their cooperation in this project and Bernhard Schenkenfelder for his support in programming the website.

Introduction

Livestock production systems are facing growing public concern for several reasons, e.g. intensification and specialization (Dörfler, 2007, Veissier et al., 2008), urbanization (van de Weerd and Sandilands, 2008) or changing human-animal-relationship (WBABMEL, 2015). Especially organic agriculture attempts to counter this trend by means of more stringent production regulations which also aim at providing animal-friendly husbandry conditions (Webster, 2001). Such provisions of good housing conditions and good management are expected to result in high animal welfare (Whay, 2007). However, to maintain high consumer trust in such products, it is necessary to establish convincing animal welfare inspections at farm level (Johnsen et al., 2001).

To ensure future-compliant livestock production systems which will be accepted by the general public, the Scientific Advisory Board on Agricultural Policy, Food and Consumer Health Protection of the Federal Ministry of Food and Agriculture, Germany (WBABMEL, 2015) suggested to establish routine self-assessments of animal welfare by farmers, using animal-based parameters. In 2016, the Austrian organic farmers' association 'BIO AUSTRIA' introduced a guideline for the self-assessment of animal welfare in cattle at farm-level (in German: 'Leitfaden Rind'; (BIO AUSTRIA, 2015)). The objective of this self-evaluation by farmers is assessing and benchmarking the animal welfare state on Austrian organic farms. Animal-based parameters are of pathological, ethological and physiological nature (Hörning, 2001, Johnsen et al., 2001) and can be regarded as direct indicators of the bodily and mental state of an animal. They reflect an animal's ability to cope with its environment (Whay, 2007) and allow a more valid estimation of the welfare state than resource-based parameters (Winckler, 2008). However, they pose the hazard of a high degree of subjectivity as regards data collection. The reliability of such an assessment, therefore, largely depends on the inter-observer reliability (IOR) of persons carrying out the measurement (Mullan et al., 2011).

While about 10,000 BIO AUSTRIA cattle farmers were encouraged to self-evaluate animal welfare, rather little is known how a resource-efficient training of a large number of farmers could take place and which level of agreement might be achieved. The aim of this study was therefore i) to establish an online training-tool for the assessment of animal-based welfare parameters in cattle, ii) to investigate which level of agreement with a gold standard may be achieved through such a tool, and iii) to find out, whether repeated trials would lead to higher agreement between test persons and a gold standard.

Material and methods

All data for the present study were collected via the website http://tierwohltraining.boku.ac.at/ which was built with WordPress Version 4.5.4. Microsoft® Excel for Mac Version 15.26 was used for data collection and processing. Statistical calculations were performed with IBM's SPSS Statistics Version 24.

From the guideline on animal welfare in cattle (BIO AUSTRIA, 2015) the following 10 animal-based parameters were selected (with classification and criteria in brackets): 1) body condition score (BCS) (regular body condition; very lean if transverse and spinous processes prominent and distinguishable, hipbones and tail head prominent and deep cavity around tail head; very fat if spinous processes not discernible, hipbones covered by visible layer of fat and tail head cavity full and folds of fatty tissue present), 2) cleanliness (clean; dirty if continuous plaques of dirt > 30 cm), 3) hairless patches (no hairless patches; hairless patches > 5 cm), 4) swellings (no swellings; swelling if visible expansion of circumference > 5 cm compared to normal), 5) lesions (no lesions; lesion if damaged skin in form of scab or wound > 2 cm), 6) ectoparasites/fungal infections (no ectoparasites present; ectoparasites present, typically mange, or ringworm present), 7) claw disorders (no claw disorder present; claw disorder if concave shape of dorsal wall, asymmetric claws, overgrown claws if > 7,5 cm from coronary band to tip of claw and inappropriate pastern angle), 8) lameness (not lame; lame if any abnormality of movement ascertainable), 9) diarrhoea (no evidence of diarrhoea; diarrhoea if loose watery manure below the tail head on both sides of the tail, area affected at least the size of a hand) and 10) resting comfort in calves (sufficient bedding if legs of lying animal are at least partly covered by bedding material; insufficient bedding).

As test persons, students of Austrian agricultural schools, students of selected lectures of the Division of Livestock Sciences at BOKU University and dairy and beef cattle farmers of the Austrian organic farmers' association 'BIO AUSTRIA' were invited by e-mail to participate in the study. Test persons had to register on the website before being given access to the different quizzes. Quizzes comprised 20–31 pictures, or 28 videos (lameness). The gold standard was set by an expert group of the Division of Livestock Sciences, BOKU. Cohen's Kappa was automatically saved to a database on completion of each quiz.

All statistical analyses were performed using nonparametric tests.

Results

152 persons registered between June 24 and August 19, 2016. 122 of them filled in the questionnaire of which 111 finished at least one quiz. Data for analysis was obtained from the latter 111 participants only. In round 1 the median agreement ranged from $\kappa = 0.40$ (hairless patches) to $\kappa = 0.90$ (cleanliness and resting comfort in calves; n = 58–100; table 1).

$\kappa \geq 0.40$ was achieved by 49 (hairless patches) – 100 % (cleanliness) of test persons with at least 80 % in 8 out of the 10 measures. 27 persons retook at least one quiz, resulting in a sample size of n = 14–24 in round 2 (table 2). Median agreement significantly improved when retaking a quiz for all parameters except hairless patches ($p = 0.146$) and lameness ($p = 0.055$). Using κ-value thresholds (0.40, 0.60 and 0.80) Fisher's exact test pointed out significant improvement in round 2 for parameters BCS (κ-level: 0.40; n = 24; $p = 0.042$), hairless patches (κ-level: 0.60; n = 20; $p = 0.033$) and diarrhoea (κ-level: 0.80; n = 20; $p = 0.018$); however, for numerous combinations Fisher's exact test could not be performed as all κ-values of the 1st and/or 2nd round ranged above/below the set threshold. Pairwise correlations of agreement values for the different parameters were low. The strongest association of $r_S = 0.395$ was obtained for the parameters hairless patches and swellings ($p = 0.001$, n = 65).

Table 1. Descriptive statistics of κ-values for the 10 animal-based parameters in round 1 (n=number of test persons per quiz).

	BCS	Cleanliness	Hairl. pat.	Swellings	Lesions	Ectopara.	Claw diso.	Lameness	Diarrhoea	Rest. comf.
Mean	0.55	0.89	0.43	0.71	0.76	0.62	0.64	0.58	0.78	0.90
Median	0.61	0.90	0.40	0.75	0.80	0.60	0.70	0.57	0.80	0.90
Stdev.	0.23	0.10	0.19	0.16	0.16	0.16	0.18	0.21	0.16	0.15
Min	-0.13	0.60	0.00	0.20	0.30	0.30	0.10	-0.07	0.40	0.00
Max	1.00	1.00	0.80	1.00	1.00	0.90	1.00	0.93	1.00	1.00
n	100	88	72	68	68	68	74	58	73	68

Table 2. Descriptive statistics of κ-values for the 10 animal-based parameters in round 2 (n=number of test persons per quiz).

	BCS	Cleanliness	Hairl. pat.	Swellings	Lesions	Ectopara.	Claw diso.	Lameness	Diarrhoea	Rest. comf.
Mean	0.59	0.93	0.47	0.79	0.86	0.75	0.83	0.66	0.89	0.96
Median	0.61	0.90	0.55	0.80	0.80	0.70	0.85	0.71	0.90	1.00
Stdev.	0.25	0.08	0.26	0.15	0.11	0.20	0.14	0.19	0.10	0.06
Min	-0.23	0.70	0.00	0.50	0.70	0.40	0.50	0.29	0.70	0.80
Max	0.94	1.00	0.90	1.00	1.00	1.00	1.00	0.93	1.00	1.00
n	24	21	20	21	19	21	22	19	20	14

Discussion

The median κ-values obtained in the first round ranged from 0.40 (hairless patches) to 0.90 (cleanliness and resting comfort in calves). Following the generally acknowledged threshold of κ = 0.4 (Brenninkmeyer et al., 2007, Viera and Garrett, 2005), at least half of the participants achieved acceptable agreement with the gold standard. The values obtained in the present study lie within the range of reported values from other studies, which have been performed under rather defined conditions along with intense training (e.g. March et al., 2007). For example, Vieira et al. (2015) reported an initial κ = 0.49 between two experienced and κ = 0.70 between two inexperienced observers for a 3-stage BCS-system for goats. For a 2-stage lameness scoring system in dairy cattle (retrospectively calculated on the basis of a 5-stage lameness scoring system), the initial inter-observer reliability between one experienced and three inexperienced observers was PABAK = 0.59 while the original 5-stage scoring system reached a value of PABAK = 0.37 (Brenninkmeyer et al., 2007). March et al. (2007) showed similar agreement with PABAK of 0.53 (2-stage, calculated retrospectively) and 0.32 (original 5-stage lameness scoring system) for an experienced and an inexperienced observer.

The lowest level of agreement was obtained for the parameter hairless patches, which may be attributed to the difficulties of assessing the size of a hairless patch on non-standardised pictures. In the present study, it was not possible to assess the amount of time spent by test persons studying information presented on the website. For this reason, it is not possible to evaluate, whether agreement with the gold standard was achieved intuitively or affected by studying the information provided on the website. Additionally, since feedback was given to the test persons after submitting each answer (false/true and an explanation why the animal was classified this way according to the gold standard), it is also possible that the participants learned while conducting the first round of quizzes.

The vast majority of test persons performing a second round was aged < 20 years, male and with no (completed) agricultural education but actively working in agriculture. Statements on the training effect, therefore, are restricted to this group and are based on a rather low sample size. A mostly significant improvement was found for all parameters, except for the parameters lameness ($p = 0.055$) and hairless patches ($p = 0.146$), for the latter again pointing at difficulties in solving the quiz task correctly. For lameness, the initial level of agreement was with κ = 0.57 (n = 19) already at a moderate level and could, although not significantly, be improved to median κ = 0.71. Other studies also showed a training effect, in particular investigated for gait scoring in cattle. March et al. (2007) reported an initial agreement of PABAK = 0.53 for gait scoring in dairy cattle (lame/not lame) which increased to PABAK = 0.75 over the course of four repetitions. Brenninkmeyer et al. (2007) calculated similar values of PABAK = 0.59 and PABAK = 0.70 as initial agreement and after 4 repetitions, respectively (lame/not lame, 4 observers).

Analysis of the training effect by using the proportion of test persons exceeding a certain κ threshold (0.40, 0.60 and 0.80, respectively) revealed less improvement with increasing κ-levels. Of special interest is the limit of κ = 0.40, as exceeding this value is generally acknowledged as acceptable agreement (March et al., 2007, Mullan et al., 2011). The parameters BCS ($p = 0.042$), hairless patches ($p = 0.141$) and lameness ($p = 1.000$) allowed to calculate the improvement achieved in round 2 at the level of κ = 0.40. For the other parameters, Fisher's exact test could not be performed on the level of κ = 0.40 as all individual values of κ in round 2, partly also in round 1, were higher than the threshold of κ = 0.40. With the present study, a training effect towards higher IOR after repeating a quiz could be observed. Information on the time lag before retaking a quiz is not available.

According to Martin and Bateson (1993), correlations between agreement values for the different parameters investigated were low. This indicates, that the participants did not consistently achieve agreement, even as regards similar parameters (e.g. hairless patches and lesions). One explanation for this finding could be the employment of different assessment methods (video for lameness and photo for other parameters). But as variance of correlation values regarding lameness was as high as for the other parameters, this factor is less likely to be of importance.

Conclusion

Acceptable agreement with the gold standard can be stated for at least half of the participants, as the medians for all ten different parameters equalled or exceeded a κ of 0.40. Regarding cleanliness and diarrhoea, all participants achieved values of $\kappa \geq 0.40$ in the first round, implying sufficient skills for these sections of welfare assessment and allowing reliable benchmarking of the farms within these parameters. When interpreting data obtained by self-evaluation, it should however be taken into consideration that for some welfare indicators inter-observer reliability may partly be questionable. Improvement of agreement was found, when the quizzes were repeated, but there was no significant improvement for hairless patches and lameness. This clearly shows limitations of the developed training-tool for these parameters. If the training effect can be achieved by older test persons (aged 20 +) needs further investigation. Correlations between agreement values for the investigated parameters were low, suggesting that animal-based parameters cannot be trained interchangeably.

The effects of a refined online training-tool (e.g. further explanations, enhanced pictures and video material, etc.) and the transferability of the online exercise to on-farm conditions should be further investigated.

References

BIO AUSTRIA (2015). Leitfaden Tierwohl Rind. BIO AUSTRIA, Linz, Austria

Brenninkmeyer C., Dippel S., March S., Brinkmann J., Winckler C., Knierim U. (2007). Reliability of a subjective lameness scoring system for dairy cows. Animal Welfare 16: 127-129.

Dörfler R. L. (2007). The Assessment of Animal Welfare in Diverse Livestock Production Systems: Ethical and Scientific Issues. Köster, Berlin

Hörning B. (2001). The assessment of housing conditions of dairy cows in littered loose housing systems using three scoring methods. Acta Agriculturae Scandinavica, Section A-Animal Science 51: 42-47.

Johnsen P. F., Johannesson T., Sandøe P. (2001). Assessment of Farm Animal Welfare at Herd Level: Many Goals, Many Methods. Acta Agriculturae Scandinavica, Section A – Animal Science 51: 26-33.

March S., Brinkmann J., Winckler C. (2007). Effect of training on the inter-observer reliability of lameness scoring in dairy cattle. Animal Welfare 16: 131-133.

Martin P., Bateson P. (1993). Measuring behaviour: an introductory guide. Cambridge University Press, Cambridge

Mullan S., Edwards S. A., Butterworth A., Whay H. R., Main D. C. J. (2011). Inter-observer reliability testing of pig welfare outcome measures proposed for inclusion within farm assurance schemes. The Veterinary Journal 190: e100-e109.

Van De Weerd H., Sandilands V. (2008). Bringing the issue of animal welfare to the public: A biography of Ruth Harrison (1920–2000). Applied Animal Behaviour Science 113: 404-410.

Veissier I., Butterworth A., Bock B., Roe E. (2008). European approaches to ensure good animal welfare. Applied Animal Behaviour Science 113: 279-297.

Vieira A., Brandão S., Monteiro A., Ajuda I., Stilwell G. (2015). Development and validation of a visual body condition scoring system for dairy goats with picture-based training. Journal of Dairy Science 98: 6597-6608.

Viera A. J., Garrett J. M. (2005). Understanding interobserver agreement: the kappa statistic. Family Medicine 37: 360-363.

WBABMEL (2015). Wege zu einer gesellschaftlich akzeptierten Nutztierhaltung. Gutachten. Berlin.

Webster A. J. F. (2001). Farm Animal Welfare: the Five Freedoms and the Free Market. The Veterinary Journal 161: 229-237.

Whay H. R. (2007). The journey to animal welfare improvement. Animal Welfare 16: 117.

Winckler C. (2008). The use of animal-based health and welfare parameters – what is it all about? In: CORE Organic project nr. 1903 - ANIPLAN. CORE Organic project nr. 1903 - ANIPLAN, no. Report from 1st ANIPLAN project workshop.

acs82_39