Acta Agriculturae Scandinavica, Section A — Animal Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/saga20

Visual assessment of body condition and skin soiling in cattle by professionals and undergraduate students using photo slides

B. Staaf Larsson , E. Petersson , M. Stéen & J. Hultgren

To cite this article: B. Staaf Larsson , E. Petersson , M. Stéen & J. Hultgren (2021) Visual assessment of body condition and skin soiling in cattle by professionals and undergraduate students using photo slides, Acta Agriculturae Scandinavica, Section A — Animal Science, 70:1, 31-40, DOI: 10.1080/09064702.2020.1849380

To link to this article: https://doi.org/10.1080/09064702.2020.1849380

Published online: 26 Nov 2020.

ACTA AGRICULTURAE SCANDINAVICA, SECTION A — ANIMAL SCIENCE
2021, VOL. 70, NO. 1, 31–40
https://doi.org/10.1080/09064702.2020.1849380

Taylor & Francis
Taylor & Francis Group

Visual assessment of body condition and skin soiling in cattle by professionals and undergraduate students using photo slides

B. Staaf Larsson [a,b], E. Petersson [c], M. Stéen [a,b] and J. Hultgren [d]

[a]Swedish Centre for Animal Welfare, Swedish University of Agricultural Sciences, Uppsala, Sweden; [b]Department of Anatomy, Physiology and Biochemistry, Swedish University of Agricultural Sciences, Uppsala, Sweden; [c]Department of Aquatic Resources, Institute of Freshwater Research, Swedish University of Agricultural Sciences, Drottningholm, Sweden; [d]Department of Animal Environment and Health, Swedish University of Agricultural Sciences, Skara, Sweden

ABSTRACT

We aimed to study to what extent body condition and skin soiling in cattle are rated differently depending on the profession, education and professional experience of assessors. Data were collected at 15 group-training sessions in the years 2009–2016. Totally, 569 persons; Swedish animal welfare inspectors, veterinarians/advisers, animal welfare scientists, other animal professionals as well as undergraduate students in animal science rated a set of 6–40 photos with respect to animal body condition, animal skin soiling, and recommended corrective action in response to perceived animal soiling. The more extensive animal science education, the less prone animal welfare inspectors were to give higher body condition scores. Students had a higher overall probability of rating the degree of soiling higher than the animal welfare inspectors and veterinarians/advisers. Students also recommended stricter corrective actions than did welfare inspectors and veterinarians/advisers, and veterinarians/advisers recommended less strict actions than animal welfare inspectors.

ARTICLE HISTORY
Received 7 October 2020
Accepted 6 November 2020

KEYWORDS
Animal welfare; body condition; cattle; cleanliness; interactive audience response; skin soiling

Introduction

Good animal welfare (AW) is to some extent maintained through national and EU legislation and official control (European Parliament, 2017). Animal welfare legislation and official control are mainly preventive, focusing on potential risks in animal housing and management (Blokhuis et al., 2010; Broom, 2017). Resource-based measures are based on observations of the animal's environment and resources such as space allowance or air quality, while management-based measures include caretaking strategies and animal handling (Keeling, 2009). However, the European Commission (2012) considers the increased use of output-based measures, which reflect AW *per se*.

Complaints about body condition and skin soiling are common AW issues in cattle husbandry (Keeling, 2009). Extreme thinness, as well as obesity, increases the risk of diseases like milk fever, retained placenta, endometritis, ketosis, abomasal displacement and dystocia in cows (Gillund et al., 2001; Roche & Berry, 2006). Thinness has been associated with low milk production (Roche et al., 2007), low conception rate (Pryce et al., 2001) and an increased risk for sole ulcer and white line disease (Green et al., 2014). Green et al. (2014) showed that cows with a body condition score <2.5 (on a scale from 1 = thin to 5 = fat) are more likely to become lame. Over-conditioning may also cause reduced milk yield (Gillund et al., 2001).

There are several reasons for assessing skin soiling in farm animals. At official Swedish AW controls, the most prevalent recorded non-compliance is soiled animals (Keeling, 2009). Improved cattle cleanliness has many benefits, for example, strengthened food safety (Hughes, 2001), increased profits through intact hides at slaughter (Nafstad, 1999), reduced mastitis incidence (Hughes, 2001) and improved animal comfort. Known complications of chronic faecal soiling are etching of the skin, infections and irritation (Nafstad, 1999). Soiled cows have an increased risk of getting mastitis (Breen et al., 2009). It is painful to remove dry lumps of manure from the skin, faecal soiling causes skin lesions (Hauge et al., 2012), and damaged hides can be an economic setback for the farmer after slaughter (Nafstad, 1999). There are several risk factors for soiling, related to building design, management and stockmanship (Radeski et al., 2015). Hughes (2001) introduced a cleanliness

CONTACT B. Staaf Larsson birgitta.staaf.larsson@slu.se

scoring system, evaluating four separate skin areas on the cow to indicate the cause of soiling.

Interrater agreement (interrater reliability) is the degree of agreement between different assessors. It can be estimated in several ways (Gwet, 2014). Percentage agreement (joint probability of agreement) is the simplest measure in a nominal or ordinal rating system, but it does not account for random agreement. Other statistics that have been proposed for nominal or ordinal data with more than two levels, correcting for the fact that agreement may happen by chance, include Fleiss' kappa (Fleiss, 1971), generalised kappa (Abraira & Pérez de Vargas, 1999) and Gwet's agreement coefficient, AC1 (Gwet, 2008, 2014). Interpretation of agreement values has been suggested as 'none to slight' for 0.01–0.20, 'fair' for 0.21–0.40, 'moderate' for 0.41–0.60, 'substantial' for 0.61–0.80 and 'almost perfect' for 0.81–1.00 (Landis & Koch, 1977). Negative values indicate systematic disagreement. In contrast to kappa, Gwet's agreement coefficient is less sensitive to trait prevalence and marginal homogeneity and does not depend upon independence between assessors, which makes it versatile. The agreement between assessors regarding body condition and skin soiling in cattle, and regarding the perceived need for corrective control actions at an official AW control, has not been studied before. Nor have factors that influence these assessments.

A fair and secure legal treatment of farm-animal husbandry requires uniform and objective official AW control (Anneberg et al., 2012; Andrade & Anneberg, 2014). As Ruddat et al. (2014) pointed out, AW assessment is challenging, especially when no gold standard is available. Assessments of animal-based measures such as body condition and skin soiling require valid and reliable scoring systems and well-trained assessors. Further research is also needed to develop training programmes for AW inspectors. A common understanding of AW requirements is important (Sørensen & Fraser, 2010; Anneberg et al., 2012). Berg and Lundmark Hedman (2020) highlighted the importance of training programs, guidelines, and checklists for the inspectors, but also for the farmers to know and understand the requirements.

The objective of this study was to investigate how different categories of professionals and students assess important animal-based measures of cattle welfare and the need for corrective actions in response to a perceived violation of legislation at an imaginary official AW control visit. A second objective was to estimate the interrater agreement for these assessments. We hypothesised that previous animal-related education and professional experience with AW result in a more uniform and consistent assessment, compared to no such education and less experience, and that long professional AW practice improves interrater agreement.

Material and methods

Data were collected in 2009–2016 during 15 group sessions organised and chaired by the first author. Sessions were as uniform as possible and typically lasted for 1.5 h. The sessions aimed to improve the participants' understanding of AW science and skills for conducting official AW control inspections, and followed a standardised protocol, with an introductory lecture on the assessment of body condition and soiling in cattle, including an explanation of rating scales.

The number of participants per session varied between 12 and 80 (mean 38) (Table 1), with a total of 569 Swedish assessors. They were categorised with respect to profession as official 'AW inspectors' (n = 281), 'veterinarians/advisers' (n = 95), 'AW scientists' (n = 27), 'students' (n = 88) or 'other profession' (n = 64). Veterinarians/advisers included county veterinary officers, practising farm-animal veterinarians, and industry or government advisers. Students were animal science undergraduates, typically at a Master or Bachelor programme at the Swedish University of Agricultural Sciences. 'Other profession' included cattle farmers, veterinary nurses and some individuals not matching any of the other categories. The participants were asked to categorise themselves with respect to their level of education in animal science, as 'full' (university degree), 'part' (courses of more than 10 ECTS university credits) (European Credit Transfer System; European Commission, 2019), or 'none' (10 ECTS or less). They were also asked to state their work experience in their current profession as 'much' (at least 3 years) or 'little' (less than 3 years) as well as gender. Gender information was collected only at the last four sessions. Apart from the introductory lecture, the participants did not receive any special training in the study.

The participants were equipped with individual wireless interactive polling keypads (TurningPoint, version 5.4.1.2, Turning Technologies, Youngstown, Ohio, USA), shown a total of 6–40 photo slides projected on a lecture-hall screen, and asked to assess; (1) animal body condition (0–20 slides; Table 1), (2) the degree of skin soiling (0–11 slides), as well as (3) recommended control actions in response to soiled animals (0–10 slides) at an imaginary official control visit to a cattle farm, henceforth referred to as the three 'assessment domains'. The photo material was collected at regular Swedish AW inspections and study visits to cattle

Table 1. Number of ratings of cattle body condition, skin soiling, and recommended corrective control action in response to perceived soiling at an imaginary official animal welfare control inspection using photo slides; 15 assessment sessions 2009–2016.

		Number of participants						Number of slides			
Month, year	Profession categories	Little	Much	Full	Part	None	Total	Body condition	Soiling	Recom-mended action	Total
		Experience[a]		Animal-science education[b]							
December, 2009	AWI	27	13	26	11	3	40	18	6	10	34
March, 2009	STU	17	0	17	0	0	17	18	6	10	34
April, 2010	AWI	22	46	44	24	0	74	17	6	10	33
February, 2011	STU	22	2	23	0	1	24	17	6	10	33
May, 2011	AWI, VET, OTH	8	31	31	6	3	43	20	11	9	40
May, 2011	AWI, VET, OTH	13	39	42	12	0	57	20	11	9	40
May, 2011	AWI, VET, OTH	26	27	40	11	5	57	20	11	9	40
May, 2011	AWI, VET, OTH	27	49	55	18	6	80	19	11	8	38
June, 2011	AWI, VET, OTH	6	18	9	12	3	25	20	11	9	40
September, 2011	AWI, VET, OTH	16	18	25	7	1	34	20	11	9	40
October, 2011	VET, OTH	3	8	4	2	5	12	20	0	0	20
February, 2012	STU	17	0	16	2	0	18	20	11	8	39
September, 2013	VET, OTH	3	23	23	3	0	28	20	10	8	38
May, 2014	SCI, OTH	9	21	25	3	1	31	0	1	5	6
November, 2016	STU	20	7	21	5	1	29	19	11	9	39
	Sum:	236	302	401	116	29	569	268	123	123	514

Note: AWI, official AW inspectors; STU, undergraduate students; VET, veterinarians or advisers; SCI, scientists; OTH, other (farmers of dairy cows or cattle farmers, animal nurses or other professionals that do not belong to another professions).
[a]Little = less than 3 years of work experience; Much = at least 3 years of work experience.
[b]Full = university degree in animal science; Part = courses of more than 10 ECTS university credits in animal science; None = maximum 10 ECTS credits in animal science.

farms, and the majority taken by the first author. Since very thin, very obese, and extremely soiled animals are relatively rare, the photos were collected during several years with different cameras and showed one or several cattle of different types, viewed from different angles, in different housing conditions. Most photos displayed cows. Photos for assessing body condition mostly showed only one animal, which in some cases was a calf. Photos for scoring skin soiling were most often taken from behind, and some of the animals were young stock. The participants were present at only one session and assessed each photo only once. At each session, all participants were shown the same slides, for practical reasons the selection of slides varied between sessions.

The participants were given approximately 30 s per slide to respond by pressing the polling keypads, and they were explicitly instructed to make all assessments independently, without consulting fellow participants. Body condition was scored on a 5-level scale (1 = very poor, 2 = moderate, 3 = good, 4 = fat, 5 = very fat; Table 2; Figure 1(a–e)) (modified after Wildman et al., 1982; Svedberg, 2006). Skin soiling was rated on a 4-level scale (1 = clean, 2 = slightly soiled, 3 = moderately soiled, 4 = very soiled; Table 2; Figure 1(f–i)), according to Svedberg (2007). The recommended action was rated on a 4-level scale (1 = no action; 2 = remark; 3 = remark with a follow-up inspection; 4 = prohibition or order; Table 2), similar to different options when conducting a Swedish AW control, where a prohibition or order can be issued with or without a fine, to force the animal keeper to take action.

Statistical analysis

Analyses were made in Stata IC, version 15 (StataCorp, College Station, Texas, USA). To facilitate comparisons between slides, all recorded scores were transformed to standardised scores on a 3-level ordinal scale -1–0–1, calculated as the deviation of each value from the overall sample mode for the slide, as −1 for original scores below the mode, 0 for original scores equal to the mode and 1 for original scores above the mode. For example, if the median recorded body condition score in a slide was 4, a recorded score of 3 or lower was transformed to −1, 4 was transformed to 0 and 5 was transformed to 1. The standardised scores were then modelled statistically using generalised ordinal logistic regression. The participants were arranged in clusters identified by profession categories. The ratings were expected to be independent between profession clusters, but not necessarily within them, which affected the standard errors of coefficients. Estimated regression coefficients were converted to odds ratios (OR) and their 95% confidence intervals were calculated.

The fixed effect of slide number (categorical; 1 through 45) was forced into all models. Initially, independent variables that represented recording session (categorical; 1 through 15), profession (AWI, VET, SCI, STU or OTH), education (short, part or full) and work experience (much or little) were tested in simple models together with slide number. Three multivariable models were then constructed, one for each assessment domain. Variables representing profession and slide number were forced in, while the remaining independent variables were

Table 2. Overall distribution of scores of cattle body condition, skin soiling, and recommended corrective control action in response to perceived soiling at an imaginary official animal welfare control inspection using photo slides (modified for visual categorisation after Wildman et al., 1982).

Assessment domain	Level	Description of level	Number (percentage) of ratings
Body condition	Very poor	Individual spinous processes had limited flesh covering, were prominent, the ends are clearly visible, and together the processes formed a definite overhanging shelf effect to the loin region. Individual vertebrae of the chine, loin, and rump regions were prominent and distinct, hooks and pin bones were sharp with negligible flesh covering, and severe depressions between hooks and pin bones were noted. The area below the tail head and between the pin bones was severely depressed causing the bone structure of the area to appear extremely sharp.	737 (7.66%)
	Moderate	Individual spinous processes were visually discernible but were not prominent. Ends of processes were sharp although they had a greater flesh covering, and the processes did not have a distinct overhanging shelf effect. Individual vertebrae of chine, loin, and rump regions were not visually distinct. Hooks and pin bones were prominent, but the depression between them was less severe. The area below the tail head and between the pin bones was depressed, but the bone structure was not devoid of flesh covering.	2509 (26.1%)
	Good	Spinous processes were discernible. Together processes appeared smooth and the overhanging shelf effect was not noticeable. Vertebrae of the chine, loin, and rump regions appeared as a rounded ridge, and hooks and pin bones were rounded and smooth. The area between the tail head appeared smooth without a sign of fat deposition.	3761 (39.1%)
	Fat	Individual spinous processes could not be visual distinguished and, together, the processes appeared flat or rounded with no overhanging shelf effect. The ridge formed by the vertebral column of the chine region was rounded and smooth, but loin and rump regions appeared flat. Hooks were rounded, and the span between the hooks was flat. Area around tail head and pin bones was rounded, with evidence of subcutaneous fat deposition.	2226 (23.1%)
	Very fat	Bone structure of the vertebral column, spinous processes, hooks, and pin bone regions was not visually apparent, and evidence of subcutaneous fat deposition was prominent. The tail head appeared to be buried in fatty tissue.	392 (4.07%)
Soiling	Clean	The animals are clean from manure on the flanks, sides and legs.	572 (12.1%)
	Slightly soiled	The animals do not have manure or dirt on the whole body, including the back and sides. Belly, flanks and legs may have a reasonable amount of manure with dry, but not old manure. No layers of manure at all.	1795 (37.9%)
	Moderately soiled	Like 'slightly soiled', but with a certain part with layers of manure on belly, sides and flanks, but no large thick areas.	982 (20.7%)
	Very soiled	The animals are heavily covered with old layers of manure, with large areas on legs and/or parts of the belly, sometimes also on the sides of the body.	1385 (29.3%)
Recommended action	No action	No action taken	1426 (30.5%)
	Remark	Remark given, but no follow-up	1535 (32.8%)
	Remark with follow-up	Remark given and follow-up inspection scheduled	1244 (26.6%)
	Prohibition or order	Prohibition or order issued	470 (10.1%)

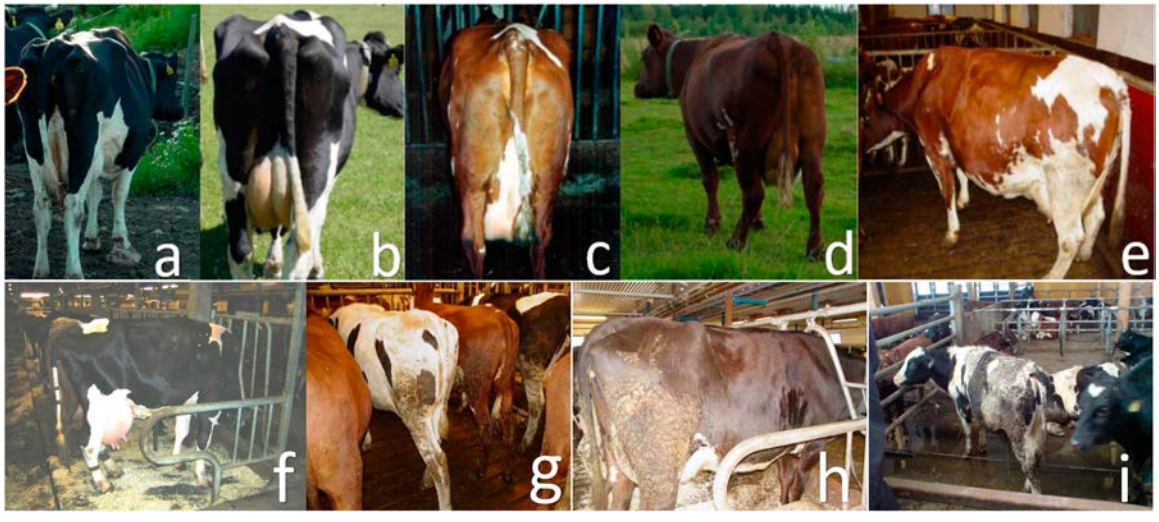

Figure 1. Photos illustrating different degrees of body condition [(a) very poor, (b) moderate, (c) good, (d) fat, (e) very fat] and skin soiling [(f) clean, (g) slightly soiled, (h) moderately soiled, (i) very soiled] in cattle, scored by study participants 2009–2016; Photographers: Jan Svedberg (a–c), Anne Larsen (d,e,g,h) and Birgitta Staaf Larsson (f,i).

tested and retained only if logically relevant. There was no indication of a confounding effect of participant gender, so this factor was excluded from further analysis. Plausible interactions were tested and included if significant at $p \leq 0.05$. Predictive margins were calculated and plotted against different predictor categories.

For each profession category and assessment domain, the agreement between participants was estimated by observed percent agreement (joint probability of agreement), a generalised kappa statistic adapted to ordinal data, multiple observers and incomplete designs (Abraira & Pérez de Vargas, 1999), and Gwet's AC1, treating slides as subjects and using the original scores. Associated 95% confidence intervals were calculated for kappa and Gwet's AC1.

Results

The slides were assessed 56–539 (mean 423) times each, generating totally 19,034 score records. The overall frequencies of the different scores are shown in Table 2 and the distribution between different levels of independent variables and outcome traits used in models are shown in Table 3.

Apart from profession and slide number, covariates included in the final multivariable models were education level and profession × education interaction in the model of body condition, and level of education, professional experience, profession × education interaction, profession × experience interaction in the model of soiling (Table 4). The effect of slide number was significant (joint $P < 0.0005$) for all three assessment domains.

The more extensive animal-related education, the less prone the AW inspectors were to give higher body condition scores (Figure 2). It was not possible to estimate the probabilities of different standardised scores of body condition for veterinarians/advisers with courses of more than 10 ECTS university credits. There were no major differences in the standardised scores of skin soiling between levels of education or professional experience. On the other hand, a statistically significant effect ($p \leq 0.05$) of profession was shown. Students had a higher overall probability of rating the degree of soiling higher than AW inspectors and veterinarians/advisers, regardless of professional experience (Figure 3). For recommended control action in response to soiling, participants with a high level of education or extensive professional experience did not differ significantly from less educated or unexperienced participants. Students recommended stricter AW control actions than did welfare inspectors and veterinarians/advisers, and veterinarians/advisers recommended less strict AW control actions than AW inspectors (Figure 4).

Table 3. Numbers and column percentages (%) of ratings for different participant categories and standardised scores from assessments of cattle body condition, skin soiling, and recommended corrective control action in response to perceived soiling at an imaginary official animal welfare control inspection using photo slides.

		Body condition ($n = 9625$)			Soiling ($n = 4734$)			Recommended action ($n = 4675$)		
Variable	Level	Below	Equal	Above	Below	Equal	Above	Below	Equal	Above
Profession	AWI	665 (56.0)	3361 (53.4)	942 (49.2)	533 (58.5)	1657 (53.2)	182 (30.9)	257 (53.4)	1549 (56.7)	659 (48.5)
	VET	179 (15.1)	1196 (19.0)	384 (20.1)	200 (22.1)	652 (20.9)	98 (16.6)	112 (23.3)	466 (17.1)	179 (13.2)
	SCI	0	0	0	0	0	0	3 (0.6)	71 (2.6)	53 (3.9)
	STU	217 (18.3)	997 (15.5)	343 (17.9)	43 (4.7)	453 (14.6)	246 (41.7)	53 (11.0)	364 (13.3)	371 (27.3)
	OTH	127 (10.7)	761 (12.1)	244 (12.8)	130 (14.3)	352 (11.3)	64 (10.8)	56 (11.6)	283 (10.4)	96 (7.1)
	Missing	30	152	47	26	89	9	15	65	23
Education level[a]	None	75 (6.4)	313 (5.0)	126 (6.7)	46 (5.1)	156 (5.1)	22 (3.8)	27 (5.7)	120 (4.5)	56 (4.2)
	Part	257 (22.1)	1349 (21.7)	426 (22.6)	212 (23.6)	677 (21.9)	112 (19.4)	93 (19.6)	624 (23.2)	264 (19.8)
	Full	831 (71.5)	4556 (73.3)	1332 (70.7)	639 (71.2)	2255 (73.0)	444 (76.8)	354 (74.7)	1949 (72.4)	1012 (76.0)
	Missing	55	229	76	35	115	21	22	105	49
Professional experience[b]	Little	549 (47.9)	2699 (44.3)	816 (44.0)	319 (36.1)	1305 (43.2)	310 (54.4)	171 (36.5)	1154 (43.5)	691 (52.6)
	Much	597 (52.1)	3395 (55.7)	1037 (56.0)	564 (63.9)	1714 (56.8)	260 (45.6)	297 (63.5)	1498 (56.5)	622 (47.4)
	Missing	72	353	107	49	184	29	28	146	68
Gender	Male	25 (15.0)	109 (12.9)	64 (20.3)	21 (30.9)	63 (13.3)	28 (13.3)	13 (22.8)	63 (19.3)	45 (13.8)
	Female	142 (85.0)	738 (87.1)	252 (79.7)	47 (69.1)	409 (86.7)	182 (86.7)	44 (77.2)	264 (80.7)	280 (86.2)
	Missing	1051	5600	1644	864	2731	389	439	2471	1056

Note: AWI, animal welfare inspector; VET, veterinarian/adviser; SCI, animal welfare scientist; STU, student; OTH, other profession.
[a]None = maximum 10 ECTS university credits in animal science; Part = courses of more than 10 ECTS university credits in animal science; Full = university degree in animal science.
[b]Little = less than 3 years of work experience; Much = at least 3 years of work experience.

Table 4. Effects of profession, education level and professional experience in three multivariable generalised[a] ordinal logistic models of standardised scores of cattle body condition, skin soiling, and recommended corrective control action in response to perceived soiling at an imaginary official animal welfare control inspection using photo slides.

			Body condition (n=9130)		Soiling (n=4610)		Recommended action (n=4375)	
Step	Variable	Level	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
−1 → 0	Intercept	–	4.56 (3.05, 6.83)	<0.0005	109 (49.7, 239)	<0.0005	1.48 (1.21, 1.81)	<0.0005
	Profession	AWI	Ref.		Ref.		Ref.	
		VET	0.94 (0.79, 1.12)	0.47	0.97 (0.68, 1.36)	0.85	0.75 (0.60, 0.95)	0.018
		SCI	–		-		1.53 (0.97, 2.42)	0.066
		STU	0.91 (0.54, 1.54)	0.73	24.1 (8.68, 67.2)	<0.0005	1.91 (1.35, 2.72)	0.0005
		OTH	0.82 (0.57, 1.17)	0.27	1.05 (0.53, 2.10)	0.89	0.75 (0.54, 1.04)	0.081
	Education level[5]	None	1.31 (1.05, 1.63)	0.016	1.08 (0.68, 1.72)	0.75	–	
		Part	Ref.		Ref.		–	
		Full	0.77 (0.66, 0.91)	0.002	1.01 (0.80, 1.27)	0.94	–	
	Profession × Education	STU × Short	0.52 (0.093, 2.91)	0.46	0.18 (0.02, 1.42)	0.10	–	
		STU × Full	1.43 (0.83, 2.48)	0.20	0.32 (0.11, 0.91)	0.033	–	
		VET × Full	1.54 (1.21, 1.95)	<0.0005	–		–	
		OTH × Short	0.90 (0.51, 1.59)	0.71	0.34 (0.11, 1.04)	0.058	–	
		OTH × Full	1.87 (1.23, 2.85)	0.004	0.43 (0.20, 0.94)	0.034	–	
	Experience[6]	Little	–		Ref.		–	
		Much	–		0.82 (0.66, 1.00)	0.055	–	
	Profession × Experience	STU × Much	–		1.29 (0.31, 5.28)	0.72	–	
		VET × Much	–		1.31 (0.85, 2.02)	0.22	–	
		OTH × Much	–		2.95 (1.72, 5.08)	<0.0005	–	
0 –> 1	Intercept	–	0.194 (0.127, 0.296)	<0.0005	0.251 (0.184, 0.342)	<0.0005	0.068 (0.053, 0.088)	<0.0005
	Profession	AWI	Ref.		Ref.		Ref.	
		VET	0.94 (0.79, 1.12)	0.47	0.97 (0.68, 1.36)	0.85	0.75 (0.60, 0.95)	0.018
		SCI	–		–		1.53 (0.97, 2.42)	0.066
		STU	0.91 (0.54, 1.54)	0.73	24.1 (8.68, 67.2)	<0.0005	2.68 (2.14, 3.36)	<0.0005
		OTH	0.82 (0.57, 1.17)	0.27	1.05 (0.53, 2.10)	0.89	0.75 (0.54, 1.04)	0.081
	Education level	Short	1.31 (1.05, 1.63)	0.016	1.08 (0.68, 1.72)	0.75	–	
		Part	Ref.		Ref.		–	
		Full	0.77 (0.66, 0.91)	0.002	1.01 (0.80, 1.27)	0.94	–	
	Profession × Education	STU × Short	0.52 (0.093, 2.91)	0.46	0.18 (0.02, 1.42)	0.10	–	
		STU × Full	1.43 (0.83, 2.48)	0.20	0.32 (0.11, 0.91)	0.033	–	
		VET × Full	1.54 (1.21, 1.95)	<0.0005	–	–	–	
		OTH × Short	0.90 (0.51, 1.59)	0.71	0.34 (0.11, 1.04)	0.058	–	
		OTH × Full	1.87 (1.23, 2.85)	0.004	0.43 (0.20, 0.94)	0.034	–	
	Experience	Little	–		Ref.		–	
		Much	–		0.82 (0.66, 1.00)	0.055	–	
	Profession × Experience	STU × Much	–		1.29 (0.31, 5.28)	0.72	–	
		VET × Much	–		1.31 (0.85, 2.02)	0.22	–	
		OTH × Much	–		2.95 (1.72, 5.08)	<0.0005	–	

Notes: OR, odds ratio; CI, confidence interval; AWI, animal welfare inspector; VET, veterinarian/adviser; SCI, animal welfare scientist; STU, student; OTH, other profession.

[a]Variables with identical values for the two steps meet the proportional odds assumption and thus did not require separate estimations.

[b]None = maximum 10 ECTS university credits in animal science; Part = courses of more than 10 ECTS university credits in animal science; Full = university degree in animal science.

[c]Little = less than 3 years of work experience; Much = at least 3 years of work experience.

As indicated by the generalised kappa and Gwet's AC1, agreement between participants within profession categories was fair to moderate, with kappa values between 0.293 and 0.372, and Gwet's AC1 between 0.308 and 0.479 (Table 5). In most cases, agreement regarding the recommended control action was slightly lower than for degree of soiling and body condition, and students had slightly lower values than AW inspectors and veterinarians/advisers, but the differences were small. For recommended AW control action, scientists had lower agreement and consistency values than the other professions. As expected, the values of percent agreement were generally somewhat higher than kappa and Gwet's AC1. Participants with at least 3 years of professional experience had a slightly better agreement for body condition than less experienced assessors, but the difference was not conclusive for soiling and control actions (Table 6).

Discussion

This study indicates that professionals and students from different backgrounds and experiences score the degree of cattle body condition and skin soiling differently, but

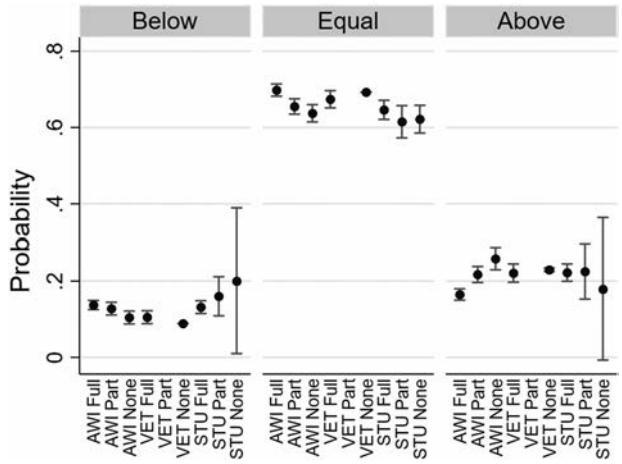

Figure 2. Predictive margins (with 95% confidence intervals) from ordinal logistic models of standardised scores of body condition, showing probabilities of scores below, equal to and above the overall median, assigned by assessors of different professional categories (AWI = animal welfare inspectors; VET = veterinarians/advisers; STU = students) and different levels of animal science education (full = university degree; part = courses of more than 10 ECTS university credits; none = maximum 10 ECTS university credits), based on scorings of photo slides 2009–2016; probabilities for veterinarians/advisers with courses of more than 10 ECTS were not possible to estimate.

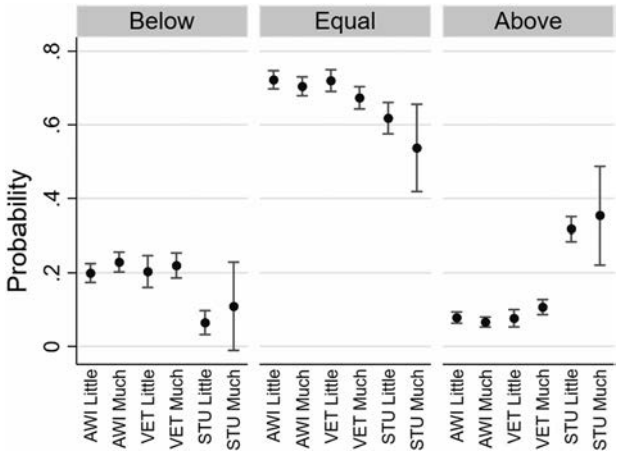

Figure 3. Predictive margins with 95% confidence intervals from ordinal logistic models of standardised scores of skin soiling, showing probabilities of scores below, equal to and above the overall median, assigned by assessors of different profession categories (AWI = animal welfare inspectors; VET = veterinarians/advisers; STU = students) with a university degree and different levels of work experience (little = less than 3 years; much = at least 3 years), based on scorings of photo slides 2009–2016; confidence intervals for probabilities below and equal to the median for students with little work experience were not possible to estimate.

persons with the same background rate these measures rather similarly.

It is unclear to what extent education and training can help secure reliable assessment of AW. Vieira et al. (2015) found positive effects of training on the reliability of visual scoring of body condition in dairy goats using sketches. Shinozaki et al. (2019) found that training appears to affect the reliability of visual assessments of capillary refill time in humans. Practical experience from Swedish dairy adviser coordination meetings indicates that regular calibration exercises are important to secure assessor skills and ensure reliable assessments (pers. comm., L. Winblad von Walter, Växa Sverige, 5 December 2018). In this study, three or more years of professional experience with AW, compared to shorter experience, was only shown to influence ratings of soiling in the 'other profession' category. Agreement between participants regarding body condition was slightly better with 3 years or more of professional experience, compared to less experience, but the results were less conclusive for soiling and suggested actions. Especially ratings of body condition differed between participants, which underlines the need for an objective scoring scale, and shows the importance of proper calibration. Further research is needed to develop best practices for training inspectors in AW assessment.

Found higher rates among students for soiling, compared to AW inspectors and veterinarians/advisers, indicate that students perceive animal soiling more severely. This may be due to differences in personal priorities because students do not encounter

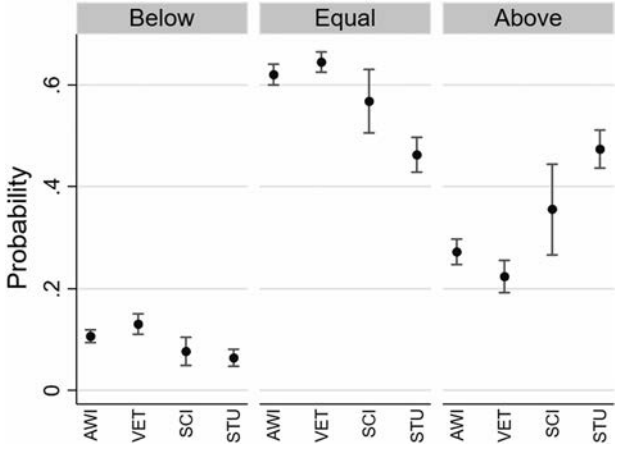

Figure 4. Predictive margins with 95% confidence intervals from ordinal logistic models of standardised scores of recommended action in response to perceived skin soiling at an imaginary official animal welfare, showing probabilities of scores below, equal to and above the overall median, assigned by assessors of different profession categories (AWI = animal welfare inspectors; VET = veterinarians/advisers; SCI = animal welfare scientists; STU = students), based on scorings of photo slides 2009–2016.

Table 5. Percent agreement, generalised kappa and Gwet's agreement coefficient (AC1) (with 95% confidence intervals), as measures of agreement between participants of different profession categories, assessing cattle body condition, degree of skin soiling, and recommended corrective control action in response to soiling using photo slides; original scores.

Statistic	Profession category	Body condition[a]	Soiling[b]	Recommended action[c]
Percent agreement	AWI	0.564	0.582	0.502
	VET	0.568	0.578	0.530
	SCI	–	–	0.477
	STU	0.529	0.533	0.498
Kappa	AWI	0.348 (0.222, 0.583)	0.304 (−0.099, 0.879)	0.306 (0.225, 0.506)
	VET	0.372 (0.275, 0.578)	0.332 (−0.044, 0.870)	0.312 (0.177, 0.568)
	SCI	–	–	0.293 (0.162, 0.541)
	STU	0.317 (0.169, 0.569)	0.293 (−0.106, 0.822)	0.310 (0.191, 0.550)
Gwet's AC1	AWI	0.476 (0.281, 0.671)	0.477 (0.331, 0.624)	0.345 (0.226, 0.463)
	VET	0.479 (0.354, 0.604)	0.466 (0.369, 0.563)	0.391 (0.274, 0.508)
	SCI	–	–	0.308 (0.239, 0.376)
	STU	0.431 (0.304, 0.558)	0.402 (0.277, 0.527)	0.338 (0.175, 0.446)

Note: AWI, animal welfare inspector; VET, veterinarian/adviser; SCI, animal welfare scientist; STU, student.
[a]5-level scale (very poor–moderate–good–fat–very fat; Wildman et al., 1982; Svedberg, 2006).
[b]4-level scale (clean–slightly soiled–moderately soiled–very soiled; Svedberg, 2007).
[c]4-level scale (no action–remark–remark with a follow-up inspection–prohibition or order).

Table 6. Percent agreement, generalized kappa and Gwet's agreement coefficient (AC1) (with 95% confidence intervals), as measures of agreement between participants with different levels of professional experience, assessing cattle body condition, degree of skin soiling, and recommended corrective control action in response to soiling using photo slides; original scores.

Statistic	Experience[a]	Body condition[b]	Soiling[c]	Recommended action[d]
Percent agreement	Little	0.542	0.526	0.486
	Much	0.574	0.552	0.501
Kappa	Little	0.327 (0.172, 0.590)	0.322 (−0.149, 0.794)	0.296 (0.188, 0.520)
	Much	0.370 (0.287, 0.564)	0.294 (−0.099, 0.837)	0.294 (0.215, 0.489)
Gwet's AC1	Little	0.448 (0.298, 0.598)	0.400 (0.353, 0.446)	0.321 (0.252, 0.390)
	Much	0.487 (0.329, 0.645)	0.432 (0.321, 0.544)	0.348 (0.227, 0.469)

[a]Little = less than 3 years of work experience; Much = at least 3 years of work experience.
[b]5-level scale (very poor–moderate–good–fat–very fat; Wildman et al., 1982; Svedberg, 2006).
[c]4-level scale (clean–slightly soiled–moderately soiled–very soiled; Svedberg, 2007).
[d]4-level scale (no action–remark–remark with a follow-up inspection–prohibition or order).

soiled animals so often, to a change in moral standards with increasing age or to cultural relativism. It may also be due to social desirability bias (Kaminska & Foulsham, 2013) or a lack of agreement on or understanding of rating scales, despite efforts to standardise ratings. Higher scores in students for suggested control actions indicate that students are also less tolerant to animal soiling. It has been shown by Margoni et al. (2018) that young adults (21–39 years) rely more on intentions and less on outcomes in judging harmful actions, compared to older people (63–90 years), which may motivate them to take stricter measures against a particular AW infringement. A survey of European citizens (European Commission, 2016) showed that young people and students were more interested in the conditions for farm animals. A stricter view on AW in young persons may clash with the standards and beliefs of more experienced assessors, which requires careful consideration when e.g. official AW control is conducted. Participant age was not recorded in this study but students were likely generally younger than the other profession categories. Oliveira et al. (2017) found veterinarians to score foot pad dermatitis in chickens at the time of slaughter lower, compared to scientists, allegedly due to prior extensive experience of severe foot lesions at slaughterhouses.

AW inspectors and veterinarians/advisers perceived the animals as cleaner than students did. In a comparable study, Bracke et al. (2008) found that veterinarians overall give a higher welfare score for animal housing systems, compared to ethologists, probably reflecting differences in education. In this study, there were no significant differences regarding recommended control actions between AW inspectors and veterinarians/advisers, nor between scientists and other profession categories.

Various factors that were not possible to control or standardise, such as photo angle, image resolution, lighting conditions (contrast, shadows), disturbing objects and irrelevant housing conditions, may have influenced scorings and the agreement between assessors. For example, a dirty housing environment may have affected assessors unconsciously, thus biasing assessments of animal skin soiling towards higher scores. Because slightly different sets of images were

used at different scoring sessions, such effects may also to some extent have influenced the estimated differences between different assessor categories. Poor image quality probably decreased agreement between assessors. Flowers et al. (2008) studied interrater agreement using a human medical wound imaging system, and reported on difficulties experienced by the assessors due to poor image quality, including low dark/light contrast and sharpness. Schmitt et al. (2008) found assessments of knee joint alignment from standardised photos of human subjects to be considerably influenced by the position of the subjects' legs.

Assessments of photo slides differ substantially from on-farm assessments of live animals. Examination in real life allows the assessor to apply different lighting conditions, perspectives and distances, which is likely to increase the reliability of the assessment. Zhu et al. (2017) showed that the visual information provided when viewing 3D images is not the same as when viewing 2D photos of human patients, which may change the clinical impression. However, photos similar to this study, are regularly used in education and training of students and professionals.

Conclusions

This study indicates that persons from different professional backgrounds, with different levels of experience, view and score skin soiling and body condition in cattle differently, while scorings by assessors with similar backgrounds agree fairly well. Undergraduate students in animal science rate cattle soiling stricter than professional assessors, and they seem prepared to take stricter corrective actions against animal soiling at an official control inspection.

Acknowledgements

We are grateful for the collaboration of the participants in the scoring sessions. The Swedish Association for the Protection of Animals supported this study by providing funding.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Swedish Association for the Protection of Animals.

ORCID

B. Staaf Larsson http://orcid.org/0000-0002-6748-6047
E. Petersson http://orcid.org/0000-0001-6443-4172
M. Stéen http://orcid.org/0000-0001-8224-0224
J. Hultgren http://orcid.org/0000-0001-9164-2321

References

Abraira, V. & Pérez de Vargas, A. (1999). Generalization of the kappa coefficient for ordinal categorical data, multiple observers and incomplete designs. *Qüestiió* 23, 561–571.

Andrade, S. B. & Anneberg, I. (2014). Farmers under pressure. Analysis of the social conditions of cases of animal neglect. *Journal of Agricultural and Environmental Ethics* 27, 103–126.

Anneberg, I., Vaarst, M. & Sørensen, J. T. (2012). The experience of animal welfare inspections as perceived by Danish livestock farmers: A qualitative research approach. *Livestock Science* 147(1–3), 49–58.

Berg, C. & Lundmark Hedman, F. (2020). Compliance with animal welfare regulations: Drivers and consequences. *CAB Reviews* 15(025), 1–10. doi:10/1079/PAVSNNR202015025. http://www.cabi.org/cabreviews.

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. doi:10.1080/09064702.

Bracke, M. B. M., Edwards, S. A., Engel, B., Buist, W. G. & Algers, B. (2008). Expert opinion as 'validation' of risk assessment applied to calf welfare. *Acta Veterinaria Scandinavica* 50, 29. doi:10.1186/1751-0147-50-29

Breen, J. E., Green, M. J. & Bradley, A. J. (2009). Quarter and cow risk factors associated with the occurrence of clinical mastitis in dairy cows in the United Kingdom. *Journal of Dairy Science* 92, 2551–2561.

Broom, D. M. (2017). Animal welfare in the European Union. Study for the PETI Committee. Brussels, Belgium: European Parliament. Accessed 30 September 2020, available at: https://op.europa.eu/sv/publication-detail/-/publication/74df7b49-ffe7-11e6-8a35-01aa75ed71a1/language-en.

European Commission. (2012). Communication from the Commission to the European Parliament, the Council and the European Economic and Social Committee on the European Union Strategy for the Protection and Welfare of Animals 2012–2015.

European Commission. (2016). Attitudes of Europeans towards Animal Welfare Special Eurobarometer 442. Accessed 15 September 2020, available at: http://ec.europa.eu/COMMFrontOffice/publicopinion/index.cfm/ResultDoc/download/DocumentKy/71348.

European Commission. (2019). ECTS - European Credit Transfer System. Accessed 14 August 2020, available at: https://ec.europa.eu/education/resources-and-tools/european-credit-transfer-and-accumulation-system-ects_en.

European Parliament. (2017). Regulation (EU) 2017/625 of the European Parliament and of the Council of 15 March 2017 on official controls and other official activities performed to ensure the application of food and feed law, rules on animal health and welfare, plant health and plant protection products.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin* 76(5), 378–382. doi:10.1037/h0031619

Flowers, C., Newall, N., Kapp, S., Lewin, G., Gliddon, T., Carville, K., Martinelli, D. & Santamaria, N. (2008). Clinician inter-rater reliability using a medical wound imaging system. *Wound Practice and Research* 16(1), 22–31.

Gillund, P., Reksen, O., Grohn, Y. T. & Karlberg, K. (2001). Body condition related to ketosis and reproductive performance in Norwegian dairy cows. *Journal of Dairy Science* 84(6), 1390–1396.

Green, L. E., Huxley, J. N., Banks, C. & Green, M. J. (2014). Temporal associations between low body condition, lameness and milk yield in a UK dairy herd. *Preventive Veterinary Medicine* 113(1), 63–71.

Gwet, K. L. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology* 61, 29–48.

Gwet, K. L. (2014). *Handbook of Inter-rater Reliability* (Gaithersburg, MD, USA: *Advanced Analytics*, LLC).

Hauge, S. J., Kielland, C., Ringdal, G., Skjerve, E. & Nafstad, O. (2012). Factors associated with cattle cleanliness on Norwegian dairy farms. *Journal of Dairy Science* 95(5), 2485–2496.

Hughes, J. (2001). A system for assessing cow cleanliness. *In Practice* 23, 517–524.

Kaminska, O. & Foulsham, T. (2013). *Understanding Sources of Social Desirability Bias in Different Modes: Evidence from Eye-Tracking* (Colchester: Institute for Social and Economic Research (ISER), University of Essex). ISER Working Paper Series, No. 2013-04.

Keeling, L. J. (2009). An analysis of animal-based versus resource-based comments in official animal welfare inspection reports from organic and conventional farms in Sweden. *Animal Welfare* 18, 391–397.

Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174.

Margoni, F., Geipel, J., Hadjichristidis, C. & Surian, L. (2018). Moral judgment in old age: Evidence for an intent-to-outcome shift. *Experimental Psychology* 65(2), 105–114. doi:10.1027/1618-3169/a000395

Nafstad, O. (1999). Skader og kvalitetsfeil på norske storfehuder [Hide damages and quality defects in cattle leather]. *Norsk Veterinærtidsskrift* 111(5), 311–319.

Oliveira, A. R. S., Lund, V. P., Christensen, J. P. & Nielsen, L. R. (2017). Inter-rater agreement in visual assessment of footpad dermatitis in Danish broiler chickens. *British Poultry Science* 58(3), 224–229.

Pryce, J. E., Coffey, M. P. & Simm, G. (2001). The relationship between body condition score and reproductive performance. *Journal of Dairy Science* 84(6), 1508–1515.

Radeski, M., Janevski, A. & Ilieski, V. (2015). Screening of selected indicators of dairy cattle welfare in Macedonia. *Macedonian Veterinary Review* 38(1), 43–51.

Roche, J. R. & Berry, D. P. (2006). Periparturient climatic, animal, and management factors influencing the incidence of milk fever in grazing systems. *Journal of Dairy Science* 89, 2775–2783.

Roche, J. R., Lee, J. M., Macdonald, K. A. & Berry, D. P. (2007). Relationships among body condition score, body weight, and milk production variables in pasture-based dairy cows. *Journal of Dairy Science* 90(8), 3802–3815.

Ruddat, I., Scholz, B., Bergmann, S., Buehring, A. L., Fischer, S., Manton, A., Prengel, D., Rauch, E., Steiner, S., Wiedmann, S., Kreienbrock, L. & Campe, A. (2014). Statistical tools to improve assessing agreement between several observers. *Animal* 8, 643–649.

Schmitt, H., Kappel, H., Moser, M. T., Cardenas-Montemayor, E., Engelleiter, K., Kuni, B. & Clarius, M. (2008). Determining knee joint alignment using digital photographs. *Knee Surgery, Sports Traumatology, Arthroscopy* 16, 776–780. doi:10.1007/s00167-008-0570-6

Shinozaki, K., Jacobson, L. S., Saeki, K., Kobayashi, N., Weisner, S., Falotico, J. M., Li, T., Kim, J., Lampe, J. W. & Becker, L. B. (2019). Does training level affect the accuracy of visual assessment of capillary refill time? *Critical Care* 23(1), 157. doi:10.1186/s13054-019-2444-3

Sørensen, J. T. & Fraser, D. (2010). On-farm welfare assessment for regulatory purposes: issues and possible solutions. *Livestock Science* 131, 1–7.

Svedberg, J. (2006). *Djurskyddstillsyn i Praktiken Nötkreatur Hull och Hullbedömningar [Animal Welfare Control in Practice Cattle Body Condition and Body Condition Scoring]* (Skara: SLU).

Svedberg, J. (2007). *Djurskyddstillsyn i Praktiken Nötkreatur Gödselförorenade Djur [Animal Welfare Control in Practice Cattle Defecated Animals]* (Skara: SLU).

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.

Wildman, E. E., Jones, G. M., Wagner, P. E. & Boman, R. L. (1982). A dairy cow body condition scoring system and its relationship to selected production characteristics. *Journal of Dairy Science* 65, 495–501.

Zhu, S., Yang, Y. & Khambay, B. (2017). A study to evaluate the reliability of using two-dimensional photographs, three-dimensional images, and stereoscopic projected three-dimensional images for patient assessment. *International Journal of Oral and Maxillofacial Surgery* 46, 394–400. doi:10.1016/j.ijom.2016.11.011