# Development and implementation of a training program to ensure high repeatability of body condition scoring of dairy cows

**E. Vasseur,[1,2] J. Gibbons, J. Rushen, and A. M. de Passillé**
Pacific Agri-Food Research Centre, Agriculture and Agri-Food Canada, Agassiz, British Columbia, Canada, V0M 1A0

## ABSTRACT

A body condition score (BCS) in dairy cattle is a subjective assessment of the proportion of body fat that she possesses and is a common measure used in animal welfare assessment. The objectives of our study were to develop and implement a training program to produce highly repeatable BCS by many assessors as part of a cross-Canada epidemiological study on dairy cow comfort and welfare. In preliminary studies, we established that without any proper standard operating procedures (SOP) to describe the practical steps of the process and good standard reference for each score, assessors provided with a BCS chart scored with each other only with substantial agreement within 0.5 points and moderate agreement on exact score (mean weighted kappa coefficient = 0.79 and 0.46, respectively). Detailed SOP were developed to assess BCS in 4 locations on a dairy farm. Assessing BCS presented more challenges in some locations (when cows exited the milking parlor, when the assessor was located outside the freestall pen) than others (when cows were headlocked at the feed bunk, when assessor was located inside the freestall pen). Additionally, training material and a training procedure were developed to ensure that future assessors would achieve almost perfect repeatability with the trainer within 0.5 points (weighted kappa coefficient >0.80). Twelve trainees followed this training and their repeatability was assessed using photographs in classroom sessions and live observations on farm over a 1-wk period. Repeatability was maintained above target agreement at periodic checks over the 6 mo of on-farm data collection. Two trainers were used as a reference standard to which all trainees were compared. This study demonstrates that to obtain reliable measures, a training program must include validated procedures to help assessors cope with a variety of farm setups. Regular repeatability checks are essential to ensure that the reference standard is maintained over time and to secure high data quality. This method to develop a training program as well as the training program implemented can be used as a model to successfully train on-farm assessors.

**Key words:** training program, interobserver repeatability, body condition score, dairy cow welfare

## INTRODUCTION

The BCS of a dairy cow is a subjective assessment of the proportion of body fat that she possesses and has been used by animal scientists and producers as a key monitoring tool in dairy cow management since its introduction in the 1970s (Roche et al., 2009). More recently, BCS has been used in on-farm animal welfare assessments (e.g., Whay et al., 2003; Welfare Quality Consortium, 2009).

The scoring chart used to measure BCS differs between countries (e.g., Bewley et al., 2010), but lowest values always reflect emaciation (thin cow) and highest values equate to obesity (fat cow; Roche et al., 2004). Research has demonstrated relationships between a cow's stored energy reserves and her health: cows that are too fat at calving are more prone to reproductive and metabolic diseases, whereas cows that are too thin 30 to 100 d postpartum may not have sufficient body reserves to support high levels of milk production during early lactation and are, therefore, more likely to enter into a state of negative energy balance (de Vries and Veerkamp, 2000). Typically, best management practices require taking corrective actions for cows with extreme BCS (e.g., DFC, 2009) and animal welfare assessment schemes monitor if a cow is too thin, too fat or in ideal condition (e.g., Welfare Quality Consortium, 2009).

The accuracy of BCS assessment is important when it is used as a decision support tool by dairy producers or in animal welfare assessments. However, considering the subjective nature of the scoring process, determining both inter- and intraassessor repeatability of BCS evaluation is important. To help achieve consistent BCS assessment, BCS educational material has been developed (e.g., Elanco Animal Health, 1996; DEFRA, 2001;

Alberta Agriculture and Rural Development, 2004). This material includes a combination of photographs and text, which detail differences in the conformation of anatomical locations that correspond to each level of BCS. However, very few studies have evaluated the effectiveness of such material in ensuring good repeatability of the scoring. Nor has much research been done on the importance of training in ensuring the accuracy of BCS assessment.

More recently, greater emphasis has been placed on the importance of formal training programs for animal welfare assessors to reduce inter- and intraobserver variation of animal-based measures and to maintain the integrity of the assessment (EFSA, 2011; Rushen et al., 2011; Gibbons et al., 2012). When training future assessors for welfare assessments, differences between people are expected due to observer-related factors such as experience and personal biases. However, with appropriate training and regular repeatability assessment, the variability in the data collected should be substantially reduced (e.g., Gibbons et al., 2012). Ideally, if different assessors receive a high standard of training with assessments at regular intervals, they should produce more accurate and reliable data (e.g., Mullan et al., 2011). Despite the recognition that training is essential to reduce variation among assessors, few studies provide detailed information on the training program used or the effectiveness of that training.

Two preliminary studies were conducted to (1) develop BCS standard operating procedures (**SOP**) using a chart used extensively in the field and (2) test the effect of training on repeatability. This paper reports the strengths and weaknesses of BCS methods in different locations of a freestall barn to provide guidelines to assessors to cope with a variety of farm setups. This paper also highlights that BCS charts, even if extensively used, may not be self-explanatory. Interobserver repeatability before and after training will be reported.

As part of a cross-Canada epidemiological study on dairy cow comfort and welfare, a training program was implemented to train assessors who were naïve to the scoring system, differed in previous experience with dairy cattle, and were geographically separated, with little direct contact. In this paper, the effectiveness of the training program to produce highly repeatable BCS between assessors will be examined.

## MATERIALS AND METHODS

The BCS chart used was the Elanco Animal Health body condition scoring chart for dairy cattle [Elanco Animal Health, 1996), based on Wildman et al. (1982) and Ferguson et al. (1994)]. This is a 14-point BCS

chart, with scores ranging from <2 to 5 in 0.25 increments, based on 10 different body parts (Figure 1).

Two preliminary studies were conducted to (1) develop BCS SOP and (2) test the effect of training on repeatability of the assessors. Both preliminary studies were conducted at the University of British Columbia Dairy Educational Research Center (Agassiz, British Columbia, Canada). Only lactating Holstein cows were assessed for BCS.

### Development of SOP

Four pairs of assessors with no experience with the chart, each including a veterinary student and an animal scientist were asked to evaluate the ease of use of the selected BCS chart at different locations on a dairy farm. Each pair of assessors was tested in 4 different locations: (1) when cows were locked at the headlock feed bunk, (2) from inside the freestall pen containing 12 to 48 cows, (3) in the milking parlor (double 12 parallel with 2 exit alleys) during milking, and (4) from outside of a freestall pen containing 12 to 48 cows, allowing the assessors to walk on 2 sides of the pen, in the feed alley, and along another side of the pen (either back or side). Using a decision Aid Form (Table 1), for each location, the pair of assessors had to follow 6 steps and by discussing together, (1) choose a position for each assessor and develop a procedure to score, (2) evaluate how well they were able to score (scale from 1 = very badly to 5 = very well), (3) record the time needed to score 40 cows, (4) describe the difficulties encountered with this strategy, (5) develop an improved strategy based on their results, and (6) test the improved strategy by redoing steps 1 to 6. As a result of this process, a BCS SOP for each location was developed containing a concise description of the sampling protocol as well as strengths and weaknesses of the use of the BCS chart at each location (Table 2). No SOP was developed especially for tie-stall farms, as cows could be observed with a high level of precision and palpated at their tie-stall.

### Effects of Training on Repeatability

The repeatability of assessors who were either given training or no training in BCS was compared. First, 2 pairs of nontrained assessors (veterinary students with little experience with cattle) were given the BCS chart. Each pair was asked to score 50 cows, both nontrained assessors scoring independently at the headlock feed bunk. No further instructions or information were provided. Range and prevalence of BCS points scored by nontrained assessors are detailed in Table 3. Interobserver repeatability between nontrained assessors was calculated for scores from live observations.

| BCS | 3.0 | 2.75 | 2.5 | 2.25 | 2.0 | < 2.0 |
|---|---|---|---|---|---|---|
| Pelvic area | V | V | V | V | V | V |
| Hook bones | rounded | angular | angular | angular | angular | angular |
| Pin bones | padded | padded | angular, fat palpable | angular, no fat palpable | angular, no fat palpable | angular, no fat palpable |
| Ribs | corrugations non visible | corrugations non visible | corrugations non visible | corrugations visible 1/2 way between tips and short ribs | corrugations visible 3/4 way between tips and short ribs | corrugations visible 3/4 way between tips and short ribs |
|  |  |  |  |  | thurl non prominent | thurl prominent |

| BCS | 3.25 | 3.5 | 3.75 | 4.0 | 4.25 | 4.5 | 4.75 | 5.0 |
|---|---|---|---|---|---|---|---|---|
| Pelvic area | U | U | U | U | U | U | U | U |
| Tailhead ligament | visible | barely visible | not visible | not visible | not visible | not visible | not visible | not visible |
| Sacral ligament | visible | visible | barely visible | not visible | not visible | not visible | not visible | not visible |
| Thurl | non flat | non flat | non flat | non flat | flat | flat | flat | flat |
| Tips short ribs | visible | visible | visible | visible | barely visible | barely/not visible | barely/not visible | barely/not visible |
| Pin bones | visible | visible | visible | visible | visible | buried | buried | buried |
| Hook bones | visible | visible | visible | visible | visible | visible | barely visible | barely/not visible |
|  |  |  |  |  |  |  |  | all bony prominences well bounded |

**Figure 1.** Chart based on Elanco Animal Health BCS for dairy cattle (Elanco Animal Health, 1996). Scores are split between thinner (BCS of 3.0 and lower; top table) and fatter cows (BCS of 3.25 and higher; bottom table). First row: 0.25 BCS point; first column: body part to look at; cell: description of the body part at each 0.25-point score; gray cell: what has changed between this score and the previous one. Pelvic area is for either a V-shaped (thinner) or U-shaped (fatter) cow.

Second, 3 trained animal scientists (with more than 4 yr of experience with cattle) acted as trainers. The 3 trainers establish standards for each BCS point of the chart. They developed a chart, which provided a strict definition for each BCS point (description of 10 body parts; Figure 1), as well as an illustrated chart composed of a photograph of the rear and the side of a cow for each BCS point. Each of the 3 trainees (veterinary students with little experience with cattle) were paired with 1 trainer and trained as follows: they were presented with (and discussed) the BCS SOP developed previously to score at the headlock feed bunk, a chart, and an illustrated chart. They then had to independently score cows with varying levels of BCS first from photographs [n = 20; compiled in a PowerPoint presentation (version 2003 and 2007; Microsoft Corp., Redmond, WA) projected in a class room] and then from live observations of 16 cows in the barn. Range and prevalence of BCS points scored by trainees (with training) from photographs and live cows are detailed in Table 3. Data were collected using a standard paper recording sheet where a summary table of the scoring system and concise definitions were included. The cow's identity was recorded using the ear tag number. Cows were viewed from the rear and right side as they stood at a headlock feed bunk. Body condition score was as-signed using both visual and tactile cues. Interobserver repeatability between trainee and trainer was calculated for scores from photographs (n = 20) and from live observations (n = 16). Once independent scores were recorded, trainers and trainees discussed the scoring until trainees' scores were similar to trainers' scores.

### Implementation of the Training Program

As part of a cross-Canada epidemiological study, assessors were trained to standardize the implementation of complete cow welfare assessment protocols (including BCS, injuries, and in-barn checklist, among others) and to achieve high repeatability between trainees (e.g., injury scores; Gibbons et al., 2012). Training to BCS was part of a 1-wk training program, followed by continued assessment during field data collection. The training program for BCS was implemented using training material developed during the preliminary studies.

### Trainers

The 2 trainers were both experienced dairy scientists with extensive experience of BCS on commercial farms. One of the 2 trainers was a trainer in the preliminary studies; the other trainer was new to the training ma-

**Table 1.** Example of a decision aid form used by each pair of assessors to develop strategies to score the body condition of cows in the milking parlor (attempt 1), in the holding area to the milking parlor (attempt 2), and in the exit alley away from the parlor (attempt 3)

| Attempt[1] | Location[2] | Position of the assessors and the cows; process to score | Self-evaluation[3] | What works[4] | What does not work[4] | Solution[5] |
|---|---|---|---|---|---|---|
| 1 | Milking parlor | Assessors stand with milkers and cows stand at milking stall; 1 assessor goes in front of cow to read ear tags | 2 | Can see either right[6] or left cow side | Cannot see cow rear[7] and BCS cannot be assessed using left side | Score cows standing in the holding area before entering the milking parlor |
| 2 | Holding area to the milking parlor | Cows are observed by assessors in parlor exit alleys while cows are standing still in the holding area to enter the parlor | 2 | Can only score cows nearest to the observer | Cannot read ear tags of cows standing in the middle of the group and not all body parts are visible; climbing fences does not help; it is too dangerous for assessors to circulate between cows while standing in waiting area | Score when cows exit the milking parlor |
| 3 | Exit alley away from the parlor | Assessor 1 identifies cows exiting milking parlor and assessor 2 assesses BCS while cows exit the parlor while positioned on a higher spot (e.g., perched on a fence) | 4 | Can see rear and right cow side | In the case of 2 exit alleys, only half of the cows can be scored | — |

[1]Attempts were from 1 = first strategy to 3 = most-improved strategy.

[2]Location in the freestall barn.

[3]Self-evaluation of how well assessors were able to score (from 1 = very badly to 5 = very well).

[4]Description of difficulties encountered and what works and what does not work.

[5]Solution: development of an improved strategy and test of this new strategy in a next attempt.

[6]Right side: being able to visually observe tips of short ribs, ribs, sacral ligament, thurl, and pelvic area. The left side should not be used to assess BCS because of the rumen.

[7]Rear: being able visually to observe tailhead ligament, hook bones (2), and pin bones (2).

**Table 2.** Strengths and weaknesses of using the BCS chart at 4 locations of a freestall farm (headlocks, inside pen, milking parlor, and outside pen) from the most (left) to the least suitable (right)

| Item | Headlocks | Inside pen | Milking parlor | Outside pen |
|---|---|---|---|---|
| Standard operating procedure (SOP) | Cows are locked at the headlock feed bunk (feed delivery or peak feeding); assessor 1 walks along the feed bunk to identify cows before entering pen; assessor 1 walks inside the pen behind the locked cows and assesses BCS of focal cows | Assessor 1 identifies the focal cow lying in her stall, encourages her to stand up, and directs the cow in the opposite direction to where assessor 2 stands; assessor 2 scores body condition of the cow as the cow stands in her stall; if scoring is not completed in the stall, assessor 2 follows the cow walking in the alley | Assessor 1 identifies all cows using ear tags as they exit the milking parlor; assessor 2 scores body condition of all animals exiting the milking parlor while positioned on a higher spot (e.g., perched on a fence); focal cows are sorted out once all cows have been identified and scored | Assessor 1 walks along sides of the pen, observes standing (still or walking) animals, and counts the number of very thin cows (BCS <2) |
| No. of assessors needed | 1 | 2 | 2 | 1 |
| Cows can be identified using ear tag | Yes | Yes | Yes | No |
| All cows of the pen can be scored in a timely manner (time needed to score 1 cow) | Yes (≤1 min/cow, once cows are locked) | Yes, but locating and scoring focal cows in large groups may be more time consuming (1 min/cow) | Yes, but scoring focal cows may take a long time and even take the entire milking (2 to 4 h); (≤1 min/cow) | Only cows standing or nearest the assessor can be scored; because ear tags cannot be read, the same cow might be scored twice (1 min/cow) |
| Cows can be locked | Yes | No | No | No |
| Proximity to the cow while scoring | <50 cm | 50 to 100 cm | 100 to 200 cm | >200 cm |
| Rear[1] and right[2] side of the cow can be observed | Yes | Yes | Yes | Either rear or either right or left side unless the cow is moving |
| All 10 body parts of the cow can be observed | Yes | Yes | No | No |
| Cows can be palpated | Yes | No | No | No |
| Level of precision of BCS | 0.25 point | 0.25 to 0.5 point | 1 point | Counting very thin cows (BCS <2) |
| Other constraints | Attracting cows at headlocks outside peak feeding time can be very time consuming | Risks associated with walking in home pen on commercial dairies (e.g., cows not habituated to being approached by unknown person and may start to run or stampede around pen) | Scoring focal cows at the milking parlor is possible only in some types of parlors (e.g., if 2 exit alleys, only half of the cows can be scored); lack of lighting, if cows run or exit too quickly | Only possible in some farms where assessors can walk along more than 1 side of the pen |

[1]Rear: being able visually to observe tailhead ligament, hook bones (2), and pin bones (2).
[2]Right side: being able to visually observe tips of short ribs, ribs, sacral ligament, thurl, and pelvic area. The left side should not be used to assess BCS because of the rumen.

**Table 3.** Effect of training on repeatability: mean, SD, and minimum (Min) and maximum (Max) weighted kappa coefficient ($K_w$) for repeatability of nontrained assessors and trainees for exact agreement and agreement within 0.25 and 0.5 points

| Measure | No. of assessors[1] | No. of BCS points[2] | BCS (Min–Max)[3] | BCS % (Min–Max)[4] | n[5] | Level of agreement[6] | Mean | SD | Min | Max |
|---|---|---|---|---|---|---|---|---|---|---|
| Without training | 4 | 11 | 2–4.5 | 1.2–23.5 | 80 | Exact | 0.46 | 0.31 | 0.24 | 0.68 |
|  |  |  |  |  |  | ±0.25 | 0.57 | 0.31 | 0.35 | 0.79 |
|  |  |  |  |  |  | ±0.5 | 0.79 | 0.20 | 0.65 | 0.93 |
| With training: photographs | 6 | 11 | 2–4.5 | 0.8–15.0 | 120 | Exact | 0.68 | 0.04 | 0.64 | 0.71 |
|  |  |  |  |  |  | ±0.25 | 0.88 | 0.02 | 0.85 | 0.89 |
|  |  |  |  |  |  | ±0.5 | 0.94 | 0.06 | 0.89 | 1.00 |
| With training: live | 6 | 10 | 2.25–4.25 | 1.0–16.7 | 96 | Exact | 0.70 | 0.03 | 0.66 | 0.73 |
|  |  |  |  |  |  | ±0.25 | 0.85 | 0.01 | 0.84 | 0.86 |
|  |  |  |  |  |  | ±0.5 | 0.89 | 0.04 | 0.87 | 0.94 |

[1]Four nontrained assessors scored cows live in the barn (repeatability measure: without training); 3 trainees scored against 3 trainers remotely using photographs and live in the barn.

[2]Number of BCS points assessed out of the 14-point BCS of the chart.

[3]Range of BCS points assessed out of the 14-point BCS of the chart.

[4]Range of the percentage of cows assigned to each BCS point.

[5]Number of cows scored.

[6]Agreement within 0.25 points with another nontrained assessor (repeatability measure: without training) or with a trainer includes exact agreement, and agreement within 0.5 points includes agreement within 0.25 points.

terial. Using the finalized BCS material (SOP, chart, and illustrated chart), the trainers scored and discussed BCS on cows until they developed uniform scoring. In addition, to ensure a high level of agreement, they tested their repeatability during 4 scoring sessions, including 2 on-farm and 2 photograph sessions. The trainers set the reference standard against which each trainee was evaluated throughout the entire training program. Trainer 1 was responsible for the training of the trainees and trainer 2 for continued assessment during field data collection. To maintain agreement between trainers, inter- and intraobserver repeatability was assessed monthly during the training period

(6-mo period) by scoring 20 cows live in the barn or remotely using photographs. Range and prevalence of BCS points scored by trainers across scoring sessions are detailed in Table 4.

### Training Program

A total of 12 trainees were involved in the epidemiological study, with varying degrees of experience working with dairy cattle, from zero experience to farm advisor. None of the trainees was part of the preliminary studies. The trainees were part of 3 teams located in Alberta, Ontario, and Quebec in Canada. The trainers

**Table 4.** Mean, SD, minimum (Min) and maximum (Max) weighted kappa coefficient ($K_w$) for inter- and intratrainer (1 and 2) repeatability across the training period for exact agreement and agreement within 0.25 and 0.5 points

| Measure | No. of scoring sessions[1] | No. of BCS points[2] | BCS (Min–Max)[3] | BCS % (Min–Max)[4] | n[5] | Level of agreement[6] | Mean | SD | Min | Max |
|---|---|---|---|---|---|---|---|---|---|---|
| Intertrainer | 6 | 13 | <2.0–4.5 | 0.4–16.3 | 119 | Exact | 0.67 | 0.09 | 0.55 | 0.83 |
|  |  |  |  |  |  | ±0.25 | 0.82 | 0.09 | 0.76 | 1.00 |
|  |  |  |  |  |  | ±0.5 | 0.96 | 0.05 | 0.88 | 1.00 |
| Intratrainer 1 | 4 | 11 | <2.0–4.5 | 1.0–18.0 | 50 | Exact | 0.62 | 0.09 | 0.53 | 0.73 |
|  |  |  |  |  |  | ±0.25 | 0.88 | 0.13 | 0.76 | 1.00 |
|  |  |  |  |  |  | ±0.5 | 0.97 | 0.05 | 0.89 | 1.00 |
| Intratrainer 2 | 4 | 11 | 2–4.5 | 0.5–26.2 | 104 | Exact | 0.80 | 0.20 | 0.60 | 1.00 |
|  |  |  |  |  |  | ±0.25 | 0.92 | 0.14 | 0.71 | 1.00 |
|  |  |  |  |  |  | ±0.5 | 0.98 | 0.05 | 0.91 | 1.00 |

[1]Trainers 1 and 2 assessed their inter- and intratrainer repeatability at 4 to 6 scoring sessions (1 mo in between scoring sessions) during the training period by scoring cows live in the barn or remotely using photographs.

[2]Number of BCS points assessed out of the 14-point BCS of the chart.

[3]Range of BCS points assessed out of the 14-point BCS of the chart.

[4]Range of the percentage of cows assigned to each BCS point.

[5]Number of cows scored.

[6]Agreement within 0.25 points between (inter-) or within (intra-) trainer includes exact agreement, and agreement within 0.5 points includes agreement within 0.25 points.

**Table 5.** Percentage of cows assigned to each BCS point across the time points of the training program by the trainees based on trainer reference scores

| Time point[1] | n[2] | <2.0 | 2 | 2.25 | 2.5 | 2.75 | 3 | 3.25 | 3.5 | 3.75 | 4 | 4.25 | 4.5 | 4.75 | 5 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | BCS | | | | | | | |
| Day 1: photographs | 200 | | | 15.5 | 13.0 | 7.0 | 9.5 | 6.0 | 21.5 | 3.5 | 12.0 | 6.0 | 6.0 | | |
| Day 1: live | 191 | | 3.7 | 13.1 | 20.4 | 9.4 | 19.9 | 15.2 | 13.6 | 2.1 | 2.6 | | | | |
| Day 2 | 172 | 2.9 | 2.9 | 12.8 | 18.6 | 7.6 | 16.9 | 22.7 | 10.5 | 1.7 | 3.5 | | | | |
| Day 7 | 233 | 0.4 | 4.7 | 26.6 | 40.3 | 8.6 | 8.2 | 8.6 | 1.7 | 0.9 | | | | | |
| Day 8 | 134 | | 6.0 | 29.9 | 47.8 | 1.5 | 8.2 | 3.0 | 2.2 | 1.5 | | | | | |
| Midway | 286 | 1.0 | 8.0 | 22.0 | 17.5 | 19.9 | 16.1 | 7.3 | 6.6 | 0.7 | 0.3 | | 0.3 | | |

[1]On d 1 of the training program, BCS was scored from photographs as well as live and on d 2, 7, and 8, BCS was scored live; BCS was scored live during the midway assessment, which was carried out 5 to 15 wk after the training program.
[2]Number of cows scored.

were located in British Columbia (Canada) but traveled to train each team. Each team received an identical training program between January and June 2011. The same trainer delivered the training, which started (d 1) with a 2-h classroom instruction session followed by a 2-h live session in 2 of the 3 research dairy units (2 tie-stalls and one freestall with headlocks, according to the location of the training in Canada). The SOP developed to score at the freestall with headlocks was also used to score at the tie-stall.

In the classroom, a presentation was given outlining the rationale, the SOP, the chart, and the illustrated chart of BCS. Uncertainties about BCS were discussed using 8 photographs. On d 1, the trainees were tested with both photographs of cows and with live cows, with varying levels of BCS points (n = 20) previously scored by the 2 trainers. The agreement scores of the 2 trainers were used as a reference standard and compared with those scores given by the trainees. Additionally, a BCS CD-ROM (Alberta Agriculture and Rural Development, 2004) was provided to assist in BCS training. During the live session on d 1, trainees assessed BCS on cows (n = 20) in 1 of the 3 research dairy units. First, the trainer and the trainees both scored the same cows (n = 20) and results were compared. Second, the trainer and trainees discussed the BCS that had low agreement during the classroom session. Finally, the trainer explained practical safety aspects to take into account when approaching animals and discussed any potential challenges that may be encountered in the field (e.g., to gain optimum visibility of different anatomical parts of the cows). On d 2, the trainees assessed BCS on cows (n = 20) in the same research dairy unit. On d 7 and 8, the trainees scored 20 cows each on 2 commercial farms. A total of 6 different commercial farms (4 tie-stalls and 2 freestalls with headlocks) were used, according to the location of the training in Canada. Only 9 trainees scored on d 8. When the target of weighted kappa coefficient ($K_w$) >0.8 within 0.5 points

of agreement with the trainer was achieved, the trainee was considered as a trained observer for BCS. Range and prevalence of BCS points scored by the trainees at each time point of the training program are detailed in Table 5.

### Midway Check

Once field data collection was in progress, the trainees were reassessed 5 to 15 wk after the initial training to ensure that they remained repeatable in their scoring. For the midway assessment, trainer 2 accompanied the trainees on farm. The trainer and trainees assessed BCS (n = 40 per farm), on 2 commercial farms. Repeatability between trainer and each trainee was calculated and discrepancies were discussed.

### Data Analysis

Descriptive results of the most suitable strategy to assess BCS, deduced from the decision aid form, were analyzed using a strength and weakness analysis of each location. The BCS were entered in Excel software (version 2003 and 2007; Microsoft Corp., Redmond, WA) and analyzed using the SAS statistical package (version 9.2; SAS Institute Inc., Cary, NC). Data were ordinal. The inter- and intraobserver repeatability was assessed using the $K_w$. The $K_w$ statistic was used to assess the extent to which the proportion of agreement within or between observers is better than chance. Individual $K_w$ were calculated between nontrained assessors, between trainees and trainers, and within and between trainers for exact agreement and agreement within 0.25 and 0.5 points. Means, standard deviations, and minimum and maximum $K_w$ were reported. The interpretation of $K_w$ values according to Landis and Koch (1977) is <0 = poor, 0.0 to 0.20 = slight, 0.21 to 0.40 = fair, 0.41 to 0.60 = moderate, 0.61 to 0.80 = substantial, and 0.81 to 1 = almost perfect. Based on the results from

preliminary studies, the target level of $K_w$ >0.8 within 0.5 points of agreement between trainer and trainer needed to be reached during training and maintained at the midway check to ensure a high level of agreement in the epidemiological study. Occurrences of agreement between trainer and trainee on exact scores and neighboring scores within 0.25 and 0.5 points were tallied for each BCS point across the time points of the training program (d 1: photographs, d 1: live, d 2 to 8, and midway check) and converted to a percentage of the total.

## RESULTS

### Development of SOP

The preliminary tests showed that, when available on a farm, a headlock feed bunk was the most suitable location to assess BCS (based on the strengths and weaknesses analysis at each location; Table 2). When cows were headlocked at a feed bunk, the assessors could be within close proximity to the cows (less than 50 cm) ensuring obstruction-free observation of the 10 body parts and allowing BCS to be assessed in 0.25-point increments. Cow identification and recording of BCS were both quicker when one assessor stood outside the pen to read the ear tag and the second assessor stood behind the animal in the alley to assess body condition. A single assessor could still read the ear tag of the animal from inside the pen. Less than 1 min per cow was required to identify and assess BCS of the cow. Scoring at the headlocks was even easier if the assessors were present when fresh feed was delivered and during peak feed time when all cows were feeding at the same time. Outside of peak feeding time, it was time consuming to move cows to the feed bunk to be headlocked.

When no headlocks were available at the feed barrier, the second-most-suitable option was to score the cows from inside the home pen. The main constraint was being able to approach cows close enough (less than 50 cm) on commercial farms where cows are not habituated to being approached by unfamiliar people in their home pen. The BCS chart used required palpation of the cows, but palpation could be impossible on many commercial farms. The best strategy developed to assess BCS inside the pen required one assessor to simultaneously record the cow's ear tag and move the cow to where the second assessor scored the cow while walking in the opposite direction. It took 1 h to score 40 cows in a 40-cow pen using this strategy; however, locating and scoring 40 focal cows in a 100-cow pen would be more time consuming.

Assessing BCS at the milking parlor was best done while the cows were exiting the milking parlor. However, the assessors needed to be very quick to score because the cows followed one another closely upon exit of the parlor, so that identification and BCS of individuals was difficult. However, with trained and confident assessors, BCS of the cows could be assessed quickly in 1-point increments, taking as little as a few seconds per cow. The total time to score 40 focal cows could be very long and even take the whole milking time (2 to 4 h) if the 40 focal cows were not in the same pen or in a very large pen and so not milked at the same time. In addition, other constraints still remained; it was not possible to palpate cows; if cows ran or exited too quickly or light levels were low, it was not possible to identify or score cows.

Finally, assessing BCS from outside the cow pen was the least-suitable option because it was very difficult to get closer than 2 m from the cows, which made it almost impossible to correctly identify individuals and accurately assess BCS. However, on some farms with small group size and where assessors could walk along more than 1 side of the pen, counting the number of very thin cows (BCS <2) was still possible.

### Effects of Training on Repeatability

The pairs of nontrained assessors encountered several difficulties in using the BCS chart provided to them (e.g., deciding which body part to look at to differentiate between 2 adjacent BCS and encouraging cows to the feed bunk) and not all cows could be scored in the 2-h period (Table 3). Agreement to within 0.5 points between nontrained assessors was substantial, whereas exact agreement was moderate. As a result of the training, agreement within 0.25 and 0.5 points between trainees and trainers was almost perfect from photographs and live observations, whereas exact agreement was substantial (Table 3). Based on the results of the preliminary study, the target to be achieved for the epidemiological study was ensuring almost perfect repeatability within 0.5 points of agreement with the trainer ($K_w$ >0.8).

### Intra- and Interobserver Repeatability for Trainers

Both intra- and intertrainer agreement within 0.25 and 0.5 points was almost perfect throughout the 6-mo period, whereas exact agreement was substantial (Table 4).

### Interobserver Repeatability
### Between Trainer and Trainee

The percentage of cows assigned to each of the 14-point BCS of the chart (prevalence of scores) across the 5 time points of the training (d 1 from photographs,

**Table 6.** Means, SD, and minimum (Min) and maximum (Max) weighted kappa coefficient ($K_w$) across the training period for exact agreement and agreement within 0.25 and 0.5 points between trainees and trainer

| Time point[1] | No. of trainees | n[2] | Level of agreement[3] | Mean | SD | Min | Max |
|---|---|---|---|---|---|---|---|
| Day 1: photographs | 12 | 200 | Exact | 0.37 | 0.29 | 0.00 | 0.87 |
| | | | ±0.25 | 0.48 | 0.32 | 0.02 | 0.92 |
| | | | ±0.5 | 0.61 | 0.32 | 0.06 | 0.95 |
| Day 1: live | 12 | 191 | Exact | 0.59 | 0.17 | 0.24 | 0.90 |
| | | | ±0.25 | 0.78 | 0.13 | 0.53 | 1.00 |
| | | | ±0.5 | 0.94 | 0.07 | 0.78 | 1.00 |
| Day 2 | 12 | 172 | Exact | 0.58 | 0.13 | 0.37 | 0.76 |
| | | | ±0.25 | 0.82 | 0.14 | 0.59 | 1.00 |
| | | | ±0.5 | 0.90 | 0.13 | 0.59 | 1.00 |
| Day 7 | 11 | 233 | Exact | 0.35 | 0.23 | 0.04 | 0.82 |
| | | | ±0.25 | 0.55 | 0.27 | 0.15 | 0.94 |
| | | | ±0.5 | 0.78 | 0.28 | 0.23 | 1.00 |
| Day 8 | 9 | 134 | Exact | 0.50 | 0.25 | 0.15 | 0.85 |
| | | | ±0.25 | 0.71 | 0.21 | 0.47 | 1.00 |
| | | | ±0.5 | 0.93 | 0.09 | 0.81 | 1.00 |
| Midway | 6 | 286 | Exact | 0.42 | 0.19 | 0.12 | 0.63 |
| | | | ±0.25 | 0.61 | 0.25 | 0.21 | 0.89 |
| | | | ±0.5 | 0.85 | 0.14 | 0.62 | 1.00 |

[1]On d 1 of the training program, BCS was scored from photo as well as live and on d 2, 7, and 8, BCS was scored live; BCS was scored live during the midway assessment, which was carried out 5 to 15 wk after the training program.

[2]Number of cows scored.

[3]Agreement within 0.25 points with the trainer includes exact agreement, and agreement within 0.5 points includes agreement within 0.25 points.

d 1 from live cows, d 2, d 7, and d 8) and at the midway check ranged from 0 to 47.8% (Table 5). The prevalence of the scores recorded was not equally distributed across the chart. On photographs and live observations in the research dairy units (d 1 and 2), the prevalence of the BCS points recorded was spread across the chart, with more than 5% of cows assigned to BCS points 2.25 to 4.5 from pictures, and to BCS points 2.25 to 3.5 when recorded live. However, in commercial facilities, 67 and 78% of cows were assigned to only 2 BCS points (2.25 to 2.5) on d 7 and 8, respectively, and 76% to 4 BCS points (2.25 to 3) during the midway check. Body condition score points at the extremes of the chart (<2.0 and above 3.5) were recorded live in less than 2% of the scores.

The results of all trainees are presented in this paper where possible (Tables 6 and 7). A few trainees failed to record BCS at some time points of the program (e.g., due to a failure to record a cow's identity). Agreement within 0.5 points between trainee and trainer increased from substantial to almost perfect on d 1 from photographs to live observations. Between d 2 and 8, agreement decreased but increased again on d 8. All trainees reached the target agreement of $K_w$ >0.8 within 0.5 points with the trainer by d 8. At the midway check, agreement remained above target, but 1 out of the 6 trainees had only substantial repeatability with the trainer ($K_w$ = 0.62). Exact agreement and agreement within 0.25 points are also reported in Table 6 and showed similar results to those reported in the prelimi-

nary studies, confirming the choice of target agreement of $K_w$ >0.8 within 0.5 points with the trainer.

When BCS points that rarely occurred were excluded (BCS points recorded in less than 5% of the scores; Tables 5 and 7), occurrences of agreement between trainee and trainer on neighboring scores within 0.5 points ranged across BCS from 46 to 100% of the scores from photographs on d 1, and from 83 to 100% when recorded from live cows (from d 1 to midway check; Table 7). Agreement between trainee and trainer on neighboring scores within 0.25 points occurred from 17 to 93% of the scores from photographs and from 43 to 100% when recorded from live cows. All BCS points together, occurrences of agreement between trainee and trainer on exact scores ranged from 28% (from photographs d 1) to 46% (d 8) of the scores. Agreement between trainee and trainer seemed to occur equally across the chart but interpretation at each BCS point is limited because the prevalence of the scores recorded was not equally distributed across the chart (too few observations at the extremes BCS points), invalidating further statistical analysis.

## DISCUSSION

Assessing BCS could be done in the 4 locations tested but scoring presented more challenges in some locations of a freestall barn compared with others. For example, scoring was easier when cows were headlocked at the feed bunk or when the assessor was located inside the

**Table 7.** Occurrence (%) of agreement between trainee and trainer on exact scores and neighboring scores within 0.25 and 0.5 points for each BCS point across the time points of the training program (n = 12 trainees)

| Time point[1] | n[2] | Level of agreement[3] | <2 | 2 | 2.25 | 2.5 | 2.75 | 3 | 3.25 | 3.5 | 3.75 | 4 | 4.25 | 4.5 | 4.75 | 5 | All BCS points[4] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Day 1: photographs | 200 | Exact | | | 13 | 54 | 21 | 68 | 8 | 33 | 14[5] | 8 | 8 | 17 | | | 28 |
| | | ±0.25 | | | 71 | 62 | 93 | 79 | 75 | 44 | 86[5] | 17 | 50 | 33 | | | 57 |
| | | ±0.5 | | | 77 | 96 | 100 | 100 | 83 | 72 | 86[5] | 46 | 50 | 67 | | | 77 |
| Day 1: live | 191 | Exact | | 57[5] | 28 | 38 | 6 | 66 | 24 | 42 | 0[5] | 80[5] | | | | | 39 |
| | | ±0.25 | | 71[5] | 88 | 67 | 83 | 84 | 86 | 50 | 100[5] | 80[5] | | | | | 76 |
| | | ±0.5 | | 100[5] | 96 | 95 | 94 | 97 | 97 | 100 | 100[5] | 100[5] | | | | | 97 |
| Day 2 | 172 | Exact | 20[5] | 40[5] | 23 | 47 | 31 | 66 | 18 | 50 | 33[5] | 67[5] | | | | | 39 |
| | | ±0.25 | 20[5] | 60[5] | 91 | 75 | 92 | 93 | 72 | 78 | 67[5] | 83[5] | | | | | 79 |
| | | ±0.5 | 80[5] | 100[5] | 100 | 94 | 100 | 100 | 85 | 94 | 100[5] | 100[5] | | | | | 94 |
| Day 7 | 233 | Exact | 100[5] | 45[5] | 15 | 45 | 45 | 47 | 45 | 100[5] | 0[5] | | | | | | 38 |
| | | ±0.25 | 100[5] | 45[5] | 76 | 61 | 85 | 74 | 85 | 100[5] | 0[5] | | | | | | 70 |
| | | ±0.5 | 100[5] | 91[5] | 85 | 95 | 90 | 100 | 100 | 100[5] | 0[5] | | | | | | 92 |
| Day 8 | 134 | Exact | | 100 | 10 | 66 | 0[5] | 55 | 0[5] | 33[5] | 0[5] | | | | | | 46 |
| | | ±0.25 | | 100 | 78 | 77 | 100[5] | 73 | 100[5] | 67[5] | 100[5] | | | | | | 79 |
| | | ±0.5 | | 100 | 93 | 98 | 100[5] | 100 | 100[5] | 100[5] | 100[5] | | | | | | 97 |
| Midway | 286 | Exact | 0[5] | 26 | 17 | 42 | 30 | 50 | 5 | 53 | 0[5] | 100[5] | | 0[5] | | | 31 |
| | | ±0.25 | 0[5] | 43 | 59 | 80 | 82 | 72 | 67 | 74 | 100[5] | 100[5] | | 0[5] | | | 69 |
| | | ±0.5 | 33[5] | 83 | 89 | 96 | 98 | 96 | 100 | 95 | 100[5] | 100[5] | | 0[5] | | | 93 |

[1]On d 1 of the training program, BCS was scored from photographs as well as live, and on d 2, 7, and 8, BCS was scored live; BCS was scored live during the midway assessment, which was carried out 5 to 15 wk after the training program.

[2]Number of cows scored.

[3]Agreement within 0.25 points with the trainer includes exact agreement, and agreement within 0.5 points includes agreement within 0.25 points.

[4]Agreement with the trainer for all BCS points together.

[5]Body condition score recorded in less than 5% of the scores.

home pen (e.g., higher precision of BCS and all 10 body parts of the cows could be observed). Scoring was more difficult when the cows exited the milking parlor or when the assessor was located outside the home pen. Based on our preliminary studies, detailed SOP were developed to assess BCS at each of the 4 locations in freestall farms. However, scoring at headlocks was the most suitable of the 4 locations tested and was then the preferred location during our on-farm epidemiological study. Assessing BCS in tie-stalls is much easier than in freestalls and is similar to scoring at the headlocks in a freestall barn (e.g., cows can be safely approached at less than 50 cm, identified easily, and palpated). Therefore, the same SOP can be used to assess BCS at the freestall with headlocks and at the tie-stall.

Our preliminary studies showed that without any proper SOP to describe the actual steps of the process (e.g., the most suitable time to lock cows is during feed delivery), as well as good references for each BCS point (e.g., detailed definition, photographs, and proper training), naïve assessors provided only with a BCS chart scored with low agreement. Consequently, the BCS chart tested, even if used extensively in the field, is not self-explanatory. After training, repeatability increased greatly and this success was likely due to the combination of 2 factors. First, reference standards were established for each point of the BCS chart by developing a chart that provides a strict definition for each BCS point (i.e., description of the 10 body parts at each BCS point) and developing an illustrated chart composed of a photograph of the rear and side of the cow for each BCS point. Second, proper training was provided, including photographs and live observation scoring, plus discussions about discrepancies in BCS between trainee and trainers, as well as an intensive one-to-one exchange between the trainee and the trainer. Similarly, Kristensen et al. (2006) reported excellent agreement ($K_w$ >0.80) within trained assessors who developed together and harmonized a measurement protocol for BCS, but a range of agreement between nontrained and trained assessors from very poor (0.17) to good (0.78).

The BCS chart used in this study allowed 14 different BCS points, with scores ranging from <2 to 5 in 0.25-point increments and required observing 10 different body parts. This is a very detailed and complex chart. By comparison, some BCS charts use 5 BCS points, ranging from 1 to 5 on in 1-point increments (e.g., Alberta Agriculture and Rural Development, 2004). For management purposes, a need exists to make feeding and health recommendations based on BCS levels at various stages of lactation. Ideally, the precision of the BCS should be sufficient to allow recommendations to be made at the cow level (e.g., targets

for BCS at drying off or at first breeding) and to allow producers to take corrective action for animals under or above recommended condition. These BCS targets are typically made on a scale with 0.25-point increments (e.g., DFC, 2009). Therefore, using a more detailed scoring chart in an animal welfare assessment may achieve another practical goal: providing results that can be used directly by producers as a management tool (i.e., allowing taking corrective actions), which is why we decided to use a more detailed chart in our epidemiological study. Using this very detailed chart, substantial repeatability was achieved on the exact score and within 0.25 points of agreement in a research dairy unit. However, lower agreements at this level of precision were achieved in commercial dairy facilities in the epidemiological study involving more assessors. Finally, if the sole goal of BCS in an animal welfare assessment is to only detect cows with extreme conditions (e.g., identifying too thin cows as a sign of illness or nutrition deficiency), it may be arguable whether such a fine level of precision of BCS is needed. For example, the Canadian standards only mention as a requirement that producers must take corrective action for cows at a BCS of 2 or lower (DFC, 2009). Therefore, assessing if Canadian producers follow their industry standards requires only to count cows at BCS of 2 or lower (as well as to check if corrective actions have been taken for those cows).

Based on the level of agreement achieved in the preliminary studies, the target of the trainees was to ensure and maintain high repeatability within 0.5 points of agreement with the trainer ($K_w$ >0.8) during the 6-mo period of data collection. This target was achieved. Previously, Ferguson et al. (1994) conducted a validation study of the same 14-point BCS chart with 0.25-point increments as used in this study by 4 well-trained classifiers. The classifiers agreed on exact scores 58% of the time and 33% of the scores differed by 0.25 points. These levels of agreement on exact score were higher than those achieved by our trainees in our study, whereas discrepancies in 0.25-point scores were similar. The $K_w$ were not reported in the study of Ferguson et al. (1994) and the description of the proportion of agreement may be misleading because it does not take into account agreement by chance. More recently, using the same BCS chart, Kristensen et al. (2006) conducted a study to estimate the agreement on exact scores among practicing dairy veterinarians attending a teaching workshop on BCS and the $K_w$ ranged from very poor (0.17) to good (0.78), with a moderate average (0.50). These levels of agreement in exact scores between assessors were similar to our findings. Consequently, comparisons between assessors of herd BCS are not warranted unless repeatability as-

sessments have been performed and a need exists to provide more estimates of the repeatability of BCS among consultants working in the field. Allowing an agreement within 0.5 points seems to be necessary to achieve excellent agreement ($K_w$ >0.80) between many assessors for assessment in uncontrolled situations and to allow comparison of herds.

The material developed during the preliminary studies was very useful and, therefore, has been kept for training the assessors for the epidemiological study. Our 1-wk training program achieved excellent agreement between trainers and trainees for BCS and this was maintained during a 6-mo period. These results corroborate those of Kristensen et al. (2006) who showed substantial improvements in agreement between assessors with limited training. This demonstrates that high repeatability of outcome measures can be achieved at a regional or national level despite the large distances between trainees and trainers, and despite differences between trainees in their prior experience with dairy cattle (e.g., injury scoring of dairy cows; Gibbons et al., 2012). The use of the trainers as the reference point to assess agreement was appropriate in this study, as the trainees were located in different parts of Canada. The high level of repeatability in the BCS assigned by trainers justifies the use of their BCS as the reference point for evaluating BCS assigned by trainees. Trainers have also been used as the reference point in other studies (for example, in assessing donkey and pig welfare; Pritchard et al., 2007; Mullan et al., 2011).

Similar to Gibbons et al. (2012), we found that a 5-d break between d 2 and 7 resulted in decreased agreement on d 7 but the agreement improved again on d 8. This highlights the importance of continual practice, particularly during the learning phase. Agreement was lower at the midway check but target agreement was still reached by 5 out of the 6 trainees. This trainee that did not reached the target agreement scored with substantial agreement with the trainer. This illustrates that repeatability should be continually checked at specific time points during data collection and it is not sufficient to carry out 1 test of repeatability at the beginning. Many countries have welfare audits being implemented at the national level, and often repeatability is not reassessed once the assessor has been trained.

It was important that, throughout the training program, trainees were exposed to a sufficient number of cows from each BCS point. The range of BCS points in the current study was similar to the 1.5 to 4.5 range in the study by Ferguson et al. (1994) and slightly higher than that in the study of Kristensen et al. (2006). However, none of these studies reported the prevalence of scores. According to Ferguson et al. (1994), BCS can be assessed to 0.25-point accuracy only between 2.5 and 4.0 (ideal condition), whereas above (too fat) and below (too thin) these values, BCS could only be differentiated to 0.5-point accuracy. In our study, BCS at the extremes of the chart (<2.0 and above 3.5) were recorded live in less than 2% of the scores, whereas the same 2 to 4 BCS (2.25 to 3) accounted for up to 78% of the scores. This deviation from normal distribution did not allow us to conduct further analysis on accuracy of BCS across the chart (e.g., whether some BCS points are more difficult to distinguish than others). However, our limited results tend to show that discrepancies in BCS between assessors occurred equally across the chart. We may suspect that the assessors were not sufficiently exposed to these extreme BCS points on live cows. However, because those extreme cows are rare in the population, the assessors should be able to differentiate extreme- from normal-condition cows and fulfill our expectations to allow benchmarking farms on cow welfare level (i.e., categorizing farms in percentages of too thin, too fat, and in ideal condition). On the other hand, from a management perspective, it may be important to distinguish these extreme scores to monitor changes at the cow level.

## CONCLUSIONS

To produce reliable data and maintain the integrity of on-farm assessment, assessors should follow a training program to achieve high repeatability with a reference standard. To help assessors to cope with a variety of farm setups, SOP need to be provided with the scoring chart and assessors need to be trained to use this chart with proper training material. Continued repeatability checks are essential to ensure that the reference standard is maintained in time and secure quality of data. This method of developing a training program as well as the training program implemented can be used as a model to successfully train on-farm assessors. Allowing agreement within 0.5 points on a 14-point BCS chart with scores ranging from <2 to 5 with 0.25-point increments seems to be necessary to achieve excellent repeatability ($K_w$ >0.80) between many assessors for assessment in uncontrolled situations and to allow comparison of herds. However, in the context of on-farm assessment, if the sole goal of BCS is to only detect cows with extreme conditions (e.g., identifying too thin cows as a sign of illness or nutrition deficiency), it may be arguable whether such fine level of precision of BCS is needed.

## ACKNOWLEDGMENTS

## REFERENCES

Alberta Agriculture and Rural Development. 2004. What's the Score? Body Condition Scoring for Livestock. CD-ROM CD 400/40-1. Alberta Agriculture and Rural Development, Edmonton, AB, Canada.

Bewley, J. M., R. E. Boyce, D. J. Roberts, M. P. Coffey, and M. M. Schutz. 2010. Comparison of two methods of assessing dairy cow body condition score. J. Dairy Res. 77:95–98.

de Vries, M. J., and R. F. Veerkamp. 2000. Energy balance of dairy cattle in relation to milk production variables and fertility. J. Dairy Sci. 83:62–69.

DEFRA (Department for Environment, Food and Rural Affairs). 2001. Condition scoring of dairy cows. DEFRA, London, UK.

DFC (Dairy Farmers of Canada). 2009. Code of practices for the care and handling of dairy cattle. DFC, Ottawa, ON, Canada.

EFSA (European Food Safety Authority). 2011. Scientific opinion on the use of animal-based measures to assess the welfare of dairy cows, EFSA panel on animal health and welfare. EFSA, Parma, Italy.

Elanco Animal Health. 1996. Body condition scoring. Bulletin AI 8478, Rev. 9/96. Elanco Animal Health, Indianapolis, IN.

Ferguson, J. D., D. T. Galligan, and N. Thomsen. 1994. Principal descriptors of body condition score in Holstein cows. J. Dairy Sci. 77:2695–2703.

Gibbons, J., E. Vasseur, J. Rushen, and A. M. de Passillé. 2012. A training programme to ensure high repeatability of injury scoring of dairy cows. Anim. Welf. 21:379–388.

Kristensen, E., L. Dueholm, D. Vink, J. E. Andersen, E. B. Jakobsen, S. Illum-Nielsen, F. A. Petersen, and C. Enevoldsen. 2006. Within- and across-person uniformity of body condition scoring in Danish Holstein cattle. J. Dairy Sci. 89:3721–3728.

Landis, J. R., and G. G. Koch. 1977. The measurement of observer agreement for categorical data. Biometrics 33:159–174.

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

Pritchard, J. C., A. R. S. Barr, and H. R. Whay. 2007. Repeatability of a skin tent test for dehydration in working horses and donkeys. Anim. Welf. 16:181–183.

Roche, J. R., P. G. Dillon, C. R. Stockdale, L. H. Baumgard, and M. J. VanBaale. 2004. Relationships among international body condition scoring systems. J. Dairy Sci. 87:3076–3079.

Roche, J. R., N. C. Friggens, J. K. Kay, M. W. Fisher, K. J. Stafford, and D. P. Berry. 2009. Invited review: Body condition score and its association with dairy cow productivity, health, and welfare. J. Dairy Sci. 92:5769–5801.

Rushen, J., A. Butterworth, and J. C. Swanson. 2011. Animal Behavior and Well-Being Symposium: Farm animal welfare assurance: Science and application. J. Anim. Sci. 89:1219–1228.

Welfare Quality Consortium. 2009. Welfare Quality® assessment protocol for cattle. Welfare Quality Consortium, Lelystad, the Netherlands.

Whay, H. R., D. C. J. Main, L. E. Green, and A. J. F. Webster. 2003. Assessment of the welfare of dairy cattle using animal-based measurements: Direct observations and investigation of farm records. Vet. Rec. 153:197–202.

Wildman, E. E., G. M. Jones, P. E. Wagner, R. L. Boman, H. F. Troutt Jr., and T. N. Lesch. 1982. A dairy cow body condition scoring system and its relationship to selected production characteristics. J. Dairy Sci. 65:495–501.