Training method for animal-based measures in dairy cattle welfare assessments

S. L. Croyle, C. G. R. Nash, C. Bauman, S. J. LeBlanc, D. B. Haley, D. K. Khosa, and D. F. Kelton[1]
*Department of Population Medicine, University of Guelph, Ontario, N1G 2W1 Canada

ABSTRACT

Quantitative assessments of animal welfare are increasingly being used in the dairy industry. It is important to have good precision and accuracy within and among assessors. This study explored the effectiveness of a 3-d training workshop for animal-based measures (ABM) of welfare in dairy cows, in which 14 people were trained to evaluate 6 ABM, specifically hock injuries (HI), lameness (LM), body condition score (BCS), and udder, flank, and leg cleanliness (collectively CLN). All scoring systems were modified to a dichotomous outcome, acceptable or unacceptable. Acceptable HI included no swelling or hair loss, unacceptable HI included swelling or scab; acceptable BCS was >2 on a 5-point scale, unacceptable BCS was ≤2. Acceptable CLN was up to minor splashing, unacceptable CLN was distinct plaques to a solid manure plaque. Lameness was evaluated using locomotion score or in-stall lameness score (SLS) in tiestalls; unacceptable LM was ≥3 (lameness score) on a 5-point scale, where 3 equals mild lameness or ≥2 of 4 behavioral in-stall lameness score indicators were detected. Classroom instruction took place on d 1 of training. Day 2 consisted of group assessment of LM (n = 25 cows), and HI, CLN, and BCS (n = 30 cows), and individual assessment of HI, CLN, and BCS (n = 20 cows) were performed. Day 3 included individual assessments of HI, CLN, and BCS (n = 33 cows), and individual video assessment of LM (n = 27 cows). An additional training video for LM was sent to trainees 3 wk after the workshop, and another follow-up assessment of LM took place via video (n = 37 cows). Repeatability and accuracy of the trainees was assessed using Fleiss's κ (FK) and Byrt's κ (BK) to examine group-level inter-rater agreement and expert-trainee agreement, respectively. The kappa systems use a scale of poor (<0), slight (0.01 to 0.20), fair (0.21 to 0.40), moderate (0.41 to 0.61), substantial (0.61 to 0.80), or almost perfect (0.81 to 1.00). At the conclusion of the workshop, FK was 0.66 for HI and 0.43 for LM, and BK mean (and range) was 0.85 (0.63 to 1.00) for HI and 0.66 (0.56 to 0.85) for LM. Each trainee achieved perfect agreement for BCS [BK mean = 1.00 (1.00 to 1.00)] and each trainee achieved almost perfect agreement for CLN [BK mean 0.90 (0.82 to 0.94)]. After the follow-up video and 3 wk of experience, trainees achieved a FK of 0.66 and a BK mean of 0.74 (0.62 to 0.89) for LM. In conclusion, multiple assessors can achieve substantial agreement for ABM with adequate training.
Key words: lameness, hock injuries, inter-rater reliability

INTRODUCTION

There is increased interest in scoring groups of cows using animal-based measures (**ABM**) such as hock injury, lameness, cleanliness, and body condition to assess welfare on dairy farms for the purposes of research, consulting, animal welfare auditing, and benchmarking (Whay et al., 2003; Rushen et al., 2011; Vasseur et al., 2013). Previous studies have explored training for various individual ABM, including hock injury scoring (Gibbons et al., 2012), gait scoring or lameness (Thomsen et al., 2008; Gibbons et al., 2014), and BCS (Vasseur et al., 2013). To date, no studies have explored training protocols for cleanliness of the leg, flank, and udder.

Assessor training is a useful tool to increase the precision and accuracy, thereby providing more consistent and valid results across farms. This allows the assessment to yield the same results with different assessors, and that these results reflect the assessment of an expert in the field (Lievens, 2001). It has been suggested that methods of scoring can be taught in a group format for injury scoring in dairy cattle, and that these individuals can achieve high levels of repeatability between raters (Gibbons et al., 2014). To our knowledge, no studies have described the process of training numerous people for multiple ABM at one time, which would be more reflective of the way assessments and training are

likely to be conducted. Additionally, no studies have described, in detail, the amount of time spent teaching and practicing during the training process, which would be valuable information to replicate training programs and achieve similar results. To more rigorously assess interobserver reliability and accuracy of the scoring, it is recommended to calculate Cohen's κ between 2 raters (i.e., rater vs. expert), Fleiss's κ when there are more than 3 raters, and Byrt's κ when the prevalence of outcomes is low, as well as reporting the raw percent agreement for comparison. A new program in Canada requires that all Canadian dairy farmers have an assessment done for key ABM. Veterinarians are expected to help farmers in their role as advisors; however, no current training format exists to prepare advisors for all the required ABM.

Our objective was to explore the effectiveness of a 3-d training workshop for ABM of welfare in dairy cows, in which 14 people could be trained to evaluate 6 ABM: hock injuries (**HI**), lameness (**LM**), BCS, and udder, flank, and leg cleanliness (**CLN**). The aim of the study was to provide a training format for future workshops to promote high accuracy and inter-rater agreement for multiple measures of animal welfare.

METHODS

Scoring Constructs

Multiple scoring systems are available for each ABM; however, the methodology for the current study was derived from Dairy Farmers of Canada's proAction Animal Care Assessment because this is the uniform scoring system implemented in a mandatory, industry-wide Canadian program (Dairy Farmers of Canada, 2015). All detailed scoring protocols discussed below were taken directly from Canadian Dairy Research Portal's Animal Comfort Tool, section "How to score injury, cleanliness, body condition, and lameness" (Canadian Dairy Research Portal, 2018). The methods have been previously validated (Domecq et al. 1995; Chapinal et al., 2009; Gibbons et al., 2014) and adapted from the Canadian Code of Practice for the Care and Handling of Dairy Cattle (National Farm Animal Care Council, 2009) and from Gibbons et al. (2012) and Vasseur et al. (2013). The training workshop described in this manuscript was used to train assessors for the National Dairy Study 2015 (Bauman et al., 2018). The workshop was designed to train 14 assessors over a 3-d period for the 6 ABM: HI, LM in freestall and tiestall facilities, BCS, udder, flank, and leg CLN. Six weeks after the original training, inter-rater agreement for LM was assessed again due to lower inter-rater agreement at the end of the 3-d workshop when compared with other ABM.

All scoring systems were modified before the start of the workshop so that the score or outcome was classified as acceptable or unacceptable, creating dichotomous outcomes for all ABM. Two-point systems have been shown to improve agreement among trainees (Knierim and Winckler, 2009) and are the method for the national proAction Animal Care Assessment Program in 2017 (Dairy Farmers of Canada, 2015). A simplified breakdown of the dichotomous scoring system for each ABM is presented in Table 1. The gait of cows housed in freestalls was evaluated on a 5-point lameness score (**LS**), where the acceptable threshold was less than 3 (Flower and Weary 2006). Cows that were housed in tiestalls were evaluated using in-stall lameness scoring (**SLS**; Leach et al., 2009). A video that describes the LS and SLS with example behaviors can be found online (http://www.nationaldairystudy.ca/videos). The SLS detection protocol has been shown to be a valid measurement for lameness when compared with gait scoring (Gibbons et al., 2014; Palacio et al., 2017).

Training Workshop Participants

All training was carried out by one of the authors (C. G. R. Nash), an experienced dairy welfare assessor who had received intensive welfare training and had professionally evaluated more than 200 herds at the time of the study using the same methods. This trainer was labeled as the expert in our study. The primary investigator (S. L. Croyle), who was individually trained by the expert before the workshop, also provided training assistance and workshop coordination. The 14 trainees were undergraduate university students (n = 3), veterinary students (n = 7), and veterinary technicians (n = 4) from the 5 Canadian Colleges of Veterinary Medicine and had a range of minimal or moderate to no previous training on evaluating these ABM. The purpose of the training was to standardize the welfare-scoring methodology to be used as one part of the comprehensive National Dairy Study (Bauman et al., 2018).

Training Methodology

A 3-d workshop was used to train and evaluate assessors for accuracy (ability to correctly identify an ABM) and inter-rater agreement. An additional training and practice video for LM was sent to the trainees 3 wk after the workshop, and a follow-up assessment was performed 6 wk after the workshop. The training method was broken into sessions A through L, described in Table 2, which highlights specific activities conducted during training period. Table 2 also provides the time allotted and the location required for each session, as well as the number of cows or the number of cow im-