8-2009

The Accuracy and Repeatability of Sow Body Condition Scoring

R. F. Fitzgerald
Iowa State University

K. J. Stalder
Iowa State University, stalder@iastate.edu

P. M. Dixon
Iowa State University, pdixon@iastate.edu

A. K. Johnson
Iowa State University, johnsona@iastate.edu

L. A. Karriker
Iowa State University, karriker@iastate.edu

See next page for additional authors

The Accuracy and Repeatability of Sow Body Condition Scoring

Abstract

The objective of this study was to estimate observer accuracy and repeatability of body condition scoring sows when scorers have different levels of prior experience. Three groups of participants (n = 10) for this study were identified as having no (NE, n = 3), some (SE, n = 4), and extensive (EE, n = 3) prior experience evaluating conformation or body condition in livestock species. Two persons having extensive prior experience with body condition scoring served as instructors (TR) during the training sessions. Twenty-five of a total 150 sows were utilized in the participant training session, and the remaining sows (n = 125) were utilized during the independent scoring process. Sows utilized in the scoring process were objectively categorized into a 5- and 9-point body condition score (BCS5 and BCS9, respectively) using last rib backfat estimates. Participants were in poor agreement with BCSbackfat as overall Kappa values were 0.23 on the BCS5 and 0.13 on the BCS9 scales. While the trainers consistently averaged the largest measures of intra- and interobserver agreement with BCSbackfat, other participants primarily in the EE and SE groups achieved similar levels of agreement. Participant BCS5 and BCS9 deviation evaluations from BCSbackfat, revealed a tendency for participants to overestimate BCS in some sows and underestimate BCS in others. While the trainers consistently averaged the largest measures of intra- and interobserver agreement with BCSbackfat, other participants primarily in the EE and SE groups achieved similar levels of agreement.

Keywords

body condition score, interobserver agreement, intraobserver agreement, sows

Disciplines

Agriculture | Animal Sciences | Statistical Methodology | Veterinary Medicine

Comments

This is a manuscript of an article published as Fitzgerald, R. F., K. J. Stalder, P. M. Dixon, A. K. Johnson, L. A. Karriker, and G. F. Jones. "The Accuracy and Repeatability of Sow Body Condition Scoring." *The Professional Animal Scientist* 25, no. 4 (2009): 415-425. doi: 10.15232/S1080-7446(15)30736-1. Posted with permission.

Authors

R. F. Fitzgerald, K. J. Stalder, P. M. Dixon, A. K. Johnson, L. A. Karriker, and G. F. Jones

The accuracy and repeatability of sow body condition scoring

Journal:	*The Professional Animal Scientist*
Manuscript ID:	draft
Manuscript Type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Fitzgerald, Robert; Iowa State University, Animal Science Stalder, Ken; Iowa State University, Animal Science Johnson, Anna; Iowa State University, Animal Science Karriker, Locke; Iowa State University, Veterinary Diagnostic and Production Animal Medicine Jones, Gordon; Western Kentucky University, Agriculture
Key words:	body condition score, interobserver agreement, intraobserver agreement, sows

RUNNING HEAD: Body condition scoring in sows

The accuracy and repeatability of sow body condition scoring

R. F. Fitzgerald,* K. J. Stalder,*[,1] A. K. Johnson,* L. Karriker,† and G. F. Jones‡

*Department of Animal Science, and †Veterinary Diagnostic and Production Animal Medicine Department, Iowa State University, Ames, Iowa 50011; ‡Department of Agriculture, Western Kentucky University, Bowling Green, Kentucky, 42101

[1]Corresponding Author: Kenneth J. Stalder, Department of Animal Science, Iowa State University, 109 Kildee Hall, Ames, Iowa 50011, USA, phone: 515-294-4683, fax: 5125-294-5698, stalder@iastate.edu

ABSTRACT

The objective of this study was to estimate observer accuracy and repeatability of body condition scoring sows when scorers have different levels of prior experience. Three groups of participants (n = 10) for this study were identified as having no (NE, n = 3), some (SE, n = 4), and extensive (EE, n = 3) prior experience evaluating conformation or body condition in livestock species. Two persons having extensive prior experience with body condition scoring served as instructors (TR) during the training sessions. Twenty-five of a total 150 sows were utilized in the participant training session, and the remaining sows (n = 125) were utilized during the independent scoring process. Sows utilized in the scoring process were objectively categorized into a 5- and 9-point body condition score (BCS_5 and BCS_9, respectively) using last rib backfat estimates. Participants were in poor agreement with $BCS_{backfat}$ as overall Kappa values were 0.23 on the BCS_5 and 0.13 on the BCS_9 scales. While the trainers consistently averaged the largest measures of intra- and interobserver agreement with $BCS_{backfat}$, other participants primarily in the EE and SE groups achieved similar levels of agreement. Participant BCS_5 and BCS_9 deviation evaluations from $BCS_{backfat}$, revealed a tendency for participants to overestimate BCS in some sows and underestimate BCS in others. While the trainers consistently averaged the largest measures of intra- and interobserver agreement with $BCS_{backfat}$, other participants primarily in the EE and SE groups achieved similar levels of agreement.

KEYWORDS: body condition score, interobserver agreement, intraobserver agreement, sows.

INTRODUCTION

Sows can enter a negative energy balance during lactation when fat and muscle body reserves are mobilized to support piglet growth through milk production. Additionally, sows that have not attained mature size have nutrient requirements to support growth and these requirements must be met through rations provided during lactation and gestation. After weaning, a limited number of days are available during gestation to replenish the sows' depleted energy stores. Accurate sow body condition estimates are essential so that pork producers can provide sows with an appropriate amount of feed needed to replenish body reserves prior to the next lactation. Thus, it is important for swine producers to have the skills necessary for accurate and repeatable quantifying of condition scores so that feed rations can be adjusted accordingly.

Young et al. (2001) suggested that body condition scores differ between producers or technicians and, based on results by Thomsen et al. (2008) evaluating lameness scoring, within producers' or technicians' own scores. Thus, the objective of this study was to estimate the accuracy and repeatability of participants with different experience levels when scoring body condition.

MATERIALS AND METHODS

The protocol and use of these animals was reviewed and approved by the Iowa State University Animal Use and Care Committee (# 4-08-6548-S) and Intuitional Review Board for human subjects (# 08-218).

Animal Description

Crossbred sows (Yorkshire x Landrace, n = 150) of parity 1 to 6 were housed in standard gestation stalls (0.75 m x 2.15 m). Using a cloth tape measure, heart girth and flank-to-flank measurements were obtained following procedures described by Iwasawa et al. (2004). Sow BW were calculated using the lactation sow BW equation listed in Sulabo et al. (2007) and ranged from 163 to 299 kg. Backfat and loin eye area were ultrasonically (Aloka 500, Corometrics Medical Systems, Wallingford, CT) estimated at the tenth and last rib by a National Swine Improvement Federation certified real-time ultrasound technician (Bates and Christian, 1994).

Twenty-five of the 150 sows were utilized in participant training, and the remaining sows (n = 125) were utilized during the scoring process. Sows utilized in the scoring process were objectively categorized into a 5- and 9-point body condition score (BCS_5 and BCS_9, respectively) using last rib backfat estimates. Backfat guidelines for BCS_5 (1, 2, 3, 4, and 5) are described in Hill et al. (1998). The BCS_9 scale was modified from the BCS_5 scale by adding half measures to the integers (i.e. 1, 1.5, 2, 2.5,…, 5). Backfat guidelines and sow distributions for BCS_5 and BCS_9 are listed in Table 1. Sows of the same BCS may or may not have been housed next to another sow of the same BCS. Variables denoted with a subscript 5 or 9 refer to that variable being calculated using the BCS_5 or BCS_9 scale, respectively.

Participant Description

Three study participant groups were recruited and selected from student populations enrolled in the Animal Science curriculum at a university. Of the 10 total students identified (male = 8, female = 2), 3, 4, and 3 participants were identified as having no (NE), some (SE), and extensive (EE) prior experience evaluating conformation or body condition in livestock species (i.e. pigs, beef cattle, dairy cattle, etc.), respectively. Ideal NE participants included students that had limited visual and physical contact with livestock and had no previous knowledge of body condition scoring methodology in any livestock species. Participants selected for the SE group had previous experience with BCS in beef cattle (Burmeister, 2006), and EE participants included students that had participated on a university-sponsored livestock evaluation team and had detailed knowledge of livestock anatomy and body condition scoring. Student knowledge of BCS was not objectively assessed prior to trial initiation. Investigators selected the students based on the students' participation on livestock evaluation teams or their ability to verbally communicate body condition scoring methodology. Two persons having extensive prior experience with body condition scoring and evaluating swine conformation served as trainers (TR) during the training sessions and as a comparative reference for this study.

Participant Training Description

One trainer having previous experience assessing body condition in sows and beef cattle was responsible for all training sessions. Students completed two, 1-hr training sessions which included one classroom and one on-farm, live animal instructional session. During the 1 hr classroom training, students were provided a diagram of a sow that illustrated anatomical locations where body condition is typically evaluated and pictures of sows representing each BCS (Fitzgerald et al., 2008). Students were instructed to observe sow body condition on the shoulder blade (scapula), spine, hip bones, and tail head. The trainer also discussed methods to objectively determine backfat (A-mode ultrasound, Lean-Meater, Renco Corp, Minneapolis, MN; and B-mode ultrasound, Aloka 500 ultrasound machine), heart girth, and flank-to-flank measurements.

Students participated in an on-farm training session utilizing 25 sows prior to scoring the additional 125 trial sows on the same day. During the on-farm session, the trainer reinforced information presented in the classroom training using live sows that had different BCS at the farm. With the guidance of the trainer, the students practiced determining BCS by verbally assigning individual sow scores.

Participant Scoring Description

Students and trainers scored sows using the BCS_9 scale from the front of the gestation stalls. Thus, participants were able to view the head, back, and both sides of the sows while the sows were in a standing position. This vantage point was established to mimic situations where a producer simultaneously scores sows and adjusts feed drops to provide more, less, or the same amount of the gestation ration. Participants were instructed to assign BCS to the 125 sows independently of other participants. Once the first round was completed for each participant, they were asked to reevaluate the same sows in a different order so that participants could not maintain identity and their associated scores. Sows were also scored by two trainers on the same day they conducted the on-farm training session.

Statistical Analysis

In this study, scores derived from B-mode ultrasonic last rib backfat estimates served as the 'official' BCS ($BCS_{backfat}$) for each sow. From participants original BCS_9, a BCS_5 was derived by taking the integer of the score. For example, sows assigned a BCS_9 of 4.5 would equal a BCS

4 on the 5-point scale. Deviations and their absolute values in participant BCS_5 and BCS_9 from the official were calculated and evaluated using mixed linear models (MIXED, SAS Inst., Inc., Cary, NC). Experience level and round of scoring were fixed effects in the model evaluating BCS deviations. Sow and participant nested within the interaction of experience level and sow served as the random effects. Both BCS_5 and BCS_9 deviations from $BCS_{backfat}$ were evaluated.

Participant bias, standard error of the difference (SED), and standard error of prediction (SEP) were calculated for student BCS_5 and BCS_9 using formulas similar to those published by Bates and Christian (1994).

$$\text{Bias} = \Sigma(score_{ijk} - official_j) / n_i$$

$$\text{SEP} = \sqrt{}(\Sigma(score_{ijk} - official_i - bias_j)^2 / (n_i\text{-}1)$$

$$\text{SED} = \sqrt{}(\Sigma(score2_{ij} - score1_{ij})^2 / n_l$$

where i = total number of sows assigned BCS over both rounds of scoring, j = number of participants, k = the k^{th} assigned BCS, and l = number of unique sows assigned a BCS (l = 125). Standard errors of prediction and SED were compared among experience levels using Analysis of Variance methods (GLM, SAS Inst., Inc.). Experience level served as the only fixed effect because participants only had one estimate of bias, SEP, and SED for the study.

The Kappa and weighted Kappa statistics were calculated using the FREQ procedure of SAS to determine intraobserver agreement between rounds of scoring and interobserver agreement between participants over both rounds of scoring. The Cicchetti-Allison weight type was utilized during the calculations of weighted Kappa. The Kappa statistic is a measure of absolute agreement in BCS for each sow; whereas the weighted Kappa statistic applies weights to scores that are not in agreement (SAS OnlineDoc 9.1, 2003). Participants' Kappa values were evaluated for differences in experience level using the random effect of participant nested within experience level. This statistic has been previously utilized to determine the agreement of lameness scores assigned by multiple raters in pigs (Main et al., 2000) and dairy cattle (Thomsen et al., 2008), and determines the level of agreement beyond that of chance (Feinstein and Cicchetti, 1990; Maclure and Willett, 1987).

Pearson correlation coefficients were calculated between BCS_5 and BCS_9 assigned by participants and ultrasonic measures using the CORR procedure of SAS.

RESULTS AND DISCUSSION

Three, 4 and 3 participants were categorized into the EE (participants 1, 2, and 3), SE (participants 4, 5, 6, and 7), and NE (participants 8, 9, and 10) groups. Participants 11 and 12 served as trainers and conducted the training sessions.

Results from this study are presented on both BCS scales. The reason for using the BCS_9 scale rather than the more traditional BCS_5 scale was to reduce the probability of participants guessing the correct BCS. Of the 125 sows, 60.0% of the sows were categorized as a BCS_5 3 compared to 28.8% using the BCS_9 scale (Table 1). However, the remaining 31.2% of the 60.0% were categorized as a 3.5 on the BCS_9 scale. Minimum and average scores on the BCS_5 were 1.0 and 2.90 and on the BCS_9 was 1.5 and 3.17, respectively. Only 3 sows were categorized as BCS 5 on both scales. In this study, sow BCS were approximately normally distributed for both scales, which is representative of commercial sow operations where producers train employees to BCS sows. However, the authors suggest that, for this study, non-normally distributed BCS may have been more advantageous as to reduce the probability that the participants assigned the correct BCS by chance alone.

Simple means for backfat, muscle, and body conformation are shown in Table 2. Tenth and last rib backfat averaged 25.9 and 18.7 mm, respectively, for all sows in the study. Tenth rib backfat increased an average of 8.6 mm for each score increase in BCS_5` and 4.5 mm for each half score increase in BCS_9. Positive linear trends for tenth ($P < 0.05$ for BCS_5 and BCS_9) and last rib ($P = 0.13$ and $P = 0.03$ for BCS_5 and BCS_9, respectively) loin eye area with BCS were not as evident as those in backfat, which may be attributed to the fact that sows were categorized into BCS based on backfat. However, flank-to-flank measurements and heart girths increased in length as BCS increased, yielding moderate Pearson correlation coefficients of 0.52 and 0.36 for

BCS_5 ($P < 0.001$) and 0.58 and 0.44 for BCS_9 ($P < 0.001$), respectively. Heart girth was found to be moderately correlated (0.69) with flank-to-flank measurements.

Deviations from the $BCS_{backfat}$

Average deviations, in this study named participant bias, between BCS assigned by participants of different experience groups and $BCS_{backfat}$ for 2 rounds of scoring are shown in Table 3. Experience level ($P < 0.001$), round of scoring ($P < 0.001$), and their interaction ($P < 0.001$) were significant sources of variation in the evaluation of deviations between participant BCS and $BCS_{backfat}$.

Mean deviations from the official over all rounds and experience levels for BCS_5 and BCS_9 were -0.01 and 0.04, respectively (Table 3). When evaluating the overall main level effect of experience level, the EE group on the BCS_9 scale and the SE group on both scales yielded average deviations significantly different than $BCS_{backfat}$. Conversely, only numerical deviations were observed for the TR ($P = 0.70$) and NE ($P = 0.44$) groups on both scales.

The EE group consistently underestimated $BCS_{backfat}$ whereas the SE group overestimated $BCS_{backfat}$ during both rounds of scoring. Interestingly, the NE and TR mean deviations were similar across both rounds and BCS scales in that both groups underestimated BCS during the first round then overestimated BCS during the second round of scoring. During the second round of scoring, all experience level groups, except EE, overestimated their first round scores.

Only a slight numerical difference was observed between the main level effect of scoring round. Overall deviations for the first and second round of scoring were 0.00 and 0.08 on the BCS_5 scale and -0.06 and 0.03 on the BCS_9 scale, respectively.

The interaction of experience level by round of scoring was a significant ($P < 0.001$) source of variation in deviations from $BCS_{backfat}$. Participants in the TR group during the first round ($P = 0.04$) and participants in the EE ($P = 0.03$) and SE ($P < 0.001$) group in the second round (BCS_9) were significantly different than zero. Evaluating only the BCS_5 scale, only deviations for participants for the SE group during the second round of scoring were significantly different from zero ($P < 0.001$). When implementing subjective BCS methods in swine production, it is important for scorers to be cognizant of their average bias so that scorers could increase or decrease their assigned BCS and feed could be adjusted more accurately.

Absolute deviations were calculated by taking the absolute value of the deviation between the participant BCS and $BCS_{backfat}$. This statistic allows for the true deviation from $BCS_{backfat}$ as compared to participant bias. For example, participants could consistently be 1 score above or below the $BCS_{backfat}$ and average a zero deviation; whereas taking the absolute value of the deviations would average a 1 score absolute deviation. Least squares means for absolute deviations from $BCS_{backfat}$ for experience level groups are shown in Table 3. The main level effect of experience group was a significant source of variation ($P < 0.02$) in absolute deviations for the BCS_9 scale only. The TR and SE groups had the numerically lowest absolute deviation on the BCS_5 scale. The TR group had lower ($P < 0.05$) absolute deviations compared to the EE and NE groups. The TR and SE group did not differ statistically on the BCS_9 scale.

The main level effect of round of scoring as well as the interaction of experience level by round of scoring, which was removed after determining significance level, was not associated ($P > 0.05$) with absolute deviations.

Upon further evaluation of participant BCS_5 and BCS_9 deviations from $BCS_{backfat}$, a tendency was observed for participants to overestimate BCS in some sows and underestimate BCS in others. From this observation, a hypothesis was formed to determine whether participants considered factors other than body condition when estimating BCS in sows. To evaluate this hypothesis, regression analyses of participant deviations were performed using tenth and last rib loin eye area and backfat, flank-to-flank measurements, and heart girth as regression terms and participants as fixed effects. Participant within sow and sow were included as random terms in the model predicting BCS deviations. Interactions of participants with each regression term were included in the model and then removed if found insignificant at $P > 0.05$.

Individual participant least square means of deviations from $BCS_{backfat}$ ranged from -0.24 to 0.30, with an overall average of 0.03 on the BCS_5 scale. Similar ranges were found on the BCS_9 scale (-0.32 to 0.24) with an overall average of -0.03. Clearly, some participants tended to overestimate BCS while others underestimated BCS. Last rib loin eye area and backfat and their interaction with participants were significant sources of variation for deviations in participant scores from $BCS_{backfat}$. Heart girth (BCS_5 and BCS_9 scale) as well as the interaction term heart girth by participant (BCS_5 scale) were observed to be associated with participant deviations. Heart girth has been shown to be associated with body weight (Iwasawa et al. 2004), suggesting that some participants may have included other factors like loin eye area, body size, BW, or any combination of the three to assist them in their evaluation and estimation of sow body condition. A low Pearson correlation coefficient of 0.22 was observed between last rib loin eye area and backfat, which may lend more evidence to the observation that participants truly evaluated muscle in addition to body condition.

Standard Error of the Difference and of Prediction

Standard errors of the difference (SED) between BCS assigned by participants during the first and second round are an estimate of the ability of individual participants to consistently assign the same BCS to sows. Standard errors of the difference for individual participants and participants categorized by prior BCS experience are shown in Table 3.

Overall, the standard error between participants' first and second round score averaged 0.59 on the BCS_5 and 0.50 on the BCS_9 scale. Prior BCS experience did not affect (P = 0.61) the ability of participants to repeatedly estimate the same BCS_9 for each sow between rounds. However, participants in the TR group averaged numerically lower SED than the EE (-0.12), SE (-0.11) and NE (-0.08) groups. This observed lack of effect of prior BCS experience on SED might be attributable to the large standard errors from too few participants in the study. Yet, participants 11 and 12 (whom served as trainers) had nearly identical SED values for both BCS scales (data not shown). Only two of the 10 students (1 participant in the NE and SE groups) had SED values equal to or less than those observed for the TR group.

Standard errors of the difference averaged 0.1 score less on the 9-point BCS scale than SED on the 5-point scale. This may be a result of a higher frequency of 1 score differences when calculating SED_5 as compared to half score differences when calculating SED_9. For example, squaring a 1 BCS difference between the participant and official scores would equal 1, whereas squaring a half BCS difference would equal 0.25.

Standard errors of prediction are an estimate of the ability of participants to estimate $BCS_{backfat}$. Standard errors of prediction for individual participants and participants categorized by prior BCS experience are listed in Table 3. Overall, participants in this study were consistently within 0.74 and 0.65 scores of $BCS_{backfat}$ on the BCS_5 and BCS_9 scales, respectively. Similar to SED, only numerical (P = 0.65 for BCS_5 and 0.32 for BCS_9) differences were observed for differences in SEP for participant experience level using both scales. Participants that served as trainers in this study averaged a numerically lower SEP (-0.06 on the BCS_5 and -0.09 on the BCS_9) than participants with extensive prior BCS experience. Similar to SED, SEP values were larger for the BCS_5 scale as compared to the BCS_9 scale.

Intra- and Interobserver agreement

Kappa and weighted Kappa values for intraobserver agreement are listed in Table 4. Overall intraobserver Kappa and weighted Kappa values on the BCS_5 scale averaged 0.43 (range of 0.25 to 0.58) and 0.51 (range of 0.29 to 0.64), respectively. Intraobserver Kappa and weighted Kappa values on the BCS9 scale averaged 0.29 (range of 0.13 to 0.40) and 0.51 (range of 0.33 to 0.61), respectively, which are considered to be in good agreement (Fleiss, 1981). All intraobserver agreement Kappa values for participants were significantly different than zero ($P <$ 0.001), which lends evidence that participants were able to show some level of repeatability beyond that of chance when body condition scoring sows. While Kappa values showed an average 32% drop in magnitude when calculated on the BCS_9 scale as compared to BCS_5 for intraobserver agreement, weighted Kappa averages showed no difference between the two scales. This may be a result of the BCS_9 scale having more categories and thus reduces the probability that participants will assign the correct BCS by chance alone.

The TR group had the numerically largest Kappa and weighted Kappa values of all groups and on both BCS scales. The EE group had the second numerically largest Kappa and weighted Kappa values on both scales, thus suggesting that some previous knowledge of evaluating body condition or conformation in sows is beneficial to the repeatability when scoring sows. This may possibly be a result of the confidence level of the participants in the TR and EE groups, although confidence level was not evaluated in this study.

Interobserver Kappa and weighted Kappa values were generally lower than intraobserver values (Table 5), which is in agreement with Thomsen et al. (2008) and Martin et al. (1987). A total of 8 interobserver Kappa values were found to be non-significant ($P > 0.05$) and none for weighted Kappa values. Interobserver agreement was calculated between each participant in the study and BCS derived from last rib backfat, and are listed in Table 5. Participants were in poor agreement with $BCS_{backfat}$ as overall Kappa values were 0.23 on the BCS_5 and 0.13 on the BCS_9 scales. However, participant agreement increased using weighted Kappa values and values were similar for both scales (0.34 for BCS_5 and 0.35 for BCS_9). Participant 11, whom served as a trainer for the live animal training sessions, had the largest Kappa and weighted Kappa values for the BCS_9 scale and ranked second to Participant 2 in both statistics on the BCS_5 scale.

Training sessions were conducted prior to any independent scoring, and were concluded after the 4-5 person training group expressed some level of verbal competency to the trainers. The unfavorable result that some participants had no agreement with other participants suggests that the participants may have expressed competency but had not yet achieved it. Relating to commercial sow production, employees may verbally express their ability to evaluate body condition to their employer, but in reality this discrepancy may yield larger variations in sow body condition in the herd. In a similar study that evaluated the effect of training on intra- and interobserver agreement during lameness evaluation scoring (using a 5-point scale), Thomsen and coworkers (2008) found that while training slightly decreased intraobserver agreement, training slightly improved interobserver agreement. The lack of absolute agreement between participants may be explained by participants assigning scores on their own scale, as evidenced by the wide range of average deviations shown in Table 3, and is in agreement with Thomsen et al. (2008).

Correlations

Pearson correlation coefficients for BCS_5 and BCS_9 with ultrasonic measures are shown in Table 6. Correlations between BCS_9 and ultrasonic measures were consistently greater than those calculated using the BCS_5 scale. Averaging over participants, the ultrasonic trait of last rib backfat yielded the greatest correlation (0.58) with BCS_9, followed by tenth rib backfat (0.51), tenth rib loin eye area (0.47), and last rib loin eye area (0.44). Similar trends were observed for the BCS_5 scale. Young and coworkers (2001) found a positive, but low relationship ($r^2 = 0.19$) of BCS_5 assigned by a farm manager and last rib backfat in 1306 sows using A-mode ultrasound. In this study, last rib backfat estimates were estimated using B-mode ultrasound.

Both tenth and last rib loin eye area were positively correlated with BCS_5 and BCS_9 for all participants. Larger loin eye area may have influenced participants to assign larger BCS to sows as compared to if BCS were assigned based on body condition alone. However, this result should not be considered negative, as both body condition and muscle are mobilized to support piglet growth through milk production.

Pearson correlation coefficients were calculated for participants' first and second round of scoring and values are listed in Table 6. Correlations between rounds of scoring within participants averaged 0.72, ranging from 0.43 to 0.82. Scores for participants in the TR group were moderately correlated (0.73).

Feeder Adjustments

Accurately sow body condition estimates enable producers to adjust feed rations to replenish or maintain body reserves. To hypothetically evaluate whether participants would have performed some corrective feed adjustment for sows outside BCS_9 3.0 and 3.5 scores, sows assigned $BCS_{backfat}$ less than 3.0 and greater than 3.5 were counted separately as well as the number of sows assigned to those categories by the participants. Participants averaged performing some level of corrective feed ration adjustments for 49.2%, ranging from 21.7 to 73.3%, of the 60 possible opportunities (30 sows of BCS_9 less than 3.0 x 2 rounds of scoring) to identify a sow of less than BCS_9 3.0 (Table 7). Similarly, of the possible 40 opportunities to identify a BCS_9 4.0 or greater, participants correctly adjusted 53.1%, ranging from 27.5 to 70.0%. Only 4 participants had both scores within a 10% range (average = 4%; e.g. participant 2 found 60.0% of sows with BCS_9 less than 3.0 and 62.5% of sows with BCS_9 4.0 or greater); whereas, the remaining participants averaged a 32% difference between correctly adjusting feeders for thin and fat sows. This evidence supports previous observations that most participants in this study scored BCS using their own scale.

In conclusion, participants in this study tended to assign BCS on their own scale and had individual specific biases towards over- or underestimating BCS in sows. Interobserver Kappa values were greater for the BCS_5 scale compared to the BCS_9 scale, but weighted Kappa values were similar for both scales suggesting that participant repeatability increased on the BCS_5 due to chance alone. While the trainer consistently averaged the largest measures of intra- and interobserver agreement with $BCS_{backfat}$, other participants primarily in the EE and SE groups achieved similar levels of agreement.

LITERATURE CITED

Bates, R. O., and L. L. Christian. 1994. The National Swine Improvement Federation guidelines for ultrasonic certification programs. Fact Sheet NSIF-FS16, Cooperative Extension Service, Purdue Univ., West Lafayette.

Burmeister, P. R. 2006. Effects of training and experience on repeatable and reliable assessment of beef cattle body condition scores. M.S. Thesis. Western Kentucky University, Bowling Green.

Feinstein, A. R., and D. V. Cicchetti. 1990. High agreement but low kappa: 1. The problems of two paradoxes. J. Clin. Epidemiol. 43:543-549.

Fitzgerald, R. F., K. J. Stalder, L. A. Karriker, C. Johnson, L. Layman, and A. K. Johnson. 2008. Developing and utilizing visual tools to assist pork producers in employee training in the evaluation of sow body condition. J. Ext. 46:2T0T7.

Fleiss, J. L. 1981. Statistical methods for rates and proportions. 2nd ed. John Wiley, New York, NY.

Hill, G., D. W. Rozeboom, N. Trottier, D. C. Mahan, L. Adeoli, T. Cline, D. Forsyth, and T. B. Richert. 1998. Tri-state swine nutrition guide. http://ohioline.osu.edu/b869/b869_44.html. Accessed January 15, 2009.

Iwasawa, T., M. G. Young, T. P. Keegan, M. D. Tokach, S. S. Dritz, R. D. Goodband, J. M. DeRouchey, and J. L. Nelssen. 2004. Comparison of heart girth or flank-to-flank measurements for predicting sow weight. p. 17 in 2004 Swine Day, Kansas State Univ., Manhattan.

Maclure, M., and W. C. Willett. 1987. Misinterpretation and misuse of the kappa statistic. J. Epidemiol. 126:161-169.

Main, D. C., J., J. Clegg, A. Spatz, and L. E. Green. 2000. Repeatability of a lameness scoring system for finishing pigs. Vet. Rec. 147:574-576.

Martin, S. W., A. H. Meek, and P. Willeberg. 1987. Veterinary epidemiology. Principles and methods. Iowa State University Press, Ames.

SAS Institute, Inc. 2003. SAS OnlineDoc 9.1. SAS Institute, Inc., Cary, NC.

Sulabo, R. C., M. D. Tokach, S. S. Dritz, E. J. Wiedemann, R. D. Goodband, J. M. DeRouchey, and J. L. Nelssen. 2007. Validation of flank-to-flank allometric equations in predicting weight of lactating sows and lactation weight change. p. 54 in Swine Day, Kansas State Univ., Manhattan.

Thomsen, P. T., L. Munksgaard, and F. A. Togersen. 2008. Evaluation of a lameness scoring system for dairy cows. J. Dairy Sci. 91:119-126.

Young, M. G., M. Tokach, R. D. Goodband, J. L. Nelssen, and S. S. Dritz. 2001. The relationship between body condition score and backfat in gestating sows. p. 5 in Swine day, Kansas State Univ., Manhattan.

LIST OF TABLES AND FIGURES

Table 1. Last rib backfat guidelines utilized to categorize sows into a 5- and 9-point body condition score and the subsequent distribution by BCS.

[1]BCS = Body condition score; BCS were scored using a 5 [BCS_5; 1, 2, …, 5] or 9 [BCS_9; 1, 1.5, 2, …, 5] point BCS scale.

	Backfat, mm				Number of sows[2]	
	BCS_5		BCS_9		BCS_5	BCS_9
BCS[1]	Minimum	Maximum	Minimum	Maximum	N	N
1.0	0	9.9	0	7.5	6	0
1.5	-	-	7.5	9.9		6
2.0	10	14.9	10	12.4	24	9
2.5	-	-	12.5	14.9		15
3.0	15	22.9	15	18.9	75	36
3.5	-	-	19	22.9		39
4.0	23	29.9	23	26.4	17	9
4.5	-	-	26.5	29.9		8
5.0	30	-	30	-	3	3

Table 2. Average tenth and last rib backfat and loin eye area, heart girth, and flank-to-flank measurements for 125 sows categorized into body condition scores[1] (BCS) using last rib backfat estimates in a study of estimating visual body condition scores.

[1]BCS were scored using a 5- (BCS_5; 1, 2, …, 5) or 9-point (BCS_9; 1, 1.5, 2, …, 5) BCS scale. Backfat guidelines for BCS_5 are described in Hill et al. (1998).

	10th rib backfat, mm		Last rib backfat, mm		10th rib loin eye area, mm		Last rib loin eye area, mm		Heart girth, mm		Flank-to-flank measurement, mm	
BCS	BCS_5	BCS_9	BCS_5	BCS_9	BCS_5	BCS_9	BCS_5	BCS_9	BCS_5	BCS_9	BCS_5	BCS_9
1.0	12.7	-	9.1	-	44.6	-	48.3	-	50.7	-	38.2	-
1.5	-	12.7	-	9.1	-	44.6	-	48.3	-	50.7	-	38.2
2.0	18.7	16.0	13.1	11.4	51.4	53.9	52.5	54.9	54.1	51.6	39.6	38.9
2.5	-	20.4	-	14.0	-	50.1	-	51.3	-	55.4	-	40.1
3.0	26.5	24.4	18.9	17.0	53.1	50.4	51.8	48.3	55.6	53.9	41.5	40.5
3.5	-	28.4	-	20.7	-	55.5	-	55.0	-	57.1	-	42.3
4.0	34.9	33.6	26.4	24.7	53.7	54.8	54.9	56.3	56.6	57.3	43.7	43.6
4.5	-	36.4	-	28.3	-	52.5	-	53.4	-	55.8	-	43.9
5.0	43.9	43.9	33.2	33.1	53.7	53.7	54.7	54.7	60.0	60.0	44.0	44.0

Table 3. Least square means of deviations[1] and absolute deviations[2], standard errors of the difference[3] (SED) and standard errors of prediction[4] (SEP) calculated from two rounds of assigning body condition scores (BCS) to 125 sows by individual study participants categorized as trainers (TR; n = 2) or as students having extensive (EE; n = 3), some (SE; n = 4), or no (NE; n = 3) prior BCS experience.

[1] Sows were assigned an official 5- (1, 2, …, 5) and 9-point (1, 1.5, 2, …, 5) BCS using last rib backfat estimates. Deviations were calculated as BCS assigned by participants minus $BCS_{backfat}$.

[2]Absolute value of the deviation.

[3]SED = $\sqrt{(\Sigma(score_{ij}2 - score_{ij}1)^2 / n_l)}$, where where i = the total number of sows assigned BCS over both rounds of scoring, j = number of number of participants, k = the kth assigned BCS, and l = the number of unique sows assigned a BCS (l = 125).

[4]SEP = $\sqrt{(\Sigma(score_{ijk} - official_i - bias_j)^2 / (n_i\text{-}1))}$, where i = the total number of sows assigned BCS over both rounds of scoring, j = number of number of participants, k = the kth assigned BCS, and l = the number of unique sows assigned a BCS (l = 125).

Statistic	Experience Level	Round of Scoring	BCS_5	SEM_5	BCS_9	SEM_9
Mean deviation						
	EE	1	-0.03	0.06	-0.09	0.05
	EE	2	-0.07	0.06	-0.11**	0.05
	SE	1	0.07	0.06	0.03	0.05
	SE	2	0.21**	0.06	0.17**	0.05
	NE	1	-0.01	0.06	-0.07	0.05
	NE	2	0.08	0.06	0.03	0.05
	TR	1	-0.04	0.06	-0.11*	0.05
	TR	2	0.08	0.06	0.03	0.05
Mean absolute deviation						
	EE	-	0.51	0.04	0.53^{a}	0.03
	SE	-	0.48	0.03	0.50^{ab}	0.03
	NE	-	0.49	0.04	0.52^{a}	0.03
	TR	-	0.48	0.04	0.46^{b}	0.03
Standard Error of the Difference						
	EE	-	0.61	0.05	0.53	0.06
	SE	-	0.62	0.04	0.52	0.05
	NE	-	0.58	0.05	0.49	0.06
	TR	-	0.53	0.06	0.41	0.07
Standard Error of Prediction						
	EE	-	0.77	0.03	0.69	0.03
	SE	-	0.73	0.03	0.65	0.02
	NE	-	0.74	0.03	0.66	0.03
	TR	-	0.71	0.04	0.60	0.04

*P< 0.1, **P< 0.05, ***P< 0.01 denotes different from zero.

Means bearing different superscripts within a heading within a column at P < 0.05

Table 4. Intraobserver agreement from two rounds of assigning body condition scores (BCS) to 125 sows by individual study participants categorized as trainers (TR; n = 2) or as students having extensive (EE; n = 3), some (SE; n = 4), or no (NE; n = 3) prior BCS experience.

[2]Experience level Least square means were not significantly different ($P < 0.05$).

		Intraobserver Kappa				Intraobserver Weighted Kappa			
		BCS_5	SE_5	BCS_9	SE_9	BCS_5	SE_5	BCS_9	SE_9
Participant	Group[2]								
1	EE	0.39	-	0.32	-	0.49	-	0.51	-
2	EE	0.48	-	0.40	-	0.60	-	0.60	-
3	EE	0.51	-	0.29	-	0.57	-	0.50	-
4	SE	0.41	-	0.25	-	0.51	-	0.51	-
5	SE	0.45	-	0.21	-	0.52	-	0.46	-
6	SE	0.25	-	0.13	-	0.43	-	0.41	-
7	SE	0.36	-	0.34	-	0.43	-	0.55	-
8	NE	0.39	-	0.22	-	0.47	-	0.46	-
9	NE	0.58	-	0.36	-	0.64	-	0.61	-
10	NE	0.27	-	0.24	-	0.29	-	0.33	-
11	TR	0.50	-	0.36	-	0.60	-	0.61	-
12	TR	0.52	-	0.34	-	0.59	-	0.57	-
Mean	EE	0.46	0.06	0.33	0.04	0.55	0.06	0.54	0.05
Mean	SE	0.37	0.05	0.23	0.04	0.47	0.05	0.49	0.04
Mean	NE	0.41	0.06	0.27	0.04	0.47	0.06	0.47	0.05
Mean	TR	0.51	0.07	0.35	0.05	0.59	0.07	0.59	0.06
P		*0.46*		*0.22*		*0.40*		*0.40*	

Table 5. Interobserver agreement for individual study participants from two rounds of assigning body condition scores (BCS) to 125 sows on a 5- and 9-point scale. Kappa and Weighted Kappa values are reported below and above the diagonal, respectively.

BCS_5	Participant												
Participant	1	2	3	4	5	6	7	8	9	10	11	12	$BCS_{backfat}$
1		0.41	0.45	0.34	0.45	0.44	0.35	0.38	0.47	0.21	0.44	0.44	0.27
2	0.28		0.42	0.48	0.51	0.45	0.38	0.42	0.50	0.35	0.48	0.44	0.42
3	0.34	0.30		0.33	0.45	0.40	0.25	0.42	0.52	0.19	0.35	0.46	0.37
4	0.20	0.35	0.19		0.50	0.44	0.46	0.47	0.38	0.41	0.54	0.37	0.33
5	0.34	0.40	0.35	0.40		0.46	0.43	0.53	0.50	0.36	0.47	0.48	0.36
6	0.29	0.29	0.28	0.29	0.34		0.42	0.38	0.43	0.27	0.44	0.39	0.39
7	0.23	0.26	0.12	0.38	0.35	0.33		0.40	0.33	0.37	0.44	0.32	0.33
8	0.27	0.30	0.34	0.39	0.48	0.26	0.32		0.47	0.31	0.45	0.46	0.31
9	0.36	0.40	0.45	0.23	0.41	0.28	0.21	0.40		0.21	0.44	0.53	0.36
10	0.08 NS	0.26	0.09*	0.37	0.30	0.18	0.32	0.24	0.10*		0.39	0.24	0.26
11	0.33	0.35	0.20	0.46	0.37	0.32	0.36	0.36	0.34	0.34		0.41	0.41
12	0.34	0.35	0.38	0.27	0.39	0.24	0.21	0.38	0.47	0.14	0.31		0.30
$BCS_{backfat}$	0.15	0.30	0.27	0.21	0.27	0.24	0.22	0.24	0.25	0.19	0.27	0.17	
BCS_9													
1		0.41	0.37	0.33	0.44	0.40	0.34	0.34	0.47	0.23	0.47	0.38	0.32
2	0.10		0.38	0.47	0.48	0.42	0.41	0.40	0.46	0.36	0.49	0.42	0.40
3	0.12	0.13		0.27	0.37	0.34	0.25	0.36	0.43	0.18	0.32	0.39	0.32
4	0.07*	0.19	0.05 NS		0.46	0.42	0.50	0.40	0.36	0.41	0.51	0.34	0.33
5	0.17	0.18	0.13	0.20		0.51	0.46	0.47	0.48	0.36	0.49	0.43	0.38
6	0.11	0.14	0.10	0.12	0.21		0.42	0.36	0.42	0.28	0.44	0.37	0.39
7	0.08	0.16	0.01 NS	0.30	0.24	0.20		0.39	0.34	0.43	0.51	0.30	0.36
8	0.10	0.12	0.13	0.16	0.28	0.12	0.20		0.45	0.30	0.44	0.42	0.30
9	0.22	0.18	0.19	0.07*	0.21	0.12	0.11	0.22		0.25	0.45	0.52	0.36
10	0.01 NS	0.15	0.01 NS	0.23	0.17	0.11	0.25	0.13	0.04 NS		0.40	0.24	0.27
11	0.20	0.16	0.06 NS	0.25	0.25	0.17	0.23	0.19	0.16	0.22		0.43	0.45
12	0.13	0.18	0.20	0.11	0.14	0.12	0.04 NS	0.17	0.28	0.08	0.16		0.35
$BCS_{backfat}$	0.10	0.15	0.12	0.09	0.14	0.15	0.15	0.11	0.12	0.07*	0.19	0.12	

NS non significant from zero, *$P<0.03$, All values are significant at the $P<0.001$ level.

Table 6. Pearson correlation coefficients for tenth and last rib backfat and loin eye area estimates[1] and rounds of scoring with participant assigned BCS using a 5- and 9- point scale

	Participant											
	1	2	3	4	5	6	7	8	9	10	11	12
Correlations for BCS_5												
BCS_9	0.951	0.950	0.928	0.933	0.923	0.958	0.897	0.923	0.928	0.901	0.940	0.925
10^{th} rib backfat	0.396	0.560	0.490	0.464	0.442	0.513	0.502	0.455	0.514	0.355	0.613	0.457
Last rib backfat	0.466	0.606	0.548	0.526	0.521	0.590	0.546	0.460	0.590	0.419	0.642	0.557
10^{th} rib Loin eye area	0.501	0.446	0.418	0.492	0.442	0.447	0.388	0.438	0.493	0.320	0.454	0.464
Last rib Loin eye area	0.450	0.397	0.445	0.483	0.436	0.410	0.338	0.379	0.485	0.285	0.457	0.433
Correlations for BCS_9												
BCS_5	0.951	0.950	0.928	0.933	0.923	0.958	0.897	0.923	0.928	0.901	0.940	0.925
10^{th} rib backfat	0.439	0.569	0.497	0.509	0.496	0.499	0.551	0.444	0.549	0.409	0.639	0.521
Last rib backfat	0.510	0.607	0.561	0.568	0.581	0.585	0.614	0.483	0.612	0.481	0.678	0.618
10^{th} rib Loin eye area	0.499	0.493	0.420	0.516	0.484	0.466	0.479	0.457	0.528	0.347	0.503	0.480
Last rib Loin eye area	0.452	0.461	0.436	0.507	0.461	0.432	0.415	0.402	0.501	0.293	0.501	0.457
Rounds of scoring	0.691	0.780	0.697	0.759	0.697	0.663	0.756	0.687	0.810	0.434	0.820	0.793

[1]Ultrasonic measures were estimated using a B-mode ultrasound device.

Table 7. Percentage of feeders[1] that participants correctly performed[2] some amount of corrective adjustment.

Participant	N of BCS_9 2.5 or less sows assigned by participants	N of BCS_9 2.5 or less sows assigned by last rib backfat.	Percentage of sows correctly assigned BCS_9 2.5 or less	N of BCS_9 4 or greater sows assigned by participants	N of BCS_9 4 or greater sows assigned by last rib backfat.	Percentage of sows correctly assigned BCS_9 4 or greater
1	33	60	55.0%	17	40	42.5%
2	36	60	60.0%	25	40	62.5%
3	44	60	73.3%	15	40	37.5%
4	19	60	31.7%	28	40	70.0%
5	29	60	48.3%	21	40	52.5%
6	36	60	60.0%	27	40	67.5%
7	15	60	25.0%	25	40	62.5%
8	27	60	45.0%	19	40	47.5%
9	42	60	70.0%	14	40	35.0%
10	13	60	21.7%	26	40	65.0%
11	27	60	45.0%	27	40	67.5%
12	33	60	55.0%	11	40	27.5%

[1]N of sows assigned BCS over two rounds of scoring (i.e. a total of 250 scoring opportunities).

[2]Assuming participants would not adjust drop feeders of BCS_9 3 or 3.5 sows