2018

Experience Improves the Reliability of Subjective Measurements of Temperament in Beef Cattle

Jaime T. Praham
University of Nebraska - Lincoln

Amy E. Tanner
Virginia Tech

Mark L. Wahlberg Wahlberg
Virginia Tech

W. S. Swecker Jr.
Maryland College of Veterinary Medicine

Ronald M. Lewis
University of Nebraska-Lincoln, ron.lewis@unl.edu

Experience Improves the Reliability of Subjective Measurements of Temperament in Beef Cattle

Jamie T. Parham
Amy E. Tanner
Mark L. Wahlberg
W. S. Swecker, Jr.
Ronald M. Lewis

Summary with Implications

Reliability of experienced and inexperienced observers when assessing the behavior of cattle when restrained in a squeeze chute (chute score), and when exiting the chute (exit score), was compared. Overall, experienced observers had higher reliability than inexperienced observers. Increasing the number of individuals scoring an animal decreased the degree of agreement. However, within an acceptable tolerance for difference in scores, such disagreement may be beneficial; it allows for subtlety in interpretations of temperament, which when averaged, may better reflect docility. Reliabilities were higher for exit score than chute score. This may reflect the complexity of the trait being evaluated, with fewer behaviors observed when cattle exit as compared to when restrained in a chute. Producers may profitably use chute and exit score to quantify docility in cattle. However, it may be worthwhile to gain experience in using the scoring system before implementing it for selection decisions.

Introduction

There are many negative effects associated with excitable temperament in cattle such as increased risk to handlers, poorer weight gain and meat eating quality, decreased tolerance to disease, and increased production costs. This has led to an increase in selection for docility. Breeding values for docility are routinely estimated in beef cattle, but the question remains as to how efficient the industry is at quantifying docility when measured subjectively.

Subjective scoring of the behavior of cattle when restrained in a squeeze chute (chute score), and when exiting the chute (exit score), has been proposed as a method to measure temperament of animals. These measurements are fast, simple and inexpensive to collect, making them attractive. However, research using these methods report inconsistent results, some proposing the use of these scoring systems while others not. Being a subjective measurement, the scores assigned to an animal are based on the opinion of the observer, which can lead to varying conclusions. Consistency of these measurements is crucial to the effectiveness of applying them in cattle enterprises to select for more docile cattle.

Previous research from this group has shown that chute and exit scores are effective methods of measuring temperament (2018 Nebraska Beef Cattle Report 74–76). The objective of this study was to determine the reliability of assessment of these scores, which could impact their value when making selection decisions.

Procedure

A 3-year study conducted at Kentland Farm, Blacksburg, VA, utilized predominately Angus (75% or more), spring-born heifer calves. Each year, 40 heifers arrived at the facility following a one week fence line weaning period at the Virginia Tech Shenandoah Valley Agriculture Research Extension Center, and placed in a single management group on grass. Details of the experimental design can be found in a previous report (2018 Nebraska Beef Cattle Report 74–76).

On each day of observation, heifers were moved calmly from a holding pen into a tub, through the alley way, and into the squeeze chute. Each heifer's head was caught and secured in a head gate and chute score (1 = docile, 6 = aggressive) was recorded by as many as 6 individuals, including the experienced observers whose assessments were analyzed in an adjoining article (2018 Nebraska Beef Cattle Report 74–76). Heart rate, temperature, and a fecal and blood sample were then taken. Upon release from the squeeze chute, an exit score (1 = docile, 5 = aggressive) was recorded by the same individuals. The heifers were evaluated repeatedly over 3 months, with some heifers scored on as many as 9 occasions.

Observers were split into experienced (E) and inexperienced (IN) groups, depending on their level of training, and their reliability was compared in two ways. First, 4 observers from the E group were selected within each year and their consistency (reliability) calculated between all pairs, all trios, or the 4 observers for chute and exit score.

There was a single individual who was present across all 3 years and who scored nearly every heifer in the study. This individual was considered the most experienced observer, and thereby the benchmark for comparison. All other observers, regardless of experience, were compared to this individual for reliability. Average reliabilities of each two way comparison were then reported separately by group (E or IN), depending on the experience level of the second person.

Statistical Analysis

Reliability of each subjective measurement was calculated using percent of agreement (PA) and intra-class correlation (ICC) functions in the R statistical package. Percent of agreement was calculated as:

$$\text{PA} = \frac{\text{Number of agreements}}{\text{Number of total observations}} \times 100$$

where PA = 0 meant no agreement and PA = 100 meant perfect agreement. The statistic was calculated with a tolerance of zero, where all observers had exactly the same score, or a tolerance of one, where all observers were within one score of each other.

Intra-class correlation was used as a second measure of reliability and described how strongly observations of the same

Table 1. Reliability of experienced observers for chute and exit score

		Percent of Agreement		Intra-class Correlation		
	N[1]	Tol = 0[2]	Tol = 1[3]	Value	Lower CI[4]	Upper CI[5]
Chute Score						
2	436	63.92	96.37	0.747	0.700	0.787
3	320	47.48	91.93	0.743	0.699	0.784
4	213	37.09	86.85	0.738	0.690	0.782
Exit Score						
2	440	82.98	99.58	0.894	0.872	0.911
3	327	74.33	99.27	0.895	0.875	0.913
4	223	68.16	99.55	0.898	0.877	0.917

[1]N = Total number of animals observed by all individuals
[2]Tolerance = 0 requires all observers to agree perfectly on a score
[3]Tolerance = 1 allows observers to disagree by one level on the scale
[4]Lower bound of the 95% confidence interval for ICC
[5]Upper bound of the 95% confidence interval for ICC

Table 2. Reliability of experienced and inexperienced observers for chute and exit score

		Percent of Agreement		Intra-class Correlation		
	N[1]	Tol = 0[2]	Tol = 1[3]	Value	Lower CI[4]	Upper CI[5]
Chute Score						
Experienced	294	63.29	95.84	0.732	0.670	0.784
Inexperienced	42	57.67	95.16	0.638	0.392	0.819
Exit Score						
Experienced	296	82.40	99.81	0.885	0.856	0.909
Inexperienced	42	82.36	99.88	0.894	0.780	0.937

[1]N = Total number of animals observed by all individuals
[2]Tolerance = 0 requires all observers to agree perfectly on a score
[3]Tolerance = 1 allows observers to disagree by one level on the scale
[4]Lower bound of the 95% confidence interval for ICC
[5]Upper bound of the 95% confidence interval for ICC

event resembled each other. An ICC of 0 represented no agreement among observers, while an ICC of 1 represented perfect agreement. Typically, an ICC of 0.70 or greater is considered to reflect strong concordance and thereby a reliable evaluation.

Results

Within the E group, consistency was summarized as the average reliabilities of groups of 2, 3 or all 4 observers, which are shown in Table 1. When the tolerance was set to zero, PA decreased as the number of observers increased for both chute and exit scores. When tolerance was set to one, allowing for slightly more subtlety among scores, the PA for both chute and exit scores were higher, as expected; still the PA decreased with an increased number of observers for chute score. The ICC for chute and exit scores were consistently around 0.74 and 0.89, respectively, which was higher than the threshold of 0.70 for reliable assessments. Furthermore, the lower bounds of the confidence interval for both ICC were at the least 0.69. Therefore, even though an increased number of observers reduced the PA, the experienced observers in this experiment were very consistent in their estimates of both chute and exit scores.

Both E and IN observers were then compared to the same individual who was present for all 3 years of the study, and considered the most experienced observer. The average reliabilities of these comparisons are given in Table 2. The PA, when tolerance was set to zero, and the ICC for chute score, was higher for the experienced (63.3 and 0.73, respectively) than inexperienced (57.6 and 0.64, respectively) observers, as expected. This was not seen in PA or ICC for exit score, with estimates being fairly similar between the E and IN groups (82.4 and 0.89, respectively). Furthermore, when the tolerance was set to one, the PA was similar between the E and IN groups for chute (95.0), and exit scores (99.8). However, confidence intervals of the IN group for both scores were wider than the E group, indicating greater variability in their scores. Thus, the amount of training or experience an observer has does impact the reliability of their assessments.

It is worth noting that under all circumstances, the reliability of exit score was higher than chute score. This may reflect the scoring systems themselves. The system for exit score is inherently less complicated than chute score, and evaluates fewer attributes of behavior. This allows exit scores to be easier to delineate than chute scores. That conclusion is supported by the evidence that when tolerance is set to one, chute and exit scores had similar PA. The change in PA from a tolerance of zero to a tolerance of one was also much larger for chute scores.

Allowing some differences (tolerance) among observers in their subjective evaluation of behavior is perhaps beneficial. A subjective scoring system, with a set number of categories, may not precisely identify all possible levels of temperament. Some cattle may not clearly fit a single score, at least in the mind of a given observer. With a tolerance of zero, regardless of the number of observers, all would necessarily have to assign an animal the same score. With a tolerance of one, and with multiple observers, greater subtlety in the evaluation may be captured.

This idea may be best illustrated by an example. Presume a threshold was set to cull cattle with a chute score of 3 or higher. If two observers assess an animal as a 3, and another as a 2, the average chute score would be a 2.7, below the threshold value. If only the first two observers' scores were allowed–effectively the situation with a tolerance of zero–this animal would have been culled from the herd. In the current study, as the number of E observers was increased, PA with tolerance of one decreased (96.4 to 86.9); however, that decrease was far less than for a PA with tolerance of zero (63.9 to 37.1). Furthermore, ICC were equal no matter how many E observers were considered. Allowing some tolerance for discrepancy in scores among trained evaluators may there-

fore allow for a more equitable assessment of temperament.

Conclusion

Chute and exit scores in cattle have been suggested as useful measurements of docility. Experienced observers were more consistent in their assessment of chute score than those who were inexperienced. However, regardless of level of experience, exit scores were consistently evaluated. Exit scores were always more reliably assessed than chute score; this may reflect the increased complexity of delineating among chute scores. With training, the reliability of chute scores became high, and approached that of exit scores. Since both scores can be assessed reliably, their use as measures of temperament could result in positive changes in docility in cattle.

Jamie T. Parham, graduate student

Amy E. Tanner, research technician, Animal and Poultry Science, Virginia Tech

Mark L. Wahlberg, associate professor emeritus, Animal and Poultry Science, Virginia Tech

William S. Swecker, Jr., professor, Virginia-Maryland College of Veterinary Medicine

Ronald M. Lewis, professor, Animal Science, Lincoln