Acclimation to handling has minimal effect on the repeatability of temperament measures

J. T. Parham[1], A. E. Tanner[2], M. L. Wahlberg[2], W. S. Swecker, Jr.[3], & R. M. Lewis[1]

[1]University of Nebraska - Lincoln, Animal Science, P.O. Box 830908, 68583, Lincoln, USA ron.lewis@unl.edu (Corresponding Author)
[2]Virginia Tech, Animal and Poultry Sciences, 3470 Litton Reaves Hall, 24061, Blacksburg, USA
[3]Virginia-Maryland College of Veterinary Medicine, 205 Duck Pond Drive, 24061, Blacksburg, USA

Summary

The key to the adoption of selecting for temperament in animals is the ability to do so in a way that is accurate, easy, and inexpensive. Chute score (CS), exit score (ES), and exit velocity (EV) have been proposed as methods to measure temperament of animals in a production setting. Before using these methods to make selection decisions, their practicality of use and efficiency must be considered. The objectives of this study were to determine whether behavioral scores or values change under repeated and routine management and, if so, assess how such change over time may affect their utility for use in selection programs. Over 3 consecutive years, a factorial design of 2 measurement protocols [frequent (F), infrequent (IN)], and 3 events, each 1 mo apart, was used. The F measurements were collected over 3 consecutive days; IN measurements were collected on d 1 within each event. Twenty predominantly Angus heifer calves, 2-wk post weaning, were randomly assigned to each protocol. Heifers were weighed, calmly moved into a squeeze chute, and their heads caught. A CS was assigned from 1 (docile) to 6 (aggressive) by up to 4 observers. Exit velocity over a distance of 2 m was obtained on release from the chute, and an ES given from 1 (docile) to 5 (aggressive) by the same observers. Data were analyzed with ANOVA using SAS. For all heifers, protocol, event, and their interaction, were compared on d 1. For heifers assigned to F, event, and day within event, were instead fitted. For both models, body weight was included as a covariate, and sire and year were fitted as random effects. Inter-observer reliabilities were calculated using Kendall's coefficients and intra-class correlation using R. Repeatabilities and heritabililties were calculated for a sire model in ASReml. Chute score decreased across events (0.61 ± 0.16; P = 0.04) and days (0.86 ± 0.20; P = 0.02) in F. Exit scores and EV changed less over time. Chute score therefore may be more indicative of acclimation to a novel environment than ES or EV. Inter-observer reliabilities were above 0.7, indicating consistent evaluation among observers, with higher values for ES than CS. Repeatability was high for all measures: 0.88 ± 0.20 for ES, 0.68 ± 0.20 for EV, and 0.62 ± 0.24 for CS. While all methods appear to offer an accurate and repeatable way to quantify and select for temperament in cattle, CS and ES may be preferred due to ease of use. Heifers became calmer with repeated gentle handling. Producers therefore may benefit from allowing cattle a few days to acclimate to new working facilities before assessing docility. This may avoid culling an animal based strictly on its initial response to a novel stimuli.

Keywords: acclimation, beef cattle, reliability, temperament

Introduction

Strong behavioral responses of cattle towards humans or any other stressor have been associated with increased risk of handler injury (Fordyce *et al.,* 1985), poorer weight gain and meat eating quality (Bates *et al*., 2014), decreased tolerance to disease (Cooke, 2014), decreased reproductive performance (Cooke, 2014), and increased production costs (Burrow & Corbet, 2000). Because of the negative outcomes associated with excitable temperament in cattle, and its moderate heritability (Burrow & Corbet, 2000), there has been increased selection for docility.

Measures of temperament in animals should be accurate, easy, and inexpensive to collect. However, before their implementation two key issues must be addressed. First, reliability of subjective scoring methods should be investigated to ensure accurate delineation of scores and thereby more precise selection decisions. Higher reliabilities signify more consistent observations and are more desirable (Bokkers *et al*., 2012), with threshold correlation coefficients for acceptable reliability above 0.7 (Martin & Bateson, 1993). Olmos & Turner (2008) reported moderate inter- and intra-observer reliabilities when assessing chute score in beef cattle, with a Kendall coefficient of 0.64, and ranges in kappa correlation coefficients from 0.49 to 0.65. Reliabilities of a qualitative behavior assessment reported by Bokkers *et al*. (2012) were quite variable even for experienced observers, indicating inconsistency in their observations of dairy cattle.

Second, cattle may acclimate to repeated handling. Curley *et al*. (2006) found that an increase in the number of times an animal goes through a working facility resulted in a decrease in temperament score. However, if the amount of that change is inconsistent among individual animals, repeatability would be lowered. Furthermore, the first observation of an animal may not be indicative of their later temperament.

If an animal's initial response is not indicative of future responses, or is unreliably evaluated at a single time point, this could negatively impact selection decisions and decrease genetic gain. Finding a method to quantify temperament that is consistent and inexpensive to implement would benefit the cattle industry. The objectives of this study were to determine whether behavioral scores or values change under repeated and routine management and, if so, assess how such change over time may affect their utility for use in selection programs.

Materials and Methods

Within each of 3 consecutive years, 40 predominantly Angus spring-born heifer calves were reared at the Virginia Tech Shenandoah Valley Agricultural Research and Extension Center with their respective dams until weaning. Upon completion of a 1 wk fence line weaning period, these calves were transported to Virginia Tech Kentland farm and placed in a single management group on grass.

Heifers were randomly assigned to one of two measurement protocols (frequent or infrequent) in groups balanced for sire. Within each year, data were collected across 3 recording periods, each 1 mo apart [event 1 (Oct.), 2 (Nov.), 3 (Dec.)]. Heifers within the frequent (F) measurement protocol were observed 3 consecutive days within each event while the heifers in the infrequent (IN) measurement protocol were observed only the first day. Day within event is designated by $d_{i,j}$, where *i* is the event and *j* is the day within an event.

On d 1 of each event, all 40 heifers were moved into a holding pen and herded to the chute system in groups of four. One at a time, each heifer's head was secured in the head gate and the side bar's left open. On a given day, up to four experienced observers simultaneously recorded a chute score (CS, 1 - Docile to 6 - Aggressive; Tulloh, 1961) within the first 15 s of restraint in the squeeze chute (Priefert Model S04). Upon release, exit score (ES, 1 - Docile to 5 - Aggressive; BIF Guidelines, 2002) was recorded by the same individuals, and exit velocity (EV, s) measured using electronic times (Polaris, FarkTek, Wylie, TX) over a 2 m distance, beginning 1 m from the head of the chute. The results from multiple observers recording heifer's simultaneously were used to establish inter-observer reliability. On d 2 and 3 of each event, the measurements were again taken on F heifers (n = 20). After d 3 of recording all 40 heifers were mixed into a single management group until the next recording period.

Statistical Analysis

Inter-Observer Reliability

Calculation of inter-observer reliability of CS and ES was based on the average assessment of 2, 3 or all 4 experienced observers who evaluated the same heifer(s) on all 9 days of the study. All reliability calculations were conducted using the irr package (Gamer *et al*., 2012) in R (R Core Team, 2013). Reliability of each subjective measurement was calculated using Kendall's coefficient of concordance (KCC) and intra-class correlation (ICC).

Acclimation to Handling

To compare the measurement protocols (F vs IN), data were analyzed using the GLIMMIX procedure in SAS (SAS Inst. Inc., Cary, NC) with two separate models. To compare the effect of measurement protocol on temperament, a 2x3 factorial model was used to estimate the effect of measurement protocol, event (1-3), and their two-way interaction. Comparisons were only made using data on the first day of each event ($d_{1,1}$, $d_{2,1}$, $d_{3,1}$). Second, to investigate changes in the F group over time, a nested model was used to estimate the effect of event (1-3), and day (1-3) within event. Year and sire were treated as random effects, and body weight as a covariate. Least squares means and standard errors (SE) were obtained using SAS with Tukey's adjustment.

Repeatability and Heritability

Although the data were small, repeatability and heritability estimations were obtained using ASReml (Gilmour *et al*., 2009) including all progeny in the dataset from one of 9 AI sires (n = 105), ranging from 3 to 23 progeny per sire. Three of the 9 AI sires were used in multiple years. A sire model was used treating event, measurement protocol, and their interaction as fixed effects with year and sire as a random effects, and body weight again as a covariate.

Correlations

Pearson (phenotypic) correlations were calculated between CS, ES, and EV collected on the first day of all events using R. Correlations were first calculated for F and IN separately with similar results. Therefore, both protocols were combined.

Results and Discussion

Intra-observer reliabilities for CS and ES are given in Table 1. Regardless of any change in scores over time, or increasing number of individuals, both scoring methods had KCC and ICC higher than 0.7, indicating good consistency among observers when scoring the same animal. However, ES had higher values than CS. This may reflect the scoring systems themselves. The system for ES is inherently less complicated than CS, and evaluates fewer attributes of behavior. The ethogram for CS evaluates multiple behaviors at once, including the degree of movement of both the body and head, vocalization, tail flicking, and breathing pattern. Conversely, the categories of ES only differ by a single adjective. This allows ES to be easier to delineate than CS. Overall, both CS and ES were reliably evaluated subjectively by experienced observers.

Event, measurement protocol, or their interaction were not different for CS, ES, or EV ($P > 0.19$). Chute score had the largest decrease across events, followed by EV and ES, as shown in Table 2. However, there was a larger numerical decrease between events within the F group (2.35 ± 0.18 to 1.80 ± 0.18 from $d_{1,1}$ to $d_{3,1}$) compared to the IN group (2.31 ± 0.18 to 2.07 ± 0.18 from $d_{1,1}$ to $d_{3,1}$) for CS. This did not occur with ES or EV.

When considering F only, there was a decrease in CS across both events and days ($P < 0.05$). Change in CS across events is shown in Figure 1a. Chute score decreased ($P = 0.06$) from event 1 to 2, with little change from event 2 to 3. Figure 1b shows the change in CS across days, with a numerical increase between $d_{1,1}$ and $d_{1,2}$. Following $d_{1,2}$, CS decreased over time, finishing on $d_{3,3}$ at a score of 1.47 ± 0.16, which was different from both $d_{1,1}$ and $d_{1,2}$ ($P < 0.05$). Such a decrease was not observed for EV and ES ($P > 0.13$), as shown in Figure 1c and 1d. Chute score therefore may be more indicative of acclimation to a novel environment than ES or EV. Allowing heifers to acclimate to a novel environment therefore may be worthwhile when evaluating their behavior in the chute to avoid premature culling decisions.

Estimated repeatability and heritability of each trait is given in table 3. All traits were considered highly repeatable. In comparison, CS and EV were moderately heritable, with lower estimates for ES. Even though the SE are large, which is expected given the amount of data available, these heritabilities of CS and ES were consistent in magnitude with literature. However, the heritability of EV was lower than reported previously (Burrow & Corbet, 2000). Since these subjective methods to evaluate behavior were repeatable, there was no evidence against their use as a suitable alternative to objective measures such as EV.

Pearson correlations of CS with ES and EV were 0.33 ± 0.05 and 0.25 ± 0.05 ($P < 0.05$), respectively. It appears the temperament of an animal while restrained in a head gate only moderately corresponds with that when released. Conversely, the correlation of ES with EV was 0.75 ± 0.04 ($P < 0.05$) indicating they capture effectively the same behavior.

Although acclimation to handling in the chute was evident, repeatability of CS was not much lower than ES. In figure 1b, most of the change in CS occurred in the first event. Following $d_{1,3}$, CS appears to remain relatively constant from $d_{2,1}$ to $d_{3,3}$, resulting in less change over time. Therefore, acclimation did not dampen the repeatability of this measure. Still, if selection decisions are based on CS the first time animals are evaluated, some may be culled unnecessarily.

In conclusion, temperament in cattle is becoming a common selection criterion due to its associations with growth, carcass quality, and well-being. Selection for docility on site requires a measurement that is accurate, inexpensive and relatively easy to record. Results from this experiment support the use of CS and ES as subjective methods to quantify temperament in animals. The use of EV is objective and is not subject to observer error, but it does require the purchase of equipment. Exit velocity and ES behave similarly and have a strong correlation. Since ES had such high reliability and repeatability, it may be the more pragmatic measure of temperament for an animal exiting the chute. Since both scores can be assessed reliably, their use as measures of temperament could result in positive changes in docility in cattle.

Importantly, heifers appear to acclimate to handling in a calm environment. Cattle are more excitable during their first handling experience. After just a few days, CS decreased substantially and remained relatively constant thereafter. Although temperament measures were highly repeatable, allowing animals to acclimate to handling may help clarify their longer term behavior within a production environment. This may avoid culling an animal based strictly on its initial response to a novel stimuli.

Table 1. Average inter-observer reliabilities of chute and exit score by number of observers.

			ICC[3]		
Observers	*n*[1]	KCC[2]	Value	Lower[4]	Upper[5]
Chute Score					
2	436	0.87	0.75	0.70	0.79
3	320	0.82	0.74	0.70	0.78
4	213	0.79	0.74	0.69	0.78
Exit Score					
2	440	0.95	0.89	0.87	0.91
3	327	0.93	0.90	0.88	0.91
4	223	0.93	0.90	0.88	0.92

[1]The total number of observations utilized
[2]Kendall's Coefficient of Concordance (KCC)
[3]Intra-class Correlation Coefficient (ICC)
[4] Lower bound of the 95% CI for ICC
[5] Upper bound of the 95% CI for ICC

Table 2. Change in scores or values across recording events for both F and IN heifers together.

	Day Within Event[1]			SEM
Measure	$d_{1,1}$	$d_{2,1}$	$d_{3,1}$	
Chute Score	2.33	1.95	1.94	0.15
Exit Score	1.95	1.97	1.89	0.16
Exit Velocity (m/s)	1.87	1.80	1.77	0.13

[1] Day within event is designated by $d_{i,j}$, where *i* is the event and *j* is the day within an event.

Table 3. Repeatability and heritability of each temperament measure.

Measure	Repeatability	Heritability
Chute Score	0.62 ± 0.24	0.29 ± 0.26
Exit Score	0.88 ± 0.20	0.17 ± 0.25
Exit Velocity (m/s)	0.68 ± 0.21	0.24 ± 0.24

(a) Mean chute score by event

(b) Mean chute score by day

(c) Mean exit score and exit velocity by event

(d) Mean exit score and exit velocity by day

Figure 1. Mean values of frequently handled cattle, separated by event and day. [[a,b] Values with different letters are significantly different ($P < 0.05$); *Difference in values only tended towards significance ($P < 0.10$).]

List of References

Bates, K. E., R. L. Weaber, J. M. Bormann, D. W. Moser, J. L. Salak-Johnson, C. C. L. Chase, R. K. Peel, H. Van Campen, G. H. Loneragan, J. J. Wagner, P. Bodhireddy, K. Prayaga, & R. M. Enns. 2014. Temperament can be an indicator of feedlot performance and carcass merit in beef cattle. Cattleman's Day, Manhattan, KS. p. 34-38.

BIF Guidelines. 2002. Guidelines for uniform beef improvement programs, 8th ed. Univ. Georgia, Athens, GA.

Bokkers, E. A. M., M. de Vries, I. C. M. A. Antonissen, & I. J. M. de Boer. 2012. Inter- and intra-observer reliability of experienced and inexperienced observers for the Qualitative Behaviour Assessment in dairy cattle. Animal Welfare 21:307-318.

Burrow, H. M., & N. J. Corbet. 2000. Genetic and environmental factors affecting temperament of zebu and zebu-derived beef cattle grazed at pasture in the tropics. Aust. J. Agric. Res. 51:155-162.

Cooke, R. F. 2014. Bill E, Kunkle interdisciplinary beef symposium: Temperament and acclimation to human handling influenced growth, health, and reproductive responses in cattle. J. Anim. Sci. 92:5325-5333.

Curley Jr., K. O., J. C. Paschal, T. H. Welsh Jr., & R. D. Randel. 2006. Technical note: Exit velocity as a measure of cattle temperament is repeatable and associated with serum concentration of cortisol in Brahman bulls. J. Anim. Sci. 84:3100-3103.

Fordyce, G., M. E. Goddard, R. Tyler, G. Williams, & M. A. Toleman. 1985. Temperament and bruising of Bos indicus cross cattle. Aust. J. Exp. Agric. 25:283-288.

Gamer M., J. Lemon, & I. Fellows Puspendra Singh. (2012). irr: Various Coefficients of Interrater Reliability and Agreement. R package version 0.84. <https://CRAN.R-project.org/package=irr>.

Gilmour, A. R., B. J. Gogel, B. R. Cullis, R. Thompson, & D. Butler. 2009. ASReml user guide release 3.0. VSN International Ltd, Hemel Hempstead, UK.

Martin, P., & P. Bateson. 1993. Measuring behaviour, an introductory guide. Cambridge University Press: Cambridge, UK.

Olmos, G., & S. P. Turner. 2008. The relationships between temperament during routine handling tasks, weight gain and facial hair whorl position in frequently handled beef cattle. Appl Anim. Behav. Sci. 115:25-36.

R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Tulloh, N. M. 1961. Behavior or cattle in yards. II. A study of temperament. Anim. Behav. 9:25-30.