University of Nebraska - Lincoln
DigitalCommons@University of Nebraska - Lincoln

Theses and Dissertations in Animal Science | Animal Science Department

Summer 7-27-2018

Subjective Measures of Temperament in Beef Heifers Are Reliable Indicators of Physiological Stress and Indicate Acclimation to Repeated Handling

Jamie Parham
University of Nebraska - Lincoln, jtparham@huskers.unl.edu

Follow this and additional works at: http://digitalcommons.unl.edu/animalscidiss

Part of the Agriculture Commons, Animal Sciences Commons, and the Animal Studies Commons

Parham, Jamie, "Subjective Measures of Temperament in Beef Heifers Are Reliable Indicators of Physiological Stress and Indicate Acclimation to Repeated Handling" (2018). *Theses and Dissertations in Animal Science*. 167.
http://digitalcommons.unl.edu/animalscidiss/167

This Article is brought to you for free and open access by the Animal Science Department at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Theses and Dissertations in Animal Science by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

SUBJECTIVE MEASURES OF TEMPERAMENT IN BEEF HEIFERS ARE RELIABLE INDICATORS OF PHYSIOLOGICAL STRESS AND INDICATE ACCLIMATION TO REPEATED HANDLING

by

Jamie T. Parham

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Animal Science

(Animal Breeding and Genetics)

Under the Supervision of Professor Ronald M. Lewis

Lincoln, Nebraska

July, 2018

SUBJECTIVE MEASURES OF TEMPERAMENT IN BEEF HEIFERS ARE RELIABLE INDICATORS OF PHYSIOLOGICAL STRESS AND INDICATE ACCLIMATION TO REPEATED HANDLING

Jamie T. Parham, Ph.D.

University of Nebraska, 2018

Advisor: Ronald M. Lewis

Associations between excitable temperament and many economically relevant traits have been established. In being heritable, temperament can be augmented through selection. Current methods to evaluate temperament in a production setting include numerous subjective and objective measurements, which some producers may find cumbersome to navigate. Those who utilize these methods may not do so efficiently if selection criteria are not indicative of an animal's response to stress, or initial evaluations are not strong indicators of future temperament. The objectives of this research were to develop a procedure for evaluation of calf behavior, indicative of physiological stress, and then determine whether stress will change under repeated and routine management as evaluated through behavioral and physiological measures. Each of three consecutive years, 20 commercial *Bos taurus* heifers, 2-wk post weaning, were randomly assigned to a factorial design of two measurement protocols [frequent (F), infrequent (IN)], and three recording periods, each 1 mo apart. The F measurements were collected over three consecutive days, and IN only on d 1, of a recording period. Heifers were calmly moved into a squeeze chute and their heads caught. Individuals assigned a chute score (CS) based on their reaction to 15 s of restraint. A fecal sample, heart rate, rectal temperature, and jugular blood sample were taken. Upon release, exit velocity (EV) over a 2 m

distance was recorded, and an exit score (ES) assigned by the same individuals. Each heifers' response to 30 s of exposure to a human stressor was then recorded both in an individual and group pen setting. An individual (IPS) and group (GPS) pen score was assigned. Scores ranged from 1 to 5 or 6, with increasing values indicative of more excitable (worsening) temperament. All subjective measures were reliably assessed. Moderate correlations existed between objective and subjective measurements of temperament, indicating they were representative of physiological stress. Furthermore, CS and IPS decreased considerably over time, especially in F heifers, as they acclimated to novel handling experiences. Producers may avoid unnecessarily culling cattle based strictly on initial response to novel stimuli by allowing acclimation to handling before assessing docility.

ACKNOWLEDGEMENTS

I have been blessed with an overabundance of individuals who have had a large impact on my life these past three years, all of which have had their own distinct influence on my education and my life.

First, I would like to thank the faculty and staff at the University of Nebraska for their instruction and assistance both inside and outside of the classroom. There was never a question that was not answered with a simple knock on a door. My education was bettered by the availability and passion that you provided. I would also like to thank my fellow graduate students for the friendships and partnerships that were gained both personally and professionally.

Next, I would like to thank the scientists, graduate students, and undergraduates who assisted in the design and collection of these data. This includes, but is not limited to, Dr. Mark Wahlberg, Dr. Amy Tanner, Dr. William Swecker, Jr., Dr. Katharine Barkley, and Lyla Pullen. I know it has been a few years since you spent early mornings, and sometimes days, with these heifers, but I hope finally seeing the results of your hard work makes them meaningful. I would also like to thank you for your input and recommendations on data analyses and manuscripts as they provided valuable opinions and conclusions. While not a part of the initial project, I would also like to thank Jessica Schmidt for her countless hours of analyzing video's of these heifers. I hope your eyesight is no worse for the wear. I enjoyed working with you both in regards to this project, as well as your own. I hope you found the experience as educational as I did.

I would especially like to thank the members of my committee. To Dr. Spangler, who provided guidance and knowledge from his years as an extension specialist, and for answering many questions about career opportunities. To Dr. Eskridge for his willingness to share years of knowledge about statistics and experimental design. I truly enjoyed my time spent in the classroom learning from you. I would especially like to thank Dr. Grandin for her willingness and ability to make herself available given a very full schedule. The phone calls and conversations pushed me to think outside of the box. You have the knack to ask very meaningful questions about the impact this research has on production, and I am grateful that you were able to provide that to this project. Last but not least, I would like to express my sincerest gratitude to Dr. Ron Lewis. It is interesting how life happens. While a master's degree under your supervision was not necessarily in the cards, you got stuck with me after all. There are no words to express how thankful I am for being given this opportunity and your continued support through this process. Thank you for these past years of guidance and the life lessons you have shown me, I will take them with me in the future.

Finally, I would not be here if not for the support and understanding of my friends and family, both new and old. You have been a constant support system not only for these last three years, but my entire life. My accomplishments go hand in hand with your continued support and encouragement. And lastly, to my fiancé Donny, there are not enough words to express my gratitude for your belief and support since day one. You are a constant reminder that anything is possible. I cannot wait to see where this next chapter takes us.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

Introduction to Temperament

Temperament is often described as an animal's behavior response to handling by humans, or to any potentially fear-eliciting situation (Burrow and Corbet, 2000; Olmos and Turner, 2008). The response of cattle to handling depends not only on their reaction to humans, but also on elements such as social context, physical environment and novelty of the situation (Gringard et al., 2000). Since these reactions are believed to be a response to fear, they are often linked with stress. Kadel et al. (2006) referred to temperament as the overall fearfulness and susceptibility of an animal to stress. In general, stressors may be grouped into two categories, the first being psychological stress such as restraint, handling or novelty, and the second being physical stress such as hunger/thirst, fatigue, injury or thermal extremes (Grandin, 1997).

Compared to other aspects of animal husbandry, the topic of behavior, more specifically temperament, is a recent area of research. However, temperament and stress are thought to affect many aspects of an animal and their quality of life. 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 (Burrow, 1997; Voisinet et al., 1997a), decreased tolerance to disease (Sheridan et al., 1994), and increased production costs (Burrow and Corbet, 2000).

Because of the negative consequences associated with excitable temperament in cattle, further research is necessary to better understand behavior. If researchers can discern the many aspects that effect cattle temperament, that knowledge can be used to decrease the amount of external stress on an animal as well as formulate ways in which temperament can be selected for. However, before temperament can be included in formal breeding programs, understanding of identifiable non-genetic sources of variation need to be established (Burrow and Corbet, 2000). The objective of this chapter is to first describe the impact of temperament on production, laying out the importance of being able to quantify and select for behavior. The second objective is to discuss methods researchers currently use to measure temperament. A discussion of results from previous research will be included, specifically highlighting relationships that are believed to exist among these methods. Finally, specific areas of research that will be addressed in this dissertation will be introduced, including investigating the reliability of subjective evaluations of temperament, understanding change in temperament over time when measured both subjectively and objectively, and lastly, understanding specific behaviors as they pertain to an animal's herd mentality. By doing so, the aim is to substantiate the objectives and importance of the research to be conducted herein.

Impact of Temperament on Production

As with any industry, economics weigh heavily on decision making processes. There are many aspects of the beef industry that can be impacted by temperament. These include the overall safety of animals and handlers, which pertains to both the feedlot and cow-calf industry, and the expense associated with injury, pregnancy rates, and losses in carcass value due to reduced meat quality.

Equipment and Safety

The temperament of an animal when approached by a novel stimulus, especially a handler, can result in injury to both the animal and the human. In a working environment, a fearful response to handling can cause animals to struggle, show agitated movements, and attempt to escape, causing these animals to have a higher risk of injury to themselves, human handlers, and other animals (Haskell et al., 2014). Also, such animals are more likely to make the process of transporting or handling a group of cattle slower and less efficient. A cross sectional survey conducted in Scotland reported that of 2,439 returned questionnaires, 1,341 handler injuries were reported by 591 (24%) respondents, all within the previous 12 months (Lindsay et al., 2004). Of these total injuries, 297 (12%) pertained to tagging of calves, 418 (17%) pertained to clipping, and 568 (23%) pertained to other cattle-handling activities. Of those reported in the survey, 82% of tagging injuries were bruises, which ranged from minor to serious internal bruising with the lower limbs being the most affected area, 3% were lacerations, and 3% were fractures. This can be compared to injuries while clipping, where 78.3% were bruising occurring mainly on the upper limbs, 6% were lacerations, and 7% were fractures. However, Lindsay et al. (2004) commented that many non-fatal injuries go unreported. While clipping of cattle arose due to concerns about contamination from dirt or feces upon arrival at an abattoir, ear tagging in calves normally occurs within the first few days of life and requires a handler to come in close contact with the cow and calf (Lindsay et al., 2004). This can create a reaction in some cattle known as maternal defensive aggression.

Maternal Defensive Aggression

Any defensive aggressiveness of a cow towards a human or other animal attempting to interfere with her calf is referred to as maternal defensive aggressiveness (Haskell et al., 2014). It can be viewed as one of the greatest threats to handler safety, as the industry has numerous opportunities for routine contact between handlers and lactating cows (Turner and Lawrence, 2007). Analysis of the annual fatal accident statistics published by the UK Health and Safety Executive (HSE) over an 11 yr period shows that 14 people, farm workers and members of the public, received fatal injuries from a lactating cow (Turner and Lawrence, 2007).

Maternal aggressiveness is a human safety as well as a calf survival issue (Haskell et al., 2014). Turner et al. (2013) found a weak indication that cows that are fearful when pregnant may produce calves with a decreased birth weight and ADG, specifically over the first 7 mo of life, compared to calmer cattle. However, maternal aggressiveness can be beneficial where calf predation is a significant risk.

Pregnancy Rates

The main objective of any cow-calf operation is to produce one calf per cow annually, making reproductive performance the main driver of profitability in cow-calf operations (Cooke et al., 2009). Excitable temperament in cattle can impair reproductive performance by stimulating neuroendocrine stress responses that disrupt the normal physiological processes associated with fertility (Curley et al., 2008). Using fixed time artificial insemination (FTAI), Cooke et al. (2011) investigated the effect of temperament score on pregnancy rates in *Bos indicus* cattle. The probability of cows becoming pregnant to FTAI was negatively associated with temperament score, with pregnancy rates in excitable cattle being lower than cattle believed to be more docile. Similar results were also reported in Nellore (Rueda et al., 2015) and *Bos taurus* (Cooke et al., 2012) cattle.

Along with reproductive performance in cattle, Cooke et al. (2012) investigated the effect of acclimation to handling on pregnancy rates in beef cattle. Acclimated heifers were processed 3 times a wk for 4 wk, and their plasma concentrations of cortisol and haptoglobin levels were compared to undisturbed heifers left to pasture. It was concluded that human handling after weaning decreased circulating concentrations of stress-related hormones and metabolites, and thereby hastened reproductive development of replacement heifers.

Meat Quality

Research in beef cattle temperament has received interest in recent years due to the connection between excitability, physiology, carcass composition, and reduction in meat tenderness (Hall et al., 2011). Comparing measurements of temperament to growth traits is common. Sant'Anna et al. (2012) found low negative estimates of genetic correlation between flight score and WW (-0.12 ± 0.07), ADG (-0.13 ± 0.08), and SC (-0.11 ± 0.07) in *Bos indicus* cattle. Similar results were reported by Reinhardt et al. (2009) in *Bos taurus* cattle. Excitable cattle had lower body weight, ADG, HCW, yield grade, quality grade, marbling score and higher mortality than docile cattle. Furthermore, as exit velocity times decreased Warner-Bratzler Shear Force values of tenderness increased (r = -0.187; Hall et al., 2011). This can be explained by increased fiber diameter and shorter sarcomere length in muscle tissue of those excitable animals, causing tougher meat (King et al., 2006). Temperament of cattle can also be linked to disease susceptibility, which can decrease ADG and affect meat quality (Reinhardt et al., 2009).

This relationship between temperament and meat quality becomes one of concern if payment based on meat eating quality instead of weight becomes more widespread (Haskell et al., 2014). Iowa State University completed an Animal Industry Report where they investigated the effects of disposition during the feedlot period on gain and carcass quality (Busby et al., 2006). Aggressive calves had 0.3 lb/d reduction in ADG compared to docile calves and their mortality rate almost doubled. Aggressive calves also had lower percentages of prime and choice carcasses compared to docile calves, resulting in lower acceptance rates for Certified Angus Beef premiums. Overall, when considering the effects of disposition on quality and yield grade, feedlot gain, death loss, and treatment costs, docile calves returned $62.19/head more than aggressive calves.

To put this into perspective, the USDA reported a total of 21.9 million cattle marketed in the US in 2017 (U.S. Department of Agriculture, 2018a, 2018b). In the study conducted by Busby et al. (2006), 758 out of 13,315 (5.7%) of cattle received an "aggressive" score. If this were to carry over to the beef industry as a whole, there would be a total of 1.25 million aggressive cattle on feed, for a total profit loss of over $77.7 million USD annually, just within the feedlot sector. However, these estimates are over a decade old and are likely to have increased.

Measurements of Temperament

The behavioral response of beef cattle to human handling and novel situations has been chosen as an adequate indicator of their temperament (Burrow, 1997; Grandin, 1997). Many methods have been introduced with which to do this, and can be delineated into restrained and non-restrained tests (Burrow, 1997). For scientific research, it is critical that each measurement of temperament in an animal is correlated to, or indicative of, their behavior during normal handling practices (Fordyce et al., 1982). Although cattle can be classified using these methods on a continuum, they can also be used to categorize more extreme animals as docile or excitable. However, measurement of behavior can be time consuming, creating interest in the development of quantifiable indicator traits to use in selection that are correlated with traits of interest (Olmos and Turner, 2008).

Restrained Tests

The most common type of restrained test is a crush test, where animals are given an individual chute score. This method, based on a six-point scale (1-docile; 6-aggressive), was introduced by Tulloh (1961). These scores are based on how an animal behaves during a specified duration in a crush or chute. These levels are quantified based on the amount of movement, vocalization, restlessness, kicking, etc., while restrained. However, once hydraulic chutes became more widely used, Grandin et al. (1995) proposed a different method, based on a four-point scale (1-calm; 4-violently shaking the squeeze chute), to account for the limited movement. Due to the hydraulics, the squeeze chute grips the animal tighter, making it difficult to differentiate between scores introduced for manual squeeze chutes. Most of today's literature will cite one of these two scales for crush tests for collecting chute scores.

A second restrained test is the balking score. This is discussed by Grandin (1993) and relates to the willingness of the animal to voluntarily enter the scale or squeeze chute without any encouragement from the handler. It is a simple measurement, giving a score of either non-balker or balker. In cases where there is a separate scale and chute area, an animal has to enter both without balking to be deemed a non-balker. It can be hypothesized that animals with docile temperament will not be as weary of such a situation and are therefore more likely to receive a non-balking score.

Non-restrained Tests

The behavior of cattle while restrained is not necessarily indicative of their behavior when in an open environment. Therefore, the use of non-restrained tests, such as pen score, are used to measure varying aspects of an animal's behavior when they have more freedom to move. Compared to restrained tests, such as chute score, there is not a widely-accepted scale used across experiments, nor is it applied commercially in industry. Instead, it varies by the objective of the experiment.

Pen scores can be assessed in two ways, individually or as a small group. To accomplish this, an animal (or group of animals) is placed into a pen, and behavioral responses to the presence of a human inside and outside of their flight zone is recorded. Hammond et al. (1996) and King et al. (2006) accomplished this using a five-point scale (1-docile; 5-very aggressive). The flight zone is the minimum distance within which a person can approach an animal before it reacts or moves away (Grandin and Deesing, 2014). It is well known that an animal's flight zone can be used to move a single animal or large group of animals in a desired direction.

Since cattle are herd animals, secluding a single animal in a pen can be a cause of stress. Individual pen scores then allow the observer to quantify the temperament of the animal when placed in a secluded situation. In comparison, several studies demonstrate the influence of group social environment on the behavioral response of farm animals (Gringard et al., 2000). Boissy and Le Neindre (1990) saw that stress response in cattle when placed in an individual pen can be reduced by the presence of peers. To further evaluate this hypothesis, a group pen score can be used to judge the reaction of one animal in the presence of peers, which can be compared to that animal's individual pen score to determine any difference in temperament.

The final type of non-restrained test can be measured in two parts. The first is a flight score. This is a subjective score given to an animal based on its behavior upon exiting the chute. Again, there is no widely adopted version of this test, but the score is often assigned based on a four-point scale (1-walk; 4-jump) (Lanier and Grandin, 2002). In conjunction with a subjective score, an exit velocity or flight speed can also be recorded. This technology was first introduced by Burrow et al. (1988) to electronically record the time taken for animals to cover a predetermined distance upon exiting the chute, or other confined area. The predetermined distance of 1.7 meters was selected as it is slightly shorter than the chute. To measure velocity, two infrared sensors are set to this standardized distance and the time it takes the animal to travel the distance, breaking both sensors, is recorded. Exit velocity is believed to be a more objective measurement of temperament in comparison to exit score or any other categorical variable.

Other Temperament Measures

Indicators of stress and discomfort are not only behavioral as they can be physiological as well (Grandin, 1997). Cortisol concentration, heart and respiratory rates are frequently used as physiological measures of stress (Andrade et al., 2001). Other measures believed to be representative of stress include body temperature and blood concentrations of glucose, non-esterified fatty acids, urea, and creatine kinase.

Two Systems. The sympathetic nervous system, also referred to as the sympathomedullary system (SMS), responds to stress by secreting catecholamines such as epinephrine and norepinpehrine into the circulatory system, resulting in increased blood pressure, heart rate, and respiration rate in animals (Burdick et al., 2011). This increase in epinephrine results in decreased eating and sleeping and is often activated alongside, or prior to, the hypothalmic-pituitary-adrenal (HPA) axis. Activation of the HPA axis results in increased secretion of glucocorticoids like cortisol. Together, these catecholamines and glucocorticoids increase cardiovascular output and catabolic effects such as metabolism of glycogen, protein, and triglycerides to provide energy, and also inhibit many body systems including reproduction and immunity.

Heart/Respiratory Rate. Elevated heart rates have been associated with pain sensations greater than routine handling, such as branding, and are highly correlated with cortisol levels (Lay et al., 1992). Respiration rates are obtained by observing the flank or rib cage movements of an animal while in a chute, often times expressed as a count over a fixed time interval (Hammond et al., 1996). Andrade et al. (2001) found that cattle that were accustomed to handling had reduced respiratory and heart rates while restrained in a chute.

Body Temperature. As mentioned, the release of catecholamines such as epinephrine and norepinephrine result in a cascade of effects including greater core body temperature (Gruber et al., 2010). Rectal temperature is also routinely monitored upon arrival at the feedlot as a measurement of well-being and possible infection (Sporer et al., 2008). Temperament has an impact on immune response in cattle. Therefore, those animals that are easily stressed have an increased risk of an inflammatory response. While Sporer et al. (2008) found no difference in rectal temperature 24 hr prior and 9.75 hr after the initiation of transport, there was a significant decrease in temperature 48 hr after the initiation of transport.

Cortisol. Increased cortisol levels have long been associated with behaviorally agitated cattle, making them a useful indicator of stress (Stahringer et al., 1989; Grandin 1997). Lay et al. (1992) concluded that dairy cattle exposed to hot iron branding had higher levels of plasma cortisol compared to those freeze branded or to a control. This conclusion came from the comparison of plasma cortisol levels taken periodically (-2.5 min to 15.5 min) after the application of branding treatment. However, since plasma cortisol is released in the blood stream, they found it a time-dependent measure that takes 10 to 20 min to reach peak values and may therefore not be as immediate a measurement as originally believed.

In a summary of short-term stress measurements, Grandin (1997) proposed different categories for cortisol levels: 1) baseline, 2) levels that occur during restraint, and 3) extreme stress. While she warned against absolute comparisons of levels across studies, a mean value of >70 ng/mL in both steers and cows possibly indicate rough handling or poor equipment.

Although plasma cortisol levels represent short-term stress on an animal, the procedure of sampling blood from animals can be stressful by itself, which could artificially raise levels (Keay et al., 2006). Conversely, cortisol can also be found in the fecal matter of animals. These fecal samples can be collected easily, without stressing animals (Möstl and Palme, 2002). However, it must be remembered that it takes time for cortisol to circulate through the digestive system. The concentration of cortisol metabolites in a fecal sample reflect cortisol production according to species-specific time intervals, which for ruminants is 10 to 12 hr. Fecal cortisol levels therefore represent pooled quantities of cortisol over an extended period of time, and can be used to measure chronic stress in animals.

Glucose. Catecholamines also stimulate glycogenolysis, which causes the mobilization of hepatic and muscle glycogen stores, resulting in elevated blood glucose concentrations (Knowles et al., 2014). Glycogenolysis is an immediate or 'emergency' energy supply for an animal, and is the principal regulatory mechanism that allows the animal to meet its physical or emotional challenges (Hemsworth and Barnett, 2000). A longer term response, called the 'stress response' invokes gluconeogenesis, or the provision of glucose from food or muscle protein, which again results in increased glucose concentrations in the blood.

Increased blood concentrations of glucose are therefore believed to be associated with stress during handling and transport. Petherick et al. (2009a) found elevated glucose levels associated with a fear of human test, where an animal was placed in a railed arena, with a stimulus person seated outside. Furthermore, glucose levels were slightly to moderately correlated with flight score in the same animals. Gruber et al. (2010) found high levels of glucose in cattle post-transportation, but were not able to make a comparison to levels before transport due to lack of blood sampling pre-transport. These levels also were not found to be correlated with measures of behavior post-transportation. Similarly, Van de Water et al. (2003) did not find any change in glucose concentrations following 15 separate transports of cattle.

Non-esterified Fatty Acids (NEFA). With increased glucogenolysis caused by the SMS, comes an increase in the concentration of NEFA (Petherick et al., 2009a). This release into the circulatory system comes as the net result of lipolysis of adipocytes, increased energy demands, or both (Gonzalez et al., 2008). In a study comparing the effect of increased social pressure caused by a reduced number of feed bunks, Gonzalez et al. (2008) found that pens with 8 or 4 heifers per bunk had higher concentrations of NEFA than those with 2 per bunk. However, Van de Water et al. (2003) saw no significant difference in NEFA before and after transport.

Urea. Any process that increases protein catabolism will tend to result in increased levels of plasma urea (Knowles et al., 2014). Therefore, an increase in cortisol may cause an increase in urea as a response to stress. Urea may also increase as a result of food deprivation.

In assessing stress due to transportation in cattle, Sporer et al. (2008) found an increase in concentrations of urea 4.5 hr after the initiation of transportation, which was significantly different than levels 45 hr after initiation.

Creatine Kinase. While not stimulated by the central nervous system, creatine kinase, also referred to as creatine phosphokinase, is an indicator of prior stress on an animal. It is present in muscle and uses creatine phosphate to convert adenosine diphosphate (ADP) into adenosine triphosphate (ATP) (Knowles et al., 2014). This metabolite is released into the blood when physical exertion results in muscle damage or bruising (Gruber et al., 2010; Knowles et al., 2014).

As a result of injury sustained due to loading, transport, and unloading, Sporer et al. (2008) found an increase in creatine kinase concentration 24 hr after the initiation of transport. These results are supported by Van de Water et al. (2003) who also found a higher concentration in creatine kinase after transportation when compared to before. The impact of handling in regards to creatine kinase is limited. However, cattle that are more temperamental are more likely to struggle in the chute and obtain bruises and muscle damage, possibly leading to increased concentrations of this metabolite.

Hair Whorl Score. While not physiological, the position of a hair whorl in relation to an animal's eye level is believed to be indicative of temperament. In humans, abnormal hair patterns in fetuses and children with early brain development disorders have been reported (Puri et al., 1995). The premise is that hair orientation occurs because of stretching of the scalp and sloping of the hair follicles as a result of cranium development during the first trimester of pregnancy. Therefore, cranial development leading to abnormal hair patterns could be linked with early brain development disorders and abnormal temperament. Since facial hair whorl position can be measured early in an animal's life, it could have value as an indicator trait in large-scale selection.

Hair whorl position can be determined in different ways. Grandin et al. (1995) categorized them as 'high' if the center was above the top of the eyes, 'low' if below the bottom of the eyes, and 'medium' if in between. Olmos and Turner (2008) used this same scale along with a separate ratio scale. This ratio scale separated the 'face' of the cow into 100 arbitrary units using horizontal lines drawn across the forehead to join the ventral aspects of the ears and between the apexes of the nostrils. Using this scale, the central point of the hair whorl was determined using image analysis software.

A higher facial hair whorl position is believed to relate to an increased level of agitation of extensively raised beef cattle with large flight zones while restrained in a crush (Grandin et al., 1995), as well as an increased reaction when approached by an unfamiliar human (Randle, 1998). Dairy cattle with two hair whorls had decreased side preference upon entering a milking parlor (Tanner et al., 1994). However, in contrast, results obtained in other studies (Randle, 1998; Brouček et al., 2004) show no differences in behavior during an open field test, introduction to a novel object, or approach by a familiar human associated with the position of hair whorl.

Subjective vs. Objective Measurement

The goal of quantifying temperament in any animal should be to use observations to select for docile animals. While producers may have inadvertently been selecting for docile cattle simply as an artifact of domestication, purposeful selection for docility is only a recent addition to breeding programs. As with any trait, in order to have selection adopted on a large scale, simple, meaningful, and inexpensive measurements of temperament need to be established (Burrow et al., 1988; Burrow and Corbet, 2000).

Many of the temperament measures previously discussed are scored on a categorical scale. This makes them very easy and inexpensive to collect. Still, they also are subjective in nature. For the purpose of comparison, objective measures such as exit velocity are not subject to inconsistency of observer and are recorded on a continuous scale, creating more accurate estimates with higher heritabilities (Hoppe et al., 2010). This is not to say that subjective measurements are not accurate, they just leave more room for error. As an illustration, consider flight scores. Unlike exit velocity, flight scores are subjective requiring no additional equipment to collect, making them simpler to implement on farm. However, exit velocity can be directly measured using infrared technology. Therefore, the expense associated with equipment purchase needs to be outweighed by the benefits of the increased accuracy of measurement; otherwise producers are more likely to adopt the inexpensive option (Burrow and Corbet, 2000).

The relationship between subjective and objective measurement needs to be investigated. Finding strong correlations between measurements would be ideal, and would suggest that simple, on farm, subjective measurements were adequate to change docility in cattle over generations. This has proven more difficult than expected. Curley et al. (2006) found that behavior assessed with subjective methodologies does not correspond to temperament as well as an objective score such as exit velocity. This was attributed to the inaccuracy previously mentioned. However, evidence still suggests that subjective measurements such as crush score are a powerful tool in quantifying animal behavior (Olmos and Turner, 2008).

Relationships of Temperament Measures

Given the plethora of measurements proposed to quantify beef cattle temperament, it is important to understand how they correspond. Some researchers report that certain temperament measurements are not repeatable, are redundant, and do not help to differentiate between excitable and docile cattle (Burrow and Corbet, 2000; Curley et al., 2006). Such results verify the need for further research to disentangle the true relationships among these measurements.

Conflicting Results

Tulloh (1961) suggested no relationship between behavior when first entering the scale (balk score) followed by entering a separate crush, or headgate, when compared with a summary chute score (ranged from 1-docile to 6-aggressive), and no relationship between balking scores upon entering the scale compared to entering the crush. Breed differences were observed in Hereford, Shorthorn, and Angus cattle for both balking score entering the crush and crush score, but not for balking score upon entering the scale. Considering this was the first use of a numerical scale to evaluate behavior in the crush, these inconsistent results were not promising. It was concluded that temperament of an individual animal cannot be effectively recorded by a single observation, but must be observed over time as day to day variations in behavior exist. However, Curley et al. (2006) reported more encouraging results. Yearling Brahman bulls were observed on three separate occasions, two months apart (d 0, d 60, and d 120). Low to moderate correlations existed between measurements of exit velocity, pen score, and chute score within d 0, d 60, and d 120 individually, but they did not persist when d 0 was compared to d 60 and d 120 for chute score. Furthermore, correlations of serum cortisol and temperament measures were strongest for exit velocity. It was concluded that temperament assessed with subjective methodologies were not as representative of lifetime temperament as exit velocity. Finally, and perhaps the most contrasting to those of Tulloh (1961), were the findings of Cooke et al. (2009). They reported positive moderate to high correlations between chute score, exit velocity, and pen score ranging from 0.45 (chute and pen score) to 0.75 (exit velocity and pen score), suggesting significant relationships among the measurements. However, in comparing these scores in a control group of cattle to those believed to be acclimated to handling, there was no difference in chute score, pen score, or exit velocity. Clearly there is still much to learn about the relationships among these measurements.

Comparing results of temperament evaluations based on crush score is difficult due to the many different numerical scales available for use. This may be the cause of the conflicting results that are found in the literature. Voisinet et al. (1997a; 1997b) were able to sort cattle into temperament categories based on chute scores with significant effects on meat quality endpoints. But King et al. (2006) concluded that chute scores have few statistically significant relationships to other temperament variables, such as exit velocity and cortisol levels, and are least effective in differentiating animals according to temperament. Because of this low association, correlations between chute score and Warner-Bratzler shear force values ranged from -0.05 to 0.14. It is important to note that these studies all used the 5-point chute scores provided by Grandin (1993). As previously discussed, the type of chute used and how much movement the cattle were allowed will have an effect on the significance of relationships.

As for the relationship of crush score with other temperament measurements, the correlation between crush score and flight speed ranges from -0.44 to 0.41 (Burrow and Corbet, 2000; Olmos and Turner, 2008), and the correlation between flight score and crush score was found to be -0.07 (Burrow and Corbet, 2000). From these results, and others suggesting moderate to low correlations, Burrow and Corbet (2000) concluded that subjective measurements of temperament were not sufficiently correlated with objective measures.

In contrast, objective measurements such as cortisol are moderately to highly phenotypically correlated with exit velocity ($r = 0.16$ to 0.41; King et al., 2006) and an overall temperament score, taken as the average of chute score, pen score, and exit score ($r = 0.49$; Cooke et al., 2009). Cortisol is also believed to be moderately phenotypically correlated with chute score, exit velocity and individual pen score ($r = 0.34$, 0.49, and 0.44, respectively; Cooke et al., 2009). However, these relationships decreased with decreasing excitability (Curley et al., 2006; King et al., 2006). From these results, it can be suggested that chute score, exit velocity and pen score are reliable measurements of stress, but more so in excitable animals.

Hair whorl scores are another area of contrasting results in today's literature. It is believed that the location of the hair whorl in relation to the facial plane of the cow can be used as an indicator trait for temperament. Grandin et al. (1995) concluded that cattle with hair whorls located above the eyes become more agitated when restrained compared to cattle with hair whorls even with or below the eyes. These results were consistent across *Bos indicus* and *Bos taurus* breeds of cattle. Other researchers suggest that the relationship between hair whorl position and temperament may be sensitive to the measurements being taken, citing no significant relationship between hair whorl position and flight speed (Olmos and Turner, 2008).

The presence of such contrasting results highlight the need for research into the relationships among temperament measurements in an attempt to delineate the complicating factors. It has been suggested that conflicting results, such as these, may be explained if the amounts of psychological and physical stress within each study were considered (Grandin, 1997). How an animal responds to a fearful situation is difficult to predict because it depends on how that animal perceives an experience. For a habituated animal, a squeeze chute may be perceived as non-threatening, but to an animal raised on range, it may trigger an extreme fear response.

Genetics of Temperament

The two temperament measurements that appear most frequently in the literature are objective flight speed and crush score. Flight speed has a moderately high heritability of 0.35 (Burrow and Corbet, 2000; Sant'Anna et al., 2013). There appears to be a relationship between the heritability of flight speed and the age at which the measurement is taken. Burrow et al. (1988) measured heritability at weaning and 18 mo to be 0.54 and 0.26, respectively, which may reflect some modification of temperament through handling experiences. Similarly, Kadel et al. (2006) reported heritabilities of flight speed post-weaning and start of finishing to be 0.30 and 0.34, respectively, with a repeatability of 0.46 between the measurement ages. Also, genetic correlations across measurement ages for flight speed and crush score were 0.98 and 0.96, respectively, while genetic correlations between flight speed and crush score post-weaning and at start of finishing were -0.37 and -0.35, respectively. Visual flight score is not as heritable with estimates ranging from 0.08 to 0.21 (Fordyce et al., 1982; Burrow and Corbet, 2000; Kadel et al., 2006; Sant'Anna et al., 2013).

Conversely, crush score is a subjective measurement, with no equivalent objective measure. Reported heritabilites range from 0.15 to 0.30 (Burrow and Corbet, 2000; Kadel et al., 2006). Kadel et al. (2006) again found a difference in heritability between post-weaning (0.19) and start of finishing phase (0.15), but not as dramatic as with flight speed, and a strong genetic correlation between the two measurements (r = 0.96). Furthermore, the genetic correlation between flight speed and chute score was 0.85 (Sant'Anna et al., 2013).

While some of these heritability estimates are low, researchers suggested that temperament was a heritable trait and that selection based on measurements of temperament could be used to increase docility in cattle. Sant'Anna et al. (2013) added that while all temperament indicator traits expressed genetic variability and therefore would respond to selection, flight speed would result in faster genetic gain. This is because flight speed is an objective measure and therefore has higher accuracy than a subjective temperament measure.

Selection for Docility

Measuring temperament in animals becomes even more appealing if we can use it to select for docile cattle. In being moderately heritable, this is possible (Le Neindre et al., 1995). A temperament scoring system was created and proposed by the Beef Improvement Federation (BIF Guidelines, 2002) as a result of the North American Limousin Foundation (NALF) identifying temperament as their number one breed priority (Hyde, 2004). In 1998, the NALF became the industry's first breed to implement a docility EPD in their national genetic evaluation, and were followed by the American Angus Association (AAA) in 2008 (Beckman, 2008).

The scoring system ranges from 1 (docile) to 6 (very aggressive) and is assigned while calves are restrained in a chute at weaning or yearling ages (NALF or AAA, respectively) (Beckman, 2008). As a threshold trait, it is reported as a probability, and expressed as a deviation from 50% (Kuehn et al., 1998).This EPD reflects the probability that the offspring will inherit genes for acceptable behavior, with greater numbers associated with calmer progeny. Heritability of docility used by NALF and AAA for genetic evaluation are 0.40 and 0.37, respectively (Beckman, 2008). However, estimates of maternal heritability are near zero for both breeds (Beckman et al., 2007; Northcutt, 2007).

Selecting for more docile cattle may, however, have a negative effect on other traits. It is possible that intense selection for docility could lead to cattle that are too docile and lose their maternal ability (Grandin, 1997; Gringard et al., 2001). It may be beneficial to stick to an intermediate docility score when making selection decisions, and most importantly taking into consideration the sector of the industry being affected.

The target or optimum docility may differ for a cow-calf operation and a feedlot environment. For a cow-calf operation, maternal defensive aggression must be considered. While selection for a decrease in maternal defensive aggression would facilitate safer handling, it could also impact the extent to which a cow will defend her calf (Turner and Lawrence, 2007). Likewise, selection for maternal behaviors to promote offspring survival could increase the aggressiveness of cattle after calving. Because of this, it is essential to understand how selection for maternal care could impact aggressiveness before making selection decisions.

For the feedlot sector of the industry, the maternal protective behavior is not of concern. Therefore, selection for increased docility, or at least acclimation to handling, would result in safer working conditions for both animals and handlers. Selection for calmer cattle in a feedlot setting could result in increased BW, ADG, meat quality, and many other important economic factors already discussed. The only pitfall of selection for docility would be to have cattle so docile they are difficult to move through the working facilities.

Breed Differences

There has long been a view that cattle that are of *Bos indicus* descent are more excitable or less docile than *Bos taurus* breeds. Burrow (1997) reported that *Bos indicus* breeds, without exception, were more difficult to handle under extensive management conditions than *Bos taurus* breeds. Further research supports this conclusion: temperament score and plasma cortisol concentrations were higher in *Bos indicus* cattle (Fordyce et al., 1982; Hammond et al., 1996; Voisinet et al., 1997b). The only evidence to the contrary comes from Burrow and Corbet (2000) who analyzed temperament in *Bos indicus* and *Bos indicus* x *Bos taurus* crossed calves. They concluded that crossbred calves had similar or worse flight speed and crush score than purebred *Bos indicus* calves.

Because *Bos indicus* cattle were believed to be more temperamental, there was initially more focus on those breeds, both globally and in warmer parts of the United States. It was believed that results, especially genetic parameters, should only be compared to or utilized for other *Bos indicus* breeds since there was no way of knowing how they would apply to continental breeds of cattle. Therefore, Müller and Keyserlingk (2006) investigated the efficacy of exit velocity and its correlation to productivity in *Bos taurus* cattle. They found similar results to the previous work done in *Bos indicus* breeds, concluding flight speed was moderately repeatable, and had a quadratic relationship with ADG. Cattle with high flight speeds had the lowest ADG over the 8-wk study period.

More specifically, Hoppe et al. (2010) found breed differences for chute and flight scores among Angus, Charolais, Hereford, Limousin and Simmental cattle ($P < 0.001$).

Charolais and Limousin had the highest chute and flight scores, followed by Simmental ($P < 0.05$). Angus had higher chute scores than Hereford ($P < 0.05$), but their flight speed scores were not different ($P > 0.05$).

This difference in temperament was explained by differences in rearing systems, with the Charolais and Limousin being raised in a traditional French system known for a strong habituation of cattle to humans. It is believed this may have masked underlying temperament traits preventing any indirect selection for docility (Grandin, 1994). This can be compared to the Angus and Hereford cattle, traditionally raised in extensive pasture conditions with minimum exposure to humans, with aggression having a higher likelihood of being indirectly selected against.

While there may be a genetic difference in temperament between breeds, temperament measures are robust for all breeds of cattle. Breed differences are therefore of little concern when considering applying these measures and interpreting results of previous research. Methodologies are still efficient in delineating levels of response to stress regardless of breed. This is beneficial, as response to handling can be compared across breeds to formulate stronger arguments for selection on docility.

Areas to Address

There is currently no consensus regarding an ideal methodology to measure temperament in an on-site assessment (Sant'Anna et al., 2013). The key to the adoption of selecting for temperament in animals is the ability to do so in a way that is inexpensive, fast, and easy to implement in a production setting. To do this, there are some questions that need to be answered. First, the reliability of the observers taking these measurements needs to be evaluated. Next, the issue of acclimation to handling needs to be addressed. Curley et al. (2006) suggest that with increased number of times an animal goes through a working facility comes a decrease in temperament score. Similarly, if these measures are being used to estimate genetic merit of an individual, the repeatability of these temperament scores need to be addressed. Additionally, any change in subjective measurements of temperament over time, or in the stress of an animal during handling, needs to be substantiated using objective measures of temperament known to be associated with stress. Finally, there has not been much research as to the effect of animal groupings on temperament. In other words, the effect of an excitable animal co-mingling with other docile animals, and vice versa, on individual behavior is not clearly defined. Cattle are herd animals, therefore it is understandable that they are more comfortable when surrounded by other members of the herd. A more detailed investigation into the intricacies of this relationship would be beneficial.

Intra- and Inter-observer Reliability

When using multiple individuals to evaluate subjective measurements of temperament such as chute and exit score, it is important to assess both their intra-observer reliability, or how consistent an observer is in categorizing the behavior across the different time periods, and their inter-observer reliability, or how consistent the different observers were in their scoring of the same animals. These statistics are important because they indicate the consistency of subjective measurements. Higher coefficients indicate more consistent observations and are therefore more desirable

(Bokkers et al., 2012). Once consistency of observations are established, conclusions can be reliably drawn from the data.

It is common in temperament studies with subjective measurements to report inter- or intra-observer reliability, or both. These statistics can be calculated in a variety of ways including percent of agreement, kappa coefficient, Kendall's coefficient of concordance, or an intra-class correlation. Percent of agreement (PA) is calculated as the number of observations in agreement divided by the total amount of observations, where PA = 0 indicates no agreement and PA = 100 indicates perfect agreement. The statistic can be calculated with a tolerance of zero, where all observers must have exactly the same score, or a tolerance of one, where all observers are within one score of each other.

To account for chance agreements, and to be able to compare reliabilities across studies, a kappa coefficient can be used. Reliability between pairs of observers or observations is calculated using weighted Cohen's Kappa (Cohen, 1960), while Fleiss' kappa coefficient (K) can be used when more than two observations are being compared (Fleiss, 1971). Kappa coefficients vary from -1 to 1 with negative values indicating agreement that is poorer than chance, zero indicating exactly chance agreement, and positive values indicating agreement that is better than chance (Fleiss and Cohen, 1973).

Intra-class correlation (ICC) describes how strongly observations resemble each other (Shrout and Fleiss, 1979). An ICC of 0 represents no agreement among observers, and an ICC of 1 represents perfect agreement. Similarly, Kendall's coefficient of concordance (KCC) ranges from 0 (no agreement) to 1 (complete agreement), and is calculated to measure agreement as an index of reliability of ordinal data.

Threshold correlation coefficients for acceptable reliability have been established, with values above 0.70 for Kendall and intra-class correlation coefficients (Martin and Bateson, 1993), and values above 0.40 for Kappa coefficients (Landis and Koch, 1977), indicating accurate evaluation. Olmos and Turner (2008) report 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. Furthermore, Vetters et al. (2013) report weighted Kappa correlation coefficients of 0.60 for ES using a 4-point scale (Lanier and Grandin, 2002).

In most of the literature, observations are taken using both experienced and inexperienced observers, which may increase the occurrence of inconsistency. Reliabilities reported by Bokkers et al. (2012) were slight to high even for experienced observers, indicating imperfection in the observations. However, there was still a significant difference in reliability between experienced and inexperienced observers, as expected. Once the reliability of subjective measurements is established, their value can be more comprehensively appraised.

Acclimation to Handling

Having excitable cattle can be detrimental to the health of the workers and the cattle themselves, with negative effects on production traits. Previous research indicates that instead of culling those animals, acclimation of cattle to human interaction is an alternative method to improve temperament (Francisco et al., 2012). This was sparked from an observation by Grandin (1993) that cattle with previous handling experiences in a livestock market settled down more quickly in an abattoir than cattle that came directly from the farm. To test this hypothesis, Francisco et al. (2012) acclimated steers by exposing them to handling processes twice a week for four weeks. Compared to a control, acclimated steers had decreased chute score, exit velocity, temperament score, and cortisol levels. Similar studies reflect this acclimation to handling or learning behavior in animals as they gained experience in a working environment (Curley et al., 2006; King et al., 2006; Behrends et al., 2009; Hall et al., 2011).

In a slightly different study, cattle had the opportunity to acclimate to the presence of people and associate that presence with food, making them less fearful and more likely to approach (Petherick et al., 2009b). Cattle were acclimated to one of three human handling and yard experiences: 1) good handling/yarding where, prior to handling, cattle were moved to a yard containing hay and then moved into the race and held with a group mate, 2) poor handling/yarding where cattle were moved to a bare yard and held there until the good group had been assessed and then moved into the race and held with a group mate for excessive amounts of time, and 3) minimal handling/yarding where cattle were moved to a yard and immediately released to access water and walk to their home paddock. Temperament was measured at the beginning of the experiment using flight speed and a "fear of humans" test, which measured the closest approach to a stimulus person, the proximity to the person, and the amount of movement of an animal around a test arena. The largest difference in temperament was seen in cattle from minimal treatment groups that were exposed to poor and minimal treatment and facilities. Animals treated well and given food prior to handling had a relatively low fear of people initially, causing a smaller difference in temperament among cattle. Flight speed decreased for all treatments, with no difference among groups at the end of the experiment, with the greatest rate of decline in the animals treated well.

Conversely, Cooke et al. (2009) acclimated animals 2 times a week for 5 months by walking among them in a pen for 15 min feeding range cubes with no change in pen scores, chute scores, exit velocities, or concentrations of plasma cortisol. In this case, acclimation did not influence temperament, as responses neither increased nor decreased with more exposure to handling.

While acclimation to handling is seen as an alternative to improve temperament, it is possible this does not always hold true. Some researchers have found that flight speed increases with repeated handling over time (Petherick et al., 2002; Müller and Keyserlingk, 2006). This increase could either be the result of stressful handling of the animals or of over handling of the animals causing an increase in their fearfulness. Furthermore, cattle may become acclimated to specific conditions or handling practices, but when exposed to a novel situation revert back to their initial temperament response (Behrends et al., 2009). This suggests that temperament measured at an animal's first working provides the best information concerning future performance and beef tenderness. Behrends et al. (2009) found that temperament measurements at weaning had stronger relationships with important production traits compared to those collected later in life.

Both the timing and severity of experience play a role in acclimation. Walking young calves through a handling system, with no invasive procedures, resulted in adult animals that were calmer (Becker and Lobato, 1997), while attempts to acclimate older cows was less successful (Cooke, 2014). Animals will not habituate to extremely adverse experiences (Grandin et al., 1986). Therefore, if unavoidable, it is better to wait until later in life to expose an animal to harsher conditions.

If temperament changes over time, an animal may need to be observed repeatedly to be accurately assessed. If an animal's temperament is the primary trait of interest in a breeding program, or is completely unmanageable, then culling of that animal based on the first observation may be justified. However, allowing acclimation to handling may be of value when individual animals have borderline acceptable temperament, especially when balanced with other breeding objectives. Cattle that excel in all other aspects of a producers breeding goal may benefit from additional observations of temperament before final selection decisions are made.

Repeated Measures and Non-Gaussian Data Analysis. Data collected from experiments are often assumed to be independent and follow a normal, or Gaussian, distribution. However, in experiments that wish to investigate behavior, this is largely not the case (Wajnberg and Haccou, 2008). More importantly, these experiments involve repeated measures, making observations dependent upon each other. Because of this, traditional ANOVA and regression methods may not be sufficient.

Repeated measures are used to study changes over time and space (Stroup, 2013). The key feature of repeated measures is the increased probability of correlation among observations on the same animal, meaning they are not independent. Therefore, linear models must account for this increased probability using a time and space source of variation within the fixed effects, or correlation among measurements. Within-subject evaluation can be completed using repeated-measures analysis of variance (RM-ANOVA) (Fisch, 2001). Specifically, RM-ANOVA can be utilized to assess whether significant changes within a phase of the experiment have occurred, or whether there were changes across time periods for an individual subject. However, a standard RM-ANOVA assumes homogeneity in the variance-covariance matrix. Should the variance-covariance matrix not be homogenous, different matrices can be instead applied, including unstructured, autoregressive, and many others depending on assumed relationships among measurements.

The normality of data can be tested using the Jarque-Bera (Skewness-Kurtosis) Test (Jarque and Bera, 1980). The Jarque-Bera test is a type of Lagrange multiplier test developed to test normality, heteroscedasticity, and serial correlation of regression residuals; it is computed from the skewness and kurtosis of a distribution. The statistic asymptotically follows the chi-squared distribution with two degrees of freedom. If the value is less than the chi-squared statistic, the null hypothesis is rejected and data are considered non-Gaussian.

The log transformation is a widely used method to address skewed data (Feng et al., 2014). Due to its ease of use and popularity, it is often included in most major statistical software packages. Comparisons of the Jarque-Bera statistic between normal and transformed data provides insights into which is more normally distributed. When using log transformed data for analyses, results will need to be back transformed to allow clear biological interpretation.

Use of Objective Measures to Establish Stress in Animals

In order to effectively select for docility it is imperative that selection criteria are truly indicative of an animal's response to stress. Subjective measurements such as chute and exit score provide fast and easy assessments of temperament, but may be subject to observer bias and inconsistency and it is unclear whether they are reliable indicators of stress.

Objective measurements have been shown to increase during stressful handling of cattle. Lay et al. (1992) concluded that dairy cattle exposed to hot iron branding had higher heart rates and levels of plasma cortisol compared to those freeze branded or to a control. Extensive research has also been conducted on the stress associated with truck transportation in beef cattle. Higher levels of plasma cortisol, urea, and creatine kinase have been found after transport (Van de Water et al., 2003; Sporer et al., 2008) as well as a higher heart rate while loading and unloading the truck (Van de Water et al., 2003).

While elevated concentrations of cortisol are a hallmark of stress in livestock species (Grandin 1997; Möstl and Palme, 2002), it is impractical to wish for producers to take blood samples on each of their animals in order to select for temperament, mainly due to the cost of the assay. Therefore, establishing a relationship between these objective and subjective measures would help corroborate selection decisions drawn from using subjective methods, making it more convenient for producers to accurately select for more docile cattle.

In addition, assessing change in these objective measures could help substantiate conclusions drawn in regards to an animal's acclimation to handling. Most literature assess the relationship among these measures in regards to transportation or pre-slaughter stress, but do not evaluate their relationship with typical handling practices. If subjective measures of temperament decrease, but individuals physiologically remain stressed, then it could be argued that the animal is not truly acclimating to its environment.

Effect of Grouping on Calf Temperament

The response to handling of any animal is not just a reaction to humans, but also depends on social context, physical environment and the novelty of the situation. As cattle are herd animals, this particularly includes the social context of the situation (Gringard et al., 2000). Isolation from herd mates can cause animals to become agitated, which can lead to aggression (Le Neindre et al., 1995). Being isolated may force animals into close contact with humans, causing them to be fearful. Le Neindre et al. (1995) found that when exposed to humans in their normal group, agitation is less frequent than when animals were secluded.

It is believed the presence of peers not only reduces stress responses of cattle to any fear-eliciting situation but also has a positive effect on learning (Grandin and Deesing, 2014). It is a common practice that horse trainers use calm, trained horses to reduce fear in younger horses when exposed to experiences, such as loading a trailer, for the first time. This idea extends into beef cattle in that cattle trained with herd mates to obtain a food reward by pressing on a plate learned faster than isolated animals (Boissy and Le Neindre, 1990).

However, little work has been done with cattle to determine whether temperament during handling can be modified by the presence of other animals (Gringard et al., 2000).

If calm herd mates are included with excitable cattle, how does this affect the temperament of the group of animals? Gringard et al. (2000) showed that the presence of peers in an adjacent pen modifies the behavior of an isolated calf when no human is present, increasing the time the animal is stationary compared to when no peers are present. This suggests that cattle do not become as agitated when they can at least see their peers. Furthermore, with peers in an adjacent pen, isolated calves were more difficult to hold in an opposing corner of a pen, separated from peers, compared to when no peers were present. When in the presence of herd mates, calves had a stronger incentive to break away from human isolation.

Building off of these findings, it is possible to compare an animal's individual pen score with that of a group pen score. Instead of animals being separated from their peers by a fence, they in this case animals would be grouped together in a single pen. Direct comparisons can then be made between an animal's behaviors while isolated versus when in a group, both in the presence of a human. These comparisons need not only be based on subjective pen scores on a numeric scale. In depth comparisons of behaviors between individually and group penned cattle have yet to be done. Detailed ethograms could be developed to assist producers in not only pinpointing key behaviors to assess in cattle, but to compare their demeanor when individually secluded versus in a group atmosphere where they are likely to feel less threatened.

Multivariate Data Analyses. Multivariate analysis is a set of statistical techniques used for analysis of data that contains more than one variable and is especially valuable when those variables are correlated (Grimnes and Martinsen, 2015). These methods provide an empirical method for information extraction. When analyzing multidimensional data, relationships among variables are fundamental to explore, and multivariate analysis provides useful techniques to understand and quantify those relationships (Olkin and Sampson, 2001). The exploration of structure and patterns among variables for complex multivariate data sets is crucial for modern data analysis.

When analyzing high-dimensional data, it is worthwhile to consider they could be replaced by a fewer number of measurements without the loss of information (Rao, 1964). One way of accomplishing this reduction is by creating a reduced set of linear transformations of the input variables (Izenman, 2013). The earliest and most popular projection method is called principal component analysis.

Principal Component Analysis. First introduced by Pearson (1901), Principal Component Analysis (PCA) is a statistical method that can be performed on a wide variety of mathematical, statistical, or dedicated computer software (Grimnes and Martinsen, 2015). It can be mathematically defined as an orthogonal linear transformation that transforms the data to a new coordinate system where the greatest amount of variance explained by any projection of the data lie on the first principal component, the second greatest variance on the second principal component, and so on.

Principal component analysis is primarily used to reduce dimensionality of data and pattern recognition (Izenman, 2013). This pattern recognition or discovery comes in the form of graphical displays of the principal component scores. Any principal component with zero or near zero variance is considered constant and can be used to detect collinearity to reduce the number of variables in the data. Similarly, outliers that alter the perceived dimensionality of the data can be identified.

Using these multivariate analyses techniques help to determine underlying relationships among not only subjective and objective measurements, but also assist in visualizing relationships among the behaviors recorded in the individual and group pen atmospheres. While not hypothesis driven, PCA may help in answering some questions about relationships among various behavior scores not evident in formal statistical analysis.

Summary

Temperament has been described as an animal's behavior response to handling by humans, or any potentially fear-eliciting situation (Burrow and Corbet, 2000; Olmos and Turner, 2008). Many studies have investigated the response of cattle to handling, finding that reaction depends not only on the presence of humans, but also on elements such as social context, physical environment and novelty of the situation. In doing so, many methods to quantify temperament have been proposed, both objective and subjective. These measurements all have their advantages and disadvantages, and the relationship among them is yet to be fully understood. However, using them together can give a relatively accurate estimate of an animal's temperament and, more importantly, a prediction of their reaction to novel situations. This is true regardless of the breed of animal.

Ideally, these methods would be combined in an easy, inexpensive way that could be used in a production setting to help producers select for more docile cattle. A consensus on a way in which to do this is yet to be determined. However, temperament is a heritable trait and successful selection for docile cattle is well underway.

The primary objective of this research, once reliability of subjective measurements is determined, is to investigate acclimation to handling in cattle. It is hypothesized that treatment groups that are worked more frequently will show a decrease in excitable temperament compared to a control. However, it is possible that these cattle become overworked, causing a possible upward shift in temperament scores towards the end of the experiment, due solely to agitation of the cattle.

The secondary objective of this research is to validate changes in subjective measurements of temperament using more objective methods previously described, as well as determine the relationship among the objective measures themselves. If changes in blood metabolites are representative of what is visually changing in these animals, then not only are the conclusions from the first objective substantiated, but the subjective methods used can be deemed representative of an animal's response to a stressful situation.

Another novel relationship to be investigated is that of individual and group pen score. While most research deals with an attempt to seclude an individual animal into a corner, both with and without the presence of peers, there is not a plethora of research available in regards to an animal's response to a human stimulus individually, and when grouped with herd mates. The objective is to evaluate the change in an animal's temperament between isolation and introduction to a herd environment. The hypothesis to be tested is that when excitable animals are introduced to a more docile group, the group as a whole will remain relatively calm, indicating the excitable animal is calmer when surrounded by other herd mates.

Overall, the literature discussed, while valuable, does have areas that have either been overlooked or include conflicting results. It is the objective of this dissertation to answer some questions about temperament and produce thought provoking findings that add to the knowledge of the subject in hopes of providing producers with a more concise and representative method for selecting for temperament.

Literature Cited

Andrade, O., Orihuela, A., Solano, J., Galina, C. S., 2001. Some effects of repeated handling and the use of a mask on stress responses in zebu cattle during restraint. Appl. Anim. Behav. Sci. 71, 175-181.

Becker, B. G, Lobato, J. E. P., 1997. Effect of gentle handling on the reactivity of zebu crossbred calves to humans. Appl. Anim. Behav. Sci. 53, 219-224.

Beckman, D. W., 2008. Docility EPD: A tool for improvement. 40th Annu. Meet. Proc., Calgary, Canada. Beef Improvement Federation. 2008, 81-87.

Beckman, D. W., Enns, R. M., Speidel, S. E., Bringham, B. W., Garrick, D. J., 2007. Maternal effects on docility in Limousin cattle. J. Anim. Sci. 85, 650-657.

Behrends, S. M., Miller, R. K., Rouquette Jr., F. M., Randel, R. D., Warrington, B. G., Forbes, R. D. A., Welsh, T. H., Lippke, H., Behrends, J. M., Carstens, G. E., Holloway, J. W., 2009. Relationship of temperament, growth, carcass characteristics and tenderness in beef steers. Meat Sci. 81, 433-438.

Boissy, A., Le Neindre, P., 1990. Social influences on the reactivity of heifers: implications for learning abilities in operant conditioning. Appl. Anim. Behav. Sci. 25, 149-165.

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

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

Brouček, J., Uhrinčat, M., Kišac, P., Hanus, A., 2006. Hair whorl position as a predictor of learning ability and locomotor behaviour in cattle. Acta. Vet. Brno. 73, 455-459.

Burdick, N. C., Randel, R. D., Carroll, J. A., Welsh Jr., T. H., 2011. Interactions between temperament, stress, and immune function in cattle. Int. J. Zool. Res. 2011.

Burrow, H. M., 1997. Measurements of temperament and their relationship with performance traits of beef cattle. Anim. Breed. Abstr. 65, 477-495.

Burrow, H. M., Seifert, G. W., Corbet, N. J., 1988. A new technique for measuring temperament in cattle. Proc. Aust. Soc. Anim. Prod. 17, 154-157.

Burrow, H. M., Corbet, N. J., 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.

Busby, W. D., Strohbehn, D. R., Beedle, P., King, M., 2006. Effect of disposition on feedlot gain and quality grade. Iowa State University Animal Industry Report. AS 652, ASL R2070.

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.

Cooke, R. F., Arthington, J. D., Araujo, D. B., Lamb, G. C., 2009. Effects of acclimation to human interaction on performance, temperament, physiological responses, and pregnancy rates of Brahman-crossbred cows. J. Anim. Sci. 87, 4125-4132.

Cooke, R. F., Bohnert, D. W., Meneghetti, M., Losi, T. C., Vasconcelos, J. L. M., 2011. Effects of temperament on pregnancy rates to fixed-timed AI in *Bos indicus* beef cows. Livest. Sci. 142, 108-113.

Cooke, R. F., Bohnert, D. W., Cappellozza, B. I., Mueller, C. J., DelCurto, T., 2012. Effects of temperament and acclimation to handling on reproductive performance of *Bos taurus* beef females. J. Anim. Sci. 90, 3547-3555.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Meas. 20, 37-46.

Curley Jr., K. O., Paschal, J. C., Welsh Jr., T. H., Randel, R. D., 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.

Curley Jr., K. O., Neuendorff, D. A., Lewis, A. W., Cleere, J. J., Welsh Jr., T. J., Randel, R. D., 2008. Functional characteristics of the bovine hypothalamic-pituitary-adrenal axis vary with temperament. Horm. Behav. 53, 20-27.

Feng, C., Wang, H., Lu, N., Chen, T., He, H., Lu, Y., Tu, X. M., 2014. Log-transformation and its implications for data analysis. Shanghai Archives of Psychiatry. 26, 105-109.

Fisch, G. S., 2001. Evaluating data from behavioral analysis: visual inspection or statistical models? Behavioral Processes. 54, 137-154.

Fleiss, J. L., 1971. Measuring nominal scale agreement among many raters. Psychol. Bull. 76, 378-382.

Fleiss, J. L., Cohen, J., 1973. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. Educ. Psychol. Meas. 33, 613-619.

Fordyce, G., Goddard, M. E., Seifert, G. W., 1982. The measurement of temperament in cattle and the effect of experience and genotype. Proc. Aust. Soc. Anim. Prod. 14, 329-332.

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

Francisco, C. L., Cooke, R. F., Marques, R. S., Mills, R. R., Bohnert, D. W., 2012. Effects of temperament and acclimation to handling on feedlot performance of

Bos taurus feeder cattle originated from rangeland-based cow-calf system. J. Anim. Sci. 90, 5067-5077.

González, L. A., Ferrett, A., Mantecca, X., Ruíz-de-la-Torre, J. L., Calsamiglia, S., Devant, M., Bach, A., 2008. Effect of the number of concentrate feeding places per pen on performance behavior, and welfare indicators of Friesian calves during the first month after arrival at the feedlot. J. Anim. Sci. 86, 419-431.

Grandin, T., 1993. Behavioral agitation during handling of cattle is persistent over time. Appl. Anim. Behav. Sci. 36, 1-9.

Grandin, T., 1994. Solving livestock handling problems. Vet. Med. 89, 989-998.

Grandin, T., 1997. Assessment of stress during handling and transport. J. Anim. Sci. 75, 249-257.

Grandin T., Deesing, M. J., 2014. Chapter 4 - Genetics and Behavior During Handling, Restraint, and Herding, In Genetics and the Behavior of Domestic Animals (Second Edition), Academic Press, San Diego, Pages 115-158.

Grandin, T., Curtis, S. E., Widowski, T. M., Thurmon, J. S., 1986. Electro-immobilization versus mechanical restraint in an avoid-avoid choice test for ewes. J. Anim. Sci. 62, 1469-1480.

Grandin, T., Deesing, M. J., Struthers, J. J., Swinker, A. M., 1995. Cattle with hair whorl patterns above the eyes are more behaviorally agitated during restraint. Appl. Anim. Behav. Sci. 46, 117-123.

Grimnes, S., Martinsen, O. G., 2015. Chapter 9 - Data and Models. In Bioimpedance and Bioelectricit Basics (Third Edition), Academic Press, Oxford. Pages 329-404.

Gringard, L., Boissy, A., Boivin, X., Garel, J. P., 2000. The social environment influences the behavioural responses of beef cattle to handling. Appl. Anim. Behav. Sci. 68, 1-11.

Gringard, L., Boivin, X., Boissy, A., Le Neindre, P., 2001. Do beef cattle react consistently to different handling situations? Appl. Anim. Behav. Sci. 71, 263-276.

Gruber, S. L., Tatum, J. D., Engle, T. E., Chapman, P. L., Belk, K. E., Smith, G. C., 2010. Relationships of behavioral and physiological symptoms of preslaughter stress to beef longissimus muscle tenderness. J. Anim. Sci. 88, 1148-1159.

Hall, N. L., Buchanan, D. S., Anderson, V. L., Ilse, B. R., Carlin, K. R., Berg. E. P.. 2011. Working chute behavior of feedlot cattle can be an indication of cattle temperament and beef carcass composition and quality. Meat Sci. 89, 52-57.

Hammond, A. C., Olson, T. A., Chase Jr., C. C., Bowers, E. J., Randel, R. D., Murphy, C. N., Vogt, D. W., Tewolde, A., 1996. Heat tolerance in two tropically adapted Bos taurus breeds, Senepol and Romosinuano, compared with Brahman, Angus, and Hereford cattle in Florida. J. Anim. Sci. 74, 295-303.

Haskell, M. J., Simm, G., Turner, S. P., 2014. Genetic selection for temperament traits in dairy and beef cattle. Front. Genet. 5, 1-18.

Hemsworth, P. H., Barnett, J. L., 2000, Human-animal interactions and animal stress. The Biology of Animal Stress. Basic Principles and Implications for Animal Welfare. Moberg, G. P., and J. A. Mench, CAB Int., Wallingford, UK. Pages 309-335.

Hoppe, S., Brandt, H. R., König, S., Erhardt, G., Gauly, M., 2010. Temperament traits of beef calves measured under field conditions and their relationships to performance. J. Anim. Sci. 88, 1982-1989.

Hyde, L., 2004. Limousin breeders improve temperament. Limousin Breeder's Edge Volume 2. North American Limousin Foundation, Englewood, CO.

Izenman, A. J., 2013. Chapter 7 - Linear Dimensionality Reduction. In Modern Multivariate Statistical Techniques, Springer Science, New York. Pages 195-236.

Jarque, C. M., Bera, A. K., 1980. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. Economics Letters. 6, 255-259.

Kadel, M. J., Johnston, D. J., Burrow, H. M., Graser, H., Ferguson, D. M., 2006. Genetics of flight time and other measures of temperament and their value as selection criteria for improving meat quality traits in tropically adapted breeds of beef cattle. Aust. J. Agric. Res. 57, 1029-1035.

Keay, J. M., Singh, J., Gaunt, M. C., Kaur, T., 2006. Fecal glucocorticoids and their metabolites as indicators of stress in various mammalian species: a literature review. J. Zoo Wildlife Medicine. 37, 234-244.

King, D. A., Schuehle Pfeiffer, C. E., Randel, R. D., Welsh Jr., T. H., Oliphint, R. A., Baird, B. E., Curley Jr., K. O., Vann, R. C., Hale, D. S., Savell, J. W., 2006. Influence of animal temperament and stress responsiveness on the carcass quality and beef tenderness of feedlot cattle. Meat Sci. 74, 546-556.

Knowles, T. G., Warriss, P. D., Vogel, K., 2014. Stress physiology of animals during transport. Livestock Handling and Transport, 4th Edition. T. Grandin, ed. CAB Int., Cambridge, MA. Pages 399-420.

Kuehn, L. A., Golden, B. L., Comstock, C. R., Andersen, K. J., 1998. Docility EPD for Limousin cattle. J. Anim. Sci. 76(Suppl. 1)/J. Dairy Sci. 81(Suppl. 1), 334 (Abstr.).

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

Lanier, J. L., Grandin, T., 2002. The relationship between Bos taurus feedlot cattle temperament and foreleg bone measurements. Proc. Western Section Am. Soc. Anim. Sci. 53, 97-98.

Lay, D. C., Friend, T. H., Bowers, C. L., Grissom, K. K., Jenkins, O. C., 1992. A comparative physiological and behavioral study of freeze branding and hot-iron branding using dairy cows. J. Anim. Sci. 70, 1121-1125.

Le Neindre, P., Trillat, G., Sapa, J., Ménissier, F., Bonnet, J. N., Chupin, J. M., 1995. Individual differences in docility in Limousin cattle. J. Anim. Sci. 73, 2249-2253.

Lindsay, S., Selvaraj, S., Macdonald, J. W., Godden, D. J., 2004. Injuries to Scottish farmers while tagging and clipping cattle: a cross-sectional survey. Occ. Med. 54, 86-91.

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

Möstl E., Palme, R., 2002. Hormones as indicators of stress. Domest. Anim. Endocrinol. 23, 67-74.

Müller, R., von Keyserlingk, M. A. G., 2006. Consistency of flight speed and its correlation to productivity and to personality in Bos taurus beef cattle. Appl. Anim. Behav. Sci. 99, 193-204.

Northcutt, S., 2007. Docility genetic evaluation research. By the numbers. American Angus Association. St. Joseph, MO.

Olkin, I., Sampson, A. R., 2001. Multivariate Analysis: Overview. In International Encyclopedia of the Social and Behavioral Sciences. Pergamon, Oxford. Pages 10240-10247.

Olmos, G., Turner, S. P., 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.

Pearson, K., 1901. Principal components analysis. The London, Edinburgh, and Doublin Philosophical Magazine and Journal of Science. 6, 559.

Petherick, J. C., Holroyd, R. G., Doogan, V. J., Venus, B. K., 2002. Productivity, carcass and meat quality of lot-fed Bos indicus cross steers grouped according to temperament. Aust. J. Exp. Agric. 42, 389-398.

Petherick, J. C., Doogan, V. J., Venus, B. K., Holroyd, R. G., Olsson, P., 2009a. Quality of handling and holding yard environment, and beef cattle temperament: 2. Consequences for stress and productivity. Appl. Anim. Behav. Sci. 120, 28-38.

Petherick, J. C., Doogan, V. J., Holroyd, R. G., Olsson, P., Venus, B. K., 2009b. Quality of handling and holding yard environment, and beef cattle temperament: 1. Relationships with flight speed and fear of humans. Appl. Anim. Behav. Sci. 120, 18-27.

Puri, B. K., El-Dosoky, A., Cheema, S., Lekh, S. K., Hall, A. D., Mortimer. A. M., 1995. Parietal scalp hair whorl patterns in schizophrenia. Biol. Psychiatry. 37, 278-279.

Rao, C. R., 1964. The use and interpretation of principal component analysis in applied research. Sankhyā: The Indian Journal of Statistics. Series A, 329-358.

Randle, H. D., 1998. Facial hair whorl position and temperament in cattle. Appl. Anim. Behav. Sci. 56, 139-147.

Reinhardt, C. D., Busby, W. D., Corah, L. R., 2009. Relationship of various incoming cattle traits with feedlot performance and carcass traits. J. Anim. Sci. 87, 3030-3042.

Rueda, P. M., Sant'Anna, A. C., Valente, T. S., Paranhos da Costa, M. J. R., 2015. Impact of the temperament of Nellore cows on the quality of handling and pregnancy rates in fixed-time artificial insemination. Livest. Sci. 77, 189-195.

Sant'Anna, A. C., Paranhos da Costa, M. J. R., Baldi, F., Rueda, P. M., Albuquerque, L. G., 2012. Genetic associations between flight speed and growth traits in Nellore cattle. J. Anim. Sci. 90, 3427-3432.

Sant'Anna, A. C., Paranhos da Costa, M. J. R., Baldi, F., Albuquerque. L. G., 2013. Genetic variability for temperament indicators of Nelore cattle. J. Anim. Sci. 91, 3532-3537.

Sheridan, J. F., Dobb, C., Brown, D., Zwilling, B., 1994. Psychoneumroimmunology: Stress effects on pathogenesis and immunity during infection. Clin. Microbiol. Rev. 7, 200-212.

Shrout, P. E., Fleiss, J. L., 1979. Intraclass correlations: uses in assessing rater reliability. Psychol. Bull. 86, 420-428.

Sporer, K. R. B., Weber, P. S. D., Burton, J. L., Earley, B., Crowe, M. A., 2008. Transportation of young beef bulls alters circulating physiological parameters that may be effective biomarkers of stress. J. Anim. Sci. 86, 1325-1334.

Stahringer R. C., Randel, R. D., Neuenforff, D. A., 1989. Effect of nalexone on serum luteinizing hormone and cortisol concentration in seasonally anestrous Brahman heifers. J. Anim. Sci. 67(Suppl.1), 359(Abstr).

Stroup, W. W., 2013. Correlated Errors, Part I: Repeated Measures: From Generalized Linear Mixed Models, Modern Concepts, Methods and Applications. Taylor and Francis Group, LLC, Florida, US.

Tanner, M., Grandin, T., Cattel, M., Deesing, M., 1994. The relationship between facial hair whorls and milking parlour side preferences. J. Anim. Sci. 72(Suppl. 1), 207.

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

Turner, S. P., Lawrence, A. B., 2007. Relationship between maternal defensive aggression, fear of handling and other maternal care traits in beef cows. Livest. Sci. 106, 182-188.

Turner, S. P., Jack, M. C., Lawrence, A. B., 2013. Precalving temperament and maternal defensiveness are independent traits but precalving fear may impact calf growth. J. Anim. Sci. 91, 4417-4425.

U.S. Department of Agriculture., 2018a. *Cattle on Feed January 26, 2018*. ISSN: 1948-9080.

U.S. Department of Agriculture., 2018b. *Cattle on Feed February 23, 2018*. ISSN: 1948-9080.

Van de Water, G., Verjans, F., Greers, R., 2003. The effect of short distance transport under commercial conditions on the physiology of slaughter calves; pH and colour profiles of veal. Livest. Prod. Sci. 82, 171-179.

Vetters, M. D. D., Engle, T. E., Ahola, J. K., Grandin, T., 2013. Comparison of flight speed and exit score as measurements of temperament in beef cattle. J. Anim. Sci. 91, 374-381.

Voisinet, B. D., Grandin, T., Tatum, J. D., O'Connor, S. F., Struthers, J. J., 1997a. Feedlot cattle with calm temperaments have higher average daily gains than cattle with excitable temperaments. J. Anim. Sci. 75, 892-896.

Voisinet, B. D., Grandin, T., Tatum, J. D., O'Connor, S. F., Tatum, J. D., Deesing, M. J.,1997b. Bos indicus-cross feedlot cattle with excitable temperaments have tougher meat and a higher incidence of borderline dark cutters. Meat Sci. 46, 367-377.

Wajnberg, É., Haccou, P., 2008. Statistical Tools for Analyzing Data on Behavioral Ecology of Insect Parasitoids, in Behavioral Ecology of Insect Parasitoids: From Theoretical Approaches to Field Applications. Blackwell Publishing Ltd, Oxford, UK.

CHAPTER TWO

SUBJECTIVE METHODS TO QUANTIFY TEMPERAMENT IN BEEF CATTLE ARE INSENSITIVE TO THE NUMBER AND BIASES OF OBSERVERS

Abstract

Associations between excitable temperament and many economically relevant traits have been established, resulting in an increased focus on docility in breeding programs in cattle. Several methods have been proposed to quantify temperament during normal handling procedures, which appear to differ in their usefulness and accuracy of measure. The objective of this study was to determine the merit of using chute score (CS), exit score (ES), and exit velocity (EV) to evaluate temperament in a production setting, by assessing (i) the impact of experience and number of observers on inter-observer reliability, and (ii) the repeatability of these measures. Over three consecutive years, a factorial design of two measurement protocols [frequent (F), infrequent (IN)], and three recording periods, each one month apart, was used. Each year, twenty commercial *Bos taurus* heifers, 2-wk post weaning, were randomly assigned to each protocol. Each day of observation, heifers were moved into a squeeze chute and their heads caught. A CS from 1 (docile) to 6 (aggressive) was assigned to loosely restrained cattle by as many as six observers of varying experience. When released from the chute, EV over a distance of 2 m was obtained, and an ES of 1 (docile) to 5 (aggressive) assigned by the same observers. Heifers were evaluated repeatedly over three months, with some heifers scored on as many as nine occasions. Inter-observer reliability was calculated using the irr package in R. Kendall's coefficient of concordance (KCC) and

intra-class correlation coefficients (ICC) were calculated for the average of two, three, or four experienced observers for all nine days of the study. Observers were then divided by experience [experienced (E), inexperienced (IE)] and compared to a benchmark observer using these same statistics. Repeatabilities were calculated using ASreml fitting protocol, event, and their interaction as fixed effects, body weight as a covariate, and sire and year as random effects. While reliability decreased with increasing number of observers for both measures, agreement was consistently higher when assessing ES compared to CS. Both measures had reliabilities indicative of acceptable agreement ($ICC_{CS} \geq 0.74$, $ICC_{ES} \geq 0.90$; $KCC_{CS} \geq 0.79$, $KCC_{ES} \geq 0.93$). However, inexperience reduced reliability of CS assignments, with more variation in scores than in E observers. All methods were highly repeatable. These findings demonstrate the usefulness of subjective methods for quantifying temperament on site. Their lack of susceptibility to individual bias in distinguishing behaviors makes them useful criteria for evaluating temperament in cattle.

Introduction

Because of the negative outcomes associated with excitable temperament in cattle (Bates et al., 2014; Cooke, 2014), and its moderate heritability (Burrow and Corbet, 2000), there has been increased selection for docility. In order to quantify temperament in beef cattle, three common measurements have been proposed based on their behavior when restrained in (chute score, CS), and exiting from (exit score, ES and exit velocity, EV), the chute. The first two methods, CS and ES, are subjective in nature and require no extra equipment to collect while EV is measured as the speed in which an animal crosses two infrared beams. Parham et al. (2018) have shown that CS can be used to delineate cattle that acclimate to repeated calm handling, resulting in more proficient selection decisions. Subjective measurements are attractive due to their ease of use. However, the efficacy of selection programs depends on the repeatability of the measurement and accuracy of observation. Since these scores are based on the subjective opinion of the observer they may be less reliable than objective measures such as EV.

The accuracy of the evaluation of temperament in animals is reflected by its reliability. Higher reliabilities signify more dependable observations and are more desirable (Bokkers et al., 2012). Threshold correlation coefficients for acceptable reliability have been established, with values above 0.70 for Kendall and intra-class correlation coefficients (Martin and Bateson, 1993), and values above 0.40 for Kappa coefficients (Landis and Koch, 1977), indicating accurate evaluation. Since objective methods such as EV are not susceptible to human subjectivity, they are typically assumed to have consistent reliability (Vetters et al., 2013). For subjective evaluations, observer reliability can be estimated among a group of individuals assessing the same animal, which is referred to as inter-observer reliability. Additionally, it can be assessed for a single individual observing the same behavior repeatedly, which is referred to as intra-observer reliability. Olmos and Turner (2008) reported moderate inter- and intra-observer reliabilities when assessing CS in beef cattle, with a Kendall coefficient of 0.64, and ranges in Kappa correlation coefficients from 0.49 to 0.65. Furthermore, Vetters et al. (2013) reported weighted Kappa correlation coefficients of 0.60 for ES using a 4-point scale (Lanier and Grandin, 2002).

The consistency of temperament evaluation is reflected by the repeatability of those measurements. Repeatability can be estimated for both subjective and objective methods. Since EV is an objective measure, it is not uncommon that its repeatability is higher than that of subjective methods. For instance, MacKay et al. (2013) reported repeatabilities of 0.48 and 0.36 for EV and CS, respectively. This estimate for EV is consistent with Kadel et al. (2006) who reported the repeatability between post-weaning and start of finishing as 0.46. All of these measurements are considered moderately repeatable, making them potentially useful as selection criteria.

Subjective evaluations of behavior are not inherently bad, and may be more robust than expected. Qualitative behavior analysis (QBA) has been used in the study of animal behavior for quite some time (Stevenson-Hinde and Zunz, 1978) and has become a popular new tool to assess behavior and welfare of animals based on common terms (Grandin and Shivley, 2015). Several QBA studies report significant agreement between observers in their assessment of behavioral expression regardless of professional background. Observers repeated their assessments with high levels of accuracy in pigs (Temple et al., 2011; Wemelsfelder et al., 2012), cattle (Rousing and Wemelsfelder, 2006; Stockman et al., 2013), sheep (Wickham et al., 2012), poultry (Wemelsfelder, 2007), and many other species. Furthermore, the qualitative assessments also were correlated with physiological measures, such as heart rate, demonstrating their usefulness in quantifying behavior (Wemelsfelder, 2007).

Genetic gain occurs most rapidly with traits that can be accurately assessed repeatedly. Chute score and ES provide producers with onsite measures that can be readily collected during routine processing of cattle. However, there are concerns about their reliability and repeatability given their subjectiveness. Therefore, before they can be used confidently to evaluate temperament, their robustness must be validated. The objective of this study was to determine the merit of using CS, ES, and EV to evaluate temperament in a production setting. This was accomplished by assessing (i) the impact of experience and number of observers on inter-observer reliability, and (ii) the repeatability of these measures, when evaluated in beef heifers post-weaning.

Methods

All procedures and protocols used in this study were approved by the Institutional Animal Care and Use Committee at Virginia Tech.

Experimental animals

Within each of three consecutive years, 40 commercial *Bos taurus* (75% Angus or more) spring-born heifer calves were reared at the Virginia Tech Shenandoah Valley Agricultural Research and Extension Center (SVAREC) in Steeles Tavern, VA (37°55′56.28″ N 79°12′50.76″ W) with their respective dams until weaning (Age = 185 ± 11 d). Upon completion of a 1 wk fence line weaning period, heifers were transported to Virginia Tech Kentland farm (37°11′60″ N 80°33′52″ W) and placed in a single management group on grass. The dataset included progeny from one of 21 sires, ranging from one to 23 progeny per sire, selected to establish divergent larger and smaller frame size offspring. Cows were bred within line to Angus sires correspondingly selected as larger or smaller based on their mature cow height estimated breeding value (Vargas Jurado et al., 2015).

Table 2.1 Ethograms utilized to score animals

Chute Score[1]	Description
1. Docile	Mild disposition. Gentle and easily handle. Stands and moves slowly during processing. Undisturbed, settled, somewhat dull. Does not pull on head gate when in chute.
2. Slightly Restless	Generally docile but moves frequently and will not remained positioned for more than a few seconds; flicks tail occasionally, blows quietly through nostrils, may be stubborn but is otherwise docile.
3. Restless	Quieter than average, but may be stubborn during processing. May try to back out of chute or pull back on head gate. Some flicking of tail.
4. Nervous	Typical temperament is manageable, but nervous and impatient. A moderate amount of struggling, movement and tail flicking. Repeated pushing and pulling on head gate.
5. Flighty (Wild)	Jumpy and out of control, quivers and struggles violently. May bellow and froth at the mouth. Continuous tail flicking. Defecates and urinates during processing.
6. Aggressive	Ranges from mildly aggressive behavior, fearfulness, extreme agitation, and continuous movement which may include jumping and bellowing while in chute to thrashing about or attacking wildly when confined in small, tight places. Pronounced attack behavior.

Exit Score[2]	Description
1. Docile	Exits chute calmly.
2. Restless	Exits chute promptly.
3. Nervous	Exits chute briskly
4. Flighty (Wild)	Exits chute wildly.
5. Aggressive	Exits chute frantically.

[1] Tulloh (1961)

[2] BIF Guidelines (2002)

Experimental design

Heifers were randomly assigned to one of two measurement protocols (frequent; infrequent) within dam frame size category (larger, smaller) and sire. Data were collected across three recording periods, each 1 mo apart [i.e., event 1 (Oct.), 2 (Nov.), 3 (Dec.)]. Heifers within the frequent (F) measurement protocol were observed three consecutive days within each event while the heifers in the infrequent (IN) measurement protocol were observed only the first day. Day within event was designated by $d_{i,j}$, where *i* was the event and *j* was the day within an event.

On day one of each event (i.e., $d_{1,1}$, $d_{2,1}$, $d_{3,1}$), all 40 heifers were moved into a holding pen. Four heifers were randomly drawn from the group and placed in a small pen that opened into an alleyway. One at a time, each heifer was calmly moved through the alley into the weigh crate. Once weighed, each heifer was then moved into the squeeze chute (Priefert Model S04) where their head was secured in the head gate and the sides of the chute left opened with no restriction on the body. Furthermore, the bottom of the chute was wide enough that heifers were easily in a standing position. On a given day up to six observers, at least three of which were considered experienced, simultaneously recorded a loosely restrained CS (Table 2.1; Tulloh, 1961) within the first 15 s of the heifer being placed in the squeeze chute. Heart rate, temperature, and a fecal and jugular blood sample were then taken. Upon release, ES (Table 2.1; BIF Guidelines, 2002) was recorded by the same individuals. A flight speed (s) was also measured using electronic timers (Polaris, FarkTek, Wylie, TX) over a 2 m distance, based on the principle developed by Burrow et al. (1988), beginning 1 m from the head of the chute. This value was converted into an EV (m/s) for analysis, where larger values corresponded with more excitable heifers. The heifers were evaluated repeatedly over three months, with some heifers scored on as many as nine occasions.

On the second and third day of each event (i.e., $d_{1,2}, d_{2,2}, d_{3,2}$, $d_{1,3}, d_{2,3}, d_{3,3}$) the same measurements were again recorded on all cattle assigned to the F protocol (*n* = 20). All measurement protocols were the same. After day three of recording all 40 heifers were mixed into a single management group until the next recording period. Across the three years, two of the 120 heifers, one from F and the other IN, were removed from the study due to lameness.

Although there was overlap, the individual observers present on a given day varied. Furthermore, these observers had different levels of experience with using these ethograms. Therefore, observers were split into experienced (E; *n* = 6) and inexperienced (IE; *n* = 10) groups, depending on their level of training, and their reliability was compared in two ways. First, four observers from the E group were selected within each year and their reliability calculated between all pairs, all trios, or the four observers for CS and ES to give an idea of the robustness of these measurements.

There was a single individual who was present across all three years and scored nearly every heifer in the study. This individual was considered the most experienced observer, and thereby the benchmark for comparison. As a second evaluation approach, 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 IE), depending on the experience level of the second person.

Statistical analysis

Summary statistics. Subjective scores were averaged across three E observers to obtain an overall score for each heifer on a given day. Those CS, ES, and EV were then summarized for all heifers on $d_{1,1}$, $d_{2,1}$, and $d_{3,1}$, and separately for the F protocol on all nine days of observation.

Although the data were small, repeatability estimates were obtained using ASreml (Gilmour et al., 2009) including all progeny in the dataset from one of nine artificial insemination (AI) sires (n = 105), with 3 to 23 progeny per sire. The remaining heifers were sired by non-AI bulls and removed from this analysis. Three of the nine AI sires were used in multiple years. A sire model was fitted for all heifers on the first day of each event. Therefore, event, measurement protocol, and their interaction were fitted as fixed effects, with body weight as a covariate, and year, sire and residual error as random effects. Dam size was shown to not define significant variation in any of the measures considered (Parham et al., 2018), and therefore was not included in the model fitted. Pedigree information consisted of all sires and their sire ancestors back through three generations.

Inter-observer reliabilities. All reliability calculations were carried out using the irr package (Gamer et al., 2012) in R. Reliability of each subjective measurement was calculated using percent of agreement, either a weighted Cohen's or Fleiss' kappa coefficient, Kendall's coefficient of concordance, and the intra-class correlation. Percent of agreement (PA) was calculated as the number of observations in agreement divided by the total number of observations and multiplied by 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. A tolerance of one allowed for slight discrepancy among individuals, providing an estimate of the extent of very similar yet not exact agreement among observers.

To account for chance agreements, and to be able to compare reliabilities across studies, a kappa coefficient was also used. Reliability between pairs of observers was calculated using weighted Cohen's Kappa (K_W; Cohen, 1960). Since only applicable when comparing two observers, and since there were more than two individuals in some comparisons, Fleiss' kappa coefficient (K) was used when comparing increasing numbers of E observers (Fleiss, 1971). Kappa coefficients vary from -1 to 1 with negative values indicating agreement that is poorer than chance, zero indicating exactly chance agreement, and positive values indicating agreement that is better than chance (Fleiss and Cohen, 1973). A value of K greater than 0.40 represents moderate agreement, and K greater than 0.60 represents substantial agreement (Landis and Koch, 1977).

Intra-class correlation (ICC) describes how strongly observations of the same recording period resembled each other (Shrout and Fleiss, 1979). An ICC of 0 represents no agreement among observers, and an ICC of 1 represents perfect agreement. Typically, an ICC of 0.70 or greater is considered to reflect strong concordance and thereby a reliable evaluation (Martin and Bateson, 1993).

Kendall's coefficient of concordance (KCC), which ranges from 0 (no agreement) to 1 (complete agreement), was calculated to measure agreement as an index of reliability of ordinal data. With a subjective scale ranging from one to five, ultimately some heifers will be given the same score. The KCC, therefore, was corrected based on the number of groups of heifers assigned the same score by an observer. This adjustment accounted for the number of heifers within a grouping. The value of this correction factor was then summed across observers (Kendall, 1948). As group sizes increase, the value of KCC would decrease.

Results

Summary statistics. Descriptive statistics for average CS, ES and EV on $d_{1,1}$, $d_{2,1}$, and $d_{3,1}$ for all heifers, and for the F heifers on all nine days of observation, are given in Table 2.2. The data summarized is for the three experienced observers. For all heifers, on average there was a decrease of 0.34 ± 0.12, 0.14 ± 0.12, and 0.35 ± 0.14 m/s between $d_{1,1}$ and $d_{3,1}$ for CS, ES, and EV, respectively. There was a larger decrease between $d_{1,1}$ and $d_{3,3}$ of 0.85 ± 0.16 and 0.49 ± 0.20 m/s for CS and EV, respectively, with less change in ES (0.30 ± 0.17), for the F heifers. The maximum CS assigned in F also decreased from a 5.00 on $d_{1,1}$ to 3.67 on $d_{3,1}$. A significant reduction in CS over time was found in F heifers, and when combined with IN heifers, in the most temperamental animals (Parham et al., 2018).

All three measures were highly repeatable. Estimates were largest for ES (0.88 ± 0.20), followed by EV (0.68 ± 0.21) and CS (0.62 ± 0.24).

Table 2.2. Descriptive statistics of chute score, exit score and exit velocity for all animals by day.

	Chute Score			Exit Score			Exit Velocity (m/s)			
Day[1]	Mean	SE	Maximum[4]	Mean	SE	Maximum[4]	Mean	SE	Minimum	Maximum
n = 118[2]										
$d_{1,1}$	2.30	0.08	5.00	1.99	0.08	4.00	2.51	0.11	0.52	7.60
$d_{2,1}$	1.95	0.08	4.67	1.97	0.08	4.33	2.36	0.12	0.43	8.97
$d_{3,1}$	1.96	0.08	4.33	1.85	0.08	4.33	2.16	0.08	0.43	4.77
n = 59[3]										
$d_{1,1}$	2.32	0.11	5.00	1.95	0.12	4.00	2.47	0.18	0.52	7.60
$d_{1,2}$	2.50	0.13	4.67	1.90	0.11	4.00	2.38	0.12	0.47	5.10
$d_{1,3}$	2.06	0.10	4.00	1.72	0.10	4.00	2.36	0.15	0.76	5.83
$d_{2,1}$	1.84	0.11	4.33	1.92	0.12	4.33	2.22	0.14	0.43	5.78
$d_{2,2}$	1.86	0.12	4.33	1.86	0.11	4.00	2.35	0.14	0.83	5.83
$d_{2,3}$	1.64	0.10	4.00	1.84	0.13	4.33	2.39	0.17	0.26	8.30
$d_{3,1}$	1.82	0.11	4.00	1.87	0.13	4.33	2.10	0.13	0.43	4.02
$d_{3,2}$	1.79	0.12	4.00	1.72	0.12	4.33	2.01	0.12	0.58	4.22
$d_{3,3}$	1.47	0.09	3.67	1.65	0.11	4.33	1.98	0.13	0.24	4.88

[1] Day is designated by $d_{i,j}$, where i is the event and j is the day within an event
[2] All frequently and infrequently handled heifers combined on a given day
[3] Values for frequently handled heifers on a given day
[4] Minimum values consistently equal to 1.00

Table 2.3. Inter-observer reliabilities of chute score and exit score for increasing number of experienced observers.

		PA[2]				ICC[5]		
Observers	n[1]	Tol = 0	Tol = 1	K[3]	KCC[4]	Value	Lower	Upper
Chute Score								
2	436	64	96	0.48	0.87	0.75	0.70	0.79
3	320	47	92	0.46	0.82	0.74	0.70	0.78
4	213	37	87	0.45	0.79	0.74	0.69	0.78
Exit Score								
2	440	83	99	0.74	0.95	0.89	0.87	0.91
3	327	74	99	0.73	0.93	0.90	0.88	0.91
4	223	68	99	0.73	0.93	0.90	0.88	0.92

[1] The total number of observations utilized

[2] Percent Agreement; Tol = 0: all observers had to give the same score, Tol = 1: all observers had to be within one score of each other

[3] All Fleiss' Kappa coefficients were different from 0 ($P < 0.001$)

[4] Kendall's Coefficient of Concordance

[5] Intra-class Correlation; Lower = Lower bound of the 95% CI for ICC, Upper = Upper bound of the 95% CI for ICC

Inter-observer reliability. The number of E observers ranged between three and four on a given day. Therefore, the average reliability between groups of two, three, or all four observers is given in Table 2.3. Overall, PA with a tolerance of zero decreased as the number of observers increased for both CS and ES, as did K and KCC. When tolerance was set to one the PA were higher, as expected; still the PA decreased with an increased number of observers for CS. Reliabilities for ES were consistently higher than CS. However, KCC and ICC values were above 0.70, and K was above 0.41, for both subjective measures regardless of the number of E observers present.

Comparison of subjective measures assigned by E and IE observers are given in Table 2.4. Reliabilities of ES are again consistently higher than CS. Furthermore, there was no difference in reliability estimates due to experience level for ES. While values of K_W were above 0.41, and KCC and ICC were close to 0.70, for IE observers when evaluating CS, the ICC confidence interval was much wider than that for E observers, with a lower bound of 0.39. To investigate this occurrence, means and SE of the difference in CS and ES comparing E and IE to the benchmark observer were calculated. For ES, the differences were -0.056 ± 0.024 and -0.038 ± 0.061 for E and IE, respectively; for CS those difference were 0.020 ± 0.041 and -0.042 ± 0.103 for E and IE, respectively.

Table 2.4. Inter-observer reliability of experienced and inexperienced observers for chute and exit score.

Experience		PA[2]				ICC[5]		
Level	*n*[1]	Tol = 0	Tol = 1	K_W[3]	KCC[4]	Value	Lower	Upper
Chute Score								
Experienced	294	63.3	95.8	0.60	0.87	0.73	0.67	0.78
Inexperienced	42	57.7	95.2	0.54	0.86	0.64	0.39	0.82
Exit Score								
Experienced	296	82.4	99.8	0.81	0.95	0.89	0.86	0.91
Inexperienced	42	82.4	99.9	0.81	0.95	0.89	0.78	0.94

[1] The total number of observations utilized
[2] Percent Agreement; Tol = 0: all observers had to give the same score, Tol = 1: all observers had to be within one score of each other
[3] Weighted Kappa correlation coefficient
[4] Kendall's Coefficient of Concordance
[5] Intra-class Correlation; Lower = Lower bound of the 95% CI for ICC, Upper = Upper bound of the 95% CI for ICC

Discussion

In this study, subjective methods of evaluating cattle behavior were reliably measured, regardless of experience and the number of observers, and highly repeatable. While reliability decreased with increasing number of observers for both CS and ES, agreement in the scores assigned remained robust. Additionally, inexperience only had a negative impact on CS, with more variation in scoring than with experienced observers.

Together these findings demonstrate the usefulness of subjective methods for quantifying temperament in a production setting, and their lack of susceptibility to individual bias in distinguishing among behaviors.

All traits were highly repeatable, meaning there was no large fluctuation in the temperament of heifers over time: all either equally acclimated to handling or remained constant in their behavior. For ES and EV there was minimal change in mean score (Table 2.2). Average ES and EV on $d_{1,1}$ were already low enough that there was little opportunity for them to reduce further. Parham et al. (2018) found no significant change over time for ES or EV, regardless of how frequently heifers were handled. However, they did show a significant decrease in CS in the group of heifers handled more frequently. As expected, heifers with higher CS on the first day of observation decreased most substantially between $d_{1,1}$ and $d_{3,3}$, with the average CS of all heifers being lower than 2.0 on $d_{3,3}$. This is reflected in Table 2.2 with the mean CS on $d_{3,3}$ being 1.47 (SD 0.68) and the maximum score assigned reducing from 5.00 to 3.67. This reduction in score may have contributed to the slight decrease in repeatability for CS compared to ES. However, because CS fell substantially during the first recording period ($d_{1,1}$ to $d_{3,1}$) and remained constant thereafter, repeatability of the measure was not unduly affected. Furthermore, since CS of all heifers fell from their starting values on $d_{1,1}$ their rankings for temperament did not substantially change resulting in the high repeatability observed.

Not only was the repeatability of ES higher than CS, but the inter-observer reliabilities were consistently higher as well. This could reflect two issues. First is the scoring systems themselves (Table 2.1). 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. The second possibility relates to the lack of change in ES over time. The repetition of observing and scoring the same behavior over time may result in greater consistency of evaluation. Therefore, a combination of simplicity and practice may have caused the difference observed in the reliabilities of CS and ES.

Regardless of the number of E individuals, both scoring methods had KCC and ICC higher than 0.70 and K higher than 0.41, indicating good consistency among observers when scoring the same heifer. With multiple observers, greater subtlety in the evaluation may be captured by taking the mean of their scores. 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. Allowing some differences among observers in their subjective evaluation of behavior is perhaps beneficial. In the current study, as the number of E observers was increased, PA with tolerance of one decreased; however, that decrease was far less than for a PA with tolerance of zero. Furthermore, ICC were equal no matter how many E observers were considered. Allowing some discrepancy in scores among trained evaluators may therefore allow for a more equitable assessment of temperament. However, increasing the number of individuals assigning scores does increase labor required and cost. Therefore, producers may wish to consider both alternatives – one or several individuals assigning scores – when deciding how to assess temperament in a production setting.

The difference in reliability between CS and ES was most clearly seen when comparing E to IE observers. For ES, the IE observer reliabilities were identical to the E, and both were again higher than reliabilities of even the E observers for CS. The SE were larger, indicating a larger amount of variation in scoring, when comparing IE to E for both methods. However, CS had the larger SE as reflected in the confidence interval reported for ICC. This larger discrepancy in reliability of CS compared to ES could be explained when comparing mean values of E and IE to the benchmark observer. Both E and IE were somewhat lower in their ES evaluations compared to the benchmark, while E observers slightly over estimated and IE underestimated the CS of heifers. Experience in using the CS ethogram clearly is more important to the accuracy of evaluation than for ES, which is more robust to inexperience.

The high repeatabilities and reliabilities for these subjective methods, especially with ES, may be substantiated by the concept of QBA. While the use of a predesigned ethogram does not explicitly adhere to Free Choice Profiling (FCP) methods (Wemelsfelder et al., 2001) where individuals create their own descriptors of behavior, they are still selecting from a list of subjective adjectives, and doing so coincides with the second step of FCP. The main difference is that ethograms are ranked from least to most severe response. Regardless, recent QBA have successfully demonstrated that despite differences in background and experience, acceptable agreement can be reached using subjective descriptions of behavior (Wemelsfelder et al., 2012). That conclusion can also be supported by these data.

Many studies have promoted the use of other forms of measuring temperament that are less subjective. Vetters et al. (2013) support the use of ES as a method to quantify temperament due to its high correlation with EV. However, the ES used in Vetters et al. (2013) was that of Lanier and Grandin (2002), which utilizes "gaits" (i.e., walk, trot, run, jump) instead of adjectives. While this methodology provides a less subjective system of quantifying an animal's behavior while exiting a chute, QBA and results from this study suggest human observers are equally able to differentiate temperaments using both types of ethograms. Vetters et al. (2013) reported K_W for ES of 0.60, while K_W for experienced individuals in this study was 0.81. While both methods resulted in substantial agreement, the more subjective ethogram for ES utilized in this study (BIF Guidelines, 2002) resulted in higher agreement among observers than that of Lanier and Grandin (2002).

Referring specifically to the subjectiveness of the ethogram, CS appears more difficult to delineate than either system for scoring ES. Regardless, experienced observers in this study were just as consistent as those recording ES in Vetters et al. (2013), with K_W of 0.60. Grandin (2014) warned that when CS were being used, it is important to know how tightly the animal was restrained to prevent movement. Furthermore, possible loss in variation of CS due to restricted movement from catching the head of the animal in the head gate and/or using the squeeze feature of the side panels has been mentioned as a deterrent for using CS to quantify behavior (Vetters et al., 2013). Animals in this study were considered to be loosely restrained. They had their heads caught in the head gate but the squeeze feature was not utilized. They were also easily able to stand upright without the chute being narrowed at the bottom. This loose restraint may explain the higher reliabilities and apparent usefulness of the ethogram compared to other studies. None-the-less, the findings of Parham et al. (2018), coupled with the high reliabilities and repeatabilities of CS observed in this study, suggest CS provides a useful evaluation of temperament in a production setting.

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, which is further substantiated by QBA. The use of EV is objective and is not predisposed to observer error, but it does require the purchase of equipment. Exit velocity and ES behaved similarly and were strongly correlated (0.81 ± 0.03, Parham et al., 2018). Since ES had such high reliability and repeatability, it may be the more pragmatic measure of temperament for an animal exiting the chute. Together these findings demonstrate the value of CS and ES to quantify temperament in a production setting. They are repeatable and can be reliably measured, regardless of previous experience and the number of individuals assigning scores. These attributes make these measures useful criteria for evaluating temperament in cattle.

Literature Cited

Bates, K. E., Weaber, R. L., Bormann, J. M., Moser, D. W., Salak-Johnson, J. L., Chase, C. C. L., Peel, R. K., Van Campen, H., Loneragan, G. H., Wagner, J. J., Bodhireddy, P., Prayaga, K., Enns. R. M., 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., de Vries, M., Antonissen, I. C. M. A., de Boer, I. J. M., 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., Corbet, N. J., 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.

Burrow, H. M., Seifert, G. W., Corbet, N. J., 1988. A new technique for measuring temperament in cattle. Proc. Aust. Soc. Anim. Prod. 17, 154-157.

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.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Meas. 20, 37-46.

Fleiss, J. L., 1971. Measuring nominal scale agreement among many raters. Psychol. Bull. 76, 378-382.

Fleiss, J. L., Cohen, J., 1973. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. Educ. Psychol. Meas. 33, 613-619.

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

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

Grandin, T., 2014. Handling Facilities and Restraint of Extensively Raised Range Cattle. Livestock Handling and Transport, 4th Edition. T. Grandin, ed. CAP Int., Cambridge, MA. Pages 94-115.

Grandin, T., Shivley, C., 2015. How farm animals react and perceive stressful situations such as handling, restraint, and transport. Animals. 5, 1233-1251.

Kadel, M. J., Johnston, D. J., Burrow, H. M., Graser, H., Ferguson, D. M., 2006. Genetics of flight time and other measures of temperament and their value as selection criteria for improving meat quality traits in tropically adapted breeds of beef cattle. Aust. J. Agric. Res. 57, 1029-1035.

Kendall, M. G., 1948. Rank correlation methods. London: Charles Griffin. 1948, 160.

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

Lanier, J. L., Grandin, T., 2002. The relationship between *Bos Taurus* feedlot cattle temperament and cannon bone measurements. Proc. West. Sect. Am. Soc. Anim. Sci. 80(Suppl. 2), 105.

MacKay, J. R. D., Turner, S. P., Hyslop, J., Deag, J. M., Haskell, M. K., 2013. Short-term temperament tests in beef cattle relate to long-term measures of behavior recorded in the home pen. J. Anim. Sci. 91, 4917-4924.

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

Olmos, G., Turner, S. P., 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.

Parham, J. T., Tanner, A. E., Barkley, K., Pullen, L., Wahlberg, M. L. Swecker, Jr., W. S., Lewis, R. M., 2018. Temperamental cattle acclimate more substantially to repeated handling. Appl. Anim. Behav. Sci. (Submitted 12 July 2018).

Rousing, T., Wemelsfelder, F., 2006. Qualitative assessment of social behavior of dairy cows housed in loose housing systems. Appl. Anim. Behav. Sci. 101, 40-53.

Shrout, P. E., Fleiss, J. L., 1979. Intraclass correlations: uses in assessing rater reliability. Psychol. Bull. 86, 420-428.

Stevenson-Hinde J., Zunz, M., 1978. Subjective assessment of individual rhesus monkeys. Primates. 19, 473-482.

Stockman, C. A., Collins, T., Barnes, A. L., Miller, D., Wickham, S. L., Beatty, D. T., Blache, D., Wemelsfelder, F., Fleming, P. A., 2013. Flooring and driving conditions during road transport influence the behavioral expression of cattle. Appl. Anim. Behav. Sci. 143, 18-30.

Temple, D., Manteca, X., Velarde, A., Dalmau, A., 2011. Assessment of animal welfare through behavioral parameters in Iberian pigs in intensive and extensive conditions. Appl. Anim. Behav. Sci. 131, 29-39.

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

Vargas Jurado, N., Tanner, A. E., Blevins, S. R., Rich, J., Mayes, R. W., Fiske, D., Swecker Jr., W. S., Lewis, R. M., 2015. Feed intake and diet selection in Angus-cross heifers of two frame sizes at two stages of growth. J. Anim. Sci. 93, 1565-1572.

Vetters, M. D. D., Engle, T. E., Ahola, J. K., Grandin, T., 2013. Comparison of flight speed and exit score as measurements of temperament in beef cattle. J. Anim. Sci. 91, 374-381.

Wemelsfelder, F., 2007. How animals communicate quality of life: the qualitative assessment of behavior. Anim. Welf. 16, 25-31.

Wemelsfelder, F., Hunter, A. E., Paul, E. S., Lawrence, A. B., 2012. Assessing pig body language: agreement and consistency between pig farmers, veterinarians, and animal activists. J. Anim. Sci. 90, 3652-3665.

Wickham, S. L., Collins, T., Barnes, A. L., Miller, D., Beatty, D. T., Stockman, C., Blache, C., Wemelsfelder, F., Fleming, P. A., 2012. Qualitative Behavioural

Assessment of transport-naive and transport-habituated sheep. J. Anim. Sci. 90, 4523-4535.

CHAPTER THREE

TEMPERAMENTAL CATTLE ACCLIMATE MORE SUBSTANTIALLY TO REPEATED HANDLING

Abstract

Temperament of cattle impacts efficiency of production systems, including animal well-being. In being heritable, temperament can be augmented through selection. Current methods to evaluate temperament in a production setting include chute score (CS), exit score (ES), and exit velocity (EV), which some producers may find cumbersome to navigate. Even those who utilize these methods may not do so efficiently if initial evaluations are not strong indicators of future temperament. The objectives of this study were to determine whether these behavioral scores or values change under repeated and routine management, and to estimate the relationships among them. Over three consecutive years, a factorial design of two measurement protocols [frequent (F), infrequent (IN)], and three recording periods, each 1 mo apart, was used. The F measurements were collected over three consecutive days, and IN measurements only on day one within a recording period. Each year, twenty commercial *Bos taurus* heifers, 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 three observers. Exit velocity over a 2 m distance 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 day one. For heifers assigned to

F, event and day within event were instead fitted. For both models, body weight was included as a covariate, with sire and year fitted as random effects. Pearson correlations among measurements were calculated on day one of each recording period separately, and for days combined. Chute score decreased across events and days in F. Heifers with higher CS on day one had the largest reduction in score. Exit scores and EV changed less over time and were highly correlated, characterizing the same behavior. Correlations between CS and ES or EV during the first recording period were close to zero, but increased as CS decreased. Chute score therefore may be more indicative of acclimation to a novel environment than ES or EV. Both CS and ES appear to offer an easy and inexpensive way to quantify temperament in cattle. Heifers became calmer with repeated gentle handling. Producers may avoid unnecessarily culling cattle based strictly on initial response to novel stimuli by allowing acclimation to handling before assessing docility.

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 (Voisinet et al., 1997a; Bates et al., 2014), decreased tolerance to disease (Sheridan et al., 1994; Cooke, 2014), decreased reproductive performance (Von Borell et al., 2007; Cooke, 2014), and increased production costs (Burrow and Corbet, 2000). Because of these negative outcomes associated with excitable temperament in cattle, and its moderate heritability (Burrow and Corbet, 2000; Sant'Anna et al., 2013), there has been increased selection for docility.

In order to effectively select for docility in cattle it is imperative that each criteria measured are indicative of behavior during normal handling practices (Fordyce et al., 1982). The three most common measurements proposed to quantify temperament are based on behavior when cattle are restrained in (chute score, CS) and exiting from (exit score, ES; exit velocity, EV) the chute. The first two methods, CS and ES, require one or more individuals to assign a specific score, often 1 to 5 or 6 based on their perception of an animal's behavior. Exit velocity records the time it takes an animal to cross two infrared beams, using specialized equipment. Although cattle can be classified using these methods on a continuum, they can also be used to categorize more extreme animals as docile or excitable.

Because subjective scoring methods such as CS and ES require no extra equipment, they are easy and inexpensive to collect. Objective measures such as EV are not subject to observer biases and are recorded on a continuous scale, resulting in more accurate estimates with higher heritabilities (Hoppe et al., 2010); however, they require extra time to set up, and the purchase of equipment. Curley et al. (2006) found that behavior assessed with subjective methodologies does not correspond to temperament as well as an objective score such as EV. Nevertheless, Voisinet et al. (1997a; 1997b) were able to sort cattle into temperament categories based on CS with significant effects on meat quality endpoints. Furthermore, pragmatically, time is always critical when handling cattle. Measuring a plethora of behaviors may become cumbersome and further agitate animals. Researchers have proposed that ES and EV are so highly correlated that they measure the same behavior (Lanier and Grandin, 2002; Vetters et al., 2013) and therefore both may not be needed. Understanding the relationship among these measurements will not only increase the understanding of behavior, but may minimize observations necessary to select for improved temperament.

Measures of temperament in animals should be fast, easy, and inexpensive to collect. However, before being used as selection criteria, the effect of repeated handling on the scores or values of these measures, and the relationships among them, must be understood. Curley et al. (2006) found that temperament score decreased as the number of times an animal went through a working facility increased. Cattle that are successfully acclimated to handling may more willingly re-enter a handling facility. However, the animal's response is not only dependent on genetic predisposition, but its previous experiences (Grandin and Shivley, 2015). Walking young calves through a handling system, with no invasive procedures, resulted in adult animals that were calmer (Becker and Lobato, 1997), while attempts to acclimate older cows was less successful (Cooke, 2014). Both the timing and severity of experience play a role in acclimation. Animals will not habituate to extremely adverse experiences (Grandin et al., 1986). Therefore, if unavoidable, it is better to wait until later in life to expose an animal to harsher conditions. Furthermore, if temperament changes over time, an animal may need to be observed repeatedly to be accurately assessed.

Producers who are concerned with docility may find it cumbersome to choose among the methods that have been proposed to quantify behavior, and therefore may be deterred from using them to make selection decisions. Even those producers who utilize these methods to select for temperament may not be doing so efficiently if cattle acclimate to handing. Understanding how these methods interrelate could provide clarity and increase the accuracy of these selection decisions. Therefore, the objectives of this study were to determine whether behavioral scores or values change under repeated and routine management, and to estimate the relationships among them.

Methods

All procedures and protocols used in this study were approved and performed in compliance with the Institutional Animal Care and Use Committee at Virginia Tech.

Experimental animals

Within each of three consecutive years, 40 commercial *Bos taurus* (75% Angus or more) spring-born heifer calves were reared at the Virginia Tech Shenandoah Valley Agricultural Research and Extension Center (SVAREC) in Steeles Tavern, VA (37°55′56″ N 79°12′50″ W) with their respective dams until weaning (185 ± 11 d in age). Upon completion of a 1-wk fence line weaning period, heifers were transported to Virginia Tech Kentland farm (37°11′60″ N 80°33′52″ W) and placed in a single management group on grass. The dataset included progeny from one of 21 sires, ranging from one to 23 progeny per sire, selected to establish divergent larger and smaller frame size offspring. Cows were bred within line to Angus sires correspondingly selected as larger or smaller based on their mature cow height estimated breeding value (Vargas Jurado et al., 2015).

Behavioral observations

Two different subjective methods were utilized to describe the behavior of each heifer. For CS, the approach of Tulloh (1961) was adapted for use.

1. Docile: Mild disposition. Gentle and easily handle. Stands and moves slowly during processing. Undisturbed, settled, somewhat dull. Does not pull on head gate when in chute.

2. Slightly Restless: Generally docile but moves frequently and will not remain stationary for more than a few seconds; flicks tail occasionally, blows quietly through nostrils, may be stubborn but is otherwise docile.

3. Restless: Quieter than average, but may be stubborn during processing. May try to back out of chute or pull back on head gate. Some flicking of tail.

4. Nervous: Typical temperament is manageable, but nervous and impatient. A moderate amount of struggling, movement and tail flicking. Repeated pushing and pulling on head gate.

5. Flighty (Wild): Jumpy and out of control, quivers and struggles violently. May bellow and froth at the mouth. Continuous tail flicking. Defecates and urinates during processing.

6. Aggressive: Ranges from mildly aggressive behavior, fearfulness, extreme agitation, and continuous movement, which may include jumping and bellowing while in chute to thrashing about or attacking wildly when confined in small, tight places. Pronounced attack behavior.

For ES, the method as described in the BIF Guidelines (2002) was used. This entailed:

1. Docile: Exits chute calmly.

2. Restless: Exits chute promptly.

3. Nervous: Exits chute briskly.

4. Flighty (Wild): Exits chute wildly.

5. Aggressive: Exits chute frantically.

Experimental design

Heifers were randomly assigned to one of two measurement protocols (frequent; infrequent) within their dam's frame size category (larger, smaller) and sire. Data were collected across three recording periods, each 1 mo apart [i.e., event 1 (Oct.), 2 (Nov.), 3 (Dec.)] starting on the second Monday or Tuesday of October, and repeated at four week intervals. Heifers within the frequent (F) measurement protocol were observed three consecutive days within each event (month) while the heifers in the infrequent (IN) measurement protocol were evaluated on only the first day of an event. Day within event was designated by $d_{i,j}$ where i was the event and j was the day within an event.

On day one of each event (i.e., $d_{1,1}$, $d_{2,1}$, $d_{3,1}$) every year, all 40 heifers were moved into a holding pen. Four heifers were randomly drawn from the group and herded into the cattle-handling facility, which consisted of a small holding pen narrowing into a curved alley that led to a weigh crate and separate squeeze chute. One at a time, each heifer was calmly moved through the alley into the weigh crate. Once weighed, each heifer was then moved into the squeeze chute (Priefert Model S04) where their head was secured in the head gate and the sides of the chute left opened with no restriction on the body. Furthermore, the bottom of the chute was wide enough that heifers were easily in a standing position. On a given day, three experienced observers simultaneously recorded a loosely restrained CS within the first 15 s of the heifer being placed in the squeeze chute. Heart rate, rectal temperature, and a fecal and jugular blood sample were then taken.

Upon release, ES was recorded by the same individuals. A flight speed (s) also was measured using electronic timers (Polaris, FarkTek, Wylie, TX) over a 2 m distance, based on the principle developed by Burrow et al. (1988), beginning 1 m from the head of the chute. This value was converted into an EV (m/s) for analysis, where larger values corresponded with more excitable heifers.

Each year, on day two and three of each event (i.e., $d_{1,2}, d_{2,2}, d_{3,2}$, $d_{1,3}, d_{2,3}, d_{3,3}$) the same measurements were again recorded on all heifers assigned to the F protocol (*n* = 20) with the exception of a blood sample on day two. However, researchers still simulated a "mock" blood sample for consistency in experience. After day three of recording all 40 heifers were mixed into a single management group until the next recording period.

Across the three years, two of the 120 heifers, one from F and the other IN, were removed from the study due to lameness.

Statistical analyses

Chute score and ES taken by the observers were averaged on each day to obtain a representative score. Based on examination of residuals, the distribution of these data appeared skewed. A natural logarithm was applied with the transformed values tested for normality using the Jarque-Bera (Skewness-Kurtosis) Test (Jarque and Bera, 1980). The log transformed data were more normal with less skewness and kurtosis. Therefore, lognormal transformed average CS and ES were analyzed using the GLIMMIX procedure in SAS (SAS Inst. Inc., Cary, NC) fitting two separate models.

To compare the effect of measurement protocol on temperament, initially a 2x2x3 factorial model was analyzed fitting protocol (F and IN), dam frame size (larger or smaller) and event (1, 2, or 3), and their two and three way interactions, as fixed effects. Year, sire, and heifer nested within the combination of year, measurement protocol and dam frame size, were treated as random effects. Comparisons were only made on the first day of each event ($d_{1,1}$, $d_{2,1}$, $d_{3,1}$). Dam frame size never explained significant variation in the response variables (P > 0.18) and thus was excluded from the model. The final mixed model fitted was:

$$y_{ijklm} = \mu + B_i + Mj + (BM)_{ij} + E_k + (BE)_{ik} + ME_{jk} + (BME)_{ijk} + W_{mik} + S_l + C(BM)_{m(ij)} \quad (1)$$

where y_{ijklm} was the response variable for heifer *m* (*m* = 1 to 20), M_j and E_k were the fixed effects of measurement protocol (*j* = F, IN) and event (*k* = 1, 2, 3), respectively. Random effects included birth year (B_i; *i* = 2008, 2009, 2010), the two way interactions of birth year with measurement protocol (BM_{ij}) and event (BE_{ik}), the three way interaction of birth year, measurement protocol, and event (BME_{ijk}), sire (S_l; *l* = 1 to 21), and the nested effect of the m^{th} heifer within the combination of year and measurement protocol [$C(BM)_{m(ij)}$]. Finally, W_{mik} was the covariate of body weight for an animal at a single event. Least squares means and SE were obtained using SAS with Tukey's adjustment for multiple comparisons. In order to express the results of CS and ES on their scale of measurement, means and SE were back-transformed to the observed scale here and in later analyses.

To measure changes in response variables over time within F, a separate model was fitted. Event, dam frame size, their interaction, and the nested effect of day within event, were fitted as fixed effects. Heifer nested within year and dam frame size combination, as well as year and sire, were treated as random effects. Again, dam frame size did not significantly explain variation in the response variables (P > 0.15) and was removed from the final mixed model fitted:

$$y_{iklmp} = \mu + B_i + E_k + (BE)_{ik} + D(E)_{p(k)} + BD(E)_{ip(k)} + W_{mk} + S_l + C(B)_{m(i)} \quad (2)$$

where all terms are defined as in model (1), with the addition of $D(E)_{p(k)}$, the nested fixed effect of the p^{th} day (p = 1 to 3) within the k^{th} event. Added random effects were the interaction of birth year *i* by day *p* nested within event *k* [$BD(E)_{ip(k)}$], and the nested effect of heifer *m* within birth year *i* [$C(B)_{m(i)}$]. Least squares means and SE were obtained using SAS with Tukey's adjustment for multiple comparisons.

A repeated measures analysis was initially conducted when fitting model 1 using PROC GLIMMIX in SAS. Treating the same response variables (CS, ES, and EV) as repeated measures on each heifer, an unstructured covariance matrix was fitted for repeated events. When evaluating the F measured heifers, days within an event also could be considered repeated measures. Therefore model 2 was fitted using an unstructured and autoregressive direct product covariance matrix for event and day, respectively, with PROC MIXED in SAS. Results from these analyses were consistent with the initial models described. Furthermore, model 2, which had the more complicated hierarchical structure, failed to converge in some cases when using repeated measures analysis. Therefore, results obtained from the factorial model are reported henceforth.

Phenotypic Pearson and Spearman rank correlations were calculated between CS, ES, and EV on $d_{1,1}$, $d_{2,1}$, and $d_{3,1}$ separately, and for all days together, in R (R Core

Team, 2013). Correlations were first calculated for F and IN separately with similar results. Therefore, both groups were combined. Confidence intervals for Pearson correlations were calculated using the r.con function in R, and used to test for differences in correlations between $d_{1,1}$, $d_{2,1}$, and $d_{3,1}$.

Data partitions

Average CS and ES for each heifer were split into threshold categories according to $d_{1,1}$ measurements. Heifers with average CS and ES that were both greater than or equal to 2.5 (*n* = 21) were considered temperamental, while heifers with both scores less than 2.5 (*n* = 54) were considered docile. This left some heifers that fell in neither category (*n* = 43), and were therefore excluded from this analysis. As a second comparison, treatment categories were created based on average CS on $d_{1,1}$. These categories consisted of heifers with CS greater than or equal to 3.0 (*n* = 27), between 2.5 and 3.0 (*n* = 21), between 2.0 and 2.5 (*n* = 27), between 1.5 and 2.0 (*n* = 24), and less than 1.5 (*n* = 19). The two models described previously were fitted within each of these two data partitions, firstly to estimate change within the F and IN measurement protocols on the first day of an event, and secondly to estimate change across days for F heifers only.

Table 3.1 Change in values across events for both frequently and infrequently handled heifers together.

Measure	*n*	Day Within Event[1]			SEM
		$d_{1,1}$	$d_{2,1}$	$d_{3,1}$	
Chute Score	118	2.27	1.89	1.87	0.07
Exit Score	118	1.84	1.86	1.76	0.04
Exit Velocity (m/s)	117	2.37	2.27	2.15	0.07

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

Results

Subjective measurements - full dataset

Mean values for all heifers on the first day of each event are given in Table 3.1. There was no effect of event, measurement protocol, or their interaction on CS, ES, or EV ($P > 0.16$). Numerically all measures decreased across events, suggesting increased docility, with the reduction in CS largest. After the first event, F had numerically lower CS compared to IN, with average CS on $d_{3,1}$ of 1.75 ± 0.07 and 2.00 ± 0.08, respectfully, compared to 2.33 ± 0.10 and 2.21 ± 0.09 on $d_{1,1}$. There were no differences in ES between the measurement protocols, with averages on $d_{1,1}$ and $d_{3,1}$ being 1.80 ± 0.06 and 1.75 ± 0.09 for F, respectively, while 1.88 ± 0.06 and 1.75 ± 0.06 for IN, respectively. Lastly, F had consistently slower EV than IN, although not significantly so.

When assessing change in measurements across days for F, there was a decrease in CS across both events ($P < 0.05$) and days ($P < 0.09$). As shown in Figure 3.1a, CS decreased ($P = 0.06$) from event 1 to 2, but did not change between event 2 and 3. Figure 3.1b shows the change in CS across days, with a numerical increase between $d_{1,1}$ and $d_{1,2}$. Following $d_{1,2}$, CS numerically decreased over time to a score of 1.38 ± 0.06 on $d_{3,3}$. The CS at $d_{3,3}$ was less than $d_{1,1}$ and $d_{1,2}$ ($P < 0.05$). These trends were not observed in EV or ES (Figure 3.1c and 3.1d).

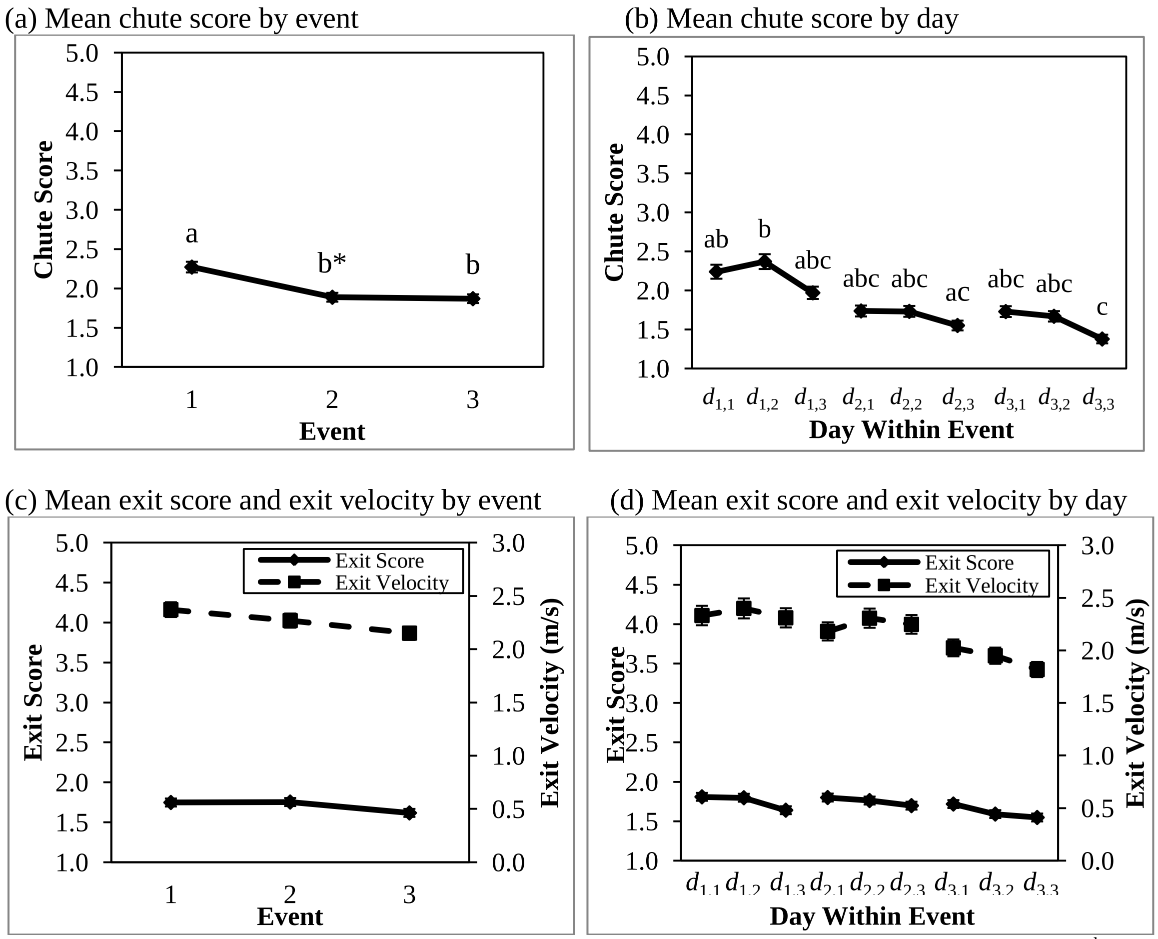


Figure 3.1. Mean values of frequently handled heifers, separated by event and day. [[a,b] Values with different letters are significantly different ($P < 0.05$); *Difference in values only tended towards significance ($P < 0.10$)]

Subjective measurements - data partitions

Within the temperamental and docile categories no effect of event, measurement protocol, or their interaction on any measurements was observed ($P > 0.17$). Among temperamental heifers, the average values of CS, ES, and EV across all events, regardless of measurement protocol, were 2.88 ± 0.18, 2.98 ± 0.14, and 3.36 ± 0.17 m/s, respectively. For docile heifers, those values were 1.65 ± 0.06, 1.54 ± 0.05, and 1.92 ± 0.08 m/s, respectively. Although the changes were not significant, there was a larger numerical decrease in CS and EV as compared to ES in the temperamental heifers from

$d_{1,1}$ to $d_{3,1}$. Within the docile category, there was no such change in scores. Heifers that began with lower temperament scores had less room to decrease.

Table 3.2. Effect of event on values of temperamental and docile heifers handled frequently.

		Event			
Measure	Threshold Category	1	2	3	SEM
Chute Score	Temperamental[1]	3.15	2.47	2.44	0.25
	Docile[2]	1.76	1.41	1.34	0.09
Exit Score	Temperamental[1]	3.01	2.88	2.88	0.16
	Docile[2]	1.45	1.48	1.33	0.06
Exit Velocity	Temperamental[1]	3.69	3.59	3.07	0.24
	Docile[2]	1.88	1.86	1.64	0.10

[1]Temperamental heifers are those with both chute and exit scores ≥ 2.5 (*n* = 10)
[2]Docile heifers are those with both chute and exit scores < 2.5 (*n* = 30)

While the effect of measurement protocol was not significant, there was a more substantial decrease in CS over time when considering F only. Table 3.2 shows the change in values across events for temperamental versus docile heifers in F. There were no significant changes in values across events in either category of heifer. However, in general there was a stronger numerical decrease in CS and EV as compared to ES, especially among temperamental heifers, but the small number of animals (*n* = 10) resulted in larger SE.

Change in CS by day for temperamental and docile heifers in F are provided in Figure 3.2. Chute score of the docile heifers essentially remained unchanged over time, but with a tendency to decrease in temperamental heifers (*P* = 0.07). Once again there was a slight numerical increase in CS from $d_{1,1}$ to $d_{1,2}$, although this was the case in both groups of animals. This increase was temporary, as CS of temperamental heifers decreased on $d_{1,3}$ and essentially remained unchanged thereafter. Final CS on $d_{3,3}$ was

2.01 ± 0.16, and less than initial CS of 3.42 ± 0.28 and 3.51 ± 0.28 on $d_{1,1}$ to $d_{1,2}$, respectively (P < 0.05), in the temperamental group.

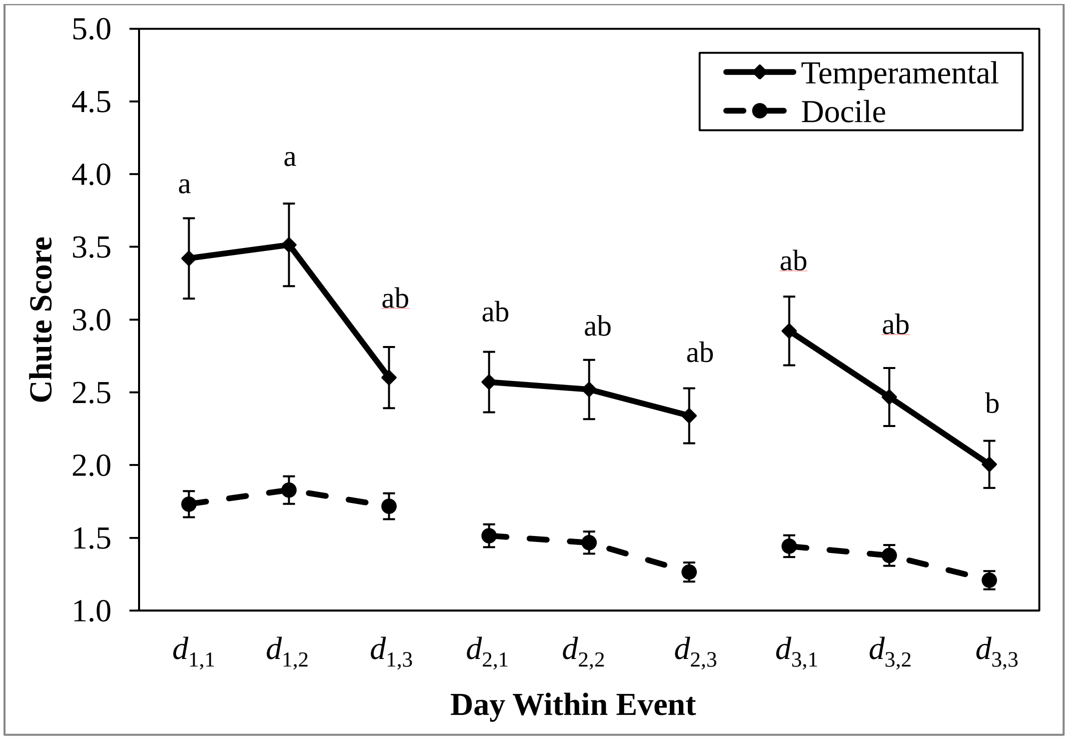


Figure 3.2. Change in chute score across days for temperamental and docile heifers handled frequently (P = 0.07). [[a,b] Means with differing superscripts differ (P < 0.05)]

Since CS appeared to be the more sensitive measure, heifers were delineated based on their average score on $d_{1,1}$. Once again, no significant variation in any measurement was defined by protocol across events. Regardless of protocol, average CS of all heifers on the first day of an event therefore are provided in Table 3.3. Chute scores decreased from $d_{1,1}$ to $d_{2,1}$ for the highest threshold category (CS ≥ 3.0; P < 0.05), with no further change from $d_{2,1}$ to $d_{3,1}$. Numerical decreases from $d_{1,1}$ to $d_{2,1}$ were smaller for all other categories where the CS on $d_{1,1}$ were lower.

Table 3.3. Effect of event on chute score of both frequently and infrequently handled heifers.

		Day Within Event[1]			
Chute Score	n	$d_{1,1}$	$d_{2,1}$	$d_{3,1}$	SEM
≥ 3.0	27	4.06[a]	2.54[b]	2.49[b]	0.25
2.5 < 3.0	21	2.76	1.90	1.98	0.17
2.0 < 2.5	27	2.24	1.85	1.94	0.14
1.5 < 2.0	24	1.62	1.62	1.67	0.11
1.0 < 1.5	19	1.24	1.41	1.34	0.07

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

[a,b] Means in a row with differing superscripts differ ($P < 0.05$)

The F protocol was again considered separately using model 2. Chute scores decreased across events for all heifers ($P < 0.07$), except for the category with CS less than 1.5, which remained unchanged. Heifers with starting scores greater than or equal to 3.0 decreased in score from $d_{1,1}$ to $d_{3,1}$ by 0.92 ($P < 0.05$). Heifers with CS between 2.0 and 2.5 and from 2.5 to 3.0 decreased by 0.51 ($P = 0.09$) and 1.00 ($P = 0.07$), respectfully; the small number of heifers ($n = 9$) with CS ranging from 2.5 to 3.0 coincided with a larger SE; thus, although the decline in score was larger it only tended toward significance.

Change in CS across days in F with CS greater than 2.0 are shown in Figure 3.3. The CS decreased the most substantially in the highest threshold group, with the largest decrease of 1.69 from $d_{1,1}$ to $d_{1,3}$ ($P < 0.05$), with no further change from $d_{2,1}$ to $d_{3,3}$. Overall, there was a decrease in CS within each event, with little or no change in CS in the month time span between events. On the final day of the study, regardless of CS on $d_{1,1}$, each category on average had a CS less than 2.0.

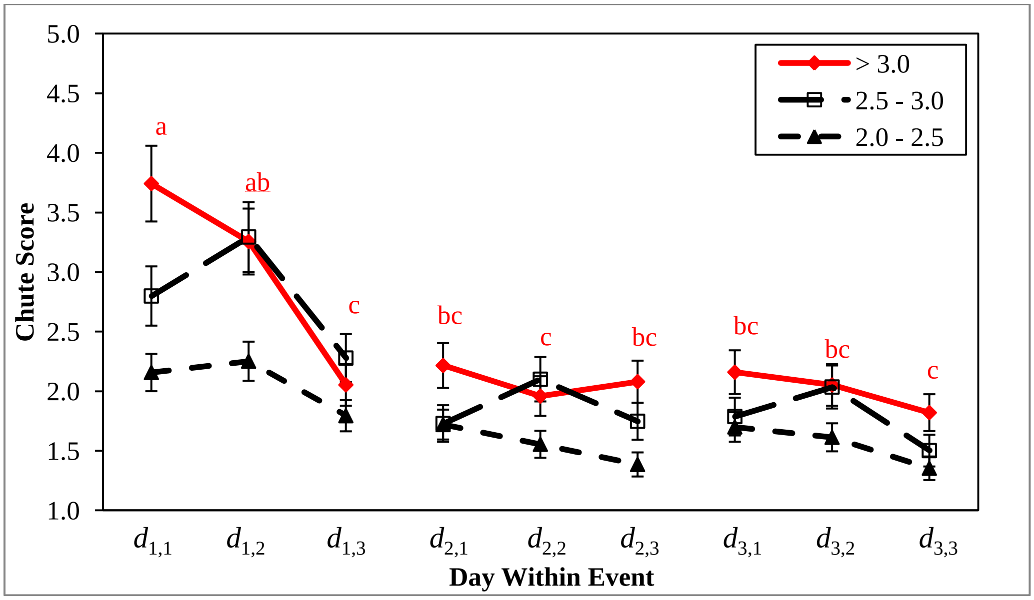


Figure 3.3. Average chute scores by day for frequently handled heifers, separated by chute score on the first day of handling. [*Effect of day within event $P < 0.05$]

Correlations

Pearson and Spearman rank correlations between CS, ES, and EV were similar for all heifers on $d_{1,1}$, $d_{2,1}$, and $d_{3,1}$ as well as all three days combined; therefore, only Pearson correlations are shown in Table 3.4. Correlations between ES and EV did not differ between days and were consistently higher than either measurement with CS, ranging from 0.76 ± 0.06 on $d_{1,1}$ to 0.87 ± 0.05 on $d_{3,1}$. The overall correlations of CS with ES and EV were moderately positive, while the correlation between ES and EV was highly positive. These correlations increased over time, with correlations on $d_{1,1}$ being much lower than those on $d_{3,1}$ ($P < 0.05$) for CS with ES and EV. Correlations between CS and ES or EV were not different from zero on $d_{1,1}$. However, on $d_{3,1}$ these correlations became moderately positive, ranging from 0.43 ± 0.08 to 0.47 ± 0.08.

Table 3.4. Pearson correlation coefficients between measurements on the first day of each event, and for all three days combined.

	[1]$d_{1,1}$		$d_{2,1}$		$d_{3,1}$		All[2]	
	CS	ES	CS	ES	CS	ES	CS	ES
ES	0.14 ± 0.09*[a]		0.30 ± 0.09[a]		0.47 ± 0.08[b]		0.31 ± 0.05	
EV	0.08 ± 0.09*[a]	0.76 ± 0.06[c]	0.28 ± 0.09[b]	0.83 ± 0.05[c]	0.43 ± 0.08[b]	0.87 ± 0.05[c]	0.27 ± 0.05	0.81 ± 0.03

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

[2] Correlations of all 3 days combined cannot be compared to $d_{1,1}$, $d_{2,1}$, or $d_{3,1}$.

[a,b] Means in a row with differing superscripts differ ($P < 0.05$)

* Correlation not different from zero ($P > 0.05$)

Discussion

Heifers in this study appeared to acclimate to repeated calm handling in the chute, particularly in the initially more temperamental cattle. Their ES and EV changed less over time. When assessed at the first management event at monthly intervals, the frequency of prior handling (F versus IN) did not affect CS. However, with repeated handling over consecutive days, heifers acclimated to their restraint in a chute. These results were consistent across different data partitions. An animal's reaction the first time released from a chute may not be a good indicator of their temperament when restrained in the chute, as correlations of CS with ES and EV were near zero on $d_{1,1}$. However, when allowed time to acclimate to handling, these correlations increased to moderate values. Conversely, phenotypic correlations between ES and EV were consistently high. Therefore, these methods may be measuring similar if not the same behaviors. Both CS and ES are commonly used subjective measurements of temperament while restrained in or exiting the chute. However, when cattle are excitable during their first handling experience, more than one observation of temperament may be beneficial before making selection decisions.

When considering both F and IN, there was no difference in ES or EV across events, regardless of heifers' characterization for temperament. This same pattern also was observed across events and days when considering only F heifers. Exit score and EV were very closely related: ES provides a subjective assessment of how cattle exit a chute while EV objectively measures speed of exit. Because the subjective evaluation of exit behavior is relatively straightforward and without need of specialized equipment, ES therefore seems the preferable metric in a production setting. Researchers have suggested ES as an easy and inexpensive replacement for EV (Lanier and Grandin, 2002; Vetters et al., 2013). This conclusion is supported by the high correlations among the two traits in this study.

Numerically, ES and EV decreased over time, and more so in temperamental heifers. This result is in contrast with Petherick et al. (2002) and Müller and Keyserlingk (2006) who reported an increase in EV over time. Their observation could have been the result of stressful or over handling of the animals resulting in increased fearfulness. Average ES and EV for all heifers on $d_{1,1}$ were 1.84 ± 0.04 and 2.37 ± 0.07 m/s, respectively, with median values even lower. Therefore, if acclimation to handling was possible, these already low starting values for ES and EV may make this more difficult to detect. Regardless, increased handling did not result in an increase in excitability for either measurement in this study.

Docile heifers becoming more docile seems less interesting than temperamental heifers that calm down over time; that is, a heifer with a CS and ES of 2.0 or less is likely already of acceptable temperament during handling since she is “generally docile” and exits the chute “promptly”. This led to our delineation between temperamental and docile groups. While no change in ES or EV was detected for temperamental heifers, such was not the case for CS.

Heifers acclimated to handling in the chute, with CS decreasing both across events and days, particularly in frequently handled and the initially more temperamental heifers. Those with CS greater than or equal to 3.0, regardless of measurement protocol, had a decreased CS from event 1 to 2. Overall, in the more frequently handled heifers, CS also decreased between the first two events. This decrease also persisted numerically for temperamental heifers from event 1 to 2, although the change was not significant. No further change in score occurred with time thereafter for these groups. This acclimation was most clearly seen across days in the frequently handled heifers, with consistent differences between $d_{1,1}$ and $d_{3,3}$ for all heifers assigned to this treatment. In temperamental heifers specifically, and those with CS greater than or equal to 3.0 on $d_{1,1}$, CS had the largest decrease during the first event, and remained relatively constant during the remaining days of the study. Most importantly, when heifers were broken into categories based on their CS on $d_{1,1}$ each CS category on average had a score of less than 2.0 on $d_{3,3}$. That score is indicative of only a slightly restless or docile heifer.

Acclimation to handling did occur for heifers when restrained in the chute, but not necessarily when exiting. On $d_{1,1}$ the correlations or relationships between temperaments in and when exiting the chute were close to zero. This aligned with average CS and ES on $d_{1,1}$ of 2.27 ± 0.07 and 1.84 ± 0.04, respectively. Heifers appeared to be more temperamental when in the chute than when exiting. While both methods measure the fearfulness of cattle due to restraint, these scoring methods are not measuring the same behaviors. As heifers were given time to acclimate to repeated handling, the correlations of CS with ES and EV increased, with overall correlations being moderate. Average CS and ES on $d_{2,1}$ were similar (1.89 ± 0.07 and 1.86 ± 0.04, respectively), reflecting this increase in relationship. Furthermore, there was a substantial increase in the covariance between CS with ES and EV over time, with little change in their variances. Essentially, heifers became as calm in the chute as they appeared while exiting.

An animal that is successfully acclimated to handling may more willingly re-enter a handling facility. The animal's response therefore is not only dependent on its inherited temperament but on its previous experiences (Grandin and Shivley, 2015). Animals will not habituate to extremely severe treatments. In a study conducted by Grandin et al. (1986), when given the choice between a commercial electro-immobilizer and squeeze-tilt table, 94% of ewes chose the squeeze-tilt table one or more times after being restrained by it. Similarly, cattle that had been accidently knocked in the head with the head gate during restraint were more likely to avoid the squeeze chute in a choice test (Grandin et al., 1994).

Heifers in this study were handled in a calm manner at a young age and were only exposed to routine handling. The most invasive procedure they experienced was blood being drawn from the jugular vein. While this study provides evidence of acclimation to multiple handling incidents, temperament is defined as an animal's reaction to novel experiences. In general, the more times an animal is handled, especially in a calm atmosphere, the less novel the incident becomes. In this study, substantial effort was put forth to ensure each handling event was identical, perhaps removing some, if not all, of the novelty. Based on this definition, the argument could be made that an animal's initial reaction to handling is the most reliable estimate of their temperament, and selection decisions therefore should be made based on that observation. If an animal's temperament is the primary trait of interest in a breeding program, or is completely unmanageable, then culling of that animal may be justified. However, allowing acclimation to handling may be of value when individual animals have borderline acceptable temperament, especially when balanced with other breeding objectives. Cattle that excel in all other aspects of a producers breeding goal may benefit from additional observations of temperament before final selection decisions are made.

Conclusion

In conclusion, temperament in cattle is a very popular selection criterion due to its impact on growth, carcass quality, as well as human and animal well-being. Selection for docility requires a measurement that is accurate, inexpensive and relatively easy to collect on-farm or on-ranch. The results of this experiment support the use of CS and ES as subjective methods to quantify temperament in cattle. Although EV is objectively recorded and thereby not subject to observer error, it requires the purchase of equipment. Since EV and ES are highly correlated, ES may be the more pragmatic measure for producers of temperament when cattle exit the chute. Importantly, heifers appear to acclimate to handling in a calm environment. Particularly in the more temperamental heifers, after just a few days, their CS decreased substantially and remained relatively constant thereafter. When cattle are excitable during their first handling experience, more than one observation of temperament may be beneficial before assessing docility. This may avoid culling an animal based strictly on its initial response to novel stimuli.

Literature Cited

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

Becker, B. G, Lobato, J. E. P., 1997. Effect of gentle handling on the reactivity of zebu crossbred calves to humans. Appl. Anim. Behav. Sci. 53, 219-224.

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

Burrow, H. M., Seifert, G. W., Corbet, N. J., 1988. A new technique for measuring temperament in cattle. Proc. of Aust. Soc. of Anim. Prod. 17, 154-157.

Burrow, H. M., Corbet, N. J., 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., Paschal, J. C., Welsh Jr., T. H., Randel, R. D., 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., Goddard, M. E., Seifert, G. W., 1982. The measurement of temperament in cattle and the effect of experience and genotype. Proc. Aust. Soc. Anim. Prod. 14, 329-332.

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

Grandin, T., Shivley, C., 2015. How farm animals react and perceive stressful situations such as handling, restraint, and transport. Animals. 5, 1233-1251.

Grandin, T., Curtis, S. E., Widowski, T. M., Thurmon, J. S., 1986. Electro-immobilization versus mechanical restraint in an avoid-avoid choice test for ewes. J. Anim. Sci. 62, 1469-1480.

Grandin, T., Odde, K. G., Schutz, D. N., Behrns, L. M., 1994. The reluctance of cattle to change a learned choice may confound preference tests. Appl. Anim. Behav. Sci. 39, 21-28.

Hoppe, S., Brandt, H. R., König, S., Erhardt, G., Gauly, M., 2010. Temperament traits of beef calves measured under field conditions and their relationships to performance. J. Anim. Sci. 88, 1982-1989.

Jarque, C. M., Bera, A. K., 1980. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. Economics Letters. 6, 255-259.

Lanier, J. L., Grandin, T., 2002. The relationship between *Bos Taurus* feedlot cattle temperament and cannon bone measurements. Proc. West. Sect. Am. Soc. Anim. Sci. 80(Suppl. 2), 105.

Müller, R., von Keyserlingk, M. A. G., 2006. Consistency of flight speed and its correlation to productivity and to personality in Bos taurus beef cattle. Appl. Anim. Behav. Sci. 99, 193-204.

Petherick, J. C., Holroyd, R. G., Doogan, V. J., Venus, B. K., 2002. Productivity, carcass and meat quality of lot-fed Bos indicus cross steers grouped according to temperament. Aust. J. Exp. Agric. 42, 389-398.

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

Sant'Anna, A. C., Paranhos da Costa, M. J. R., Baldi, F., Albuquerque., L. G., 2013. Genetic variability for temperament indicators of Nelore cattle. J. Anim. Sci. 91, 3532-3537.

Sheridan, J. F., Dobb, C., Brown, D., Zwilling, and B., 1994. Psychoneumroimmunology: Stress effects on pathogenesis and immunity during infection. Clin. Microbiol. Rev. 7, 200-212.

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

Vargas Jurado, N., Tanner, A. E., Blevins, S. R., Rich, J., Mayes, R. W., Fiske, D., Swecker Jr., W. S., Lewis, R. M., 2015. Feed intake and diet selection in Angus-cross heifers of two frame sizes at two stages of growth. J. Anim. Sci. 93, 1565-1572.

Vetters, M. D. D., Engle, T. E., Ahola, J. K., Grandin, T., 2013. Comparison of flight speed and exit score as measurements of temperament in beef cattle. J. Anim. Sci. 91, 374-381.

Voisinet, B. D., Grandin, T., Tatum, J. D., O'Connor, S. F., Struthers, J. J., 1997a. Feedlot cattle with calm temperaments have higher average daily gains than cattle with excitable temperaments. J. Anim. Sci. 75, 892-896.

Voisinet, B. D., Grandin, T., Tatum, J. D., O'Connor, S. F., Tatum, J. D., Deesing, M. J., 1997b. Bos indicus-cross feedlot cattle with excitable temperaments have tougher meat and a higher incidence of borderline dark cutters. Meat Sci. 46, 367-377.

Von Borell, E., Dobson, H., Prunier, A., 2007. Stress, behaviour and reproductive performance in female cattle and pigs. Horm. Behav. 5, 130-138.

CHAPTER FOUR

SUBJECTIVE METHODS OF QUANTIFYING TEMPERAMENT IN HEIFERS ARE INDICATIVE OF PHYSIOLOGICAL STRESS

Abstract

In order to effectively select for docility, it is imperative that selection criteria are truly indicative of an animal's response to stress. Subjective measurements such as chute (CS) and exit score (ES) provide fast and easy assessments of temperament, but are subject to observer bias and inconsistency and it is unclear whether they are true indicators of stress. Cortisol and other metabolites have been associated with increased stress in cattle, especially in regards to transportation. The objectives of this study were to identify relationships between objective and subjective methods of evaluating temperament, and determine whether objective measures of stress change under repeated and routine management. Over three consecutive years, a factorial design of two measurement protocols [frequent (F), infrequent (IN)], and three recording periods, each 1 mo apart, was used. The F measurements were collected over three consecutive days, and IN measurements only on day one within a recording period. Each year, 20 commercial *Bos taurus* (75% Angus) heifers, 2-wk post weaning, were randomly assigned to each protocol. While restrained, behavior was measured subjectively using a CS, along with body temperature, heart rate, and a fecal and blood sample. These samples were then analyzed for levels of fecal and serum cortisol, glucose, creatine kinase, non-esterified fatty acids, and blood urea nitrogen. Upon release, an ES and exit velocity were also collected. Objective data were analyzed with ANOVA using SAS. For all heifers, protocol, event, and their interaction, were compared on day one. For heifers assigned to F, event and day within event were instead fitted. For both models, body weight, ambient temperature, and wind speed were included as covariates, with sire and year fitted as random effects. Pearson correlations among measurements were calculated for the first day of each recording period combined. Chute and ES, as well as exit velocity, were moderately and positively correlated (r = 0.24 to 0.33) with body temperature, heart rate, glucose concentration, and serum cortisol levels. The only metabolite with significant decrease in concentration over time was creatine kinase, following a large spike on the third day of collection. More temperamental heifers had numerically higher body temperatures and heart rates, and serum concentrations of glucose and cortisol. Heifers that appeared more reactive in the chute and while exiting, did appear more stressed metabolically, making CS and ES acceptable measures of stress during routine handling practices and acceptable methods of selecting for temperament in beef cattle.

Introduction

Temperament in cattle has a documented impact on growth, carcass quality, reproduction, and well-being (Fordyce et al., 1985; Von Borell et al., 2007; Bates et al., 2014; Cooke, 2014). These negative outcomes, and its moderate heritability (Sant'Anna et al., 2013), have led to an increase in selection for docility in cattle. While subjective methods such as chute score (CS) and exit score (ES) provide quick, easy, and inexpensive methods in which to quantify temperament during normal handling practices, in order to effectively select for docility it is imperative that these criteria are truly indicative of an animal's response to stress.

Subjective measurements of temperament are scored on a categorical scale and are therefore subject to observer bias and inconsistency. Furthermore, Curley et al. (2006) suggested that temperament score decreased as the number of times an animal went through a working facility increased, meaning that initial assessments of temperament may not be representative of future performance. Both instances could lead to inaccurate assessments of temperament, and incorrect selection decisions. Previous work by Parham et al. (2018a, 2018b) has shown that while both CS and ES can be reliably assessed by numerous observers, heifers acclimate to handling when restrained in a chute.

Comparatively, objective measurements result in more accurate estimates with higher heritabilities (Hoppe et al., 2010). Objective measurements have been shown to increase during stressful handling of cattle. Lay et al. (1992) concluded that dairy cattle exposed to hot iron branding had higher heart rates and levels of plasma cortisol compared to those freeze branded or handled, but not hot branded. Extensive research has also been conducted on the stress associated with truck transportation in beef cattle. Higher levels of plasma cortisol, urea, and creatine kinase have been found after transport (Van de Water et al., 2003; Sporer et al., 2008) as well as a higher heart rate while loading and unloading the truck (Van de Water et al., 2003). While most of these metabolites increased shortly after the onset of transportation, creatine kinase was at its peak 24 h after the start of transportation, indicating cattle may have experienced muscle damage and physical stress (Sporer et al., 2008).

While elevated concentrations of cortisol are a hallmark of stress in livestock species (Grandin, 1997; Möstl and Palme, 2002), it is impractical for producers to take blood samples on each of their animals in order to select for temperament. Ideally, strong relationships would exist between objective and subjective measurements of temperament, making it more convenient for producers to accurately select for more docile cattle. Cortisol, heart rate, and body temperature are believed to be moderately correlated with chute score (Cooke et al., 2009, Gruber et al., 2010), while a post-transportation assessment of behavior showed no correlation with cortisol or glucose levels (r = -0.06 and 0.09, respectively), but a moderate correlation with creatine kinase (Gruber et al., 2010). It was also suggested that these relationships decreased with decreasing excitability (Curley et al., 2006; King et al., 2006).

Producers who are concerned with temperament are unlikely to make selection decisions based on objective measurements such as cortisol level and heart rate in their animals. Subjective methods provide an option that is fast, inexpensive, and can be implemented during normal handling procedures. However, to be successful, they have to delineate cattle that are more easily stressed than others. Most literature assess the relationship among these measures in regards to transportation or pre-slaughter stress, but do not evaluate their relationship with typical handling practices. Parham et al. (2018b) show that cattle acclimate to repeated handling, and suggest that selection decisions be made after animals had time to adjust to new working facilities, but only use subjective measures of temperament. Therefore, the objectives of this study were to identify relationships between objective and subjective methods of evaluating temperament, and determine whether objective measures of stress change under repeated and routine management.

Methods

All procedures and protocols used in this study were approved and performed in compliance with the Institutional Animal Care and Use Committee at Virginia Tech.

Experimental animals

Within each of three consecutive years, 40 commercial *Bos taurus* (75% Angus or more) spring-born heifer calves were reared at the Virginia Tech Shenandoah Valley Agricultural Research and Extension Center (SVAREC) in Steeles Tavern, VA (37°55′56″ N 79°12′50″ W) with their respective dams until weaning (185 ± 11 d in age). Upon completion of a 1-wk fence line weaning period, heifers were transported to Virginia Tech Kentland farm (37°11′60″ N 80°33′52″ W) and placed in a single management group on grass. The dataset included progeny from one of 21 sires, ranging from one to 23 progeny per sire, selected to establish divergent larger and smaller frame size offspring. Cows were bred within line to Angus sires correspondingly selected as larger or smaller based on their mature cow height estimated breeding value (Vargas Jurado et al., 2015).

Experimental design and data collection

Heifers were randomly assigned to one of two measurement protocols (frequent; infrequent) within their dam's frame score category (larger, smaller) and sire. Data were collected across three recording periods, each 1 mo apart [i.e., event 1 (Oct.), 2 (Nov.), 3 (Dec.)]. Heifers within the frequent (F) measurement protocol were observed three consecutive days within each event (month) while the heifers in the infrequent (IN) measurement protocol were evaluated on only the first day of an event. Day within event was designated by $d_{i,j}$ where i was the event and j was the day within an event. Two heifers, one from F and the other IN, were removed from the study due to lameness issues.

On day one of each event (i.e., $d_{1,1}$, $d_{2,1}$, $d_{3,1}$) all 40 heifers were moved into a holding pen. Four heifers were randomly drawn from the group and herded into the cattle-handling facility. One at a time, each heifer was calmly moved through the alley into the weigh crate. Once weighed, each heifer was then moved into the squeeze chute (Priefert Model S04) where their head was secured in the head gate and the side bars left open. On a given day, a chute score (CS), exit score (ES), and exit velocity (EV, m/s) were recorded for each heifer as described by Parham et al. (2018b). While restrained in the chute several measurements were collected. A heart rate and body temperature were recorded, and a fecal sample collected in pre-labeled aluminum foil, which was snap frozen in liquid nitrogen and stored at -20˚C until analyzed. Additionally, after the heifer was haltered and its head secured, blood samples were taken from the jugular vein into lithium-heparinized evacuated and plain serum blood tubes. These samples were stored on ice until centrifugation at 2,000 x g for 10 minutes at 4˚C (within five to 24 h of collection). Serum was then stored at -20˚C until analyzed.

On day two and three of each event (i.e., $d_{1,2}$, $d_{2,2}$, $d_{3,2}$, $d_{1,3}$, $d_{2,3}$, $d_{3,3}$) the same measurements were again recorded on all heifers assigned to the F protocol (n = 20 each year) with the exception of a blood sample on day two. However, researchers still simulated a "mock" blood sample for consistency in experience. After day three of recording all 40 heifers were mixed into a single management group until the next recording period.

Weather information

Weather conditions at Kentland farm were recorded using WeatherSTEM technology. Current conditions were archived each hour including ambient temperature (˚C), soil temperature (˚C), wind speed (km/h), and wind direction (deg.). These files were merged with existing data, rounding the time each heifer was placed in the chute to the nearest hour.

Physiology

Fecal analysis. Frozen fecal samples were thawed at room temperature and weighed into 0.5 g samples and mixed with 5 mL of 80% methanol. Samples were then centrifuged for 15 min at 2,500 x *g*, and the supernatant removed and stored at -20˚C. Samples obtained were analyzed in duplicate for cortisol metabolites using the Siemens Coat-A-Count Insulin Radioimmunoassay (Siemens Medical Solutions Diagnostics, Los Angeles, CA) according to manufacturer's instructions.

Laboratory analytical error was assessed as the ratio of the absolute value of the difference between duplicates and their mean. If that value exceeded 0.10 (or 10%), the analyses were repeated until the standard was achieved. Such also was the case for the serum analyses.

Serum analysis. Serum cortisol concentrations were measured in duplicate using the Siemens Coat-A-Count Insulin Radioimmunoassay (Siemens Medical Solutions Diagnostics, Los Angeles, CA). Concentrations of the serum chemistry metabolites of blood urea nitrogen (BUN), creatine kinase (CK), and glucose were measured in duplicate using the QuantiChrom Urea Assay, EnzyChrom Creatine Kinase Assay, and QuantiChrom Glucose Assay Kits (BioAssay Systems, Hayward, CA), respectively, using a clear bottom 96-well plate and plate reader according to manufacturer's instructions. Specifically, in regards to CK, if calculated activity was higher than 300 u/l, the sample was diluted in 0.9% saline and repeated, per manufacturer's instructions. Serum concentrations of non-esterified fatty acids (NEFA) were determined using the Wako HR series NEFA-HR(2) microtiter assay (Wako Diagnostics, Richmond, VA).

Statistical analyses

To compare the effect of measurement protocol, initially a 2x2x3 factorial model (model 1) was analyzed fitting protocol (F and IN), dam frame size (larger or smaller), event (1, 2, or 3), and their two and three way interactions, as fixed effects. Year, sire, and heifer nested within the combination of year, measurement protocol and dam frame size were treated as random effects. Since some objective metabolites may be impacted by the environment in which they were taken, the ambient temperature and wind speed for each day in one hour intervals were included as covariates.

Comparisons of F and IN were only made on the first day of each event ($d_{1,1}$, $d_{2,1}$, $d_{3,1}$). Dam frame size never defined significant variation in the response variables ($P > 0.16$) and thus was excluded from the model. However, body weight was instead included as an additional covariate. The final model fitted therefore included protocol, event, and their interaction as fixed effects. Year, sire, and heifer nested within the combination of year and measurement protocol, were treated as random effects. While all three covariates were not significant for all response variables, they have biological implications on the response variables and were included to ensure consistency in the model fitted. Change in metabolites over time was analyzed using the GLIMMIX procedure in SAS (SAS Inst. Inc., Cary, NC). Least squares means and SE were obtained using SAS with Tukey's adjustment for multiple comparisons.

To measure changes in response variables over time within F, a separate model (model 2) was fitted. Event, dam frame size, their interaction, and the nested effect of day within event, were fitted as fixed effects. Heifer nested within year and dam frame size combination, as well as year and sire, were treated as random effects. Again, dam frame size did not define significant variation in the response variables ($P > 0.13$) and was removed from the final model fitted. Covariates were identical to those used in model 1. Least squares means and SE were obtained using SAS with Tukey's adjustment for multiple comparisons.

A repeated measures analysis was initially conducted as in Parham et al. (2018b). Results from these analyses were consistent with the initial models described. Therefore, results obtained from the simpler factorial models are reported henceforth.

Data partitions. Based on conclusions drawn from Parham et al. (2018b), data were partitioned based on average CS on $d_{1,1}$. These categories consisted of heifers with CS greater than or equal to 3.0 (n = 27), between 2.5 and 3.0 (n = 21), between 2.0 and 2.5 (n = 27), between 1.5 and 2.0 (n = 24), and less than 1.5 (n = 19). The two models described previously were fitted, firstly to estimate change within the F and IN measurement protocols on the first day of an event, and secondly to estimate change across days for F heifers only.

Relationship between measurements. Pearson correlations were calculated between CS, ES, and EV with objective measurements of stress for all heifers on the first day of each event ($d_{1,1}$, $d_{2,1}$, and $d_{3,1}$) in R (R Core Team, 2013). Correlations were first calculated for F and IN separately with similar results between measurement protocols. Therefore, both groups were combined.

To better visualize the relationships among the recorded data, principal component analyses (PCA) were conducted to detect relationships between CS, ES, and EV with weather patterns as well as objective measurements using a combination of the "stats" and "factoextra" packages in R.

Results

Table 4.1. Change in objective measurements for frequently and infrequently handled heifers (n = 118) on the first day of each event

	Day Within Event			
Measurement	$d_{1,1}$	$d_{2,1}$	$d_{3,1}$	SEM
Body Temperature, ˚C	39.17	39.24	39.59	0.13
Heart Rate, bpm	131.93	124.11	128.39	5.10
BUN[1], mg/dl	39.35	35.47	27.61	4.78
CK[1], units/l	16.83[a]	10.28[b]	10.53[ab]	1.85
Glucose, mg/dl	119.30	119.33	118.90	16.75
NEFA[1], mmol/l	0.33	0.35	0.37	0.05
Serum Cortisol, ng/ml	38.27	46.28	50.98	4.38
Fecal Cortisol, ng/ml	10.72	11.56	12.33	3.15

[1] BUN = Blood Urea Nitrogen; CK = Creatine Kinase; NEFA = Non-esterified Fatty Acids

[a,b] Days within a row with differing superscripts differ ($P < 0.05$)

Objective measurements - full dataset

Objective measurements for all heifers on the first day of each event are provided in Table 4.1. There was no effect of measurement protocol, event, or their interaction on any of the measurements ($P > 0.14$), with the exception of CK, which decreased from the first to second event ($P < 0.05$). Numerically, heart rate, BUN, and glucose levels decreased between $d_{1,1}$ and $d_{3,1}$ while body temperature, NEFA, serum cortisol, and fecal cortisol increased.

When assessing change in measurements across days for F, there was no significant change in the metabolites ($P > 0.12$) with the exception of CK. The concentration of CK decreased from 35.12 ± 5.68 units/l for event 1 to 15.00 ± 5.68 units/l on event 2 ($P < 0.05$) and 13.15 ± 5.68 units/l on event 3 ($P < 0.05$), with no difference between event 2 and 3 ($P = 0.92$). These differences are further delineated in Table 4.2, showing change in each measurement across days. When only considering $d_{1,1}$ and $d_{3,1}$ within the F heifers, again there was a numerical decrease in HR, BUN, CK, and glucose between days, with an increase in serum cortisol. However, body temperature and NEFA slightly increased while fecal cortisol levels actually decreased between $d_{1,1}$ and $d_{3,1}$. When comparing the first and last days of observation, there was a numerical decrease in heart rate, BUN, serum cortisol, and fecal cortisol. While there was no significant difference in CK between $d_{1,1}$ and $d_{3,3}$, CK differed among days ($P < 0.05$) because of a substantial increase in its levels on $d_{1,3}$, which was greater than all other days of collection. Interestingly, CK levels consistently increased between the first and last day of an event. While not significant, the same pattern was true for NEFA.

Table 4.2. Change in objective measurements for frequently handled heifers (n = 59) by day

Measurement	Day									SEM
	$d_{1,1}$	$d_{1,2}$	$d_{1,3}$	$d_{2,1}$	$d_{2,2}$	$d_{2,3}$	$d_{3,1}$	$d_{3,2}$	$d_{3,3}$	
Body Temperature, ˚C	39.23	39.05	38.82	39.14	39.18	38.92	39.36	38.98	39.19	0.21
Heart Rate, bpm	124.35	137.58	129.24	122.59	125.95	123.87	122.68	121.04	109.59	9.10
BUN[1], mg/dl	37.03		32.06	34.19		33.45	28.62		27.17	4.49
CK[1], units/l	19.91[a]		52.07[b]	10.25[a]		17.68[a]	6.23[a]		18.00[a]	6.48
Glucose, mg/dl	119.23		108.99	114.73		91.82	111.55		133.06	16.61
NEFA[1], mmol/l	0.38		0.51	0.38		0.47	0.39		0.51	0.06
Serum Cortisol, ng/ml	39.43		41.62	42.82		43.87	43.61		38.26	5.36
Fecal Cortisol, ng/ml	12.10	10.97	10.24	11.47	8.97	7.68	10.74	8.97	11.06	3.06

[1] BUN = Blood Urea Nitrogen; CK = Creatine Kinase; NEFA = Non-esterified Fatty Acids

[a,b] Days within a row with differing superscripts differ ($P < 0.05$)

Table 4.3. Chute score threshold comparison of change in objective measures for frequently and infrequently handled heifers on the first day of each event

		Day Within Event			
Measurement	n[1]	$d_{1,1}$	$d_{2,1}$	$d_{3,1}$	SEM
Body Temperature, ˚C					
1.0 - 1.5	19	39.94	39.14	39.02	0.22
1.5 - 2.0	24	39.53	39.38	38.98	0.28
2.0 - 2.5	27	39.58	39.48	39.11	0.16
2.5 - 3.0	21	39.61	39.16	38.97	0.18
> 3.0	27	39.94	39.52	39.19	0.19
Heart Rate, bpm					
1.0 - 1.5	19	116.56	109.64	119.46	5.99
1.5 - 2.0	24	136.11	122.08	130.49	9.55
2.0 - 2.5	27	129.99	122.61	132.86	7.69
2.5 - 3.0	21	127.20	136.58	140.39	8.67
> 3.0	27	133.83	134.78	141.30	9.23
BUN[1], mg/dl					
1.0 - 1.5	19	37.76	37.60	30.00	7.28
1.5 - 2.0	24	36.77	33.82	24.25	5.48
2.0 - 2.5	27	41.57	37.03	30.83	4.15
2.5 - 3.0	21	37.24	34.96	29.14	3.82
> 3.0	27	38.45	34.57	27.47	3.93
CK[1], units/l					
1.0 - 1.5	19	12.37	9.07	9.97	1.82
1.5 - 2.0	24	18.57[a]	9.53[b]	10.26[b]	1.91
2.0 - 2.5	27	16.47	10.62	10.61	1.76
2.5 - 3.0	21	16.82	10.94	11.90	3.47
> 3.0	27	15.27	14.37	10.36	4.43
Glucose, mg/dl					
1.0 - 1.5	19	111.82	108.03	120.66	20.59
1.5 - 2.0	24	115.58	119.34	114.59	13.88
2.0 - 2.5	27	111.05	121.77	104.78	13.95
2.5 - 3.0	21	120.71	110.03	126.72	17.52
> 3.0	27	135.96	134.12	129.85	18.23
NEFA[1], mmol/l					
1.0 - 1.5	19	0.32	0.36	0.42	0.07
1.5 - 2.0	24	0.34	0.37	0.45	0.06
2.0 - 2.5	27	0.28	0.31	0.38	0.05
2.5 - 3.0	21	0.31	0.33	0.38	0.06
> 3.0	27	0.27	0.34	0.36	0.05

Table 4.3 continued

Measurement	n^1	Day Within Event $d_{1,1}$	$d_{2,1}$	$d_{3,1}$	SEM
Serum Cortisol, ng/ml					
1.0 - 1.5	19	31.88	42.56	39.26	5.48
1.5 - 2.0	24	39.30	44.33	37.90	5.28
2.0 - 2.5	27	40.67	46.63	45.58	7.55
2.5 - 3.0	21	44.82	45.57	46.23	3.59
> 3.0	27	53.26	52.37	48.75	5.59
Fecal Cortisol, mg/ml					
1.0 - 1.5	19	12.90	11.58	11.06	3.84
1.5 - 2.0	24	12.52	12.66	9.87	2.50
2.0 - 2.5	27	12.38	11.57	11.93	3.22
2.5 - 3.0	21	13.04	11.88	10.02	3.05
> 3.0	27	11.47	11.01	9.61	4.28

[1] BUN = Blood Urea Nitrogen; CK = Creatine Kinase; NEFA = Non-esterified Fatty Acids

[a,b] Rows with differing superscripts differ ($P < 0.05$)

Objective measurements - data partitions

No variation in any measurement was defined by protocol across events when heifers were separated based on their CS on $d_{1,1}$. Average values of both F and IN heifers on the first day of each event are therefore given in Table 4.3. The most temperamental heifers on $d_{1,1}$ numerically had higher average body temperatures and heart rates as well as glucose and serum cortisol levels. However, they consistently had the lowest fecal cortisol concentrations. Alternatively, heifers with CS between 1.0 and 1.5 had numerically lower heart rates, CK, and serum cortisol concentrations than other groups.

Table 4.4. Chute score threshold comparison of change in objective measures for frequently handled heifers across days

Measurement	n^1	Day									SEM
		$d_{1,1}$	$d_{1,2}$	$d_{1,3}$	$d_{2,1}$	$d_{2,2}$	$d_{2,3}$	$d_{3,1}$	$d_{3,2}$	$d_{3,3}$	
Body Temperature, ˚C											
1.0 - 1.5	6	39.88	39.60	39.13	39.09	38.89	38.88	39.01	38.58	38.58	0.26
1.5 - 2.0	12	39.39	39.28	38.76	39.16	38.98	38.68	38.95	38.57	38.86	0.29
2.0 - 2.5	18	39.44	39.07	38.82	39.37	39.01	38.51	38.92	38.69	38.96	0.16
2.5 - 3.0	9	39.82[a]	39.18[ab]	39.05[b]	39.07[ab]	39.04[ab]	38.92[ab]	39.11[ab]	38.74[b]	39.21[ab]	0.22
> 3.0	14	39.87[a]	39.39[ab]	39.33[ab]	39.43[ab]	39.18[ab]	38.95[b]	39.24[ab]	38.96[b]	39.39[ab]	0.18
Heart Rate, bpm											
1.0 - 1.5	6	109.69	142.40	112.09	107.58	127.10	133.88	112.79	110.64	108.02	15.75
1.5 - 2.0	12	131.98	141.45	129.34	114.98	123.97	119.29	122.72	116.10	99.08	16.03
2.0 - 2.5	18	113.23	124.45	127.16	117.18	118.37	113.95	121.12	122.21	110.51	9.67
2.5 - 3.0	9	124.09	138.12	119.96	136.56	122.46	124.49	126.18	122.57	115.99	8.89
> 3.0	14	126.89	140.96	134.19	130.75	130.86	130.63	128.54	124.61	121.92	14.29
BUN[1], mg/dl											
1.0 - 1.5	6	31.75		31.45	36.36		33.98	30.68		32.56	7.79
1.5 - 2.0	12	38.46		35.72	36.22		35.34	23.03		26.37	4.82
2.0 - 2.5	18	39.60		32.01	35.08		35.26	29.04		28.10	4.07
2.5 - 3.0	9	35.18		29.07	33.73		30.64	31.60		27.60	3.55
> 3.0	14	38.27		31.31	32.59		35.46	27.90		24.88	3.81
CK[1], units/l											
1.0 - 1.5	6	15.80		29.35	7.56		10.89	10.84		7.86	6.82
1.5 - 2.0	12	18.32[ab]		54.96[b]	8.61[a]		16.16[ab]	9.31[a]		18.84[ab]	8.50
2.0 - 2.5	18	15.04[ab]		44.36[b]	8.48[a]		17.38[ab]	10.53[a]		20.69[ab]	9.01
2.5 - 3.0	9	10.38[a]		70.28[b]	13.36[a]		17.67[a]	9.54[a]		19.38[a]	8.48
> 3.0	14	16.44[a]		56.33[b]	22.29[ab]		21.32[ab]	11.52[a]		30.62[ab]	8.70

[1] BUN = Blood Urea Nitrogen; CK = Creatine Kinase; NEFA = Non-esterified Fatty Acids

[a,b] Rows with differing superscripts differ ($P < 0.05$)

Table 4.4 continued.

Measurement	n^1	Day $d_{1,1}$	$d_{1,2}$	$d_{1,3}$	$d_{2,1}$	$d_{2,2}$	$d_{2,3}$	$d_{3,1}$	$d_{3,2}$	$d_{3,3}$	SEM
Glucose, mg/dl											
1.0 - 1.5	6	119.23		92.79	99.79		76.05	139.46		124.18	28.26
1.5 - 2.0	12	111.38		119.91	120.83		86.57	98.03		119.10	15.74
2.0 - 2.5	18	112.69		96.98	111.43		85.05	92.54		140.22	14.79
2.5 - 3.0	9	122.97		114.10	95.53		95.78	128.97		139.43	24.39
> 3.0	14	119.22		111.45	130.98		100.01	130.14		144.33	21.59
NEFA[1], mmol/l											
1.0 - 1.5	6	0.44		0.49	0.39		0.30	0.46		0.46	0.15
1.5 - 2.0	12	0.39		0.63	0.40		0.51	0.47		0.61	0.10
2.0 - 2.5	18	0.27		0.45	0.32		0.52	0.42		0.54	0.07
2.5 - 3.0	9	0.32		0.51	0.39		0.48	0.40		0.53	0.09
> 3.0	14	0.29		0.47	0.39		0.51	0.39		0.48	0.08
Serum Cortisol, ng/ml											
1.0 - 1.5	6	34.80		46.23	45.35		40.27	36.93		31.33	10.09
1.5 - 2.0	12	36.39		44.01	36.72		46.32	35.69		29.84	8.27
2.0 - 2.5	18	39.56		37.44	45.19		39.47	43.47		34.92	6.95
2.5 - 3.0	9	43.30		39.05	40.25		36.50	45.25		44.83	5.90
> 3.0	14	50.04		42.39	48.28		46.68	43.79		43.40	6.26
Fecal Cortisol, ng/ml											
1.0 - 1.5	6	15.04	14.81	13.57	12.46	10.72	12.26	10.83	8.72	10.21	2.51
1.5 - 2.0	12	14.74	13.03	12.55	11.99	10.42	8.88	8.73	11.07	10.80	2.61
2.0 - 2.5	18	11.79	11.69	9.86	11.77	9.11	7.31	10.62	7.68	10.97	2.94
2.5 - 3.0	9	12.95	10.45	10.15	10.80	7.82	7.80	10.31	9.22	10.71	3.35
> 3.0	14	11.61	10.70	10.09	11.13	7.76	7.23	10.44	8.20	10.17	3.71

[1] BUN = Blood Urea Nitrogen; CK = Creatine Kinase; NEFA = Non-esterified Fatty Acids; [a,b] Rows with differing superscripts differ ($P < 0.05$)

The F protocol was again considered separately for each CS threshold group. Creatine kinase tended to decrease from event 1 to 2 ($P < 0.10$) in F heifers with CS between 1.5 and 3.0. More detailed changes in measurements across days are given in Table 4.4. Heifers with CS on $d_{1,1}$ of 2.5 and greater decreased in body temperature between $d_{1,1}$ and $d_{3,2}$ ($P < 0.05$). Furthermore, other than in the most docile heifers, CK on $d_{1,3}$ was again higher than other days of collection. As noted previously, there also was a numerical increase in CK between the first and last day of an event. Numerical trends across days were more indistinct. However, the most temperamental heifers appeared to have higher body temperatures, heart rates, CK, and serum cortisol levels, with lower concentrations of fecal cortisol.

Table 4.5. Pearson correlations of chute score, exit score, and exit velocity with each of the objective measurements for frequent and infrequently handled heifers on the first day of each event

	Chute Score		Exit Score		Exit Velocity	
Measurement	Correlation	SE	Correlation	SE	Correlation	SE
Body Temperature, ˚C	0.244	0.052	0.261	0.052	0.235	0.052
Heart Rate, bpm	0.248	0.052	0.327	0.051	0.316	0.051
BUN[12], mg/dl	0.035	0.053	0.024	0.053	0.071	0.054
CK[12], units/l	0.075	0.053	0.030	0.053	0.078	0.054
Glucose, mg/dl	0.268	0.051	0.262	0.052	0.238	0.052
NEFA[1], mmol/l	-0.183	0.052	-0.148	0.053	-0.143	0.053
Serum Cortisol, ng/ml	0.321	0.051	0.304	0.051	0.287	0.052
Fecal Cortisol[2], ng/ml	-0.078	0.054	0.068	0.054	0.029	0.054

[1] BUN = Blood Urea Nitrogen; CK = Creatine Kinase; NEFA = Non-esterified Fatty Acids

[2]Correlations not different from zero ($P > 0.05$)

Relationship between subjective and objective measurements

Pearson correlations of CS, ES, and EV with the various objective measurements on the first day of each event are shown in Table 4.5. Correlations of all three measurements with BUN, CK, and fecal cortisol levels were not different from zero while correlations with body temperature, heart rate, glucose, and serum cortisol were moderately positive. Interestingly, the correlations of CS, ES, and EV with NEFA concentrations were consistently negative.

The estimated values of the Pearson correlations are supported by the results of the PCA analysis shown in Figure 4.1. The first and second principal components together described 35.2%, 35.4%, and 34.9% of the total variation when investigating associations between weather and metabolites with CS, ES, and EV, respectively. Eigen values for each PCA are reported in Table 4.6. All three analyses had similar results. For CS (Figure 4.1.a), ambient and body temperature had the largest positive loadings, while NEFA had the largest negative loading, for the first PC. For the second PC, relative humidity and wind speed had the largest positive and negative loadings, respectively. The same trend was observed for the other behavioral measures (Figures 4.1.b and 4.1.c), although the sign of the loadings switched. The behavioral measures had, in general, heavy loadings in the third PC, which explained less than 13% of the variation. Still, the behavioral measures clustered nearer the objective measurements with which they had larger Pearson correlations: body temperature, glucose, and serum cortisol. Their loadings also were opposite in sign to NEFA to which they were negatively correlated.

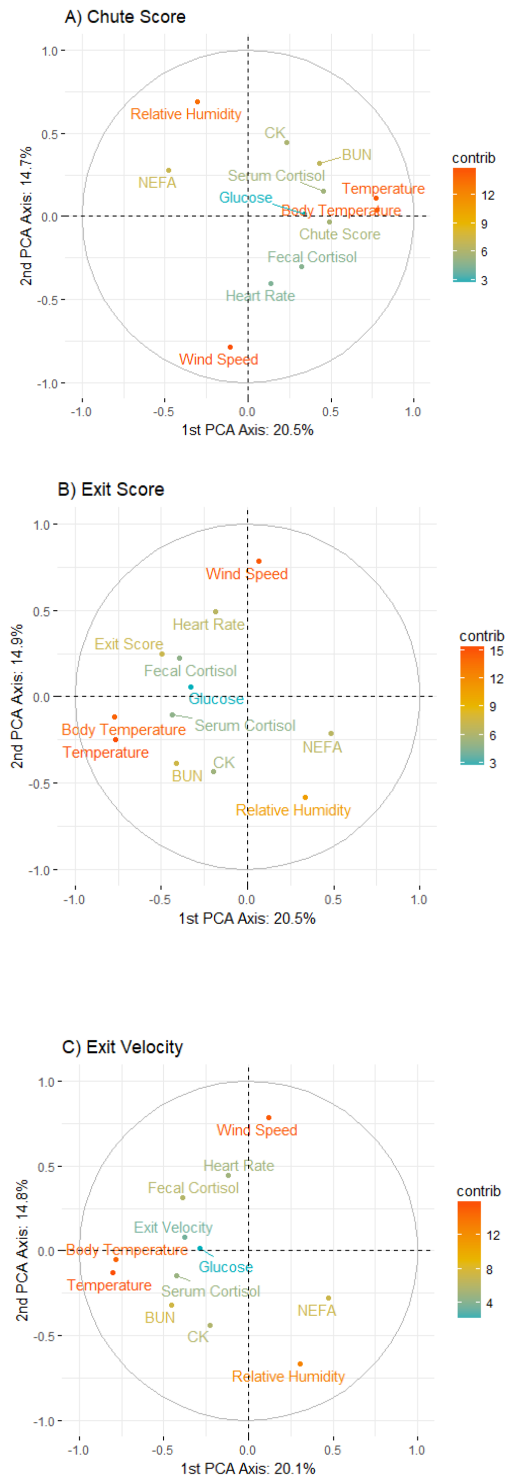


Figure 4.1. Principal component analysis results and contributions of frequently and infrequently handled heifers on the first day of each event for all subjective and objective measures.

Table 4.6. Eigenvectors for the first three principal components for analyses ran between chute score, exit score, or exit velocity with objective metabolites and weather data. Dominant loadings are highlighted in bold.

	Chute Score			Exit Score			Exit Velocity		
Variable	PC 1	PC 2	PC 3	PC 1	PC 2	PC 3	PC 1	PC 2	PC 3
Chute Score	0.31	-0.02	**0.51**	-	-	-	-	-	-
Exit Score	-	-	-	-0.32	0.19	0.42	-	-	-
Exit Velocity	-	-	-	-	-	-	-0.24	0.06	**-0.50**
Body Temperature	**0.50**	0.03	-0.01	**-0.49**	-0.09	-0.17	**-0.50**	-0.04	0.12
Heart Rate	0.09	-0.31	0.28	-0.12	0.37	0.42	-0.08	0.33	**-0.50**
BUN[1]	0.27	0.24	-0.28	-0.26	-0.29	0.04	-0.29	-0.24	0.05
Creatine Kinase	0.15	0.33	0.10	-0.12	-0.33	0.13	-0.15	-0.33	-0.10
Glucose	0.21	0.01	**0.37**	-0.21	0.04	**0.47**	-0.19	0.01	-0.44
NEFA[1]	**-0.31**	0.21	0.14	**0.31**	-0.16	0.05	**0.30**	-0.21	0.03
Serum Cortisol	0.29	0.12	0.38	-0.28	-0.08	0.21	-0.27	-0.11	-0.21
Fecal Cortisol	0.21	-0.23	**-0.38**	-0.25	0.17	-0.06	-0.25	0.24	0.07
Temperature	**0.49**	0.08	-0.31	**-0.49**	-0.19	**-0.33**	**-0.52**	-0.10	0.31
Wind Speed	-0.07	**-0.59**	0.12	0.04	**0.59**	-0.07	0.08	**0.59**	-0.01
Relative Humidity	-0.19	**0.52**	0.13	0.21	**-0.44**	**0.46**	0.20	**-0.50**	**0.37**
Variation Defined, %	20.5	14.7	12.6	20.5	14.9	12.5	20.1	14.8	12.4

[1] BUN - Blood Urea Nitrogen; NEFA = Non-esterified Fatty Acids

Discussion

In this study, CS, ES, and EV appear to be acceptable methods to assess stress in an animal during normal handling practices as they were moderately positively correlated with body temperature, heart rate, glucose, and serum cortisol. While correlations with creatine kinase and fecal cortisol were not different from zero, this may be due to a delay in the expression of these metabolites.

These findings support previous research measuring change in subjective measures over time. Parham et al. (2018b) proposed that these heifers acclimated to repeated calm handling, particularly in more temperamental animals. While there was no significant change in most objective measures assessed at the first management event at monthly intervals, nor any difference between measurement protocols, this conclusion is consistent with changes found in subjective measures. However, when assessing only frequently handled heifers, there was a decrease in CK following $d_{1,3}$ for heifers with CS greater than 1.5 on $d_{1,1}$. Frequently handled heifers also had a reduced heart rate, and serum and fecal cortisol concentrations, between the first and last day of observation. Heart rate and serum cortisol were consistently highest in temperamental heifers and lowest in the most docile group. Furthermore, heifers considered the most temperamental had higher body temperatures and glucose concentrations with the lowest fecal cortisol concentrations, while the most docile heifers had lower concentrations of CK. Overall, these results support the conclusions of Parham et al. (2018b) that cattle acclimate to repeated calm handling. Because of this, when cattle are excitable during their first handling experience, more than one observation of temperament may be beneficial before making selection decisions.

Moderate correlations of CS and EV with body temperature, heart rate, glucose, and serum cortisol concentrations are consistent with literature (Cooke et al., 2009; Petherick et al, 2009; Gruber et al., 2010). Furthermore, since EV and ES were so highly correlated (r = 0.81 ± 0.03, Parham et al., 2018b) it is understandable that they would be correlated with ES as well. Since increased heart rate and plasma cortisol levels have been associated with stressful situations such as transport (Sporer et al., 2008; Van de Water et al., 2003), it follows that heifers that exhibited behavioral symptoms of stress during confinement via a subjective measurement also responded physiologically; that is, they had increased concentrations of serum cortisol and glucose, along with elevated body temperatures and accelerated heart rates. However, Van de Water et al. (2003)

found no significant decrease or increase of glucose levels before and after transportation of 15 different loads of cattle.

Conversely, correlations of temperament measures with fecal cortisol and CK were not different from zero. While serum cortisol peaks approximately 10 to 20 minutes following stress (Lay et al., 1992), fecal cortisol takes longer to be metabolized and excreted in ruminants, reaching peak levels approximately 10 to 12 hours later (Möstl and Palme, 2002); consequently, fecal cortisol provides a longer-term measure of stress. In this study, fecal samples were taken approximately 24 hours following handling. Therefore, it is understandable that fecal cortisol levels were not correlated with subjective measures of temperament. On the first day, the heifer would not have been subjected to any stress, providing a basal measurement of cortisol. On the second and third day, heifers had likely reverted back to base levels of excreted cortisol by the time of sampling.

Similarly, CK activity increases due to tissue damage and bruising in animals and is often used as an indicator of physical stress and muscle damage (Mpakama et al., 2014). Although Gruber et al. (2010) found the correlation between CS and CK to be 0.17, levels of this metabolite were measured in blood collected post-transportation while the CS was assessed on the day before transport. Behavior in the chute and during transport could have resulted in tissue damage or bruising in animals after the CS was assessed, resulting in an increase in CK and therefore a stronger correlation the following day. In this study, the correlation of CS, ES, and EV with CK measured on the same day was not different from zero, again because the sample was taken before circulating levels were able to respond to any tissue damage or bruising. It is more likely strong correlations would exist between CS on the day prior to the increase in CK concentrations, as in Gruber et al. (2010).

There is a clear response in CK following handling in this study, albeit delayed, which supports the preceding idea. Blood samples were only taken on the first and last day of an event; therefore, when comparing all heifers on the first day of each event, it appears as though no significant change in CK occurred. However, the second collection shows a numerical increase in concentration within each event. This increase is largest during the first event, with $d_{1,3}$ being the highest concentration for all days. Parham et al. (2018b) showed an increase in CS from $d_{1,1}$ to $d_{1,2}$ in the frequently handled heifers, with $d_{1,2}$ having the highest average CS for all nine days of observation. Therefore, this increase in creatine kinase on $d_{1,3}$ could be the result of increased tissue damage resulting from an increase in excitability in heifers during the previous day. While not as large, this increase from the first to third day of an event persisted for all three events. Furthermore, CK levels for the least excitable category of heifers increased on a much smaller scale relative to others. This low CS would represent a heifer that did not struggle or resist while restrained, and therefore would be much less likely to sustain any bruising or tissue damage. Finally, CK levels on $d_{1,1}$ and $d_{3,3}$ were similar when all heifers were analyzed together. Therefore, they did not appear to have any more tissue damage at the end of the study than they did at the beginning, which could be the result of calmer disposition while restrained in the chute.

The CK levels obtained in the current study unfortunately cannot be directly compared to those elsewhere due to a difference in methodology, even though concentrations are in the same units. Manufacturer's instructions (EnzyChrom Creatine Kinase Assay, BioAssay Systems, Hayward, CA) stated that if concentrations of CK were above 300 units/l, the sample was to be diluted and reanalyzed. Such was the case here. Therefore, concentrations of CK in this study are lower than those previously reported in literature.

While CK was the only metabolite to show significant changes over time, other objective measurements did provide evidence of acclimation to handling and to a reduction in stress. The lack of significant change in any of the metabolites when comparing all heifers, regardless of measurement protocol, are consistent with subjective measurements of excitability (Parham et al., 2018b). However, acclimation was more apparent in the frequently handled heifers, especially those considered more excitable on the first day of observation. On $d_{3,3}$ measurements of body temperature and heart rate for all frequently handled heifers were consistent with previously reported values of cattle in a resting state (38.73 ± 0.06 ˚C and 101 ± 15 bpm, respectively; Sporer et al., 2008; Van de Water et al., 2003). Furthermore, serum cortisol also was lowest on $d_{3,3}$, but still higher than the 6 ng/ml baseline level for Angus cross heifer calves proposed by Grandin (1997). Levels that typically occur during restraint in a head gate are higher, ranging from 24 ng/ml in weaned British or European calves to 63 ng/ml in *Bos indicus* influenced steers and heifers (Grandin, 1997). Based on these values, heifers in the current study were no more stressed than calves of British or European influence at weaning (46 ng/ml;

Grandin, 1997). Using these serum cortisol levels, it can be concluded that these heifers were not subjected to any adverse conditions, which could have inhibited their ability to acclimate to handling.

Although serum cortisol levels were consistent with those considered normal during handling practices, they reflected physiological status at the time of the perceived stressful incident. Fecal cortisol levels provided a baseline comparison of an animal's temperament 10 hours prior to the handling experience. Fecal cortisol levels were consistently lower than serum cortisol, as expected. However, more interesting was that more excitable heifers had higher serum cortisol but lower fecal cortisol concentrations, with the more docile heifers being the opposite. Hourly measures of fecal cortisol concentrations would help clarify this observation. However, it could be speculated that excitable heifers may circulate and expel higher concentrations of cortisol at a faster rate than docile heifers. Therefore, at the time of measurement, temperamental heifers may have already excreted the higher concentrations of cortisol seen in the blood, leaving them with lower basal concentrations compared to their docile counterparts.

While increased levels of these blood and fecal metabolites, heart rate and body temperature have been associated with stress, they mainly have been characterized in cattle in the most stressful aspects of handling such as transportation and pre-slaughter. The difficulty in other cases, like routine handling procedures, is defining the baseline between animals that are calm and stressed. It is well documented that genetic factors influence physiological measures of stress (Grandin and Deesing, 2014). Grandin (1997) provides a comprehensive list of studies conducted with varying breed and sex combinations, and their impact on cortisol levels. More recently, Sporer et al. (2008) found different concentrations of plasma cortisol levels between Angus, Friesian, and Belgian Blue x Friesian cross cattle. Furthermore, CK levels were also found to be different between Bonsmara and Brahman cattle (Mpakama et al., 2014). Other than defining concentrations and levels of measurements at rest, there is not comprehensive information available on how to determine which animals are stressed based on objective measures such as body temperature, heart rate, and other blood metabolites. The presence of breed and sex differences suggests defining baseline values may be difficult but such information would still be valuable.

Conclusion

In conclusion, temperament in cattle is a popular selection criterion due to its impact on many economically relevant factors in the beef industry. Selection for docility in a production setting requires measurements that are accurate, inexpensive, and relatively easy to collect, making CS and ES ideal methods of assessing temperament. In this study, heifers that exhibited behavioral symptoms of stress during confinement based on subjective measurements also physiologically responded with increased concentrations of serum cortisol and glucose, along with elevated body temperature and accelerated heart rates. Lastly, heifers that were more excitable in the chute had delayed increases in CK levels, suggesting possible tissue damage as a result of increased movement in the working facilities. Therefore, CS and ES appear to capture an animal's response to stress. Furthermore, changes in these objective measures were consistent with changes in CS previously documented by Parham et al. (2018b), supporting the conclusion that cattle that are excitable during their first handling experience may acclimate to repeated, calm handling.

Literature Cited

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

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.

Cooke, R. F., Arthington, J. D., Araujo, D. B., Lamb, G. C., 2009. Effects of acclimation to human interaction on performance, temperament, physiological responses, and pregnancy rates of Brahman-crossbred cows. J. Anim. Sci. 87, 4125-4132.

Curley Jr., K. O., Paschal, J. C., Welsh Jr., T. H., Randel, R. D., 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., Goddard, M. E., Tyler, R., Williams, G., Toleman, M. A., 1985. Temperament and bruising of Bos indicus cross cattle. Aust. J. Exp. Agric. 25, 283-288.

Grandin, T., 1997. Assessment of stress during handling and transport. J. Anim. Sci. 75, 249-257.

Grandin, T., Deesing, M. J., 2014. Genetics and behavior during handling, restraint, and herding. Genetics and the Behavior of Domestic Animals (Second Edition). pp. 115-158.

Gruber, S. L., Tatum, J. D., Engle, T. E., Chapman, P. L., Belk, K. E., Smith, G. C., 2010. Relationships of behavioral and physiological symptoms of preslaughter stress to beef longissimus muscle tenderness. J. Anim. Sci. 88, 1148-1159.

Hoppe, S., Brandt, H. R., König, S., Erhardt, G., Gauly, M., 2010. Temperament traits of beef calves measured under field conditions and their relationships to performance. J. Anim. Sci. 88, 1982-1989.

King, D. A., Schuehle Pfeiffer, C. E., Randel, R. D., Welsh Jr., T. H., Oliphint, R. A., Baird, B. E., Curley, Jr., K. O., Vann, R. C., Hale, D. S., Savell, J. W., 2006. Influence of animal temperament and stress responsiveness on the carcass quality and beef tenderness of feedlot cattle. Meat Sci. 74, 546-556.

Lay, D. C., Friend, T. H., Bowers, C. L., Grissom, K. K., Jenkins, O. C., 1992. A comparative physiological and behavioral study of freeze branding and hot-iron branding using dairy cows. J. Anim. Sci. 70, 1121-1125.

Möstl E., Palme, R., 2002. Hormones as indicators of stress. Domest. Anim. Endocrinol. 23, 67-74.

Mpakama, T., Chulayo, A. Y., Muchenje, V., 2014. Bruising in slaughter cattle and its relationship with creatine kinase levels and beef quality as affected by animal related factors. Asian Australas. J. Anim. Sci. 27, 717-725.

Parham, J. T., Tanner, A. E., Wahlberg, M. L., Grandin, T., Lewis, R. M., 2018a. Subjective methods to quantify temperament in beef cattle are insensitive to the number and biases of observers. Appl. Anim. Behav. Sci. (Submitted 16 July 2018).

Parham, J. T., Tanner, A. E., Barkely, K., Pullen, L., Wahlberg, M. L., Swecker, Jr., W. S., Lewis, R. M., 2018b. Temperamental cattle acclimate more substantially to repeated handling. Appl. Anim. Behav. Sci. (Submitted 12 July 2018).

Petherick, J. C., Doogan, V. J., Venus, B. K., Holroyd, R. G., Olsson, P., 2009. Quality of handling and holding yard environment, and beef cattle temperament: 2. Consequences for stress and productivity. Appl. Anim. Behav. Sci. 120, 28-38.

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

Sant'Anna, A. C., Paranhos da Costa, M. J. R., Baldi, F., Albuquerque., L. G., 2013. Genetic variability for temperament indicators of Nelore cattle. J. Anim. Sci. 91, 3532-3537.

Sporer, K. R. B., Weber, P. S. D., Burton, J. L., Earley, B., Crowe, M. A., 2008. Transportation of young beef bulls alters circulating physiological parameters that may be effective biomarkers of stress. J. Anim. Sci. 86, 1325-1334.

Van de Water, G., Verjans, F., Geers, R., 2003. The effect of short distance transport under commercial conditions on the physiology of slaughter calves; pH and colour profiles of veal. Livest. Prod. Sci. 82, 171-179.

Vargas Jurado, N., Tanner, A. E., Blevins, S. R., Rich, J., Mayes, R. W., Fiske, D., Swecker Jr., W. S., Lewis, R. M., 2015. Feed intake and diet selection in Angus-cross heifers of two frame sizes at two stages of growth. J. Anim. Sci. 93, 1565-1572.

Von Borell, E., Dobson, H., Prunier, A., 2007. Stress, behaviour and reproductive performance in female cattle and pigs. Horm. Behav. 5, 130-138.

CHAPTER FIVE

DETAILED ANALYSIS OF PEN SCORE FOR ITS USEFULNESS IN DELINEATING BEHAVIORS INDICATIVE OF TEMPERAMENT IN CATTLE

Abstract

A chute score (CS) and exit score (ES) are used commonly as subjective methods to evaluate temperament in cattle production systems. A pen test, which allows behavior to be observed in a non-restrained setting, may be an effective additional method to evaluate temperament by allowing more variation among animals to be expressed. However, the value of pen scores in assisting producers in evaluating temperament is equivocal. The overall objective of this study therefore was to validate the usefulness of a pen score in delineating specific behaviors indicative of temperamental cattle. This was achieved by first assessing the reliability of pen score, along with its relationship with other objective and subjective measures of stress, and then to investigate the behaviors expressed during confinement both individually and in a group setting. Any changes in behaviors over time also was assessed. In three consecutive years, a factorial design of two measurement protocols [frequent (F), infrequent (IN)], and three recording periods, each 1 mo apart, was used. The F measurements were collected over three consecutive days, and IN measurements only on day one within a recording period. Each year, 20 commercial *Bos taurus* heifers, 2-wk post weaning, were randomly assigned to each protocol. Behavior was measured using a CS, ES, and exit velocity (EV), along with body temperature, heart rate, and a fecal and blood sample. These samples were then analyzed for levels of various metabolites. Following routine handling, each heifers'

response to 30 s of exposure to a human stressor was recorded both individually and in groups of four. An individual (IPS) and group (GPS) pen score was assigned by three experienced observers from 1 (docile) to 6 (aggressive). Video recordings were evaluated, and their associations with IPS and GPS categories assessed. Data were analyzed with ANOVA using SAS. For all heifers, protocol, event, and their interaction, were compared on day one. For heifers assigned to F, event and day within event were instead fitted. For both models, body weight was included as a covariate, with sire and year fitted as random effects. Behaviors delineated using video recordings were analyzed using the same model, with IPS and GPS also included as a fixed effect. Reliability of IPS and GPS were determined using a kappa (K) coefficient. Both IPS and GPS were reliably assigned, with coefficients greater than threshold values indicative of acceptable assessment (K = 0.64 and 0.44 for IPS and GPS, respectively) Furthermore, both scores were moderately positively correlated with objective measures of temperament (r = 0.28 to 0.37). Individual and GPS were more highly correlated with ES, possibly because both are considered non-restrained measures of temperament. Analyses of videos suggest that flight zone size and the fastest pace reached had a strong association with IPS and GPS ($P < 0.001$). Those scores also were indicative of the time animals spend stationary compared to trotting or running ($P > 0.05$). Lastly, these heifers acclimated to repeated handling in an individual pen setting across both days and events ($P < 0.05$), with more substantial acclimation in F heifers. However, average GPS on the first day of the study identified docile heifers therefore further acclimation to handling within groups was not evident ($P > 0.14$). Individual pen score is therefore a reliable measure of temperament that is indicative of an animal's response to stress, and may be useful when attempting to make selection decisions. However, more temperamental heifers became calmer with repeated gentle handling.

Introduction

Temperament in cattle has a documented impact on growth, carcass quality, reproduction, and well-being (Fordyce et al., 1985; Von Borell et al., 2007; Bates et al., 2014; Cooke, 2014). These negative outcomes, and its moderate heritability (Sant'Anna et al., 2013), have led to an increase in selection for docility in cattle.

In order to effectively select for docility in cattle it is imperative that each criteria measured are indicative of behavior during normal handling practices (Fordyce et al., 1982). The three most common measurements proposed to quantify temperament are based on behavior when cattle are restrained in (chute score, CS) and exiting from (exit score, ES and exit velocity, EV) the chute. Another, less common, non-restrained measure of temperament is pen score, which can be assessed on animals individually and in small groups. Le Neindre et al. (1995) proposed a method in which an animal was isolated individually in a pen with a handler who had 2 min to direct it into a corner, hold it there for 30 s, and then stroke it. Based on an animal's response, a subjective score was assigned from 1 (calm) to 5 (very excited). This method has been used to compare temperament of Angus and Simmental cattle with heritabilities of 0.61 ± 0.17 and 0.55 ± 0.15, respectively (Gauly et al., 2001). These tests, however, had no association with weight gain in cattle. Increased concern about handler safety while attempting to stroke the animal was brought to light (Kilgour et al. 2006) and the method was then curtailed to the handler simply standing in the middle of the pen for 30 s, with no attempt made to restrain the animal (Turner et al., 2011). This method was instead referred to as an "isolation score" and rated on a scale of 1 to 6.

The benefit, and perhaps drawback, of using a non-restrained test such as pen score to assess temperament is that it allows for a larger amount of variation in the expression of behaviors. As an example, cattle in close proximity to people daily will have smaller flight zones than genetically similar cattle raised on pasture (Grandin and Deesing, 2014). This variation in flight zone size has an impact on how stressed an animal appears when secluded in a pen with a human, and the behaviors they express in trying to remove the present threat. This could result in differences in expressions of behaviors, including the pace at which animals move away from the human, and whether they attempt to escape perhaps colliding into fences. While a subjective ethogram is a useful tool in helping to delineate these behaviors, animals will seldom fall into a single category, which could lead to inconsistency in scoring among observers.

Subjective measurements are attractive due to their ease of use. However, the efficacy of selection programs depends on the reliability of the observation and whether they are indicative of an animal's response to stress. Since these scores are based on the subjective opinion of the observer they may be less reliable than objective measures such as EV or cortisol. However, CS and ES can be reliably assessed by numerous individuals regardless of their prior experience (Parham et al., 2018a). Furthermore, these scores are moderately correlated with objective measurements including body temperature, heart rate, and serum concentrations of glucose and cortisol (Parham et al., 2018b). Therefore, CS and ES provide fast, easy, and inexpensive measures of temperament in an animal. While correlations of pen score with CS and exit velocity range from 0.24 to 0.75 (Cooke et al., 2009; Turner et al., 2011), their associations with objective measures of stress are less well known. Cooke et al. (2009) reported a correlation between pen score and plasma cortisol of 0.44. Furthermore, reliability of pen scores has yet to be determined.

If pen score provides reliable measurement of temperament in an animal, producers who utilize it in selection decisions may not do so efficiently if an animal's initial response to a stressor is not indicative of future responsiveness. This is referred to as acclimation to handling. Curley et al. (2006) suggested that an animal’s temperament score decreased as the number of times it was exposed to a working facility increased. Cattle successfully acclimated to handling may more willingly re-enter a handling facility, resulting in calmer cattle that require less time to process and are safer to handle.

In combination, these issues require a detailed analysis of the usefulness of pen score in delineating specific behaviors indicative of temperamental cattle. This objective was achieved by first assessing the reliability of pen score, along with its relationship with other objective and subjective measures of stress. Secondly, the behaviors expressed during confinement both individually and in a group setting were investigated, and any change in these behaviors over time assessed. If pen score was useful in delineating temperament, it could be utilized by producers wishing to make phenotypic selection decisions on temperament. Furthermore, producers interact with their cattle in many different situations, both individually and as a group. If these types of interactions with cattle are common, producers may also benefit from delineating simple, more objective, behaviors indicative of pen scores to use when assessing temperament.

Methods

All procedures and protocols used in this study were approved by the Institutional Animal Care and Use Committee at Virginia Tech.

Experimental animals

Within each of three consecutive years, 40 commercial *Bos taurus* (75% Angus or more) spring-born heifer calves were reared at the Virginia Tech Shenandoah Valley Agricultural Research and Extension Center (SVAREC) in Steeles Tavern, VA (37°55′56″ N 79°12′50″ W) with their respective dams until weaning (185 ± 11 d in age). Upon completion of a 1-wk fence line weaning period, heifers were transported to Virginia Tech Kentland farm (37°11′60″ N 80°33′52″ W) and placed in a single management group on grass. The dataset included progeny from one of 21 sires, ranging from one to 23 progeny per sire, selected to establish divergent larger and smaller frame size offspring. Cows were bred within line to Angus sires correspondingly selected as larger or smaller based on their mature cow height estimated breeding value (Vargas Jurado et al., 2015).

Experimental design and data collection

Heifers were randomly assigned to one of two measurement protocols (frequent; infrequent) within their dam's frame score (larger, smaller) and sire. Data were collected across three recording periods, each 1 mo apart [i.e., event 1 (Oct.), 2 (Nov.), 3 (Dec.)]. Heifers within the frequent (F) measurement protocol were observed three consecutive days within each event (month) while the heifers in the infrequent (IN) measurement protocol were evaluated on only the first day of an event. Day within event was designated by $d_{i,j}$ where i was the event and j was the day within an event.

On day one of each event (i.e., $d_{1,1}$, $d_{2,1}$, $d_{3,1}$), all 40 heifers were moved into a holding pen. Four heifers were randomly drawn from the group and herded into the cattle-handling facility, which consisted of a small holding pen narrowing into a curved alley that led to a weigh crate and separate squeeze chute. One at a time, each heifer was calmly moved through the alley into the weigh crate. Once weighed, each heifer was then moved into the squeeze chute (Priefert Model S04). While restrained, a chute score (Tulloh, 1961), heart rate and rectal temperature was recorded (Parham et al., 2018c). A blood sample was collected via jugular venipuncture, and a fecal sample taken.

Upon release from the chute, an ES (BIF Guidelines, 2002) and exit velocity (EV; Burrow et al., 1988) were recorded, and the heifer was calmly walked down to a 12 x 6 m pen. Once individually enclosed, a video camera recorded the next 30 s of exposure to a human stressor, who walked into and then stood in the center of the pen. During this time, individual pen scores (IPS) were assigned independently by three experienced observers using an ethogram based on King et al. (2006):

1. Docile: walks slowly, can be approached slowly, not excited by humans.
2. Slightly Restless: aware of humans, head up, moves away from approaching human, runs fence line, stops and looks around.
3. Restless: constantly runs along fence line, head up.

4. Nervous: agitated, runs along fence line, head up, looking for a way of escape, and will run if humans come closer, stops before hitting gates and fences, avoids humans.
5. Very Nervous: runs, head high and very aware of humans, may run into fences and gates, flighty.
6. Wild (Aggressive): excited, runs into fences, runs over anything in its path.

Each heifer was then moved into a 12 x 8 m pen separated from the individual pen by an alleyway with no direct and limited visual contact. Heifers remained until four heifers had been moved through the working facilities and placed together. Once in a group of four, the same ethogram was used to assign a group pen score (GPS) to each heifer individually, based on their reaction to a human stressor. In this case, the human walked to the center of the pen, paused, and continued diagonally in the direction of the group of heifers before returning to the center of the pen. This interaction was also video recorded for 30 s. Observers assigning scores were located behind or on a separate fence, located behind the video camera.

On day two and three of each event (i.e., $d_{1,2}, d_{2,2}, d_{3,2}$, $d_{1,3}, d_{2,3}, d_{3,3}$) the same measurements were again recorded on all heifers assigned to the F protocol (*n* = 20). After day three of recording all 40 heifers were mixed into a single management group until the next recording period.

Across the three years, two of the 120 heifers, one from F and the other IN, were removed from the study due to lameness.

Video analyses

Continuous data were evaluated by one trained person who viewed the individual and group pen score video recordings of each heifer, or group of heifers, using Observer software (Noldus, The Netherlands) for detailed behaviors using individual and group ethograms provided in Appendix A and Appendix B, respectively. Individual heifers could not be clearly distinguished in some group pen videos. Therefore, ethograms for evaluating heifers gathered in the pen were related back to the behavior of the group as a whole. For clarification, since the individual evaluating the videos was not an observer during the study, this individual will be referred to as the viewer.

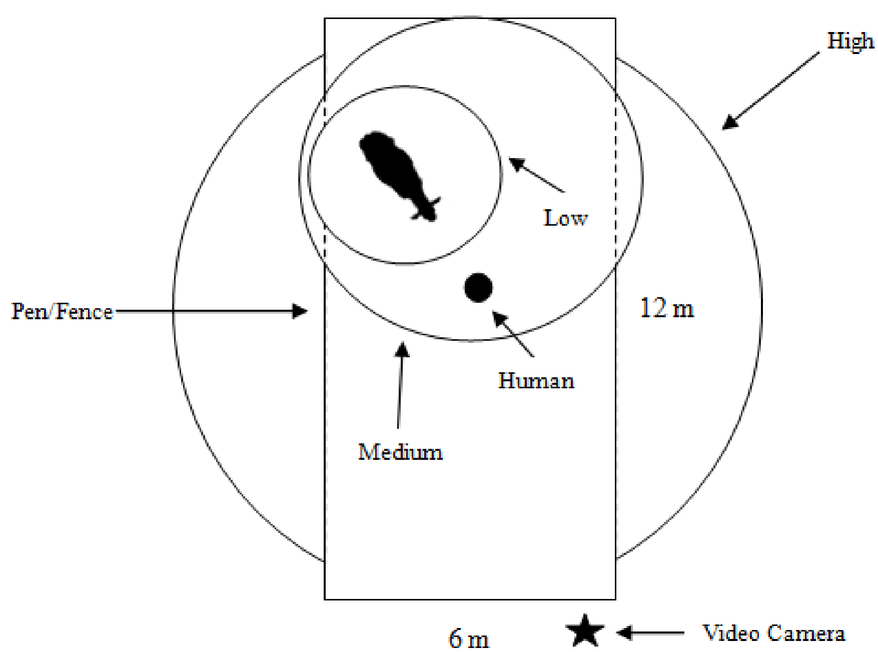


Figure 5.1. Diagram of the individual pen environment

Video analysis of individual pen. First, the viewer recorded several overall behaviors based on the video as a whole. The temperament of the heifer was further delineated as to whether or not the human felt safe to enter the pen with the animal, or entered the pen but did not stay. The reaction of the heifer to the human was recorded as initiating flight or fight instincts, or if the heifer appeared neutral to the individual's presence. Similarly, the viewer recorded a flight zone size as either low, medium, or high as shown in Figure 5.1.

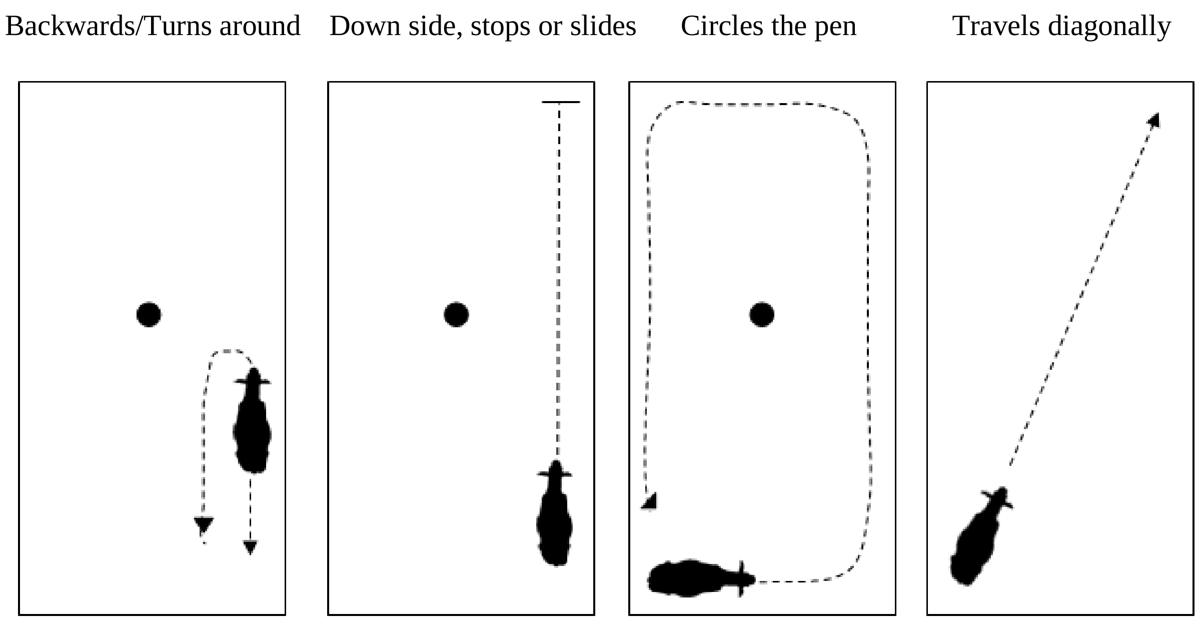


Figure 5.2. Diagram depicting direction of movement ethogram

The viewer also recorded the number of times each heifer threatened or charged the human, and the number of times she head-butted the fence (either actively or coincidently), hit the fence, attempted to escape, or flicked her tail.

Lastly, several timed activities were recorded if its duration was longer than one second. These activities included the pace of the animal (stationary, fidgets, walks, trots, runs/canters) and their direction of movement (Figure 5.2). From this information, the fastest pace reached by each heifer was recorded. While each video was expected to last for 30 s, timed activities were standardized to ensure consistent comparisons. This was done by dividing each separate timed activity by the total time of the video, then multiplying that by 30 s.

Video Analysis of Group Pen. In order to compare group behavior to individual responses, the group pen scores for each heifer were averaged to obtain an overall score. The flight zone size of the group was assigned at the initial approach of the human based on the reaction of a majority of the group. Also, because the individual walked into the middle of the pen, paused, and then walked towards the group of cattle, their location was recorded both at the start of the video as well as after the human returned to the center of the pen.

The amount of time spent moving also was recorded for each animal and each pace. Since these evaluations were summarized on a group basis, time spent by each heifer for a pace were averaged. Since the length of each video was not exactly 30 s, these times were adjusted accordingly.

Other timed activities for the group included the duration heifers were willing to remain separate from one another as well as where the majority of each group focused their attention, be it on the human in the pen, the observers standing near the video camera, or elsewhere. Again, all timed activities were standardized to 30 s.

Statistical analyses

Statistical analyses for these data were completed in three steps. Data collected during the study were analyzed initially. This included the reliability of the assessments, their relationships with other subjective and objective measurements of temperament, and finally, for any change in IPS or GPS over time. Secondly, the validity of the detailed ethograms to delineate temperament was assessed by comparing video analyses based on categories of IPS and GPS assigned on a given day. Lastly, these detailed ethogram behaviors were also assessed for any change over time.

Inter-observer reliability. Reliability of both individual and group pen score for the three observers was calculated using Fleiss' kappa coefficient and an intra-class correlation. All reliability calculations were carried out using the irr package (Gamer et al., 2012) in R.

Relationship between measurements. Pearson correlations were calculated between IPS and GPS on a given day with all observations including CS, ES, EV and all objective measurements of behavior for heifers on the first day of each event ($d_{1,1}$, $d_{2,1}$, and $d_{3,1}$) separately, and then combined in R (R Core Team, 2013). Correlations were first calculated for F and IN separately with similar results between measurement protocols. Therefore, both groups were combined. Correlations of IPS and GPS with CS, ES, and EV appeared to increase over time and are reported separately.

Acclimation to handling. To compare the effect of measurement protocol, initially a 2x2x3 factorial model (model 1) was analyzed fitting protocol (F and IN), dam frame size (larger or smaller) and event (1, 2, or 3), and their two and three way interactions, as fixed effects. Year, sire, and heifer nested within year, measurement protocol and dam frame size combination, were treated as random effects. Comparisons were only made on the first day of each event ($d_{1,1}$, $d_{2,1}$, $d_{3,1}$). Dam frame size never explained significant variation in the response variables ($P > 0.12$) and thus was excluded from the model. The final fixed effect model fitted therefore included protocol, event, and their interaction, with the addition of body weight as a covariate.

To measure changes in response variables over time within F, a separate model (model 2) was fitted. Event, dam frame size, their interaction, and the nested effect of day within event, were fitted as fixed effects. Heifer nested within year and dam frame size combination, as well as year and sire, were treated as random effects. Again, dam frame size did not define significant variation in the response variables ($P > 0.24$) and was removed from the final model fitted. However, body weight was included as a covariate. Change in pen scores over time was analyzed using the GLIMMIX procedure in SAS (SAS Inst. Inc., Cary, NC) fitting two separate models. Least squares means and SE were obtained using SAS with Tukey's adjustment for multiple comparisons.

The IPS and GPS recorded by the three observers were averaged on each day to obtain a representative score. Based on examination of residuals compared in SAS using model 1, the distribution of these data appeared skewed. A natural logarithm was applied and residuals of the transformed values again analyzed. The log transformed data resulted in less skewed residuals, and were therefore utilized for all analyses. However, in order to express the results on their scale of measurement, means and SE were back-transformed to the observed scale.

A repeated measures analysis was initially conducted as in Parham et al. (2018c). Results from these analyses were consistent with the initial models described. Therefore, results obtained from the simpler factorial models are reported henceforth.

Video analyses - categorical ethograms. Several of the ethograms associated with the individual and group pen evaluations were categorical in their measure: flight zone size (low, medium, high), fastest pace (stationary, fidget, walk, trot, canter/run), fight or flight (neutral, fight, flight), whether the observer felt safe to enter the pen (yes or no), and the group's location before the human entered the pen, and after they reached the center. Since observation in categories were counted, these data were analyzed using ordered contingency tables.

To determine whether ethograms were performing as designed – that is scores coincided with the intended behaviors captured in the videos – heifers were categorized and compared based on their IPS on a given day being less than 2.0 (n = 413), equal to 2.0 but less than 3.0 (n = 132), equal to 3.0 but less than 4.0 (n = 55), equal to 4.0 but less than 5.0 (n = 34), and greater than or equal to 5.0 (n =19). The GPS were much lower on average, with no average GPS being greater than 4.0. Therefore, GPS categories were compared using average GPS less than 2.0 (n = 119), equal to 2.0 but less than 3.0 (n = 38), and 3.0 or greater (n = 9). When assessing change in response for IPS, F and IN were compared by summing each category for their three days of observation ($d_{1,1}$, $d_{2,1}$, and $d_{3,1}$), and for F by evaluating response across all nine days of observation. Since a group of heifers could contain both F and IN animals, change in response was simply done by comparing group totals for the first day of each event. Due to this confounding, comparison of behavior for F across all nine days were not included. Contingency tables and P-values were obtained using the "coin" (Hothorn et al., 2008) and "rcompanion" (Mangiafico, 2018) packages in R, which account for one or both of the variables being ordered.

Video analyses - continuous ethograms. All other IPS ethograms were analyzed independently for change over time with the same models previously described using PROC GLIMMIX in SAS. However, to investigate the validity of ethograms, either IPS category was included as yet another fixed effect in the model.

As mentioned, a single group could include any combination of F or IN heifers on the first day of each event. Therefore, the model used to compare GPS ethograms was reduced to include the fixed effect of GPS category and event, with body weight as a covariate. Year, sire, and group nested within year were then treated as random effects. There was no comparison of ethograms across all nine days of the study due to specific F and IN heifers being intermixed on the first day of an event.

Results

Inter-observer reliability

Individual pen score had an average kappa coefficient and intra-class correlation of 0.64 and 0.92, respectively. Reliabilities of GPS were slightly lower, with an average kappa coefficient and intra-class correlation of 0.44 and 0.77, respectively. However, for both IPS and GPS, kappa coefficients and intra-class correlations were above reported threshold values for acceptable reliability, namely > 0.40 (Landis and Koch, 1977) and > 0.70 (Martin and Bateson, 1993), respectively, indicating accurate evaluation.

Table 5.1. Correlations of individual and group pen score with objective measurements of temperament

		Individual Pen Score		Group Pen Score	
Measure	*n*	r	SE	r	SE
Temperature, ˚C	350	0.37	0.05	0.30	0.05
Heart Rate, bpm	351	0.28	0.05	0.29	0.05
BUN[1], mg/dl	351	0.04	0.05	0.02	0.05
CK, units/l	351	0.10	0.05	0.14	0.05
Glucose, mg/dl	351	0.28	0.05	0.28	0.05
NEFA, mmol/l	351	-0.23	0.05	-0.13	0.05
Plasma Cortisol, ng/ml	350	0.28	0.05	0.30	0.05
Fecal Cortisol[1], ng/ml	344	-0.03	0.05	-0.02	0.05

[1]Correlations not different from zero ($P > 0.05$)

Relationship between measurements

Correlations for IPS and GPS with objective measures of temperament are provided in Table 5.1. Pen scores were moderately positively correlated with body temperature, heart rate, glucose concentration, and serum cortisol; there was a low positive correlation with creatine kinase. However, correlations with blood urea nitrogen and fecal cortisol were not different from zero. Finally, slight to moderate negative correlations existed for both IPS and GPS with NEFA concentrations. These correlations are consistent with correlations of these same objective measures with CS, ES, and EV previously reported by Parham et al. (2018b).

Correlations for IPS and GPS with CS, ES, and EV for $d_{1,1}$, $d_{2,1}$, and $d_{3,1}$ separately and combined are given in Table 5.2. Overall, strong correlations existed between IPS and GPS with ES and EV, as they are both non-restrained measures of temperament, with moderate correlations between IPS and GPS with CS. The lowest correlations existed for IPS and GPS with CS on the first day of observation, but increased over time. This pattern was not present for ES and EV, although it tended to be lower on $d_{1,1}$ compared to $d_{2,1}$, and $d_{3,1}$.

Table 5.2. Correlations of individual and group pen score with chute score, exit score, and exit velocity

	Individual Pen Score			Group Pen Score		
Day[1]	CS	ES	EV	CS	ES	EV
$d_{1,1}$	0.26 ± 0.09[a]	0.60 ± 0.07	0.45 ± 0.08[a]	0.15 ± 0.09*[a]	0.42 ± 0.08	0.33 ± 0.09
$d_{2,1}$	0.41 ± 0.09[ab]	0.70 ± 0.07	0.55 ± 0.08[ab]	0.41 ± 0.09[b]	0.54 ± 0.08	0.41 ± 0.09
$d_{3,1}$	0.54 ± 0.08[b]	0.67 ± 0.07	0.63 ± 0.07[b]	0.47 ± 0.08[b]	0.53 ± 0.08	0.45 ± 0.08
All[2]	0.42 ± 0.05	0.65 ± 0.04	0.54 ± 0.05	0.36 ± 0.05	0.50 ± 0.05	0.40 ± 0.05

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

[2] Correlations of all 3 days combined cannot be compared to $d_{1,1}$, $d_{2,1}$, or $d_{3,1}$

[a,b] Means in a column with differing superscripts differ ($P < 0.05$)

* Correlation not different from zero ($P > 0.05$)

Acclimation to handling

When comparing measurement protocols for change in IPS over time, there was an interaction of measurement protocol and event ($P = 0.03$). Therefore, Figure 5.3 provides the mean IPS for F and IN handled heifers on the first day of each event. Individual pen scores did not differ between F and IN on $d_{1,1}$, but decreased in the F group from $d_{1,1}$ to $d_{2,1}$ ($P = 0.03$) and remained constant from $d_{2,1}$ to $d_{3,1}$ ($P = 0.99$). However, IPS on $d_{2,1}$ and $d_{3,1}$in F were not different from those for IN on those same days, although their values were numerically smaller. Final IPS on $d_{3,1}$ for the F and IN groups was 1.35 ± 0.05 and 1.75 ± 0.07, respectively.

Conversely, there was no effect of event, measurement protocol, or their interaction on GPS ($P > 0.26$). Mean GPS for all heifers on the first day of each event were 1.72 ± 0.05, 1.58 ± 0.04, and 1.51 ± 0.04 for $d_{1,1}$, $d_{2,1}$, and $d_{3,1}$, respectively.

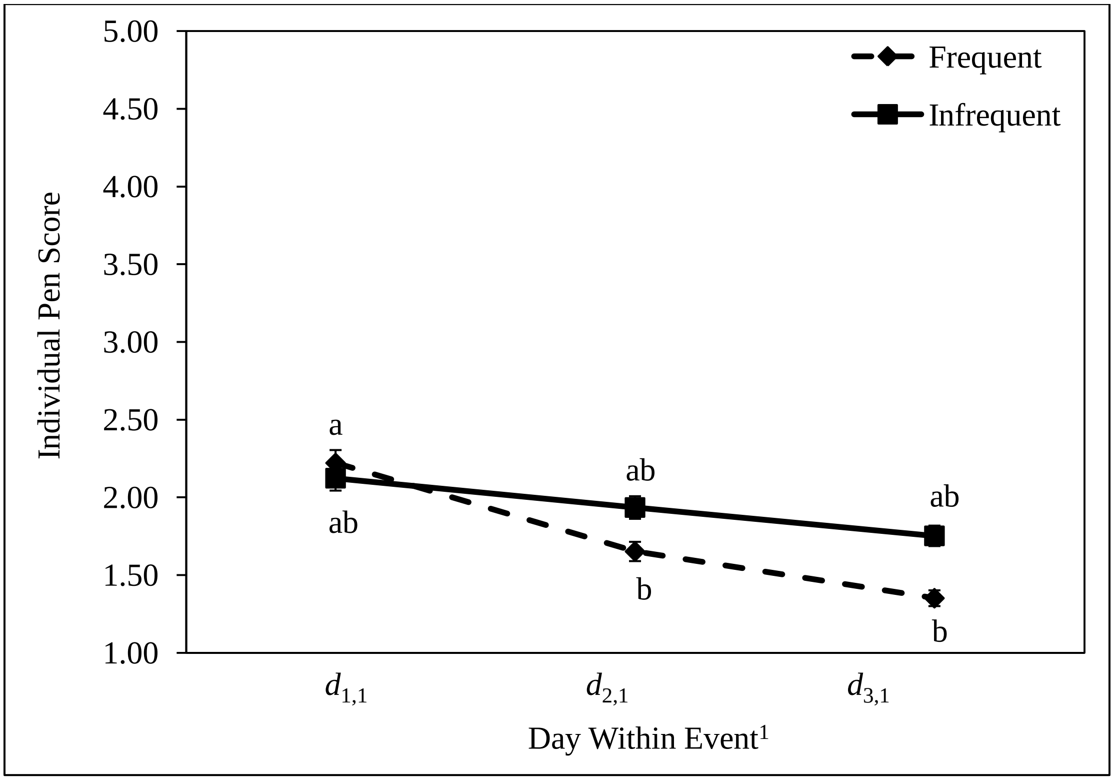
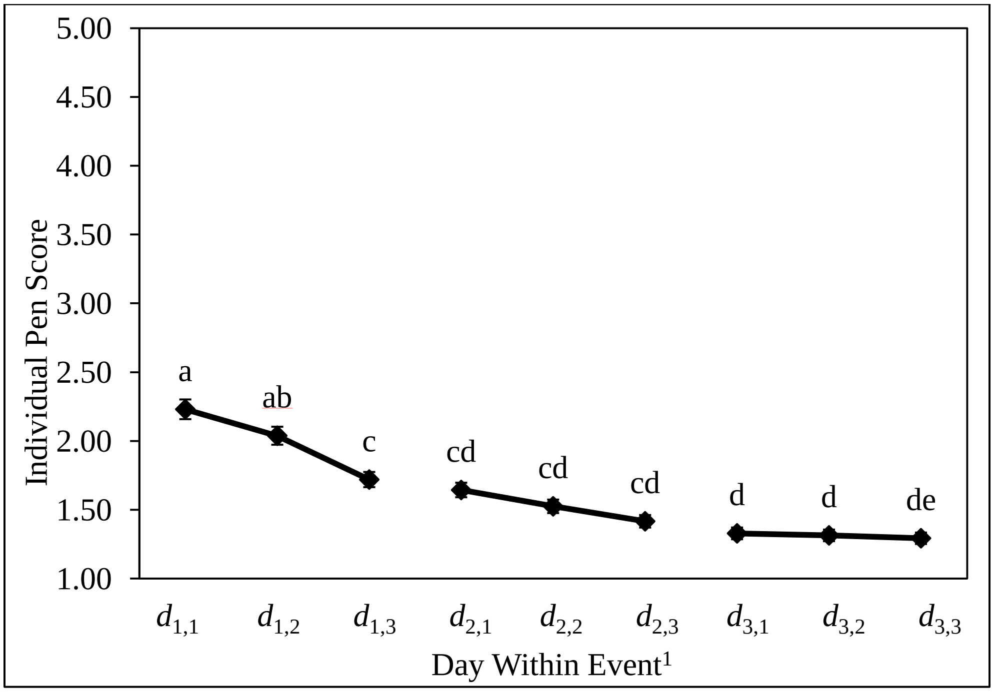


Figure 5.3. Change in individual pen score over time for frequently and infrequently handled heifers [[a,b] Means with differing letters differ ($P < 0.05$); [1]Day within event is designated by $d_{i,j}$, where i is the event and j is the day within an event]

When assessing change in measurements across days for F, there was a decrease in IPS across both events ($P < 0.01$) and days ($P < 0.01$). Individual pen score decreased ($P = 0.01$) from 1.98 ± 0.06 on $d_{1,1}$ to 1.53 ± 0.05 on $d_{2,1}$, but only tended to decrease from $d_{2,1}$ to 1.31 ± 0.04 on $d_{3,1}$ ($P = 0.07$). Figure 5.4a shows the change in IPS across days, with a numerical decrease between $d_{1,1}$ and $d_{1,2}$, but a significant difference between $d_{1,1}$ and all other days. Following $d_{1,2}$, IPS numerically decreased during the second event ($d_{2,1}$, $d_{2,2}$, and $d_{2,3}$), with that difference becoming more substantial when compared to the last three days of observation ($P < 0.05$). By $d_{3,3}$, IPS reduced to 1.29 ± 0.04, which was less than at $d_{1,1}$ through $d_{2,1}$ ($P < 0.05$). These trends were not observed in GPS ($P > 0.10$), with mean values for event 1 to 3 of 1.68 ± 0.06, 1.48 ± 0.06, and 1.39 ± 0.05, respectively. Mean GPS across days are provided in Figure 5.4b. A low GPS of 1.69 ± 0.06 on $d_{1,1}$ left little room to decrease over time. However, GPS on $d_{3,3}$ was numerically lower, with a value of 1.32 ± 0.05.

(a) Individual Pen Score



(b) Group Pen Score

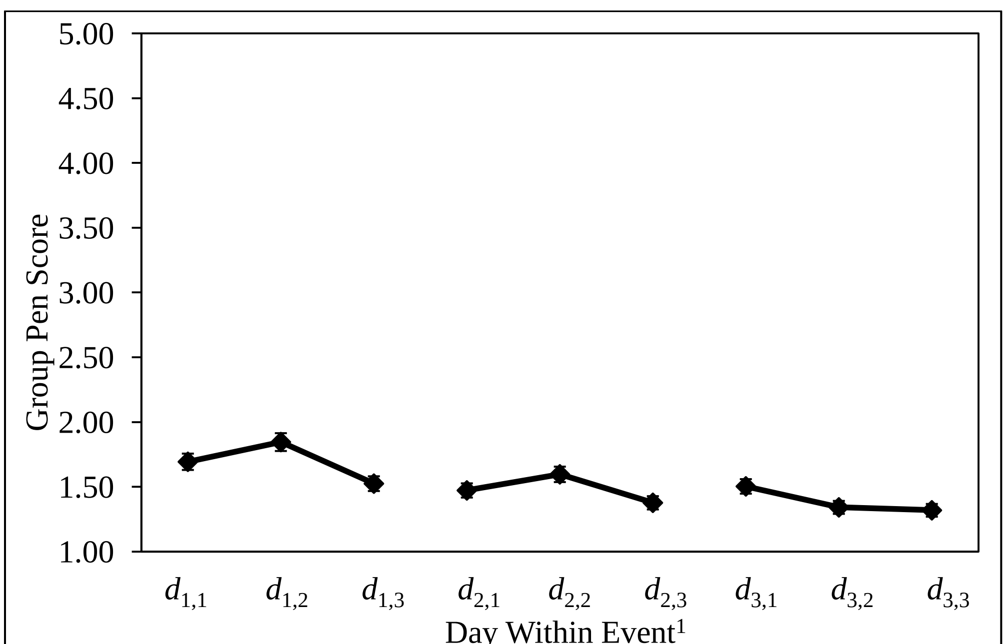


Figure 5.4. Change in pen scores across days for frequently handled heifers. [[ab] Data points with differing letter assignments differ ($P < 0.05$); [1]Day within event is designated by $d_{i,j}$, where i is the event and j is the day within an event]

Table 5.3. Proportion of heifers in each individual pen category by ethogram[1]

		IPS Category[2]				
		1	2	3	4	5+
Ethogram	Category	*n* = 413	*n* = 132	*n* = 55	*n* = 34	*n* = 19
Flight Zone Size	Low	0.70	0.08	0.02	0	0
	Medium	0.30	0.86	0.40	0	0.05
	High	0	0.05	0.58	1.00	0.95
Fight or Flight	Neutral	0.78	0.14	0	0	0
	Flight	0.22	0.85	0.98	0.79	0.42
	Fight	0	0.01	0.02	0.21	0.58
Fastest Pace	Stationary	0.12	0.01	0	0	0
	Walk	0.54	0.05	0	0	0
	Trot	0.33	0.80	0.55	0.21	0.11
	Canter	0.01	0.14	0.45	0.79	0.89
Human Entry	Safe to Enter	1.00	1.00	0.84	0.53	0.21
	Unsafe	0	0	0.16	0.47	0.79

[1]All chi-square different than expected ($P < 0.001$)

[2] Individual Pen Score Categories: IPS < 1.9 = 1; 2.0 < IPS < 2.9 = 2; 3.0 < IPS < 3.9 = 3; 4.0 < IPS < 4.9 = 4; IPS > 5.0 = 5+

Validity of video ethograms

Contingency tables for IPS relative to those ethograms involving categories are provided in Table 5.3 as proportions of the total number of observations in each category. All chi-square tests were significant ($P < 0.001$) meaning observed proportions in each category were different than expected. Heifers with an IPS of 1 had a low flight zone size, responded neutrally to a human stressor, were more likely to walk around the pen, and were calm enough in demeanor that the human felt safe to enter the pen with them. Heifers appeared less calm as IPS increased in value. Flight zone sizes increased to medium, heifers became more "flighty" than neutral, were more likely to trot around the pen, and in some cases were flighty enough that the human did not feel safe enough to enter the pen. Lastly, those heifers with IPS of 4 or greater had high flight zone sizes, appeared more aggressive, expressing their fight instinct, were more likely to canter/run in the pen, and the human was much less likely to enter the pen with them.

Table 5.4. Proportion of heifers in each group pen category by ethogram

		GPS Category[2]		
		1	2	3+
Ethogram	Category	*n* = 119	*n* = 38	*n* = 9
Flight Zone Size[1]	Low	0.49	0.08	0
	Medium	0.48	0.53	0.33
	High	0.03	0.39	0.67
Fastest Pace[1]	Stationary	0.01	0	0
	Walk	0.23	0	0
	Trot	0.52	0.21	0
	Canter	0.25	0.79	1.00
Location, Beginning	Opposite, Far	0.71	0.81	0.78
	Direct, Far	0.07	0.11	0.11
	Opposite, Near	0.01	0	0
	Inner Area	0.21	0.08	0.11
Location, Middle	Opposite, Far	0.71	0.63	0.44
	Direct, Far	0.16	0.26	0.33
	Opposite, Near	0.04	0.08	0.11
	Inner Area	0.08	0.03	0.11

[1]Chi-square different than expected ($P < 0.001$)

[2] Group Pen Score Categories: GPS < 1.9 = 1; 2.0 < GPS < 2.9 = 2; GPS > 3.0 = 3+

Contingency tables for categorical ethograms by GPS are given in Table 5.4, also as proportions of the total number of observations in each category. Average GPS did not seem to impact the location of the group at the beginning of the video, or when the human reached the center of the pen. Regardless, heifers were most commonly in the opposite, far corner to the human. However, there was an association between GPS and flight zone size and the fastest pace reached by the group ($P < 0.001$). Groups with average pen scores between 1.0 and 2.0 were more likely to have a low or medium flight zone size, but compared to IPS, their most common fastest pace was a trot, not a walk. Conversely, groups with average pen scores greater than or equal to 3.0 had high flight zone sizes and always cantered around the pen.

When comparing F and IN on $d_{1,1}$, $d_{2,1}$, and $d_{3,1}$ only as well as F over all nine days, IPS category had an effect on all ethogram response variables ($P < 0.05$), with the exception of the time spent turning around. Therefore, least squares means and SE within each IPS category are reported for $d_{1,1}$, $d_{2,1}$, and $d_{3,1}$ only. Table 5.5 provides mean values for each of the counted behaviors by IPS category. All counted behaviors increased in frequency with increasing IPS, with heifers with IPS greater than 5.0 consistently having the highest values. Interestingly, the number of times heifers head butt the fence, hit the fence, and flick their tails significantly increased ($P < 0.05$) when comparing an IPS between 3.0 and 4.0 to an IPS of 4.0 and higher. Likewise, the number of threats is highest and most likely to occur in heifers with an IPS greater than 5.0.

Table 5.5. Least squares means of the number of times each behavior was observed in each individual pen score category for $d_{1,1}$, $d_{2,1}$, and $d_{3,1}$ only.

	IPS Category[1]				
Behavior	1	2	3	4	5+
Escape	0.02 ± 0.06^{a}	0.19 ± 0.09^{a}	0.25 ± 0.13^{ac}	0.87 ± 0.18^{b}	0.94 ± 0.23^{bc}
Head Butt Fence	0.01 ± 0.04^{a}	0.01 ± 0.06^{ab}	0.25 ± 0.08^{b}	1.19 ± 0.11^{c}	1.64 ± 0.15^{c}
Hit Fence	0.01 ± 0.10^{a}	0.09 ± 0.14^{a}	0.79 ± 0.19^{b}	3.52 ± 0.27^{c}	3.59 ± 0.35^{c}
Tail Flick	0.96 ± 0.38^{a}	1.87 ± 0.54^{a}	5.52 ± 0.77^{b}	14.88 ± 1.09^{c}	17.27 ± 1.42^{c}
Threats	0.00 ± 0.03^{a}	0.01 ± 0.04^{a}	0.04 ± 0.06^{a}	0.44 ± 0.08^{b}	1.43 ± 0.10^{c}

[ab] Rows with differing superscripts differ ($P < 0.05$)

[1] Individual Pen Score Categories: IPS < 1.9 = 1; 2.0 < IPS < 2.9 = 2; 3.0 < IPS < 3.9 = 3; 4.0 < IPS < 4.9 = 4; IPS > 5.0 = 5+

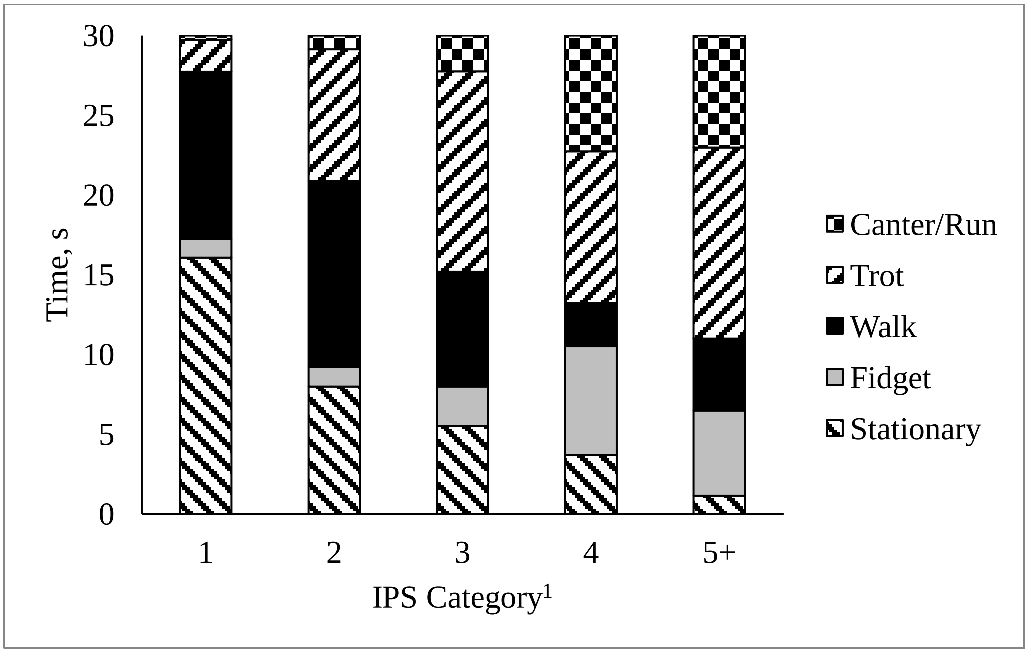


Figure 5.5. Least squares means of the time (s) heifers within each individual pen score category spent exhibiting each pace on $d_{1,1}$, $d_{2,1}$, and $d_{3,1}$ only. [[1] Individual Pen Score Categories: IPS < 1.9 = 1; 2.0 < IPS < 2.9 = 2; 3.0 < IPS < 3.9 = 3; 4.0 < IPS < 4.9 = 4; IPS > 5.0 = 5+]

Least squares means for the amount of time spent exhibiting each pace by individual pen score category on $d_{1,1}$, $d_{2,1}$, and $d_{3,1}$ only were standardized to 30 s and shown in Figure 5.5. Heifers with an IPS less than 2.0 spent more time standing still than any other IPS category ($P < 0.05$). Heifers with IPS greater than 1.0 but less than 3.0 spent less time fidgeting than those with IPS greater than 3.0 ($P < 0.05$), and spent more time walking than those with an IPS between 4.0 and 5.0 ($P < 0.05$). Time spent trotting was least for heifers with an IPS less than 2.0, followed by those with IPS between 2.0 and 3.0; heifers with IPS greater than 3.0 spent the most time trotting ($P < 0.05$). Finally, heifers with IPS greater than 4.0 spent more time cantering/running than those with an

IPS between 3.0 and 4.0 ($P < 0.05$), and both spent more time than those with an IPS less than 3.0 ($P < 0.05$).

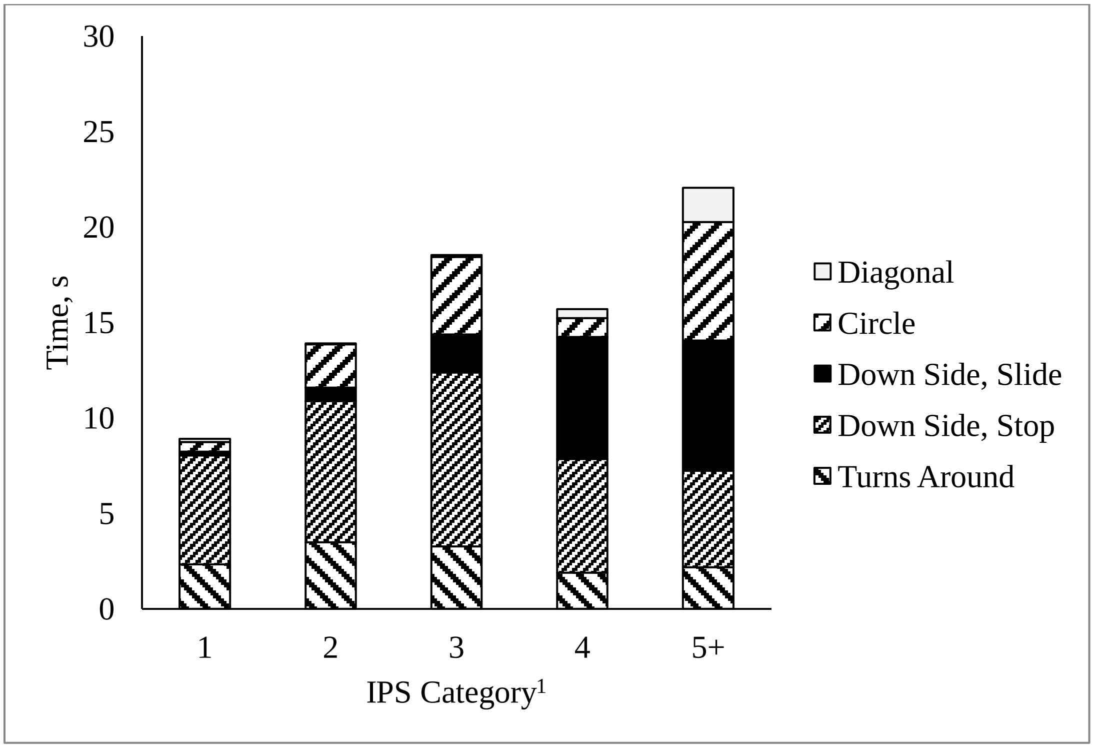


Figure 5.6. Least squares means of the time (s) each individual pen score category spent exhibiting each behavior for $d_{1,1}$, $d_{2,1}$, and $d_{3,1}$ only. [[1] Individual Pen Score Categories: IPS < 1.9 = 1; 2.0 < IPS < 2.9 = 2; 3.0 < IPS < 3.9 = 3; 4.0 < IPS < 4.9 = 4; IPS > 5.0 = 5+]

Least squares means for the amount of time spent exhibiting each direction of movement by individual pen score category on $d_{1,1}$, $d_{2,1}$, and $d_{3,1}$ only are provided in Figure 5.6. Expectedly, heifers with higher IPS spent more time in motion, with the exception of heifers with an IPS between 4.0 and 5.0. While the time heifers spent turning around is not significantly different among IPS categories, this is the only direction of movement where this was true. Heifers with an IPS less than 3.0 spent most of their time in motion moving down the side of the pen and stopping in the corner, suggesting they were moving at a slow enough pace to stop themselves without sliding. This can be compared to heifers with an IPS greater than 4.0 that, instead of stopping, slid into the corner of the pen. Comparatively, the time spent moving down the side and stopping numerically increased as IPS category increased from 1.0 to 3.0, with a clear difference between heifers with an IPS less than 2.0 and those between 3.0 and 4.0 ($P < 0.05$). However, this movement decreased in heifers with IPS greater than 4.0, as they instead spend significantly more time sliding than all other IPS categories ($P < 0.05$). Time spent circling as well as moving diagonally through the pen is greater in heifers with IPS greater than 5.0 compared to any other category ($P < 0.05$). It can be hypothesized that since these extreme heifers did not usually have a human in the pen with them, they were not deterred from running through the middle of the pen as was the case with other heifers.

There was no significant effect of GPS category on where the majority of each group of heifers focused their attention, averaging 11.63 ± 2.66, 5.44 ± 1.81, and 11.04 ± 2.95 s on the human in the pen, the camera, or neither, respectively, for all three categories. The GPS category did, however, have a significant impact on whether a group of heifers rejoined each other following separation in the pen, as shown in Figure 5.7. Heifers with a GPS greater than 3.0 rejoined more often than those with a GPS less than 3.0 ($P < 0.05$). Although not significant, this was reflected in the willingness of the group to remain separated in the pen. Groups with average pen scores less than 3.0 were more likely to remain separated than those with average GPS greater than 3.0. Overall, the time each group spent together is not different among GPS category.

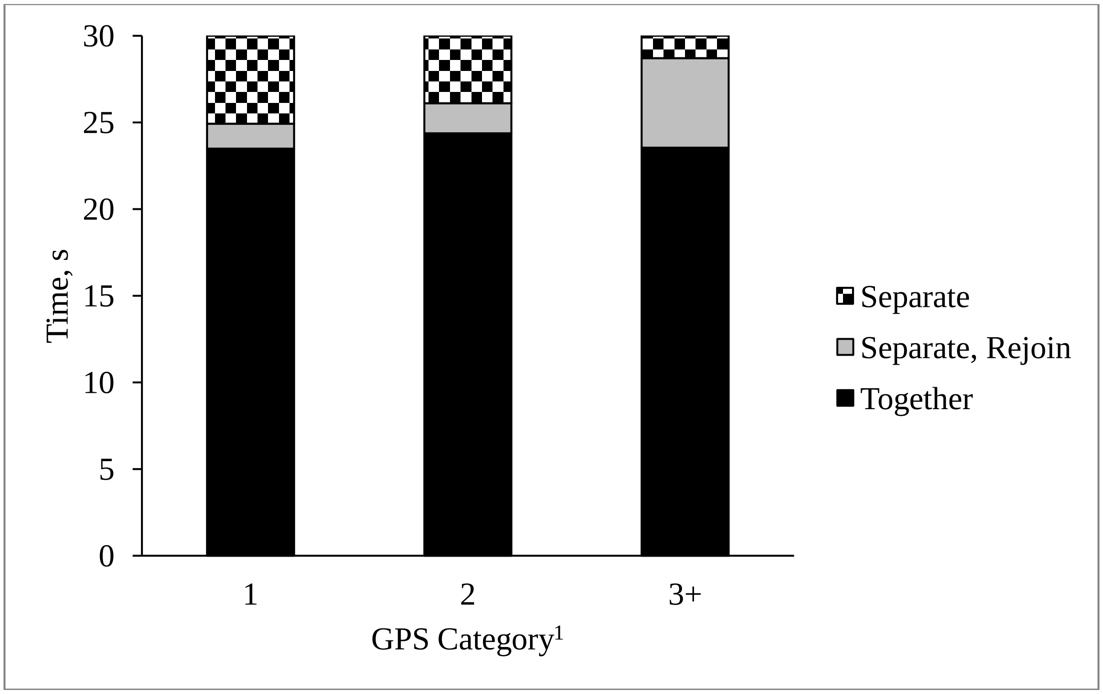


Figure 5.7. Least squares means of the time (s) heifers within each group pen score category spent separated or together on $d_{1,1}$, $d_{2,1}$, and $d_{3,1}$ only. [[1]Group Pen Score Categories: GPS < 1.9 = 1; 2.0 < GPS < 2.9 = 2; GPS > 3.0 = 3+]

Finally, least squares means for the amount of time spent exhibiting each pace is provided in Figure 5.8 on $d_{1,1}$, $d_{2,1}$, and $d_{3,1}$ only by group pen score category. The effect of GPS category was significant for all paces except for the amount of time spent fidgeting. Overall, groups with an average pen score less than 2.0 spent the most amount of time standing still and walking, but the least amount of time trotting and cantering compared to other categories ($P < 0.05$). Those with average GPS greater than 2.0 did not differ in the amount of time they were stationary, walking, or trotting. However, groups with average pen scores greater than 3.0 did spend the most time cantering compared to other categories ($P < 0.05$).

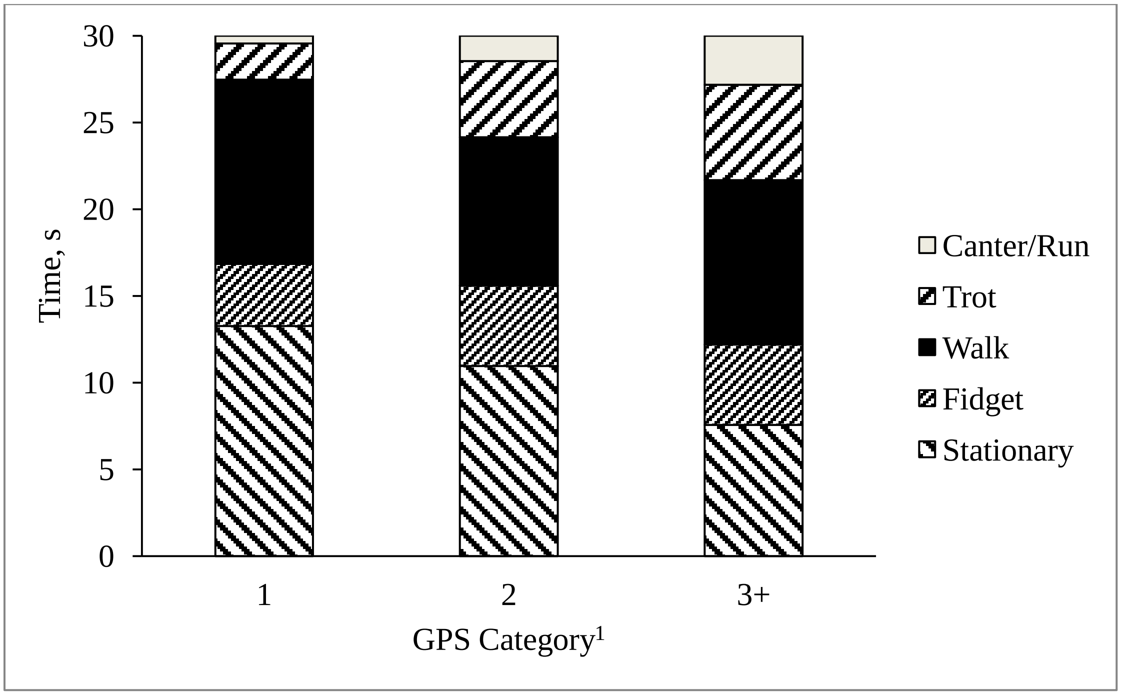


Figure 5.8. Least squares means for the amount of time spent exhibiting each pace by group pen score category on $d_{1,1}$, $d_{2,1}$, and $d_{3,1}$. [[1]Group Pen Score Categories: GPS < 1.9 = 1; 2.0 < GPS < 2.9 = 2; GPS > 3.0 = 3+]

Analysis of ethograms for differences in behavior and change over time

Contingency tables comparing F and IN measurement protocols by summing heifers in each category over $d_{1,1}$, $d_{2,1}$, and $d_{3,1}$ are given in Table 5.6. There was no difference between F and IN for fastest pace or whether the human felt safe enough to enter the pen (*P* > 0.18). However, flight zone size and the attitude of the heifer were different than expected for the two groups (*P* = 0.03) based on chi-square values. Frequently handled heifers had low to medium flight zone sizes, while approximately half of heifers in IN had medium sized flight zone sizes. Lastly, heifers in F were more likely to seem neutral when in the presence of a human stressor compared to IN, which had more heifers in the flight category.

Table 5.6. Proportion of heifers, by measurement protocol, in each individual pen score ethogram category across $d_{1,1}$, $d_{2,1}$, and $d_{3,1}$ of the study.

		Measurement Protocol	
Ethogram	Category	Frequent	Infrequent
Flight Zone Size[1]	Low	0.42	0.29
	Medium	0.42	0.51
	High	0.16	0.20
Fight or Flight[1]	Neutral	0.51	0.36
	Flight	0.44	0.63
	Fight	0.04	0.02
Fastest Pace	Stationary	0.04	0.04
	Walk	0.35	0.25
	Trot	0.40	0.52
	Canter/Run	0.20	0.20
Human Entry	Safe to Enter	0.93	0.97
	Unsafe	0.07	0.03

[1]Chi-square different than expected ($P < 0.05$)

Contingency tables comparing IPS ethograms of F heifers across days are given in Table 5.7. Based on chi-square statistics, flight zone size, attitude of the heifer, and the fastest pace reached were different than expected ($P < 0.001$). Over time, the proportion of heifers with medium and high flight zone sizes decreased, resulting in an increase in the proportion of heifers with low flight zone sizes. Similarly, the proportion of heifers with a flight or fight instinct decreased while the proportion of heifers that appeared neutral increased. Lastly, the proportion of heifers that had a fastest pace of either stationary or walk increased, as a result of a decrease in the amount of heifers that had fastest pace of either trot or canter/run.

Table 5.7. Proportion of frequently handled heifers, by day, in each individual pen score ethogram category.

Ethogram	Category	Day Within Event[2]								
		$d_{1,1}$	$d_{1,2}$	$d_{1,3}$	$d_{2,1}$	$d_{2,2}$	$d_{2,3}$	$d_{3,1}$	$d_{3,2}$	$d_{3,3}$
Flight Zone Size[1]	Low	0.19	0.36	0.47	0.47	0.54	0.53	0.64	0.78	0.76
	Medium	0.63	0.52	0.39	0.36	0.37	0.38	0.23	0.14	0.19
	High	0.18	0.13	0.14	0.17	0.08	0.09	0.13	0.09	0.05
Fight or Flight[1]	Neutral	0.26	0.54	0.51	0.59	0.61	0.60	0.72	0.81	0.76
	Flight	0.68	0.43	0.44	0.36	0.34	0.36	0.26	0.19	0.24
	Fight	0.05	0.04	0.05	0.05	0.05	0.04	0.02	0	0
Fastest Pace[1]	Stationary	0.02	0.02	0.05	0.02	0.10	0.11	0.11	0.24	0.19
	Walk	0.16	0.27	0.34	0.43	0.41	0.36	0.49	0.55	0.52
	Trot	0.42	0.55	0.53	0.43	0.44	0.47	0.34	0.16	0.24
	Canter/Run	0.40	0.16	0.08	0.12	0.05	0.05	0.06	0.05	0.05
Human Entry	Safe to Enter	0.93	0.91	0.90	0.93	0.93	0.93	0.94	0.93	0.98
	Unsafe	0.07	0.09	0.10	0.07	0.07	0.07	0.06	0.07	0.02

[1]Chi-square different than expected ($P < 0.001$)

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

Although heifers had a decrease in their IPS across events, when analyzing change in counted behaviors over time, there was no effect of event, measurement protocol, or their interaction when comparing F and IN heifers ($P > 0.15$). The exception was a decrease in the amount of time spent running down the side and sliding from $d_{1,1}$ to $d_{2,1}$ ($P = 0.04$). Mean values for each ethogram for the first day of each event are provided in Table 5.8.

Table 5.8. Least squares means and SE for individual pen score ethograms on $d_{1,1}$, $d_{2,1}$, and $d_{3,1}$.

	Day Within Event[1]			
	$d_{1,1}$	$d_{2,1}$	$d_{3,1}$	SE
Counted Behavior				
Escape Behavior	0.36	0.52	0.48	0.10
Head Butts Fence	0.70	0.55	0.61	0.07
Hits Fence	1.58	1.64	1.57	0.16
Tail Movement	8.86	7.99	7.41	0.60
Threats	0.35	0.42	0.38	0.04
Pace (s)				
Stationary	4.62	5.81	4.40	1.42
Fidget	3.04	2.45	2.68	0.72
Walk	4.16	5.77	6.03	0.93
Trot	6.93	6.56	7.13	0.47
Canter/Run	3.26	2.67	2.67	0.44
Direction of Movement (s)				
Turns Around	3.18	3.53	2.98	0.69
Down Side, Stop	6.26	6.62	7.05	0.96
Down Side, Slide	3.91[a]	2.59[b]	3.18[ab]	0.31
Circle	1.65	3.21	3.54	0.61
Diagonal	0.51	0.53	0.52	0.17

[ab] Rows with differing superscripts differ ($P < 0.05$)
[1]Day within event is designated by $d_{i,j}$, where i is the event and j is the day within an event

Table 5.9. Least squares means and SE for individual pen score ethograms for the frequently handled heifers across days.

	Day Within Event[1]									SE
	$d_{1,1}$	$d_{1,2}$	$d_{1,3}$	$d_{2,1}$	$d_{2,2}$	$d_{2,3}$	$d_{3,1}$	$d_{3,2}$	$d_{3,3}$	
Counted Behavior										
Escape Behavior	0.36	0.33	0.28	0.45	0.34	0.26	0.27	0.37	0.29	0.08
Head Butts Fence	0.50	0.43	0.39	0.42	0.43	0.52	0.50	0.47	0.50	0.08
Hits Fence	1.18	1.06	1.13	1.22	1.08	1.09	1.12	1.14	1.23	0.15
Tail Movement	9.14	8.91	8.68	8.51	7.59	7.60	7.77	7.74	7.41	0.64
Threats	0.39	0.37	0.47	0.51	0.40	0.44	0.38	0.45	0.50	0.07
Pace (s)										
Stationary	5.41[a]	6.25[a]	9.76[ab]	9.19[ab]	9.36[ab]	10.70[ab]	8.16[ab]	11.46[ab]	13.09[b]	2.11
Fidget	2.56	2.73	1.97	1.97	2.09	1.60	1.85	0.81	1.70	0.53
Walk	4.40	6.53	5.39	6.03	5.65	5.47	5.20	3.23	4.38	1.15
Trot	8.01	7.57	7.16	6.58	6.40	6.83	7.09	5.99	6.43	0.66
Canter/Run	2.16	1.26	1.30	1.65	1.52	1.76	1.81	1.85	1.80	0.34
Direction of Movement (s)										
Turns Around	3.87[ab]	4.40[a]	1.55[b]	2.53[ab]	2.10[ab]	2.06[ab]	2.26[ab]	1.18[b]	2.30[ab]	0.70
Down Side, Stop	5.85	6.42	9.78	6.83	7.47	5.95	6.04	6.36	6.05	1.18
Down Side, Slide	3.62	2.38	2.53	2.38	2.38	2.87	3.22	2.85	2.50	0.43
Circle	2.10	2.32	3.34	3.77	2.78	3.34	3.16	2.21	2.42	0.81
Diagonal	0.69	1.04	1.05	0.73	1.15	0.78	0.84	0.57	0.57	0.36

[ab] Rows with differing superscripts differ ($P < 0.05$)

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

When comparing F heifers only, IPS ethogram values did not often change over events or days ($P > 0.08$), except for the amount of time heifers remained stationary, and the amount of time they spent turning around. Mean values across days for all IPS ethograms are in Table 5.9. The amount of time spent standing still increased from $d_{1,1}$ to $d_{3,1}$, while the amount of time spent turning around decreased from $d_{1,2}$ to $d_{3,2}$, but were not different from $d_{3,3}$.

Table 5.10. Proportion of heifers in each group pen category by ethogram

		Day Within Event[1]		
Ethogram	Category	$d_{1,1}$	$d_{2,1}$	$d_{3,1}$
Flight Zone Size	Low	0.33	0.30	0.38
	Medium	0.57	0.53	0.50
	High	0.10	0.17	0.12
Fastest Pace	Stationary	0	0	0
	Walk	0.10	0.10	0.12
	Trot	0.40	0.43	0.48
	Canter	0.50	0.47	0.40
Location, Beginning	Opposite, Far	0.90	0.73	0.77
	Direct, Far	0.03	0.07	0.12
	Opposite, Near	0	0.03	0
	Inner Area	0.07	0.17	0.12
Location, Middle	Opposite, Far	0.67	0.80	0.62
	Direct, Far	0.23	0.10	0.23
	Opposite, Near	0.03	0.07	0.08
	Inner Area	0.07	0.03	0.08

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

Contingency tables comparing the proportion of groups within each GPS ethogram category on $d_{1,1}$, $d_{2,1}$, and $d_{3,1}$ are given in Table 5.10. None of the proportions were different than expected ($P > 0.39$) based on chi-square statistics. Furthermore, there was no change in any of the ethograms when comparing all heifers on the first day of each event, as shown in Table 5.11.

Table 5.11. Least squares means and SE for all heifers on $d_{1,1}$, $d_{2,1}$, and $d_{3,1}$ for different group pen score ethograms.

	Day Within Event[1]			
Ethogram	$d_{1,1}$	$d_{2,1}$	$d_{3,1}$	SE
Group attention				
On Human	11.37	11.99	11.55	2.72
On Camera	6.21	4.83	5.28	1.87
Other	10.82	10.03	12.29	3.04
Willingness to Separate				
Separate	2.34	3.22	4.62	1.38
Separate, Rejoin	3.64	2.75	1.94	1.09
Together	23.70	23.95	23.32	1.73
Pace (s)				
Stationary	9.29	10.54	11.63	1.85
Fidget	3.73	5.21	3.55	0.90
Walk	9.99	8.43	9.73	0.86
Trot	4.05	3.71	3.78	0.48
Canter/Run	1.61	1.44	1.46	0.23

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

Discussion

In this study, IPS and GPS appear to be acceptable methods to assess stress in an animal during normal handling practices as they were reliably assessed by multiple observers and moderately positively correlated with body temperature, heart rate, glucose, and serum cortisol. Correlations of IPS and GPS with ES and EV were moderately positive across all days of collection; however, correlations with CS increased over time, but overall remained moderately positive. Detailed evaluations of these ethograms showed that animals with higher IPS and average GPS had larger flight zone sizes and reached faster paces for longer periods of time. Furthermore, animals with lower IPS appeared more neutral to the presence of a human, moved around less often, and were less prone to hit the fence or flick their tail; groups with higher average GPS were less willing to separate from their peers.

Flight zone size and the appearance of fight/flight attitude, measured independently, appeared to differ among F and IN groups. This was consistent with IPS over time, as the interaction of measurement protocol and event was significant. Frequently handled heifers appeared to acclimate more substantially to handling, especially across days. With lower starting values, there was no significant change in GPS over time. This possibly was due to calmer temperament from heifers being in a group setting. Overall, while both pen scores can be assessed easily and reliably, IPS resulted in stronger responses to stress and is therefore more useful to producers wishing to make selection decisions. However, heifers acclimate to repeated calm handling.

Therefore, when cattle are excitable during their first handling experience, more than one observation of temperament may be beneficial before making selection decisions.

Individual pen scores and GPS assigned during the study were reliably assessed, with reported values being higher than thresholds for acceptable reliabilities (Landis and Koch, 1977; Martin and Bateson, 1993). In fact, IPS and GPS reliabilites were higher than those for CS assigned by the same individuals (0.46 and 0.74 for kappa and intraclass correlations, respectively; Parham et al., 2018a). Furthermore, these subjective methods were also moderately positively correlated with body temperature, heart rate, glucose, and serum cortisol, all of which are objective measurements known to be associated with stress in animals (Van de Water et al., 2003; Sporer et al., 2008). These correlations are consistent with those reported for CS and ES by Parham et al. (2018b) within the same animals, indicating IPS and GPS were acceptable measurements of stress in these heifers. Similarly, when compared to CS, ES and EV, overall correlations with IPS and GPS were moderately positive overall, but increased from $d_{1,1}$ to $d_{3,1}$ with CS. Lastly, correlations were consistently higher for ES than CS. This is perhaps due to the fact ES, IPS, and GPS are all considered non-restrained methods of evaluating temperament.

All correlations and reliabilities were larger for IPS than GPS. There are two possible explanations to this occurrence. First, cattle are herd animals and therefore the presence of social partners reduces heifers' behavioral signs of disturbance towards fear-eliciting stimulation (Boissy and Le Neindre, 1990). When secluded in an individual pen calves spent more time standing still and were easier to handle with peers present in an adjacent pen than without (Gringard et al., 2000). This reduction in response may lead to lower correlations with measures known to be indicative of stress. It also leads to less variability in the behaviors expressed, making it more difficult to delineate among GPS categories, especially when attempting to evaluate multiple heifers at once.

The amount of variation expressed in a non-restrained test can either be a benefit or a drawback; specific behaviors may help to delineate categories but also could cause confusion if the behavior of an animal does not neatly correspond with a given score. Therefore, detailed ethograms were developed to analyze pen behavior in a more intricate manner using video recordings, with the hope that suggestions could be made to assist producers who wish to select for temperament based on pen score or in some other non-restrained environment.

Both IPS and GPS were impacted by the individual or group's flight zone size. For IPS most heifers assigned a score of 1.0 had low, those assigned a 2.0 mainly had medium, and almost all heifers with an IPS greater than 4.0 had high, flight zone sizes. These flight zones will also impact a heifer's response to human presence. Most heifers with an IPS of 1.0 or 2.0 remained neutral to the presence of a human, while those with IPS of 5.0 or more were more likely to be aggressive. The relationship of these measures with average GPS for all four heifers is not as clear, as the group could be a mix of a number of different IPS combinations. However, groups with an average pen score of 1.0 had smaller flight zone sizes than those with average pen scores greater than 3.0.

Another evaluation tool that resulted in consistent delineation between IPS and GPS categories was the fastest pace reached by the heifer(s) over the 30 s duration of the video. As expected, heifers with lower IPS were more likely to only walk around the pen compared to those with IPS greater than 3.0, whose maximum pace was always either a trot or a canter. Similar patterns were present for GPS categories, as those groups with average pen scores greater than 2.0 never walked or remained stationary, and those with average GPS greater than 3.0 always cantered/ran at some point during the test. When expressing these patterns as a duration of time, similar patterns existed. Heifers with lower IPS spent more time walking and standing still compared to heifers with IPS greater than 4.0 that spent more time trotting and cantering around the pen. Interestingly, the heifers with higher IPS also appear to have spent more time fidgeting.

When assessing an animal's IPS, the number of times a heifer attempted to escape, head butt or hit the fence, flicked their tail, or threatened the human in the center of the pen increased with increasing IPS, and may therefore provide more objective methods for delineating temperament. Heifers assigned a higher IPS also remained in motion for longer periods of time. Two useful observations that helped delineate temperaments were whether the heifers continuously circled the pen, as that occurred more often in more temperamental heifers, and heifers' self-control when moving down the side of the fence. The frequency heifers slid on the gravel into a corner of the pen was different between heifers with an IPS greater or less than 4.0. Those with IPS less than 4.0 had enough control to bring themselves to a complete stop before sliding and/or hitting the fence line while heifers with IPS greater than 4.0 had too much momentum to do so, and therefore slid into the corner, at times hitting the fence.

As previously mentioned, cattle are herd animals. Therefore, when assessing average GPS the willingness of the group to separate, measured by how much time they spend together and apart, may help to delineate categories. While there was no difference in the amount of time spent together, groups of heifers with average GPS greater than 3.0 were more likely to rejoin each other following being separated than those with lower average GPS that often remained separated.

Frequencies of counted behaviors, and timed durations of pace, direction of movement, and willingness to separate, did not change over events or days. Similarly, there was no change in flight zone size, or the fastest pace reached, in the group pen tests suggesting a lack of acclimation to repeated handling. Those results were consistent with analysis of change in individually assigned GPS on the day of observation. Average GPS on the first day of observation (1.69 ± 0.06) was indicative of docile heifers, overall. This is perhaps an indication of reduction in behavioral signs towards fear-eliciting stimulation due to being placed in the presence of peers. This does not imply GPS on the first day of the study is a good indicator of overall temperament. Instead, it could be argued that the presence of peers masked the actual temperament of excitable animals.

Heifers did, however, acclimate to repeated handling in an individual pen setting. Frequently handled heifers decreased in IPS more substantially on the first day of each event compared to IN. Similarly, F had more heifers with a low flight zone size compared to IN, as the proportion of heifers in F with medium and high flight zone sizes decreased from $d_{1,1}$ to $d_{3,3}$. Similarly, the proportion of heifers who remained neutral to the presence of a human stressor followed the same pattern. While the fastest pace reached by each individual heifer did not differ between F and IN, within F, the proportion of heifers that remained stationary or walked increased while the number who trotted or ran decreased. When assessing change in average IPS across days in F, the largest decrease in temperament occurred during the first event, and essentially remained constant for the remainder of the study. This observation is consistent with change in CS within F as reported by Parham et al. (2018c).

In this study, IPS may prove more useful than GPS simply because variations in behavior are not being masked due to heifers being placed in a group setting. Analyzing multiple animals together may not only be cumbersome, especially during a 30 s period, but the increased comfort of the animal may mask responses that would have otherwise been expressed individually. The ability of an animal to fully convey their response to stress has a large implication on the effectiveness of a subjective ethogram. Grandin (2014) warns that when CS are being used, it is important to know how tightly the animal was restrained to prevent movement. Furthermore, possible loss in variation of CS due to restricted movement when catching the head of the animal in the head gate, and/or using the squeeze feature of the side panels, has been discussed as a deterrent for using CS to quantify behavior (Vetters et al., 2013). Individual pen scores are considered a non-restrained test and therefore leave the animal completely free to move and express as they choose, within the confines of the pen. This provides added benefit when attempting to evaluate the temperament of an animal. Heifers that acclimated to handling in this study also acclimated to handling in the chute, but to a larger degree (Parham et al., 2018c). It could be hypothesized that this larger decrease in response is a consequence of the lack of restraint when evaluating IPS compared to CS.

When comparing the usefulness of CS and IPS to measure temperament of an animal, it is important to consider which is easier and safer to implement in a production setting. It is likely that most, if not all, producers will place their cattle in a chute during their first year of life, meaning CS would provide an easy method of quantifying temperament that requires no extra time or effort to utilize. Comparatively, it is less likely that producers normal handling practices require that they are individually secluded in a pen their cattle. It would require more time, resources, and effort to record an IPS on each animal. Furthermore, as proposed by Kilgour et al. (2006), IPS introduces a concern for handler safety that is not present when animals are restrained in a chute. As an illustration, the human in this study did not feel safe enough to enter the pen with a heifer 10% of the time. Although there is more variation in response when using a non-restrained test such as pen score, safety and ease of use should be considered when deciding between methodologies. Ultimately, that decision is in the hands of the producer.

Conclusion

Pen scores collected on heifers either individually or as a group can be reliably measured and are indicative of an animal's response to stress during normal handling practices. However, because cattle are a herd species, responses to stress were muted when temperament was analyzed in a group setting. Therefore, if wishing to make selection decisions, IPS may prove more useful. Producers who are unsure of how to use subjective ethograms to evaluate pen scores can also consider estimating each animal's flight zone size, how much time they spend running compared to walking, or the fastest speed an animal reaches during evaluation. Similarly, counting how many times animals hit or head butt the fence, attempt to escape, or flick their tail may also help delineate temperamental and docile heifers. Overall, these simple behaviors provide producers with yet another method to evaluate temperament in a production setting that is easy, fast, and inexpensive. However, cattle do acclimate to repeated exposure to a human stressor in an individual pen setting. Therefore, when cattle are excitable during their first handling experience, more than one observation of temperament may be beneficial before assessing docility.

Literature Cited

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

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

Boissy, A., Le Neindre, P., 1990. Social influences on the reactivity of heifers: implications for learning abilities in operant conditioning. Appl. Anim. Behav. Sci. 25, 149-165.

Burrow, H. M., Seifert, G. W., Corbet, N. J., 1988. A new technique for measuring temperament in cattle. Proc. of Aust. Soc. of Anim. Prod. 17, 154-157.

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.

Cooke, R. F., Arthington, J. D., Araujo, D. B., Lamb, G. C., 2009. Effects of acclimation to human interaction on performance, temperament, physiological responses, and pregnancy rates of Brahman-crossbred cows. J. Anim. Sci. 87, 4125-4132.

Curley Jr., K. O., Paschal, J. C., Welsh Jr., T. H., Randel, R. D., 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., Goddard, M. E., Seifert, G. W., 1982. The measurement of temperament in cattle and the effect of experience and genotype. Proc. Aust. Soc. Anim. Prod. 14, 329-332.

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

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

Gauly, M., Mathiak, H., Hoffmann, K., Kraus, M., Erhardt, G., 2001. Estimating genetic variability in temperamental traits in German Angus and Simmental cattle. Appl. Anim. Behav. Sci. 74, 109-119.

Grandin, T., 2014. Handling Facilities and Restraint of Extensively Raised Range Cattle. Livestock Handling and Transport, 4th Edition. T. Grandin, ed. CAP Int., Cambridge, MA. Pages 94-115.

Grandin, T., Deesing, M. J., 2014. Genetics and behavior during handling, restraint, and herding. Genetics and the Behavior of Domestic Animals (Second Edition). pp. 115-158.

Gringard, L., Boissy, A., Boivin, X., Garel, J. P., 2000. The social environment influences the behavioural responses of beef cattle to handling. Appl. Anim. Behav. Sci. 68, 1-11.

Hothorn, T., Hornik, K., A. van de Wiel, M., Zeileis, A., 2008. Implementing a Class of Permutation Tests: The coin Package. J. Stat. Softw. 28, 1-23.

Kilgour, R. J., Melville, G. L., Greenwood, P.L., 2006. Individual differences in the reaction of beef cattle to situations involving social isolation, close proximity of humans, restraint and novelty. Appl. Anim. Behav. Sci. 99, 21-40.

King, D. A., Schehle Pfiffer, C. E., Randel, R. D., Welsh Jr., T. H., Oliphint, R. A., Baird, B. E., Curley Jr., K. O., Vann, R. C., Hale, D. S., Savell, J. W., 2006. Influence of animal temperament and stress responsiveness on the carcass quality and beef tenderness of feedlot cattle. Meat Sci. 74, 546-556.

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

Le Neindre, P., Trillat, G., Sapa, J., Ménissier, F., Bonnet, J. N., Chupin, J. M., 1995. Individual differences in docility in Limousin cattle. J. Anim. Sci. 73, 2249-2253.

Mangiafico, S., 2018. rcompanion: Functions to Support Extension Education Program Evaluation. R package version 1.13.2. <https://CRAN.R-project.org/package=rcompanion>.

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

Parham, J. T., Tanner, A. E., Wahlberg, M. L., Grandin, T., Lewis, R. M., 2018a. Subjective methods to quantify temperament in beef cattle are insensitive to the number and biases of observers. Appl. Anim. Behav. Sci. (Submitted 16 July 2018).

Parham, J. T., Tanner, A. E., Wahlberg, M. L., Swecker, Jr., W. S., Lewis, R. M., 2018b. Subjective methods of quantifying temperament in heifers are indicative of physiological stress. Appl. Anim. Behav. Sci. (To be submitted).

Parham, J. T., Tanner, A. E., Barkely, K., Pullen, L., Wahlberg, M. L., Swecker, Jr., W. S., Lewis, R. M., 2018c. Temperamental cattle acclimate more substantially to repeated handling. Appl. Anim. Behav. Sci. (Submitted 12 July 2018).

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

Sant'Anna, A. C., Paranhos da Costa, M. J. R., Baldi, F., Albuquerque., L. G., 2013. Genetic variability for temperament indicators of Nelore cattle. J. Anim. Sci. 91, 3532-3537.

Sporer, K. R. B., Weber, P. S. D., Burton, J. L., Earley, B., Crowe, M. A., 2008. Transportation of young beef bulls alters circulating physiological parameters that may be effective biomarkers of stress. J. Anim. Sci. 86, 1325-1334.

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

Turner, S. P., Navajas, E. A., Jyslop, J. J., Ross, D. W., Richardson, R. I., Prieto, N., Bell, M., Jack, M. C., Roehe, R. 2011. Associations between response to handling and growth and meat quality in frequently handled Bos taurus beef cattle. J. Anim. Sci. 89, 4329-4248.

Van de Water, G., Verjans, F., Geers, R., 2003. The effect of short distance transport under commercial conditions on the physiology of slaughter calves; pH and colour profiles of veal. Livest. Prod. Sci. 82, 171-179.

Vargas Jurado, N., Tanner, A. E., Blevins, S. R., Rich, J., Mayes, R. W., Fiske, D., Swecker Jr., W. S., Lewis, R. M., 2015. Feed intake and diet selection in Angus-cross heifers of two frame sizes at two stages of growth. J. Anim. Sci. 93, 1565-1572.

Vetters, M. D. D., Engle, T. E., Ahola, J. K., Grandin, T., 2013. Comparison of flight speed and exit score as measurements of temperament in beef cattle. J. Anim. Sci. 91, 374-381.

Von Borell, E., Dobson, H., Prunier, A., 2007. Stress, behaviour and reproductive performance in female cattle and pigs. Horm. Behav. 5, 130-138.

APPENDIX A: INDIVIDUAL PEN ETHOGRAMS

Ethogram	Type	Description
Individual Pen Score	Overall	1. Docile: walks slowly, can be approached slowly, not excited by humans. 2. Slightly Restless: aware of humans, head up, moves away from approaching human, runs fence line, stops and looks around. 3. Restless: constantly runs along fence line, head up. 4. Nervous: agitated, runs along fence line, head up, looking for a way of escape, and will run if humans come closer, stops before hitting gates and fences, avoids humans. 5. Very Nervous: runs, head high and very aware of humans, may run into fences and gates, flighty. 6. Wild (Aggressive): excited, runs into fences, runs over anything in its path.
Comfort of Human	Overall	1. Stays for duration: Human feels it is safe to remain in pen with animal for duration of observation. 2. Enters but does not stay for duration: Human feels it is safe to enter pen with animal but feels unable to remain for duration of observation. 3. Never enters pen: Human does not feel safe enough to enter into pen with animal.
Heifer to Observer	Overall	1. Neutral: Heifer is not disturbed by the presence of the human. May move around pen. 2. Attempt to escape (Flight): Heifer's eyes oriented towards human, moves away from threat. May look for a way of escape by sticking head out of or underneath fence. 3. Attack (Fight): Heifer is threatened to a point where she pursues the human standing inside or outside of the pen.
Flight Zone Size	Overall	1. Low: Heifer stands still, minimal movement. Human is clearly not within the heifer's flight zone. 2. Medium: Human may be in heifer's flight zone, or on the perimeter of it. Heifer clearly reacts to the presence of the human, but responds in a temperate manner (mildly or moderately) to distance herself from human. Not as reactive as High. 3. High: (Given this score automatically if human does not enter pen) Human is clearly in the heifer's flight zone, which may be larger than the pen. Wildly moves about pen. May charge human.

Ethogram	Type	Description
Pace	Timed	1. Stationary: No movement. 2. Fidgets: Small movements back and forth through nervousness. 3. Walk: Four time movement. Each leg moves on its own and in a set order. 4. Trot: Two time movement. Two diagonal pairs of legs in motion at the same time. 5. Run/Canter: The canter is a three time rhythm. There is either left lead or right lead canter. The run is a four beat rhythm. All four feet are never on the ground at the same time. The run progresses out of canter. 6. Out of view of camera.
Stationary Location	Timed	1. Inner area of pen: Heifer stands away from the fence line of the pen; Closest to human. 2. Long side of pen: Heifer stands along the longer length of pen. 3. Short side of pen: Heifer stands along the shorter length of pen. 4. Corner of pen: Heifer stands in corner of pen; Farthest away from human.
Direction of Movement	Timed	1. Backwards/Turns around: Heifer changes direction of movement. 2. Down single side of pen, stops: Heifer moves along the fence but is able to stop before coming to the corner. 3. Down single side of pen, slides: Heifer moves along the fence with such momentum that she slides (feet skid) in the gravel. May run into the fence (corner). 4. Circles the pen: Heifer continuously moves about the pen in a circular or erratic pattern. May be half or full circles. 5. Travels through pen, diagonally: (Occurs most often when human is not in pen) Heifer starts in one corner of pen, and travels through the middle instead of going down the side.

Ethogram	Type	Description
Counted Behaviors	Counts	1. Number of threats (attempt to charge but turns away in different direction). 2. Number of charges. 3. Number of times head-butts fence (actively or coincidentally due to sliding). 4. Number of times hits fence (hip or shoulder check). 5. Number of times attempts to escape (puts head through gate, etc.). 6. Number of tail movements (quantified as a deviation from vertical).

APPENDIX B: GROUP PEN ETHOGRAMS

Ethogram	Type	Description
Group Pen Score	Overall	1. Docile: walks slowly, can be approached slowly, not excited by humans. 2. Slightly Restless: aware of humans, head up, moves slowly away from approaching human. 3. Restless: runs along fences, stands in corner if humans stay away. 4. Nervous: runs along fences, head up and will run if humans come closer, stops before hitting gates and fences, avoids humans. 5. Very Nervous: runs, stays in back of the group, head high and very aware of humans, may run into fences and gates. 6. Wild (Aggressive): excited, runs into fences, runs over anything in its path.
Stationary Location	Overall	1. Opposite, far corner: Group stands in farthest corner from camera, next to gate. 2. Direct, far corner: Group stands in far corner, in front of camera. 3. Inner area of pen: Group stands away from the fence line of the pen. 4. Opposite, near corner: Group stands on the side nearest the camera, in the opposite corner. 5. Direct, near corner: Group stands in closest corner, directly in front of camera.
Group Flight Zone Size	Overall	1. Low: Group of heifers stand still, minimal movement. Human is clearly not within their flight zone until much later in approach. 2. Medium: Human may be in the group's flight zone, or on the perimeter of it, during approach. Group clearly reacts to the presence of the human, but responds in a temperate manner (mildly or moderately) to distance themselves from human. Not as reactive as High. 3. High: Human is clearly in the group's flight zone upon initial approach, which may be larger than the pen. Wildly move about pen. May charge human

Ethogram	Type	Description
(Average) Pace	Timed	1. Stationary: No movement. 2. Fidgets: Small movements back and forth through nervousness. 3. Walk: Four time movement. Each leg moves on its own and in a set order. 4. Trot: Two time movement. Two diagonal pairs of legs in motion at the same time. 5. Run/Canter: The canter is a three time rhythm. There is either left lead or right lead canter. The run is a four beat rhythm. All four feet are never on the ground at the same time. The run progresses out of canter. 6. Out of view of camera.
Attention	Timed	1. On human: The majority (3/4) of heifers in the group are attentive to the human. 2. On camera/other human: The majority of heifers (3/4) in the group are attentive to the camera. 3. Other: Heifers attention is either elsewhere, or not a majority (2/4).
Willingness to Separate	Timed	1. Together: All heifers in the group travel/are stationary together. 2. Separate but rejoin: One or more of the heifers are split from the group, with immediate attempt to rejoin. 3. Separate: One or more of the heifers are split from the group, with no movement to rejoin.

CHAPTER SIX

SYNTHESIS AND CONCLUSIONS

Introduction

Temperament of cattle has a large impact on the beef industry. Due to its effect on human and animal safety, producers have been inadvertently selecting against temperamental cattle for years. However, with increased understanding of how temperament effects pregnancy rates, growth, meat quality, and other economically relevant traits, this selection has become more formalized in recent years, leading to an increase in the amount of research being developed to better understand cattle behavior. This research has led to a multitude of information being provided to producers and others in academia about many different methods in which to delineate cattle behavior, often times with conflicting results. This is most likely due to differences in environment, previous handling experiences, and genetics.

Temperament is moderately heritable in beef cattle. Therefore, if accurately selected upon, it can create real change in a relatively short amount of time. Some breed associations including the American Angus Association, American Salers Association, American Simmental Association, and North American Limousin Foundation release expected progeny differences (EPD) for docility in their animals. However, these EPD are only as accurate as the data being collected. Inconsistency between producers and the repeatability of the trait can have an impact on how accurate these assessments of temperament are. Furthermore, not all producers will use genetic information when selecting sires, and even less when selecting replacement heifers for their herds.

Therefore, it is important to provide them with methods that are fast, easy, and inexpensive to implement during normal handling practices that are truly representative of that animal's overall disposition. Because of this, the objectives of this dissertation research were to 1) develop a procedure for the evaluation of calf behavior, indicative of physiological stress and 2) determine whether stress will change under repeated and routine management as evaluated through behavioral and physiological measures. It is my hope that the previous chapters have formally answered those questions. Thus, this chapter will serve to first summarize those results, creating a larger picture of their meaning, then further discuss the genetic component of selection for docility and how these results can be utilized in that effort. I include a short section on what I would have done differently based on what was learned from these data, and finish with ideas for moving forward and an overall conclusion and recommendation for producers wishing to select for temperament in their cattle.

Reliability and Repeatability

Before assessing any of the subjective measurements of behavior, their repeatability as well as the reliability of their collection had to be determined. For this study, chute and exit score were highly repeatable. From a genetics standpoint, measurements of an animal that are repeatable are beneficial as the first measurement on an animal is representative of future measurements. When this is the case, a smaller number of records are required on an animal to estimate a breeding value with higher accuracy, also requiring less resources and time.

In their formal comparison, chute, exit, individual, and group pen scores of each heifer was the average of three experienced observers on a given day. When assessing their inter-observer reliability, or the agreement among those three observers, it was determined that both chute and exit score were reliably measured not only by those three experienced individuals, but when adding in a fourth as well. In addition, inexperience only had a major impact on the assessment of chute score, with exit score the consistently more reliable measure. This is believed to be a reflection of the complexity when comparing the ethograms for chute and exit score. The ethogram for chute score 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 exit score only differ by a single adjective. This allows exit score to be easier to delineate than chute score. However, regardless of complexity, using simple adjectives was sufficient enough for individuals to accurately delineate behavior when restrained in and exiting from the chute.

Furthermore, both individual and group pen scores were reliably measured. While agreement among observers was not as strong as it was with exit score, reliabilities were higher for individual pen score than chute score. This was not the case for group pen score, which had the lowest reliability calculations. This could be due to the design of the study. Both individual and group pen scores were assigned based on 30 s of observation. This is more than enough time to assign a single pen score, but perhaps is overwhelming when trying to assess multiple heifers in a group setting in the same amount of time. Observers could have felt rushed, only having 7.5 s to evaluate each of the four heifers in the group setting. Inconsistency could have also stemmed from the three observers viewing the heifer's behavior in a different order. It is unlikely that the observers could take in all four animal's actions at once. They would have instead been watching one heifer, assigning their score, and moving to the next. This would have undoubtedly been done in a different order for each individual, meaning they saw different variations of behavior, leading to slightly different scores assigned to the same animal. Regardless, all subjective measurements were reliably assigned by experienced observers.

These conclusions impact research and production in a few ways. First, concerns about the use of subjective methods over more objective measures and blood metabolites can be combated, with the caveat that the subjective measures are consistent with objective metabolites believed to be indicative of stress, which was shown. It may be true that an exit velocity is more accurately measured than an exit score because it is not subject to observer biases and inconsistency. However, the results of the current analyses show that with experience, all subjective measurements of temperament are reliably measured.

That conclusion is supported by recent research efforts dealing with qualitative behavior assessment (QBA) of animals. Using QBA, individuals are able to use their own adjectives to describe how an animal is behaving or feeling at the time. While it may seem anthropomorphic, results from using this methodology have shown very consistent and promising results across multiple individuals indicating that humans are quite accurate when using less objective methods of evaluating behavior. Second, these results support the use of chute, exit, and pen score to delineate cattle based on their temperament. This is beneficial to producers who wish to use a methodology that is fast and easy to implement while routinely handling cattle instead of worrying about the setup of equipment to measure exit velocity or taking a blood sample or rectal temperature. These subjective methods are just more time and energy efficient to collect, and often less stressful on the cattle.

Acclimation to Handling

Once reliability was established, the change in subjective measurements of temperament over time and the relationships among each measurement was investigated. The relationship between exit score and exit velocity is strong as they both measure temperament as an animal is released from restraint. Therefore, both measures of temperament are not necessary in a production setting. However, the relationship between how an animal behaves in the chute versus exiting was negligible on the first day of observation, suggesting these methods evaluate completely different behaviors. This relationship grew stronger as heifers became as calm in the chute as they appeared while exiting. While not as strong, the same pattern was evident when comparing individual and group pen scores to chute and exit scores, with stronger associations existing between pen and exit scores compared to pen and chute scores.

A possible explanation for this pertains to the different types of measurements being compared. Chute score is referred to as a restrained evaluation of temperament, because each animal is assessed while being held in a chute. However, exit score and pen score are considered non-restrained tests as they do not limit the animal's expression of behavior. Stronger correlations existed between pen scores and exit score than between pen scores and chute scores, possibly for this reason. Restrained and non-restrained tests are measuring different aspects of temperament. Although both exit score and pen scores are non-restrained, their correlation is not as high as that of exit score and exit velocity. Therefore, while related, they are not necessarily measuring the same behavior, where exit score and exit velocity undoubtedly are.

When investigating change in behavior over time, there were no differences in exit score, exit velocity, or group pen score. In regards to chute score, cattle did not appear to differ in temperament when comparing more or less frequently handled heifers to the same extent as frequently handled heifers over successive days. However, pen scores did suggest frequently and infrequently handled heifers acclimated at different rates to confinement with a human stressor over time, with those handled more frequently having a larger decrease in score across the first day of each event. Furthermore, the group of heifers handled more frequently saw the most extreme drop in both chute score and individual pen score within the first three days of the study, with scores numerically decreasing, and essentially leveling out for the remaining six days. This is most likely due to the relatively docile scores already assigned by the end of the first event, and heifers not having much farther to decrease on either scale.

Average subjective measurements of these heifers on $d_{1,1}$ were indicative of relatively docile cattle. This could be a reflection of the calmer starting temperament of a research based herd of cattle as compared to a range herd. Research cattle, even as calves, come from cows who have become accustomed to being around humans, which is not likely the case in range cattle. This is possibly reflected in these data. Regardless, when analyzing change in behavior over time, it is not important whether heifers of acceptable temperament calm down, but whether those who have a negative reaction to handling acclimate to their surroundings. When heifers were delineated based on their initial reactions in the chute, there was a much more drastic decrease in response over time compared to those with calmer starting temperament. On average, all heifers were of acceptable temperament at the end of the research trial.

While these cattle appeared to acclimate to handling while restrained in a chute and individually in a pen, these subjective observations needed to be substantiated using objective measurements known to be associated with stress. Moderate correlations existed for temperature, heart rate, glucose, and serum cortisol with chute score, exit score, individual and group pen score, as well as exit velocity. Therefore, it does appear these measurements are accurately assessing physiological stress of an animal. Comparisons of fecal to serum cortisol concentrations show what would be considered "basal" to "stressed" cortisol levels, respectively. While cortisol concentrations are higher in serum compared to fecal matter, they are not above what are considered acceptable levels of cortisol among animals when being routinely handled. This further supports the idea that these cattle were, on average, of acceptable temperament at the start of the research trial.

Overall, the changes in objective measures of stress response corroborated the conclusions drawn from the subjective methods in regards to acclimation to handling. Again, these conclusions impact production in a number of ways. Most importantly, this further supports the use of methods such as chute and exit score in assessing cattle temperament on site. While exit velocity was also correlated with objective measures, it appears a redundant measure of an animal's temperament upon begin released from restraint. As mentioned, chute and exit score provide producers with a fast, easy, and non-invasive method to measure temperament that does not add any extra time, labor, or equipment to assess while exit velocity requires the purchase of equipment. Furthermore, depending on management and facilities individual pen score could provide an additional method of measuring temperament, albeit with some additional time requirements and safety concerns.

Another point of consideration for producers is the evidence that cattle did acclimate to handling both in the chute and when secluded individually in a pen. These cattle are therefore more willing to re-enter the same handling facility. This suggests that an animal's response is not only dependent on its inherited temperament but also on previous experiences. When cattle are handled at a young age in a calm atmosphere, they may become easier to manage. This is not to say that all cattle will acclimate: there may be cases where culling is necessary. However, it can be suggested that when cattle are excitable during their first handling experience, more than one observation of temperament may be beneficial before assessing docility. This may prevent culling an animal based strictly on its initial response to novel stimuli.

As a final observation, the most interesting occurrence when viewing these data is the increase in chute score and group pen score of the most frequently handled heifers between the first and second day of observation. This increase was evident when analyzing all heifers in the study, and remained for all except the most temperamental heifers when delineating based on chute score. The heart rate of these heifers also increased on the second day, along with an increase in creatine kinase concentrations on the third day. Creatine kinase is present in the blood as a result of tissue damage or bruising. This increased circulation of this metabolite in the blood on the third day could be a result of increased movement in the handling facility, and when restrained in the chute, resulting in some tissue damage in temperamental animals.

Overall, it appears that when cattle are handled for two consecutive days, they are more temperamental on the second day of handling. Previous personal experiences in working cattle would support this conclusion. However, by the third consecutive day the behavioral response was less drastic, suggesting they were accustomed to the handling process. In regards to this, it is interesting to consider the number of cattle that were handled on the first as compared to the second day, and the time that would be required to work those heifers. On the first day, all forty heifers were handled, which would take twice the amount of time as the second day, when only twenty heifers were worked. This increase in response from the first to second day may be a reflection of heifers unwillingness to be out of their pastures for such an extended period of time yet again. However, on the third day, the same twenty heifers were handled once more, taking perhaps even less time than the second day, possibly resulting in less residual irritation.

Individual versus Group Pen Assessment

Since both individual and group pen scores were recorded, detailed ethograms were developed in an attempt to further delineate behaviors ranging from how many times an animal flicks their tail to the amount of time spent walking in the pen. It was our hope that some of these detailed behaviors would prove useful in delineating temperament of animals to further assist producers in their assessment of temperament in their cattle. Assessing flight zone size and the fastest pace reached by animals in both a group and individual setting appear to be effective when wishing to assess temperament. In some cases, a groups' willingness to remain separated from each other in a pen was also an indicator of temperament of the group as a whole. However, many other behaviors taken as an average of the group lose a majority of the variation that is present when assessing animals individually. Cattle are just not as easily stressed when in the presence of their peers due to their herd mentality. This is supported by the fact that on average, no group of cattle had an average group pen score greater than 4.0, and only one group's pen score average was equal to 4.0.

When assessing cattle individually, however, there is a large amount of behavioral variation expressed. The number of times an animal attempts to escape, head butts or hits the fence, flicks its tail, or threatens can be recorded with pen scores with larger numbers occurring in more temperamental cattle. Furthermore, temperamental cattle appear more aggressive towards a human when secluded individually in a pen, moving around more often and at faster paces; conversely, docile cattle appear undisturbed, and are more likely to stand still or walk around the pen.

Conclusion

In conclusion, subjective measurements of temperament provide a more realistic and beneficial methodology to measure temperament in a production setting. Exit score and exit velocity are highly correlated, but measure completely different behaviors than chute score. Both chute and exit score are repeatable, and can be reliably measured, more so if individuals have experience in using the ethograms and delineating the behaviors described within. Furthermore, individual and group pen scores were also reliably assigned, making them useful tools to make selection decisions. However, it is important to consider that some cattle may acclimate to gentle, repeated handling when completed at a young age. If an animal's temperament is the primary trait of interest in a breeding program, or is completely unmanageable, then culling of that animal may be justified. However, allowing acclimation to handling may be of value when individual animals have borderline acceptable temperament, especially when balanced with other breeding objectives. Cattle that excel in all other aspects of a producer’s breeding goal may benefit from additional observations of temperament before final selection decisions are made.

How Does This Impact Selection Decisions?

Temperament is a moderately heritable trait in beef cattle, and can therefore be efficiently selected for. Two methods of selection for temperament have been proposed, first by using docility EPD when selecting sires, and second by making independent threshold culling decisions on females. Whether implemented separately or together, research has shown that both methods would result in change in temperament within a herd. Depending on the method and end goal of production, there are many things that should be considered and warrant further discussion.

First, this research shows that chute and exit score are repeatable. However, this statement is followed by the idea that an animal's first reaction to novel stimuli is not indicative of future performance. These are obviously conflicting ideas, but further explanation may provide clarity. First, repeatability of these methods were estimated using the first day of each event. Regardless of how the data were analyzed, exit score did not change over time, likely due to the generally low starting exit score for all heifers. This would explain its high repeatability. Conversely, chute scores appeared to decrease over time, following the second day of observation. However, once the third day was complete, there was no further decrease in chute score, which may have contributed to raising the repeatability of the measure. In other words, giving cattle time to acclimate to their surroundings helps to obtain a more consistent measure of temperament while restrained. Furthermore, only the most temperamental cattle had substantial decreases in chute score over time, simply as a reflection of the ethogram. If heifers had a starting chute score of two, they did not have much room on the scale to acclimate compared to a heifer with a starting score of four. The high repeatability suggests that in regards to estimating breeding values for sires or dams, small numbers of observations will still result in an accurate estimate. However, the idea of giving an animal time to acclimate is more useful for producers who have animals of borderline acceptable temperament that may excel in other areas of their breeding program.

It is worth noting, however, that the definition of temperament is an animal's response to novel stimuli. Once given time to acclimate to an experience such as handling through the same facilities, the experience is no longer novel. Therefore, if a heifer is acclimated to being worked through a specific chute, but then is asked to load into the back of an unfamiliar trailer, she may react in a similar manner to the first time she was handled, as it is a new novel experience.

Taking these points into consideration, how a producer selects for temperament may differ depending on their production goals and end customer. A cow-calf operation produces calves each year, but these calves could be sold to a feedlot, used as replacement heifers, backgrounded and sold as yearlings, or even raised and sold as bulls. All of these outcomes would benefit from differing levels of docility. The starkest contrast would be to compare calves to be sold in a feedlot to those heifers kept as replacements. It has been suggested that selecting for increased docility in females results in a decrease in maternal defensive aggression. Therefore, cows that deliver and rear a calf in range conditions may not benefit from being more docile, even though it would facilitate safer handling. Conversely, this maternal protective behavior is not of a concern for the feedlot sector of industry. Therefore, selection for increased docility, or at least acclimation to handling, would result in safer working conditions for both animals and handlers, and have a positive impact on many other important economic factors already discussed. Producers would want to be careful to not select for cattle so docile they are difficult to move through the working facilities, be it on their own operation or in a feedlot. Overall, some sort of intermediate temperament could be the best selection practice for producers.

One method to select for temperament that may assist in this, but is not currently available, is a separate selection index for docility. Selection indices provide producers with one value to compare across sires that incorporates multiple sources of information. An index for docility could be developed that incorporates the many different facets that temperament effects in the industry. This idea came about due to the lack of information in literature that places a dollar amount on the impact of temperament on the beef industry. It was an initial goal to begin the process of developing such an economic index to help producers better understand the impact of selecting for temperament in their cattle. However, one of the difficulties in developing new indexes is to ensure that the traits that are being incorporated are not present in any other existing indexes, referred to as double counting. In the case of docility, it is likely that its effect on meat quality and growth are already inadvertently incorporated into the dollar beef ($B) index, or any index that deals with meat quality. The same is likely true for its impact on reproduction, immune function, and other known impacted traits. Where docility is not incorporated is its impact on human safety and equipment costs. However, development of an index is only as accurate as the information used to develop one. Unfortunately, incidences of worker injury due to an animal are not as readily documented as one would hope. Coming by an honest representation of that information is difficult. Even starting at the feedlot sector trying to obtain information on animal and human well-being as well as docility of cattle was not successful. Therefore, this idea was forgone for other research topics. However, if available, a selection index for docility that incorporates new information would be beneficial to the beef industry, especially producers hoping to select for docility.

Hindsight is 20-20

As with any project, learning the results and issues that arose from the existing design and execution always leave room for improvement or things that could have been done differently. For this project, the first realization pertains to the selection of sires and randomization based on the frame size of the animal. While these divergent lines were created for a separate university project, the frame size was consistently insignificant and was removed from all data analyses in regards to docility. If the resources were available, it would have been more efficient to select sires for this project based on their docility EPD. This way, selection for divergent lines could have resulted in more variation in temperament responses that what was obtained from these research based heifers.

Not only could docility EPD have been taken into consideration when selecting sires, but a more thorough selection of sires across years would have been beneficial in estimating genetic parameters. In the first year of the study, there were quite a few heifers selected that were sired by clean up bulls unrelated to prominent AI bulls used; this decreased the total number of heifers available for use in genetic evaluation from an already small number. In the second year of the study, more of an effort was placed on creating genetic connectedness across years, using some of the same sires from the first year of the study. By the third year, all forty heifers were from AI sires, and had multiple half-sibs in the project. Although the selection of sires became better over time, this was not an initial consideration in the study.

When estimating heritability and repeatability, all geneticists will ask for more animals to obtain more precise estimates with smaller standard errors. It really is the case for this project. Although the estimates of heritability and repeatability were consistent with those found in literature, the small number of animals resulted in large standard errors. Since chute and exit score are highly repeatable, that means that accurate estimates can be obtained using a small number of observations on an animal. Therefore, if conducting this study again, genetic analyses would of course benefit from using more animals, but more importantly, it would benefit from using even more sires.

Another possible area of criticism in regards to the design of this study is that all heifers were handled by the same individual in a calm manner across all days of the study. Because of this, special attention was made to ensure any acclimation conclusions were prefaced with the statement that these heifers were repeatedly handled in a calm manner. This is not likely representative of how all cattle are handled in the industry. My first response to this comment would be that all cattle should be handled this way in every area of production; however, that is quite the stretch of imagination. Therefore, if conducting this study again, I would wish to implement a criss-cross experimental design in which each year heifers were randomly assigned to either a calm or excited treatment protocol initially. As an illustration, suppose 100 heifers were available for use in this study; 50 would be assigned to the calm and excited treatment protocols, separately. Those in the calm treatment protocol would be handled identical to the current study, while those in the excited treatment would be handled in a louder and rougher environment. However, after the initial three days of observation, treatment protocols would then be switched for the next three days for half of the heifers in each category (n = 25 from calm to excited and n = 25 from excited to calm), and then back to the original protocol for the final three days. As a control, the other half (n = 25) of each treatment protocol would remain in their respective groupings for all nine days of the study. Statistical analyses would then account for the design of the study, quantifying any effect on temperament response from prior treatment assignment, which may be useful in the summary of results.

Designing a study such as this would not only quantify the impact of rougher compared to calmer handling, but would also compare the effect of previous handling experiences on current temperament measures.

Learning and Moving Forward

While the ideas in the previous section would be beneficial, they were not implemented in the current experimental design. However, there are still many things that can be taken away from this research trial. Most importantly, selection for temperament in cattle is possible. Not only indirectly, as producers have been doing for years, but also directly using measurements such as chute score, exit score, and individual pen score. When comparing methodologies, they appear to be measuring different behaviors. However, if producers are only wishing to use one measurement to select for temperament, I would suggest the use of chute score. Although heifers acclimated more substantially in an individual pen setting, and non-restrained tests give animals' full range to express variations in behavior response, placing each individual animal into a pen by itself and assessing temperament is not as easily implemented into normal handling practices as chute score, and is less safe for the animal and handler. Furthermore, strong correlations between IPS and CS suggest only one measurement is necessary to quantify temperament. One instance where behavior in a pen could be assessed routinely would be if producers calve in a dry lot or pasture close to their residence. However, it could be argued that maternal defensive aggression may complicate the interpretation of temperament.

Although more difficult to delineate, I find CS the more useful measure and the one that producers are probably more concerned with in practice. The main contact that producers have with the animal occurs when it is restrained in the chute, be it giving shots, pouring on fly spray, or artificially inseminating. This is also when the animal is most likely to injure itself by flailing around and sticking legs or other extremities where they do not belong.

Using chute score as a measurement of temperament can help producers pinpoint threshold behavior that is unacceptable to them, hopefully while ensuring cattle are not too docile and therefore difficult to work through the handling facilities. However, for this to be successful, it is imperative that producers fully understand the behaviors they are delineating. Accurate selection decisions do not occur without understanding of the different phenotypes being selected upon. This is where extension professionals come into the picture. It is the mission of all extension personnel to educate the industry on methodologies and ideas that will benefit their operation. Providing them with educational information in regards to selection for temperament that they can incorporate into extension programming would further promote understanding and utilization of this technology.

Bringing this idea to fruition, during my doctoral career, I was given the opportunity to work alongside an undergraduate student in the design, approval, recruitment, implementation, analysis, and summary of a research project aimed at improving the reliability of chute score evaluations of beef cattle. Participants differed in age and genders, and ranged in their level of beef cattle experience. Individuals were shown video clips of heifers restrained in a chute and were asked to evaluate their behavior using the same ethogram available during the initial study. Based on responses to a survey, individuals were then randomly assigned to one of three treatments consisting of no training, being asked to watch a 20 min video describing the ethogram and providing examples, and watching the same video along with practicing applying the ethogram to 10 videos of cattle restrained in a chute. One week later, individuals returned and completed their respective treatments and re-watched another set of video recordings. It was concluded that age, sex, and prior cattle handling experience had no significant impact on the increase in the reliability between the first and second session. Both methods of training significantly increased interobserver reliability when compared to the control. However, there was no difference in reliability between the two treatments. It was proposed this was due to frustration with the duration of the video and additional practice videos. Therefore, it is our goal to decrease the length of the training video presentation and provide it, along with optional sample videos showing application of the ethogram, to beef cattle producers across the state of Nebraska and elsewhere. This could be achieved using the University of Nebraska's beef cattle extension website and the annual Nebraska Beef Cattle Report, allowing far-reaching communication of these materials to the beef cattle industry.