TEMPERAMENT EVALUATION IN BEEF CATTLE: UNDERSTANDING EVALUATOR BIAS WITHIN SUBJECTIVE MEASUREMENTS OF DOCILITY SCORE, TEMPERAMENT SCORE, AND QUALITATIVE BEHAVIOR ASSESSMENT

A Thesis
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Jordan Kelley Hieber

In Partial Fulfillment of the Requirements
for the Degree of
MASTER OF SCIENCE

Major Department:
Animal Sciences

May 2016

Fargo, North Dakota

North Dakota State University
Graduate School

Title

TEMPERAMENT EVALUATION IN BEEF CATTLE:

UNDERSTANDING EVALUATOR BIAS WITHIN SUBJECTIVE

MEASUREMENTS OF DOCILITY SCORE, TEMPERAMENT SCORE,

AND QUALITATIVE BEHAVIOR ASSESSMENT

By

Jordan Kelley Hieber

The Supervisory Committee certifies that this ***disquisition*** complies with North Dakota State University's regulations and meets the accepted standards for the degree of

MASTER OF SCIENCE

SUPERVISORY COMMITTEE:

Dr. Lauren Hanna
Chair

Dr. Carl Dahlen

Dr. Britt Heidinger

Approved:

6/10/2016
Date

Dr. Gregory Lardy
Department Chair

ABSTRACT

An experiment was conducted to quantify evaluator bias and determine the impacts bias may have on genetic evaluations. Three subjective evaluation methods were recorded (n = 806) at weaning, including docility score (**DS**; 1 to 6), temperament score (**TS**; 1 to 5), and qualitative behavior assessment (**QBA**; 12 attributes). Using Principal Component Analysis, a temperament index was calculated using QBA scores for each animal based on the first principal component as an additional temperament score. The final model utilized fixed effects of date of evaluation, sex, and evaluator (across only) in a traditional animal model to calculate estimated breeding value (EBV) and heritability for each temperament evaluation. Comparisons of EBV were based on correlation coefficients and quartile placements of animals within and across evaluators. Less than 30% of individuals had EBV that were 3 quartiles apart for all evaluation methods, indicating that evaluator interpretation has some impact on genetic evaluations.

ACKNOWLEDGEMENTS

There are a lot of people to which I owe thanks, all of whom have helped to get me where I am today. First and foremost, I would like to thank my advisor Dr. Lauren Hanna for giving me the opportunity to continue my education, and for guiding and mentoring me throughout my graduate career. I would also like to thank my committee members Drs. Carl Dahlen and Britt Heidinger for their time and support in helping me succeed. Also, this project would not have been possible without the financial contributions of the North Dakota Agricultural Experiment Station.

I would like to thank those who helped out with my research project including Dr. Bryan Neville and the rest of the staff at the Central Grasslands REC, Dr. David Riley, Dr. Sarah Wagner, as well as the undergraduate student workers. To Haipeng Yu, my "brother" for the past two years", thank you for all of your help, especially with all of the lab work and entering and auditing data. I also cannot forget all of the amazing faculty and staff within Hultz Hall that have given countless words of encouragement and guidance.

To my fellow graduate students, specifically Mattia Gunkelman, Blaine Novak, Lexie Siomka, and Laura Bachmeier, as well as Dani Black and Dr. Casie Bass, thank you for all of the laughs, cheering me on along the way, and being shoulders to lean on. To Justin and Mellissa Crosswhite, I do not know if you will ever know how much your presence in my life means to me or what you have done for me the past four years.

I owe a huge thank you to the people who have been there for me outside of academia including the Chinn Family, the Folstad Family, as well as Skip Anderson and Kelli Armbruster. All of you have provided me with an escape when I needed a break and have supported me long before I started on my graduate school endeavor. Your friendship means the world to me!

To my closest friends Kelli Chinn, Laken Polega, and Trevor Barnes, thank you for always being there for me, for laughing until we cried, and all of the crazy adventures. I will never be able to thank you for everything you guys have done for me and I look forward to many more years of friendship.

Lastly, I owe the largest thank you to my family. To my parents, as I have gotten older, I have gotten a better understanding about the sacrifices that you made in order for me to chase my dreams. To my grandparents, I will never be able to repay you for everything that you have done for me. I know that at the end of the day, you are all my loudest cheerleaders and biggest supporters; I can only hope that I can continue to make you proud!

DEDICATION

To my family and close friends, for their continuous support and countless words of encouragement.

To my great papa, great grandma, and grandpa, who are, I know, watching over me and cheering me on from above.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

DS .... docility score

EBV.. estimated breeding value

EPD .... expected progeny differences

ID .... identification

km .... kilometer

mm .... millimeter

MMD.. movement-measuring device

NW... northwest

PC1.. first principal component

PC2.. second principal component

QBA .... qualitative behavior assessment

TI.. temperament index

TS .... temperament score

VAS.. visual analog scale

LIST OF SYMBOLS

$\hat{\sigma}_a^2$ .... estimated additive genetic variance

$\hat{\sigma}_p^2$ .... estimated phenotypic variance

c^2 .... proportion of phenotypic variance due to permanent environment effects

h^2 .... heritability

$\hat{\sigma}_{pe}^2$ .... estimated permanent environmental variance

LIST OF APPENDIX TABLES

Table Page

CHAPTER 1. INTRODUCTION AND LITERATURE REVIEW

Introduction

Cattle temperament, often defined as the reaction of that animal to human handling (Burrow and Dillon, 1997), is an important factor that producers are taking into consideration within their production systems. Temperament is known to have many influences on production aspects such as reproduction, immunity, and carcass characteristics (King et al., 2006; Cooke et al., 2009a, 2009b, 2011; and Burdick et al., 2011). When producers select cattle with calmer temperaments, they also can improve the performance of those production traits, and, therefore, the efficiency of their overall operation. When it comes to evaluating cattle temperament, there are several types of evaluations that are either objective or subjective in nature. With every producer having their own interpretation of subjective temperament scales, evaluator bias may exist. This bias can exist across or within herds and may ultimately impact selection response.

The **primary objective** of this project is to determine the impact of evaluator bias in subjective measures of cattle temperament on genetic selection criteria. To accomplish the primary objective, the following **secondary objectives** include:

- to identify systematic environmental effects for the statistical model, including evaluators, for docility score, temperament score, qualitative behavior assessment attributes, and temperament index;
- to determine the impact of evaluator on genetic parameter estimates for docility score, temperament score, qualitative behavior assessment attributes, and temperament index; and
- to determine the impact of evaluator on breeding value estimation for docility score, temperament score, and temperament index.

Cattle Temperament

Cattle temperament can be defined as continual responses, both behavioral and physiological, observed in the presence of a stressor or environmental challenge (Sutherland et al., 2012). A study by Riley et al. (2014) looked at behavioral characteristics involved in temperament including:

1) aggressiveness, an animal's willingness to hit an evaluator;
2) nervousness, an animal's response to new situations;
3) flightiness, related to the overall movement of the animal;
4) gregariousness, the willingness for an animal to sort away from its counterparts; and
5) overall temperament, a subjective overall evaluation of an animal's temperament.

By understanding cattle temperament, producers can have a better understanding of how it affects their production characteristics and make selection decisions in their herds accordingly.

Evaluating Temperament in Cattle

There are several different cattle temperament evaluations that have been developed through research and observation that can be measured on-farm including, but not limited to, exit velocity, movement-measuring device (MMD), chute score, balking, hair whorl position, docility score, temperament score, and qualitative behavior assessment (QBA).

Objective Measurements

Objective measurements are values taken during temperament evaluations that are concrete and repeatable from person to person. One form of an objective measure is exit velocity, which is also called flight speed. Exit velocity is the rate at which an animal exits a working chute. This is measured using an infrared or infrared-like sensor system to determine the time it takes an animal to get across a fixed distance where the time is then converted to speed (meters/second). In this method, faster animals are considered more temperamental and slower animals are considered more calm (Burrow et al., 1988). Anecdotal evidence suggests that measures such as exit velocity may not measure the temperament of an animal completely, but rather its reaction to getting away from a stressful event. This has not been well proven in literature, yet there is a method that measures an animal's movement during a stressful event. An electronic movement-measuring device (MMD) measures an animal's movement while on the scale as a voltage. As the voltage increases or decreases, the MMD will record a peak. The number of peaks is correlated to the amount of movement the animal makes and as the number of peaks increases or decreases, the animal moves more or less, respectively (Waynert et al., 1999). Stookey et al. (1994) found that when heifers could not see their conspecifics, their MMD scores were more elevated and, therefore, were more temperamental compared with heifers that maintained visual contact with their conspecifics.

Subjective Measurements

Subjective measurements are values given to an animal at a specific point in time, based on an evaluator's perception. Chute score is evaluated on a scale from 1 to 5 when an animal is in a working chute without its head caught (Grandin, 1993). Scales used for chute score include: 1 – an animal that is calm and does not move, 2 – slightly restless, 3 – restless and shakes the working chute, 4 – vigorously shaking the working chute continuously, and 5 – struggling violently, rearing, and twisting of the body. While chute score is a common evaluation method, it does have its shortcomings. For example, Burrow and Corbet (2000) showed that cattle with excitable temperaments actually freeze when restrained instead of showing their natural behavior during temperament evaluations that use restraints, indicating chute score may not accurately capture all types of behavioral responses. In 1993, Grandin also collected balking ratings as animals entered a working chute and a weigh scale. Cattle were either classified as non-balkers or balkers: (1) non-balker: animal entered the chute or scale voluntarily when the gate opened or only a light tap was required and (2) balker: a tail twisting or hard slap on the rump was required to get an animal to enter the chute or scale.

Other physical characteristics have also been considered. For example, the location of a hair whorl on an animal's forehead has been associated with temperament. The center of the hair whorl is used to determine its height location, where "high" is considered if the center of the whorl is above the eyes, "middle" is considered when the center fell between the top and the bottom of the eyes, and "low" is considered if the center is below the bottom of the eyes (Grandin et al., 1995). Grandin and colleagues (1995) found that cattle with "high" hair whorl locations were more excitable than those with "middle" or "low" whorl locations. Lanier et al. (2001) did a study of combining height location based on Grandin et al. (1995) and lateral whorl location as on the right, left, or in the middle of the facial centerline. The facial centerline was a path from the poll to the center of the nose that was two and a half centimeters wide and the center of the whorl was used to determine vertical and horizontal location. It was again found that those with hair whorls closer to the poll had more excitable temperament scores and that 62% of the cattle with whorls off the centerline had a temperament score of 2 compared to only 52% having whorls on the centerline (Lanier et al., 2001). In 2008, Olmos and Turner's research using exit velocity and chute score indicated that the hair whorl position and temperament relationship is sensitive to the type of temperament measurement used and, therefore, using hair whorl position as an indicator trait for selecting against more temperamental cattle is limited.

According to the Beef Improvement Federation (2010), docility score is evaluated on a scale from 1 to 6 when an animal is in a working chute with its head restrained in the head gate. Scores are broken down as follows: 1 - a calm, easily handled animal that does not pull back on the head gate and exits the chute calmly; 2 – still calmer than average but may pull back on the head gate, flicks its tail, and exits the chute promptly; 3 – nervous and impatient animal with a moderate amount of struggling against the head gate and tail flicking, and exits the chute briskly; 4 – flighty, jumpy, struggles violently, and continuously flicks its tail while possibly being vocal, and exits the chute wildly; 5 – similar to a 4, but with added aggression that exits the chute frantically; and 6 - excited, very aggressive, and thrashes about. Docility score is the method of choice for breed associations, where some have incorporated an expected progeny difference (EPD) value for selection purposes. According to Northcutt and Bowman (2010), a $62.19 per head net return was realized during the Iowa Tri-County Steer Classic Futurity (TCSCF) for cattle in the most docile category compared with cattle in the most aggressive category. A drawback with docility is that it has yet to be determined if just using contemporary group information in the analysis adjusts for differences in evaluators across herd or ranches when estimating EPDs.

Temperament score was adapted from the breeding program of the Agropecuária Jacarezinho® and is evaluated when an animal leaves the working chute and enters a working pen. Its scale is from 1 to 5, with the intermediate score (3) being removed to avoid the option of evaluators choosing a median value (Sant'Anna and Paranhos da Costa, 2013). A value of 1 is an animal that walks slowly and allows close approximation with the observer and a value of 5 is an animal that runs the entire time of the observation, jumps against the fence, and tries to attack the observer. Both docility score and temperament score are very easy for producers to incorporate into their everyday working environments as they are very straightforward measurements and can be recorded while working animals.

The QBA attributes were adapted from the cattle QBA assessment described in the Welfare Quality Protocol assessment system (Welfare Quality® Consortium, 2009). The original 20 attributes were reduced down to 12, adopting some of the adjectives to maintain an equal amount of positive and negative expressions; those 12 behavioral attributes include active, relaxed, fearful, agitated, calm, attentive, positively occupied, curious, irritated, apathetic, happy, and distressed (Sant'Anna and Paranhos da Costa, 2013). Similar to temperament, QBA is scored when an animal leaves a working chute and enters a working pen. Evaluators are provided a single page per animal of all of the attributes, with each attribute having a corresponding 126 mm visual analog scale (VAS) to indicate the level of expression associated with that attribute for a given calf. During the evaluation, evaluators interpret the animal for each attribute and mark the VAS accordingly to indicate the level of expression, where a mark at the far left indicates no expression and a mark at the far right indicates full expression. The score itself is then the distance of the mark from the far left side of the VAS (in mm). Sant'Anna and Paranhos da Costa (2013) was the first study to use this method and was only done in Nellore cattle. There is little known about this form of temperament evaluation and its usefulness, particularly in *Bos taurus* breeds, therefore further investigation on this method is warranted.

Subjective temperament evaluations developed through research rely on the evaluator's interpretation of that animal's behavior as well as the scale used. This can result in human error or bias that effects temperament assessments and their use for selection (Curley Jr. et al., 2006). An evaluator's interpretation of a given method depends on their previous cattle experience, where someone with extensive experience working with livestock may be more in tune to an animal's body language compared to someone who has little to no experience. Overall, objective temperament methods are quantitative and do not have any subjective bias, but may require more expenses for the equipment that is used and may not adequately capture the complexity of temperament for selection purposes. As for subjective temperament methods, they can easily be incorporated into an everyday working environment, but rely on human interpretation of that animal's behavior.

Research has shown that flight speed is a highly heritable trait (Burrow and Corbet, 2000), however other temperament evaluations (subjective measures) tend to have low heritabilities (Fordyce et al., 1982; Burrow, 2002). Flight speed has been reported to have a heritability range of 0.26 to 0.54 (Burrow et al., 1988; Burrow and Corbet, 2000; Burrow, 2001; Nkrumah et al., 2007; Rolfe et al., 2011). Petherick et al. (2002) suggested that flight speed's heritability and repeatability is due to the fact that it largely measures the genetic fearfulness of an animal. Heritability has been reported for various versions of subjective chute measurements with several scales utilizing different breeds of *Bos taurus* and *Bos indicus* type cattle. Focusing on those measures that were previously discussed, Hoppe et al. (2010) looked at chute score with 5 different *Bos taurus* breeds and found a heritability range of 0.11 to 0.33, and Barrozo et al. (2012) looked at temperament score (1-4 scale; 1 was a calm animal and 4 was an aggressive animal) and reported a heritability of 0.18 ± 0.02. The average value of heritabilities reported on various measures of temperament was 0.27, indicating a moderate to high heritability that selection can be applied to.

When evaluating temperament, the sequence order in which calves are evaluated should be taken into consideration. Although it is thought that a temperamental animal amongst calmer cohorts would agitate the calmer animals and, therefore, impact their temperament score, this is yet to be proven in literature. The Australian cattle industry has published contradictory anecdotal reports as to whether or not grouping has an influence on an individual's temperament (Petherick et al., 2002). It has been suggested that the presence of 1 or 2 temperamental animals amongst a group of calmer animals will result in the rest of the group becoming temperamental and, as well, the opposite has also been suggested. Petherick et al. (2002) found that amongst a group of mixed animals, both temperamental and calm, those individuals with good temperaments maintained their good temperament, as did the individuals with poor temperaments. These results indicate that grouping animals of different temperaments is unlikely to lead to long-term changes to the temperament of the animals. Yet, the animals grouped in this study were only together for 100 days and were also adults when they were grouped, therefore, there is a possibility that younger animals' temperament could change due to the influence of others over a long period of time.

Relationship of Temperament with Production Characteristics

By selecting livestock with a more docile, production-desired temperament, it is expected to keep stress to a minimum and, therefore, improve their productivity (Boissy et al., 2005). These less excitable animals tend to have faster growth rates and greater average daily gains (Burrow and Dillon, 1997; Voisinet et al., 1997; Petherick et al., 2009), increased efficiency of feed conversion (Petherick et al., 2002) and carcasses of greater quality grade and greater marbling (King et al., 2006) compared with more excitable cattle. Consequently, cattle with a calmer temperament are less likely to be discounted when marketed on a value based carcass basis for being dark cutters or having bruises (Fordyce et al., 1985; Apple et al., 2005).

Increased stress due to handling situations has been shown to negatively affect the reproductive system (Burdick et al., 2011). Cooke et al. (2009a, 2009b, 2011) have done research on the effects of acclimating cattle with human interaction and handling on performance, temperament, and physiological responses. They found that cows that possess a more excitable temperament had reduced pregnancy rates compared to cows that were calmer (Cooke et al., 2009a). These results may be partially due to altered neuroendocrine stress responses associated with temperament that interrupts the physiological processes required for proper reproductive function (Dobson et al., 2001).

In 2011, Burdick et al. reported that more temperamental cattle with increased stress had a reduced adaptive immune function. A study using Brahman steers found that more excitable steers had reduced *in vitro* lymphocyte proliferation and Immunoglobulin G concentrations compared to their calmer counterparts (Oliphint, 2006). In that same study, Oliphint (2006) found that when comparing calmer calves to more temperamental calves, the temperamental calves had a reduced response to vaccination. In addition, response to handling can have negative effects on management and production as cattle that are more temperamental could potentially harm themselves as well as handlers (Burrow, 1997). Cattle in many production settings are handled by humans on a regular basis and, therefore, should be more relaxed during human interaction, especially those that are handled from an early age (Krohn et al., 2001). These animals are more likely to be calmer during future handling events, thereby reducing the amount of risk to both animal and handler.

Sex of the animal can also influence temperament. In general, it is thought that intact males are more aggressive than castrated males. In a 1993 study done by T. Grandin, bulls and steers were scored for chute score and balking score during handling. It was found that 40% of the steers received a rating of 1 (calm) while 25% of bulls received a rating of 1 during all temperament-rating restraint sessions, thereby indicating that steers are significantly calmer than bulls. Voisinet et al. (1997) found that sex served as a significant source of variation in average daily gain and average temperament score, where heifers had more excitable temperament scores compared to male contemporaries without any influence from observer or temperament scoring system. Riley et al. (2014) looked at five attributes of cattle temperament and found that heifers had significantly greater average scores compared to steers and bulls, and that bulls had the lowest average temperament scores even though they only differed from steers for nervousness.

Along with sex, breed and genetics can influence cattle and their temperament. A study done by Hearnshaw and Morris (1984) attempted to assess the temperament difference between *Bos indicus* (Brahman, Braford, and Africander bulls) x Hereford, *Bos taurus* (Simmental, and Friesian bulls) x Hereford, and traditional Herefords. Hearnshaw and Morris (1984) found that during evaluation for seven behavioral responses (tail swishing, straining back, backwards and forwards movement, paddling with back feet in an attempt to escape, kicking, kneeling, and jumping), the *Bos taurus* sired calves were significantly calmer compared to the Brahman sired calves. The calmer temperament would lead to these calves being less of a risk to both themselves, when by themselves or in a group, and handlers during their lifetime.

Little is known or reported on evaluator bias that may be present in subjective scoring methods or how it may impact estimates of genetic merit that are used by producers for selection. The following experiment aims to enhance our understanding of this topic area using three subjective methods of docility score, temperament score (described by Sant'Anna and Paranhos da Costa, 2013), and QBA.

CHAPTER 2. MATERIALS AND METHODS

Animals

All cattle were managed according to the Federation of Animal Science Guide for the Care and Use of Agricultural Animals in Agriculture Research and Teaching (FASS, 2010). All procedures were reviewed and approved by the Institutional Animal Care and Use Committee of North Dakota State University.

Eight hundred and six weaned age calves were used over a two-year period (Year 1: n = 423 and Year 2: n = 383). Calves were produced by the cow herd at the North Dakota State University Central Grasslands Research Extension Center (CGREC), located approximately 14 km NW of Streeter, ND. This cow herd consists of approximately 425 Angus-based females (mature cows and heifers) that are bred to either Angus or Hereford bulls. In the first year, dams born prior to 2012 had unknown breed type, but were all mated to Angus bulls. This produced calves that were considered 50% Angus and 50% Unknown breed type (50A50Un; n = 341). Dams born in 2012, however, were sired by Angus bulls, but their dams were still of unknown breed type. When mated to Angus bulls, dams born in 2012 produced calves that were 75% Angus and 25% Unknown (75A25Un; n = 81).

In the second year, dams born prior to 2012 were bred to Angus or Hereford bulls producing calves that were 50% Angus and 50% Unknown (n = 245) or 50% Hereford and 50% Unknown (50H50Un; n = 14), respectively. Dams born in 2012 were also bred to either Angus or Hereford bulls resulting in calves of 75% Angus and 25% Unknown breed type (75A25Un, n = 14) or 50% Hereford, 25% Angus, and 25% Unknown breed type (50H25A25Un; n = 55). Dams born in 2013 were sired by Angus bulls, but out of dams of unknown breed type. When mated to Angus or Hereford bulls, calves of either 50% Hereford, 25% Angus, and 25%

Unknown breed type (50H25A25Un, n = 18) or 75% Angus and 25% Unknown breed type (75A25Un, n = 37) were produced, respectively.

Collection Procedure

Data collected included temperament evaluations of three subjective scoring systems for docility score, temperament score, and qualitative behavior assessment (QBA; Table 1). For temperament score, the intermediate score (3) was removed from the scale to avoid the option of evaluators choosing the median value (as described by Sant'Anna and Paranhos da Costa, 2013). For QBA, evaluators were provided a single page per calf of all of the attributes, with each attribute having a corresponding 136 mm visual analog scale (VAS) to indicate the level of expression associated with that attribute for that given calf. The QBA score was then the distance of the mark from the far left side of the VAS (in mm) measured with a digital fractional caliper (General Tools & Instruments, New York, NY). We opted to use a 136 mm VAS in the current study instead of the 125 mm and 126 mm VAS used in previous research as there is no standardized VAS size and, therefore, it is expected that that the length of the scale will not influence the outcome. Other reasons that we opted to go with 136 mm is it created an easy length for printing score sheets and it will allow for more variation of each of the attributes to be measured.

Calves were evaluated as they were brought through the working pens based on management group (e.g., young dams vs. old dams). As calves came through the handling facility, they first entered the Silencer chute (Moly Manufacturing Inc., Lorraine, KS), which is where their weaning weight and docility score were recorded (Figure 1). After a calf left the chute, it immediately entered a four-platform standing scale that measured weight distribution

Table 1. Descriptions of temperament evaluation scales used for evaluation on calves at weaning.

Evaluation method	Location evaluated	Evaluation scale
Docility	Silencer chute	1 – a calm, easily handled calf that does not pull back on the head gate and exits the chute calmly 2 – still calmer than average but may pull back on the head gate, flicks its tail, and exits the chute promptly 3 – nervous and impatient animal with a moderate amount of struggling against the head gate and tail flicking, and exits the chute briskly 4 – flighty, jumpy, struggles violently, and continuously flicks its tail while possibly being vocal, and exits the chute wildly 5 – similar to a 4 but with added aggression that exits the chute frantically 6 - excited, very aggressive, and thrashes about
Temperament	Working pen	1 - animal walks slowly while allowing close approximation with the observer 2 – trots or runs for a few seconds while allowing moderate approximation with the observer 4 – runs the entire time of the observation, looks for an escape with constant tail movement, and does not allow close approximation with the observer 5 - runs the entire time of the observation, jumps against the fence, and tries to attack the observer
Qualitative Behavior Assessment	Working pen	12 attributes, each attribute has corresponding 136 mm line, evaluator places mark on line based on calf's expression of attribute, mark on the left is no expression and mark on the right is full expression. Definition of attributes: ▪ Active – quick in physical movement (lively); disposed to action (energetic) ▪ Relaxed – set or being at rest or at ease ▪ Fearful – full of fear ▪ Agitated – disturbed, excited, angered ▪ Calm – tranquil, peaceful ▪ Attentive – watching something carefully; paying attention ▪ Positively Occupied ▪ Curious – showing a desire to learn or know more about something or someone ▪ Irritated – being bothered, irked, aggravated, annoyed ▪ Apathetic – showing little or no feeling or emotion ▪ Happy – showing feelings of pleasure and enjoyment ▪ Distressed – showing extreme unhappiness or pain

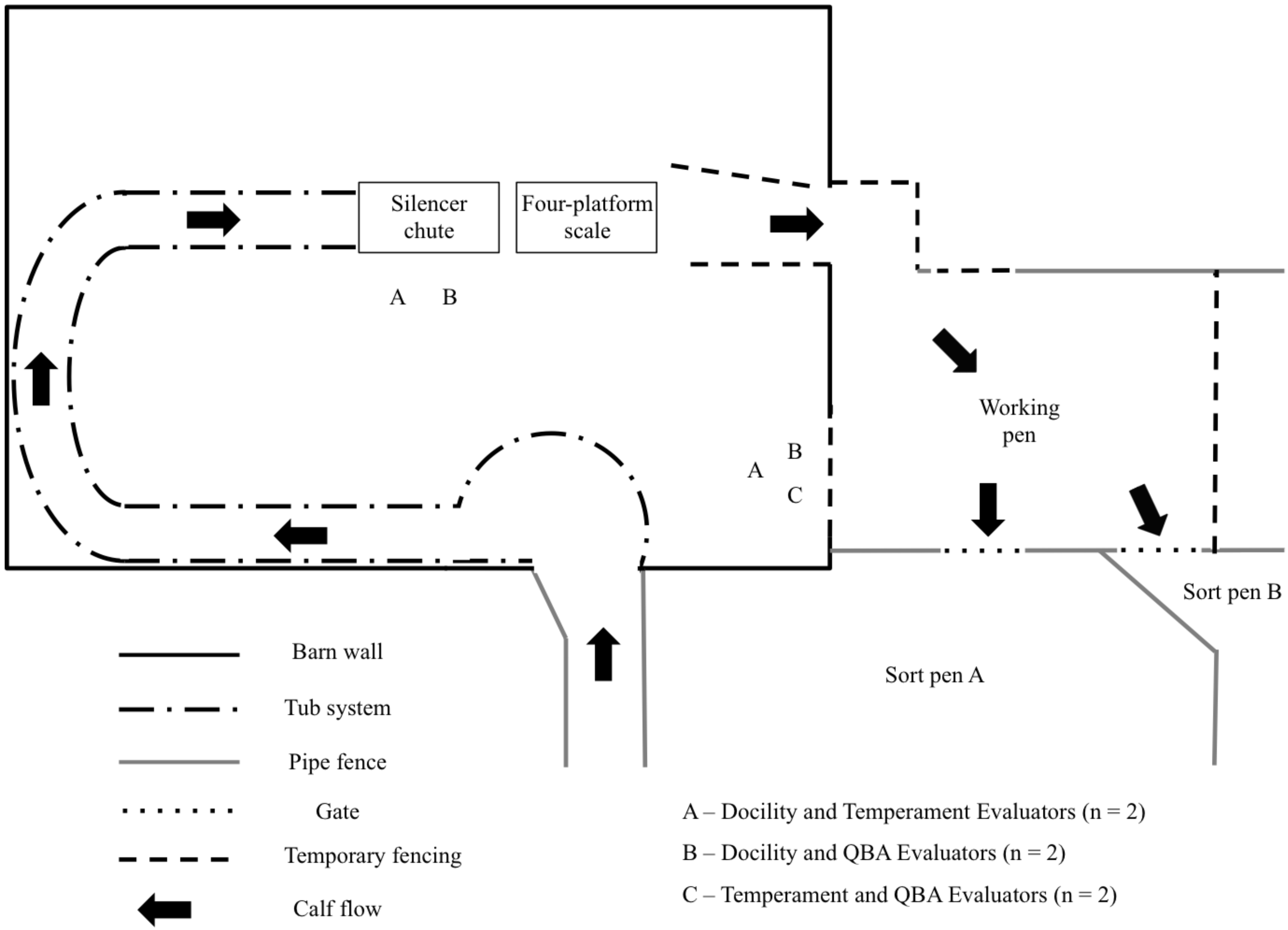

Figure 1. Experimental setup used for temperament evaluation at North Dakota State University Central Grasslands Research Extension Center.

eight to ten times per second. The calf remained on the scale for at least 45 seconds. Once released from the four-platform scale, calves entered a working pen where they were evaluated for temperament score and QBA (Figure 1). A cattle handler was present in the working pen and slowly walked toward and moved each calf so that evaluators could score specific attributes on temperament and QBA. Following evaluation in the working pen, calves were sorted into pens based on the management needs.

A separate study will utilize genotypes to understand the impact of evaluators on genomic predictions at a later time. To achieve this objective, blood was collected via jugular venipuncture while the calf was restrained in the chute. As temperament score and QBA are assessed after the silencer chute, there was concern that blood collection before temperament evaluation may also influence scoring. To determine if this was true, calves were split into 2 groups at random in the first year, where one group of calves was evaluated for temperament score and QBA blood collection and the other group of calves was evaluated before blood collection. Analysis of Year 1 data showed no impact ($P > 0.05$; Table 2), of blood collection timing on temperament score or QBA. In Year 2, therefore, blood collection preceded assessment of temperament score and QBA for all calves. To assess the impact of evaluator on temperament scoring, each evaluator was assigned two of the three subjective scoring methods (n = 6 evaluators total per year and n = 4 evaluators per method per year).

Statistical Analysis

Four calves were removed from further analyses. Three calves from 2014 were removed; two because they were intact bulls (tag identification (ID) 14039 and 14093) and one because it did not come through the chute during evaluations (tag ID 14193). The fourth calf (tag ID

15138) was from 2015, where records showed that calf ID was dead at birth, but temperament and weaning weights were recorded on a calf with that ID during collection. As no picture of the calf with the tag ID was available, it was assumed that there was an error with collection and its records were dropped. This left 802 calves with records for analyses.

Table 2. Impact of blood collection status nested within day of evaluation for temperament traits recorded on calves at weaning during Year 1 of data collection.

		***P*-value**[2]	
Evaluation	**ANOVA *P*-value**	**Day 1**	**Day 2**
Temperament score	0.198	0.072	0.992
QBA[1] attributes			
Active	0.369	0.183	0.642
Relaxed	0.336	0.144	0.838
Fearful	0.569	0.924	0.290
Agitated	0.274	0.135	0.553
Calm	0.364	0.197	0.549
Attentive	0.190	0.072	0.762
Positively occupied	0.520	0.380	0.464
Curious	0.460	0.217	0.875
Irritated	0.723	0.595	0.545
Apathetic	0.904	0.999	0.653
Happy	0.277	0.110	0.898
Distressed	0.525	0.302	0.638
Temperament index	0.478	0.486	0.320

[1]QBA refers to qualitative behavior assessment.
[2]Linear contrasts of day 1 score before blood collection versus day 1 score after blood collection and day 2 before versus day 2 after; range of records available was 808 to 856.

Following Sant'Anna and Paranhos da Costa (2013), an additional temperament measurement was calculated from the QBA scores using a Principal Component Analysis in SAS (SAS Institute, Inc., Cary, NC). A Principal Component Analysis identifies relations among combined data variables and then determines indexes that are principal components that describe the data variation (Richardson, 2009). Sant'Anna and Paranhos da Costa (2013) showed that the first principal component (PC1) accounted for 49.47% of the variation in the data for the Nellore cattle, where each animal's generated first principal component score became the temperament index (TI). The same approach was investigated for use in this study following Sant'Anna and Paranhos da Costa (2013) as well as recommendations described in Jolliffe (2014).

Within and across Year 1 and Year 2 data, simple means were generated using mean procedures of SAS (SAS Institute, Inc., Cary, NC). Boxplots were generated to check for data outliers, and Pearson and Spearman Rank correlations were generated for QBA attributes using the correlation procedure of SAS. The statistical model for each temperament trait (n = 15) within and across evaluators (n = 4 per trait; n = 8 evaluators over the 2-year period) was assessed for significant or important fixed and random effects contributing to variation using mixed model procedures of SAS (SAS Institute, Inc., Cary, NC). Within a year, there was a total of 6 evaluators, where 2 of the evaluators from the first year could not return for the second year and, therefore, comparable evaluators replaced them in the second year. Fixed effects evaluated across Year 1 and Year 2 data were date of evaluation (n = 4), sex (n = 2), evaluator (n = 8), breed type (n = 4) as well as relevant interactions. All fixed effects were tested with sequence nested within date of evaluation included as a fixed covariate, whereas animal was a random effect. When testing evaluator in the model, a repeated measures design was used with variance-covariance structures tested to capture correlations among the residuals for a given animal. The final model was tested across each trait (n = 15) within evaluator (n = 4) as well as across each trait with evaluator as a fixed effect. Least squares means were generated for significant effects, where experiment-wise error rate was controlled using Tukey-Kramer adjustment.

The final model was used in ASReml (Gilmour et al., 2015) with pedigree to calculate estimated breeding values (EBV) and genetic parameter estimates of additive genetic variance and heritability for a given evaluator. Across evaluators, a permanent environment effect was also estimated by including a second random animal effect as well as its ratio with phenotypic variance (denoted as c^2). To determine the impact of evaluators on the prediction of genetic merit, estimated breeding value (EBV) comparisons were made based on 1) Pearson and Spearman Rank correlation coefficients, and 2) changes in quartile rankings of animals within a method across evaluators, similar to that described by Hulsman Hanna et al. (2014).

CHAPTER 3. RESULTS AND DISCUSSION

Principal Component Analysis

Pearson and Spearman Rank correlation coefficients among the 12 QBA attributes across evaluators are presented in Table 3 and within evaluators are presented in Appendix Tables A1 to A5. Across evaluators, it was found that all but four Pearson correlation coefficients were significant ($P < 0.05$); those correlations that were not significant included curious vs. irritated, curious vs. distressed, irritated vs. happy, and happy vs. distressed. Within evaluators, an overall large proportion of the correlations were also found to be significant ($P < 0.05$); when looking at all of the evaluators, there were 39 significant Pearson correlations and 43 significant Spearman Rank correlations out of 66 total correlations. Evaluators 5 and 8 each had the largest number of non-significant correlations, 14 and 15 respectively, and for the overall non-significant correlations, there were several similarities. For those non-significant correlations, evaluators 5 and 8 show similarities in correlations that were not significant for agitated vs. happy, positively occupied vs. irritated, apathetic vs. irritated, happy vs irritated (Appendix Tables A1 to A5).

For Pearson correlation coefficients, non-significant correlation included fearful vs. positively occupied for evaluators 5 and 6, agitated vs. curious for evaluators 1 and 5, curious vs. irritated for evaluators 5 and 6, irritated vs. apathetic for evaluators 5 and 8, irritated vs. happy for evaluators 5, 6, and 8, and happy vs. distressed for evaluators 4, 5, and 8. For Spearman Rank correlation coefficients, non-significant correlations included positively occupied vs. fearful for evaluators 5 and 6, positively occupied vs. agitated for evaluators 5, 6, and 8, curious vs. active for evaluators 1 and 6, curious vs. fearful for evaluators 1, 5, and 6, curious vs. agitated for evaluators 1, 5, and 8, irritated vs. positively occupied for evaluators 5 and 8, apathetic vs.

Table 3. Pearson and Spearman Rank correlation coefficients for QBA attributes measured across evaluator[1].

	Active	**Relaxed**	**Fearful**	**Agitated**	**Calm**	**Attentive**	**Positively occupied**	**Curious**	**Irritated**	**Apathetic**	**Happy**	**Distressed**
Active		-0.517	0.445	0.545	-0.500	0.131	-0.119	-0.122	0.422	0.220	0.011	0.196
Relaxed	-0.539		-0.548	-0.617	0.878	0.163	0.508	0.444	-0.447	0.366	0.530	-0.284
Fearful	0.497	-0.572		0.684	-0.547	0.129	-0.177	-0.212	0.565	-0.254	-0.215	0.512
Agitated	0.546	-0.609	0.726		-0.632	0.127	-0.164	-0.187	0.664	-0.105	-0.214	0.433
Calm	-0.521	0.877	-0.566	-0.623		0.200	0.523	0.469	-0.473	0.315	0.542	-0.296
Attentive	0.105	0.171	0.125	0.159	0.208		0.584	0.532	0.245	0.181	0.568	0.212
Positively occupied	-0.148	0.491	-0.157	-0.100	0.507	0.566		0.693	0.039	0.403	0.796	0.059
Curious	-0.209	0.458	-0.220	-0.160	0.486	0.543	0.697		-	0.313	0.714	-
Irritated	0.412	-0.417	0.545	0.645	-0.433	0.322	0.129	0.065		-0.117	-	0.667
Apathetic	0.088	0.408	-0.232	-0.112	0.358	0.208	0.476	0.343	-0.043		0.519	-0.195
Happy	-0.085	0.527	-0.229	-0.173	0.529	0.542	0.767	0.722	0.096	0.571		-
Distressed	0.164	-0.181	0.407	0.378	-0.178	0.342	0.172	0.174	0.608	-0.160	0.180	

[1] QBA refers to qualitative behavior assessment. Pearson correlations are above and to the right of the diagonal and Spearman correlations are below and to the left of the diagonal. Significant correlations ($P < 0.05$) are displayed in the table. Correlations between two traits with "-" are non-significant ($P > 0.05$).

attentive for evaluators 1 and 8, happy vs. active for evaluators 4, 6, and 8, and happy vs. agitated for evaluators 5 and 8. Even though some evaluators had significant correlation coefficients for these pairs of QBA attributes, they varied from minimal (e.g., less than 0.10) to large (e.g., greater than 0.70) relationships.

To identify the relationship among the individual QBA attributes and create a smaller subset of values to use for temperament evaluations, a principal component analysis was performed. Significant correlation coefficients confirmed that linear relationships among these QBA attributes existed, which are needed for principal component analysis. For the five evaluators that scored QBA attributes, the PC1 explained 41.86 to 45.9 % of the variation in the data while the second principal component (PC2) explained 10.47 to 29.29 % of the variation in the data. Following the Kaiser criterion (Kaiser, 1960), we would only retain factors with eigenvalues greater than one, which was between two and three factors. In the scree plots found in Figure 2, the eigenvalues are plotted in decreasing size, where the sharp bend resembling an elbow indicate principle components that do not account for a significant amount of variation and divides the major or important factors from the trivial or minor factors (Ledesma and Valero-Mora, 2007). In our case, this point across evaluators is between the second and fourth principal components, resembling what was found with the Kaiser criterion. Eigenvalues less than one contribute minimal variation (typically less than 1%) and indicate principal components that are not useful for further analysis. Due to PC1 explaining the majority of the variation in the data, as well as knowing that PC1 maximizes variance (Shalizi, 2016), scores of individuals on PC1 were then used as a temperament index (TI) to summarize the QBA data (as described in Sant’Anna and Paranhos da Costa, 2013).

The 12 attributes of QBA are broken down into “positive attributes” and “negative attributes” where relaxed, calm, positively occupied, curious, apathetic, and happy are positive, and active, fearful, agitated, attentive, irritated, and distressed are negative attributes. Considering the loading plots (Figure 3), PC1 presented higher positive loading for attributes of calm and relaxed for 2 of the evaluators, yet also had higher negative loadings for those same attributes for the other 3 evaluators. The second principal component presented higher positive loadings for attributes of happy for 2 of the evaluators and apathetic for 2 of the evaluators yet

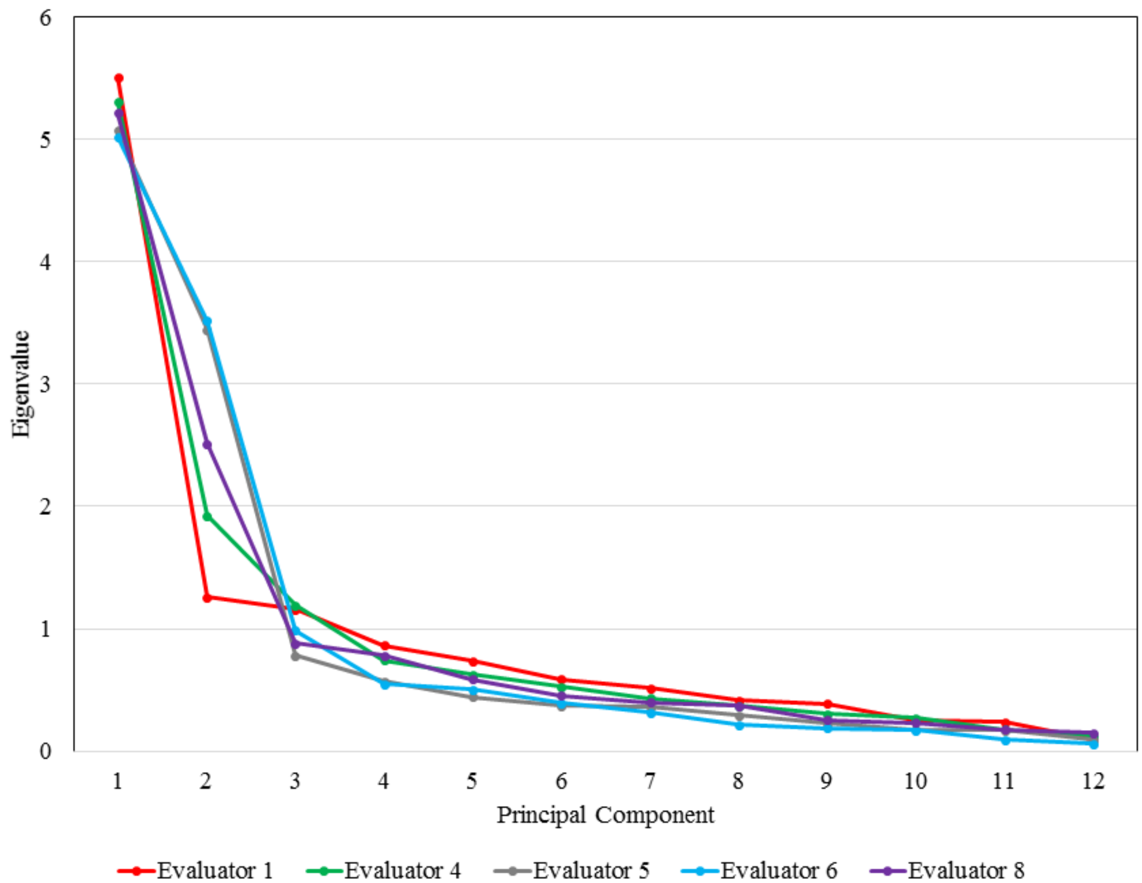

Figure 2. Principal component analysis scree plots. Eigenvalues less than one contribute minimal variation (typically less than 1%) and indicate principal components that are not useful for further analysis.

higher negative loadings for relaxed for 2 of the evaluators and apathetic for 1 other evaluator. When looking at evaluators individually, our results are both consistent and inconsistent with Sant'Anna and Paranhos da Costa (2013), indicating that evaluators can have an important influence on this approach. Results that were consistent for 3 of the 5 evaluators, which had higher positive loadings for agitated, fearful, and active for PC1, as well as higher negative loadings for relaxed and calm for PC2. The two remaining evaluators were inconsistent, having had higher positive loadings for calm and relaxed and higher negative loadings for active and agitated for PC1. Compared to Sant'Anna and Parahos da Costa (2013) overall, their negative attributes had positive loadings for this study's PC1 while their positive attributes had negative

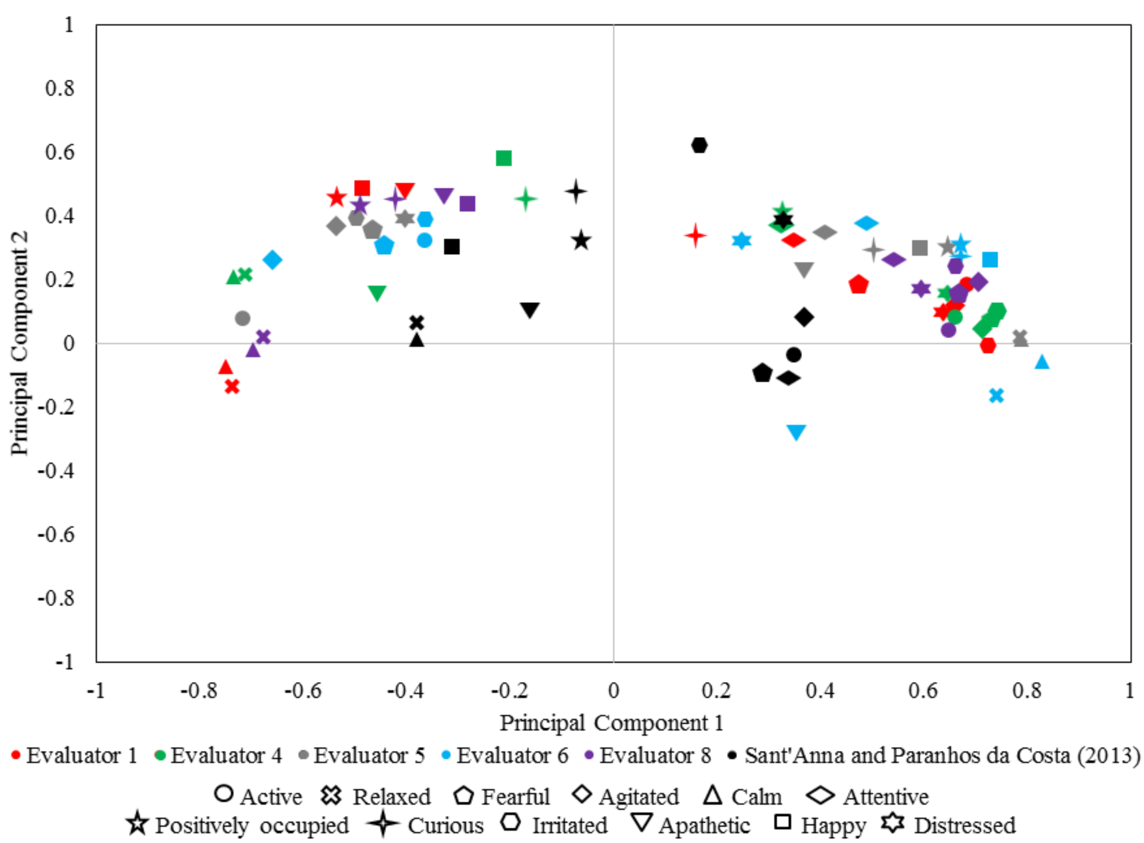

Figure 3. Principal component analysis loading plot results for behavioral attributes on the first and second components plotted against Sant'Anna and Paranhos da Costa (2013).

loadings. For PC2, their negative attributes were split with 3 being positive loadings and 3 being negative loadings while evaluators 3 and 5's negative attributes were positive loadings.

When looking at our loading plot results on a per evaluator basis, it was clear that 3 evaluators were similar, but different from the other 2 evaluators. For evaluators 1, 4, and 8, their negative attributes were positive loadings and for their positive attributes, 1 attribute for evaluators 1 and 4 were positive loadings and the remaining attributes were negative loadings for PC1. As for PC2, evaluator 4, 5 and 8's negative attributes were positive loadings, whereas 4 and 5 of evaluator 1 and 6's negative attributes were positive loadings, respectively. On the other hand, evaluators 5 and 6 had positive attributes that were positive loadings for PC1, where 5 and 4 of their negative attributes were negative loadings for PC1, respectively. For PC2, all 6 positive attributes were positive loadings for evaluator 5 and only 3 for evaluator 6 were positive loadings yet all 6 negative attributes were positive loadings for both evaluators. Based on our loading results, it is clear that 3 of our 5 evaluators scored differently than the other 2 and with that, it is expected that their EBV predictions will be different to coincide with these results.

The scatter plot for individual scores of animals of PC1 and PC2 is shown in Figure 4 for each evaluator. The scatter plot was divided into four groups (quadrants) where group I (first quadrant) contains individuals that scored positive for both PC1 and PC2, group II (second quadrant) contains individuals that scored positive for PC1 and negative for PC2, group III (third quadrant) contains individuals that scored negative for both PC1 and PC2, and group IV (fourth quadrant) contains individuals that scored negative for PC1 and positive for PC2. Sant'Anna and Paranhos da Costa (2013) used the quadrants to group and classify the temperament of the cattle used in their experiment. They found that temperament improved (became more desirable) from group I to group IV and, therefore, they named the groups as very bad temperament, bad temperament, good temperament, and very good temperament, respectively. Due to two of the evaluators scoring differently from the other three, this did not hold true in our population and, therefore, we could not model our groups the same as Sant'Anna and Paranhos da Costa (2013).

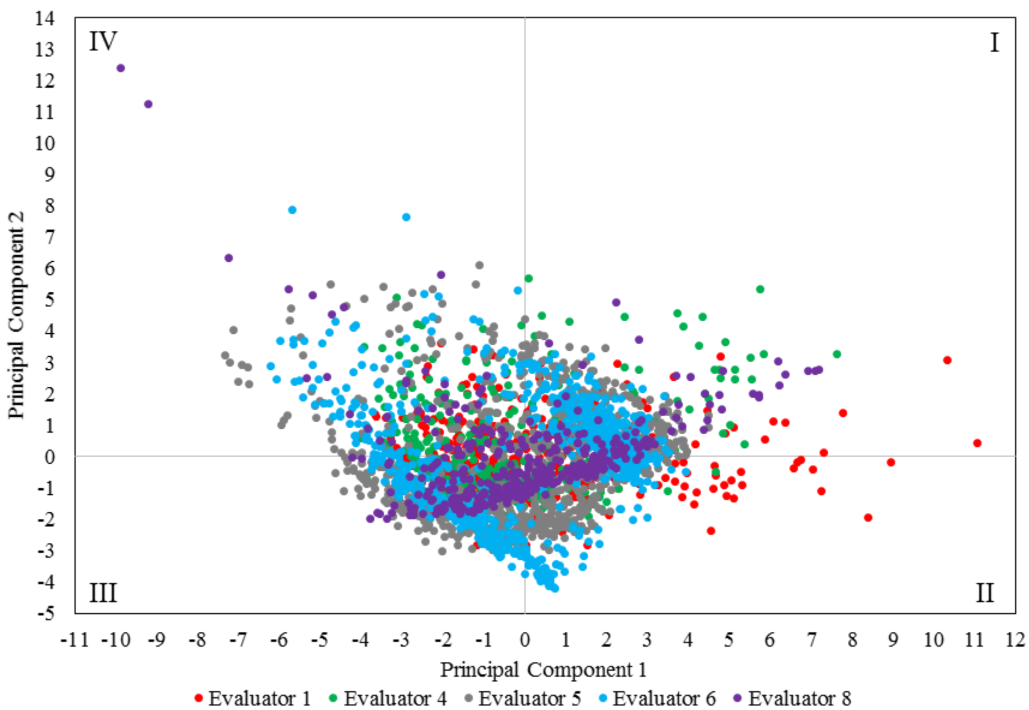

Figure 4. Scatter plot for individual animals on the first and second components. The roman numerals I to IV represent the quadrants that a particular animal's principal component scores fall into.

Temperament Score Characteristics

Summary statistics, including the minimum, maximum, means and standard deviations for docility score, temperament score, each of the QBA attributes, and temperament index are presented in Table 4. No outliers were identified in the data based on these statistics or boxplots produced.

Table 4. Summary statistics for temperament traits measured across evaluators for Year 1 and Year 2 calves[1].

	2014			**2015**			**Overall**		
Evaluation	**Min**	**Max**	**Mean ± SD**	**Min**	**Max**	**Mean ± SD**	**Min**	**Max**	**Mean ± SD**
Docility score[2]	1.000	5.000	1.840 ± 0.770	1.000	6.000	1.910 ± 0.780	1.000	6.000	1.870 ± 0.780
Temperament score[3]	1.000	4.000	1.800 ± 0.890	1.000	5.000	1.960 ± 0.890	1.000	5.000	1.880 ± 0.890
QBA[4] attributes									
Active	0.000	135.060	43.320 ± 34.460	0.000	136.000	57.170 ± 37.370	0.000	136.000	49.920 ± 36.530
Relaxed	0.000	136.000	88.400 ± 32.280	0.000	136.000	55.080 ± 38.300	0.000	136.000	72.530 ± 39.000
Fearful	0.000	121.810	15.410 ± 15.970	0.000	134.920	23.430 ± 23.020	0.000	134.920	19.230 ± 20.050
Agitated	0.000	127.960	21.710 ± 19.460	0.000	136.000	31.600 ± 29.530	0.000	136.000	26.430 ± 25.270
Calm	0.000	136.000	93.980 ± 30.180	0.000	136.000	58.630 ± 39.030	0.000	136.000	77.140 ± 38.910
Attentive	0.000	134.440	70.810 ± 25.780	0.000	135.250	37.680 ± 25.690	0.000	135.250	55.010 ± 30.600
Positively occupied	0.000	133.790	51.630 ± 30.460	0.000	133.590	15.090 ± 18.760	0.000	133.790	34.220 ± 31.410
Curious	0.000	132.380	50.950 ± 30.530	0.000	133.760	14.32 ± 20.980	0.000	133.760	33.480 ± 32.130
Irritated	0.000	115.710	21.530 ± 19.650	0.000	135.600	20.920 ± 24.680	0.000	135.600	21.240 ± 22.190
Apathetic	0.000	133.900	59.580 ± 40.800	0.000	136.000	35.030 ± 48.450	0.000	136.000	47.880 ± 46.250
Happy	0.000	132.840	57.390 ± 35.960	0.000	105.000	9.670 ± 15.800	0.000	132.840	34.650 ± 36.930
Distressed	0.000	107.730	15.080 ± 16.360	0.000	135.620	13.220 ± 20.670	0.000	135.620	14.190 ± 18.560
Temperament index[5]	-7.240	11.030	0.560 ± 2.120	-9.900	8.860	-0.610 ± 2.290	-9.900	11.030	0.000 ± 2.280

[1] Sample size for 2014 = 420, 2015 = 382, and across both years = 802. Minimum (Min), maximum (max), mean and standard deviation (SD) are reported.
[2] Docility score: scale of 1 to 6 with 1 = calm and 6 = excited.
[3] Temperament score: scale of 1 to 5, with no intermediate score; 1 = animal walks slowly while allowing close approximation with the observer and 5 = runs the entire time of the observation, jumps against the fence, and tries to attack the observer.
[4] QBA refers to qualitative behavior assessment, measured on a 136 mm visual analog scale. A low value (towards zero) indicates no or little expression, where a high value (towards 136) would indicate high or maximum expression.
[5] Temperament index: the first principal component score generated from QBA scores using a Principal Component Analysis in SAS (SAS Institute, Inc., Cary, NC.).

Grandin (1993) looked at docility (referred to as temperament for the study) on *Bos taurus* cross bulls and steers, and then broke down temperament categories into calm, calm or restless, mixed ratings, and behaviorally agitated. Overall, 51% of their bulls and 64% of their steers had better average docility scores (1.600 ± 0.100) compared to our average score of 1.870 ± 0.780 across both years, which is understandable since we were evaluating heifers and steers.

Compared to Sant'Anna and Paranhos da Costa's (2013) average QBA results, it was found that fearful was very similar whereas active, relaxed, agitated, calm, and attentive scores were lower in the current study (showed less expression) and positively occupied, happy, and curious were greater in the current study (showed more expression). Sant'Anna and Paranhos da Costa's (2013) research was conducted on *Bos indicus* cattle, while the current study was conducted on *Bos taurus* cattle. The trends that we found in this study are consistent with previous research conducted on both *Bos indicus* and *Bos taurus* and their crosses such as *Bos indicus* x Hereford crosses, Nellore-Angus crosses, and *Bos taurus* x Hereford crosses. Hearnshaw et al. (1979) found that calves that were either quarter or half Brahman had significantly poorer temperaments than those of British cross influence. Hearnshaw and Morris (1984) found significant differences in temperament between breeds of *Bos indicus* and *Bos taurus* during the study as a whole, as well as across years. Looking directly as the temperament scores of *Bos indicus* sired calves, sire breed was significant unlike when looking at *Bos taurus* sired calves where temperament scores were non-significant.

Barrozo et al. (2011) evaluated temperament score of Nellore cattle and reported a greater average temperament score compared with our values in the current experiment. Sant'Anna and Paranhos da Costa (2013) also evaluated temperament score as well as flight speed, chute score, and movement score. All average scores increased from group IV to group I, and movement score was the only evaluation method that the average score was not significant for all groups as group III and IV were similar. The average score reported for temperament score for the very good temperament group (group IV) was 1.840 ± 0.460, where the scores continually increased for the more temperamental groups. While these were the calmest cattle of their study, they were the only group that had a lower average score compared to what was found in this experiment of 1.880 ± 0.890. Both of these studies used *Bos indicus* calves, and therefore it is expected that they would score higher than the *Bos taurus* calves of this study (Barrozo et al., 2011; Sant'Anna and Paranhos da Costa, 2013).

Statistical Modeling

Breed Type

Considering that some breeds are known to differ for their temperament characteristics, breed type of the calves included in the study was evaluated for significance. Of the 15 traits evaluated, breed type was significant ($P \leq 0.05$) amongst 10 of them when included with other relevant fixed and random effects (see subsequent discussion). Of those 10, only nine of them had significant pairwise comparisons (Table 5). After reviewing the significant pairwise comparisons across traits, there was only one pairwise comparison of breed type that was significantly different ($P < 0.05$), which was the comparison of calves that were 50% Angus 50% Unknown and calves that were 75% Angus and 25% Unknown. As the breed pedigree of the older dams (i.e., dams born prior to 2012) were completely unknown and part of the breed pedigree was known for younger dams (i.e., dams born in 2012 or 2013), the breed type effect included was mostly likely picking up differences among scores due to calves from old dams compared to calves from young dams, which is not an actual breed effect. Because of this, the

Table 5. Sample sizes, least squares means (LSmeans), and standard errors (S.E.) for breed type effect of temperament traits measured across evaluators[1].

	Breed type[2]							
	50A50Un		50H25A25Un		50H50Un		75A25Un	
Evaluation method	**Number of records**	**LSmeans ± S.E.**	**Number of records**	**LSmeans ± S.E.**	**Number of records**	**LSmeans ± S.E.**	**Number of records**	**LSmeans ± S.E.**
Docility score[3]	2336	1.956 ± 0.021[a]	121	1.990 ± 0.086[a]	52	1.951 ± 0.132[a]	680	1.926 ± 0.038[a]
Temperament score[4]	2327	1.853 ± 0.026[a]	128	2.100 ± 0.108[a,b]	52	1.940 ± 0.166[b]	680	2.100 ± 0.049[b]
QBA[5] attributes								
Active	2344	48.189 ± 0.723[a]	128	51.896 ± 3.171[a,b]	52	55.921 ± 4.895[b]	684	54.103 ± 1.356[b]
Relaxed	2342	73.545 ± 0.890[a]	128	71.106 ± 3.900[a,b]	52	64.602 ± 6.016[b]	684	66.966 ± 1.666[b]
Fearful	2340	18.986 ± 0.483[a]	128	21.853 ± 2.109[a,b]	52	23.451 ± 3.255[b]	684	22.450 ± 0.904[b]
Agitated	2334	25.455 ± 0.657[a]	127	30.282 ± 7.786[a,b]	52	33.381 ± 4.297[b]	684	30.617 ± 1.192[b]
Calm	2342	78.236 ± 0.907[a]	127	74.819 ± 3.980[a,b]	52	70.305 ± 6.141[b]	684	70.893 ± 1.700[b]
Attentive	2338	54.621 ± 0.560[a]	128	58.027 ± 2.330[a]	52	59.503 ± 3.592[a]	684	55.287 ± 1.035[a]
Positively occupied	2338	34.255 ± 0.464[a]	128	36.074 ± 1.863[a]	52	32.751 ± 2.869[a]	679	32.360 ± 0.848[a]
Curious	2342	32.935 ± 0.545[a]	128	30.720 ± 2.227[a]	52	33.747 ± 3.431[a]	683	30.206 ± 1.001[a]
Irritated	2341	19.996 ± 0.569[a]	128	19.849 ± 2.494[a,b]	52	22.560 ± 3.850[b]	683	24.690 ± 1.066[b]
Apathetic	2342	46.200 ± 0.560[a]	128	43.729 ± 2.311[a,b]	52	41.392 ± 3.562[b]	684	41.820 ± 3.562[b]
Happy	2344	32.964 ± 0.524[a]	128	35.992 ± 2.167[a]	52	32.102 ± 3.340[a]	683	30.415 ± 0.967[a]
Distressed	2343	13.959 ± 0.439[a]	128	14.489 ± 1.911[a,b]	52	18.272 ± 2.950[b]	684	16.849 ± 0.822[b]
Temperament Index[6]	2307	-0.035 ± 0.049[a]	127	0.022 ± 0.203[a]	52	0.223 ± 0.312[a]	676	0.024 ± 0.091[a]

[1] The model used to evaluate breed type included date of evaluation, sex, and evaluator as additional fixed effects, sequence nested within date of evaluation as a fixed covariate, repeated measures by animal and random animal effect.
[2] 50A50Un refers to a breed type of 50% Angus and 50% Unknown. 50H25A25Un refers to a breed type of 50% Hereford, 25% Angus, and 25% Unknown. 50H50Un refers to a breed type of 50% Hereford and 50% Unknown. 75A25Un refers to a breed type of 75% Angus and 25% Unknown.
[3] Docility score: scale of 1 to 6 with 1 = calm and 6 = excited.
[4] Temperament score: scale of 1 to 5, with no intermediate score; 1 = animal walks slowly while allowing close approximation with the observer and 5 = runs the entire time of the observation, jumps against the fence, and tries to attack the observer.
[5] QBA refers to qualitative behavior assessment, measured on a 136 mm visual analog scale. A low value (towards zero) indicates no or little expression, where a high value (towards 136) would indicate high or maximum expression.
[6] Temperament index: the first principal component score generated from QBA scores using a Principal Component Analysis in SAS (SAS Institute, Inc., Cary, NC.).
[a,b] Means with different superscripts within rows are different ($P \leq 0.05$).

significance of breed type was considered an artifact and dropped from the model. Year was also not evaluated in the model due to confounding effects with evaluator, as some evaluators were present in one year, but not the other.

Across Evaluators

Covariance structure was tested with fixed effects of date of evaluation, sex, and evaluator, a fixed covariate of sequence nested within date of evaluation, and a repeated statement and random statement based on calf ID. Autoregressive 1 structure was found to be the most parsimonious modeling structure for the error term, which was then used to evaluate the model effects. The final model to produce least square means consisted of date of evaluation, sex, and evaluator as fixed effects, fixed covariate of sequence nested within date of evaluation, and repeated and random effect of calf. Least squares means were generated for fixed effects, where experiment-wise error rate was controlled using Tukey-Kramer adjustment.

Date of evaluation was a significant effect (Table 6) for all traits ($P < 0.002$). Looking across both years, temperament score was the only evaluation method in which date of evaluation was not significant across the four days of evaluation. In Year 1, agitated was the only trait that resulted in similar means compared to the other 14 traits. In Year 2, there were more similarities including active, relaxed, fearful, calm, attentive, positively occupied, curious, and temperament index. Date of evaluation was also similar for agitated, yet agitated was the only trait that was similar within year and significant across years. Café et al. (2011) evaluated temperament on Brahman and Angus in two different geographical locations in Australia. In New South Wales, flight speed was evaluated on 14 occasions and crush score was evaluated on 17 occasions where day of measurement was a significant effect in both breeds for flight speed ($P < 0.001$) and crush score ($P < 0.001$). Although there is no literature to directly support our

Table 6. Sample sizes, least squares means (LSmeans), and standard errors (S.E.) for date of evaluation effect of temperament traits measured across evaluator.

	Date of evaluation[1]							
	1		2		3		4	
Temperament evaluation	**Number of records**	**LSmeans ± S.E.**	**Number of records**	**LSmeans ± S.E.**	**Number of records**	**LSmeans ± S.E.**	**Number of records**	**LSmeans ± S.E.**
Docility score[2]	820	2.211 ± 0.038[a]	855	1.921 ± 0.037[b]	742	1.915 ± 0.040[b]	772	1.757 ± 0.039[c]
Temperament score[3]	819	1.814 ± 0.049[a]	850	1.954 ± 0.049[a]	742	1.972 ± 0.052[a]	776	1.899 ± 0.051[a]
QBA[4] attributes								
Active	824	50.083 ± 1.372[a]	856	46.735 ± 1.355[b]	748	51.885 ± 1.448[a]	780	49.472 ± 1.410[a,b]
Relaxed	823	79.880 ± 1.621[a]	856	86.754 ± 1.601[b]	747	62.137 ± 1.745[c]	780	60.016 ± 1.700[c]
Fearful	823	20.855 ± 0.873[a]	854	17.640 ± 0.863[b]	748	21.575 ± 0.941[a]	779	19.857 ± 0.918[a]
Agitated	821	25.115 ± 1.211[a]	850	23.124 ± 1.197[a]	747	29.153 ± 1.281[b]	779	29.564 ± 1.247[b]
Calm	823	90.255 ± 1.686[a]	855	93.093 ± 1.666[b]	748	60.728 ± 1.803[c]	779	63.592 ± 1.757[c]
Attentive	821	76.676 ± 0.959[a]	854	69.584 ± 0.947[b]	747	37.442 ± 1.012[c]	780	36.636 ± 0.983[c]
Positively occupied	820	54.865 ± 0.773[a]	854	44.518 ± 0.762[b]	745	18.318 ± 0.796[c]	778	17.901 ± 0.777[c]
Curious	823	52.569 ± 0.907[a]	854	45.351 ± 0.896[b]	748	16.007 ± 0.949[c]	780	15.133 ± 0.926[c]
Irritated	823	26.388 ± 1.066[a]	854	18.415 ± 1.053[b]	748	25.270 ± 1.132[a]	779	14.586 ± 1.103[b]
Apathetic	823	52.613 ± 0.982[a]	855	48.962 ± 0.969[b]	748	37.310 ± 1.024[c]	780	41.441 ± 0.999[d]
Happy	823	59.378 ± 0.869[a]	856	52.777 ± 0.857[b]	748	11.798 ± 0.921[c]	780	6.176 ± 0.899[d]
Distressed	824	21.423 ± 0.771[a]	855	13.366 ± 0.762[b]	748	17.228 ± 0.837[c]	780	7.003 ± 0.816[d]
Temperament index[5]	807	1.165 ± 0.081[a]	838	0.419 ± 0.080[b]	742	-0.742 ± 0.086[c]	775	-0.902 ± 0.084[c]

[1] 1 and 2 refers to October 6 and 7, 2014, respectively, and 3 and 4 refers to September 29 and 30, 2015, respectively.

[2] Docility score: scale of 1 to 6 with 1 = calm and 6 = excited.

[3] Temperament score: scale of 1 to 5, with no intermediate score; 1 = animal walks slowly while allowing close approximation with the observer and 5 = runs the entire time of the observation, jumps against the fence, and tries to attack the observer.

[4] QBA refers to qualitative behavior assessment, measured on a 136 mm visual analog scale. A low value (towards zero) indicates no or little expression, where a high value (towards 136) would indicate high or maximum expression.

[5] Temperament index: the first principal component score generated from QBA scores using a Principal Component Analysis in SAS (SAS Institute, Inc., Cary, NC.).

[a,b,c,d] Means with different superscripts within rows were different ($P \leq 0.05$).

Table 7. Sample sizes, least squares means (LSmeans), and standard errors (S.E.) for sex effect of temperament traits measured across evaluator.

	Sex			
	Heifer		Steer	
Temperament evaluation	**Number of records**	**LSmeans ± S.E.**	**Number of records**	**LSmeans ± S.E.**
Docility score[1]	1476	1.963 ± 0.025[a]	1705	1.939 ± 0.023[a]
Temperament score[2]	1479	1.935 ± 0.030[a]	1700	1.886 ± 0.029[a]
QBA[3] attributes				
Active	1488	51.257 ± 0.926[a]	1712	47.831 ± 0.878[b]
Relaxed	1486	70.768 ± 1.114[a]	1712	73.626 ± 1.052[b]
Fearful	1488	20.604 ± 0.606[a]	1708	19.359 ± 0.572[a]
Agitated	1483	27.425 ± 0.821[a]	1706	26.053 ± 0.778[a]
Calm	1487	75.563 ± 1.127[a]	1710	78.271 ± 1.086[b]
Attentive	1484	56.164 ± 0.691[a]	1710	54.029 ± 0.648[b]
Positively occupied	1485	33.643 ± 0.559[a]	1705	34.158 ± 0.525[a]
Curious	1486	31.828 ± 0.656[a]	1711	32.701 ± 0.615[a]
Irritated	1485	22.508 ± 0.722[a]	1711	19.821 ± 0.683[b]
Apathetic	1487	44.164 ± 0.706[a]	1711	45.998 ± 0.663[b]
Happy	1487	32.155 ± 0.629[a]	1712	32.909 ± 0.589[a]
Distressed	1488	15.788 ± 0.543[a]	1711	13.727 ± 0.511[b]
Temperament index[4]	1466	0.007 ± 0.059[a]	1688	-0.037 ± 0.055[a]

[1] Docility score: scale of 1 to 6 with 1 = calm and 6 = excited.
[2] Temperament score: scale of 1 to 5, with no intermediate score; 1 = animal walks slowly while allowing close approximation with the observer and 5 = runs the entire time of the observation, jumps against the fence, and tries to attack the observer.
[3] QBA refers to qualitative behavior assessment, measured on a 136 mm visual analog scale. A low value (towards zero) indicates no or little expression, where a high value (towards 136) would indicate high or maximum expression.
[4] Temperament index: the first principal component score generated from QBA scores using a Principal Component Analysis in SAS (SAS Institute, Inc., Cary, NC.).
[a,b] Means with different superscripts within rows were different ($P \leq 0.05$).

results, as most studies that include date of evaluation are evaluating the same animals across all days such as Café et al. (2011), Hulsman Hanna et al. (2013) evaluated the temperament of Nellore-Angus cross steers at weaning during two different seasons within a year (spring and fall) depending on when the calves were born. Season by birth year was found to be significant for all steers from all 3 cycles of population ($P < 0.001$), which is consistent with our findings.

Sex was also found to be significant amongst seven of the temperament traits, all of which were QBA attributes (Table 7). Of those seven significant QBA attributes, heifers had significantly higher means than steers for four of them, including active, attentive, irritated, and distressed. Active, irritated, and distressed are attributes that can be grouped together as more temperamental descriptors and, when considering them in that facet, the results of this project are consistent with the results found by both Voisinet et al. (1997) and Riley et al. (2014). Steers had significantly higher means than heifers for the remaining three significant attributes of relaxed, calm, and apathetic. These findings are also consistent with the work of Voisinet et al. (1997) and Riley et al. (2014) as these three attributes can be grouped together as more mellow traits and, therefore, coinciding with previous research that steers are calmer than heifers. Since steers have been found to be calmer than heifers, it makes sense that they had significantly higher means for those 3 attributes compared to the heifers. Although sex was not significant across all traits evaluated, it has been proven in literature to be an important fact in temperament evaluations (Grandin, 1993; Gauly et al., 2001), therefore, sex was kept as a fixed effect for all traits, regardless of significance.

Including evaluator as a fixed effect was significant for all traits (Tables 8 to 11), which supports our hypothesis that evaluators differ for how they interpret behaviors on subjective scales. As 2 evaluators in the first year could not return for the second year, comparable evaluators for the second year were used based on experience level. It was interesting to find, therefore, a significant difference between those specific evaluators, where evaluator 8 replaced evaluator 1 and evaluator 7 replaced evaluator 3 (Tables 9 to 11). This suggests that although their experience level may be the same, their interpretations of behavior for those given scales was quite different. There were evaluators within single traits that were similar ($P > 0.05$)

Table 8. Number of records per evaluator for temperament traits.

	Evaluator[1]							
	2	3	7	1	4	8	5	6
Temperament evaluation	**Number of records**	**Number of records**	**Number of records**	**Number of records**	**Number of records**	**Number of records**	**Number of records**	**Number of records**
Docility score[2]	801	419	382	418	801	368	-	-
Temperament score[3]	802	420	382	-	-	-	790	793
QBA[4] attributes								
Active	-	-	-	420	802	382	802	802
Relaxed	-	-	-	419	801	382	802	802
Fearful	-	-	-	420	800	381	801	802
Agitated	-	-	-	419	796	382	798	802
Calm	-	-	-	420	800	382	801	802
Attentive	-	-	-	419	801	382	799	801
Positively occupied	-	-	-	418	797	382	799	801
Curious	-	-	-	420	799	382	802	801
Irritated	-	-	-	419	801	382	802	800
Apathetic	-	-	-	420	802	382	801	801
Happy	-	-	-	419	802	382	802	802
Distressed	-	-	-	419	802	382	802	802
Temperament index[5]	-	-	-	412	781	381	791	797

[1] Evaluators are grouped by evaluation type instead of numerically.
[2] Docility score: scale of 1 to 6 with 1 = calm and 6 = excited.
[3] Temperament score: scale of 1 to 5, with no intermediate score; 1 = animal walks slowly while allowing close approximation with the observer and 5 = runs the entire time of the observation, jumps against the fence, and tries to attack the observer.
[4] QBA refers to qualitative behavior assessment, measured on a 136 mm visual analog scale. A low value (towards zero) indicates no or little expression, where a high value (towards 136) would indicate high or maximum expression.
[5] Temperament index: the first principal component score generated from QBA scores using a Principal Component Analysis in SAS (SAS Institute, Inc., Cary, NC.).

Table 9. Least squares means (LSmeans) and standard errors (S.E.) for evaluator effect of docility score measured across evaluator.

	Evaluator					
Temperament evaluation	**1**	**2**	**3**	**4**	**7**	**8**
Docility score[1]	1.596 ± 0.033[a]	1.828 ± 0.024[b]	2.009 ± 0.033[c]	1.480 ± 0.024[a]	2.577 ± 0.034[d]	2.219 ± 0.034[e]

[1] Docility score: scale of 1 to 6 with 1 = calm and 6 = excited.

[a,b,c,d,e] Means with different superscripts within rows were different ($P \leq 0.05$).

Table 10. Least squares means (LSmeans) and standard errors (S.E.) for evaluator effect of temperament score measured across evaluator.

	Evaluator				
Temperament evaluation	**2**	**3**	**5**	**6**	**7**
Temperament score[1]	1.976 ± 0.030[a]	1.848 ± 0.038[b]	1.536 ± 0.030[c]	1.961 ± 0.030[a]	2.231 ± 0.039[d]

[3] Temperament score: scale of 1 to 5, with no intermediate score; 1 = animal walks slowly while allowing close approximation with the observer and 5 = runs the entire time of the observation, jumps against the fence, and tries to attack the observer.

[a,b,c,d] Means with different superscripts within rows were different ($P \leq 0.05$).

Table 11. Least squares means (LSmeans) and standard errors (S.E.) for evaluator effect of QBA[1] and temperament index measured across evaluator.

	Evaluator				
Temperament evaluation	**1**	**4**	**5**	**6**	**8**
QBA attributes					
Active	22.472 ± 1.223[a]	34.805 ± 0.931[b]	34.055 ± 0.931[b]	86.891 ± 0.931[d]	69.498 ± 1.270[c]
Relaxed	93.878 ± 1.596[e]	62.084 ± 1.148[b]	77.985 ± 1.147[d]	74.253 ± 1.147[c]	52.784 ± 1.608[a]
Fearful	10.215 ± 0.900[a]	22.480 ± 0.637[d]	14.358 ± 0.637[b]	17.957 ± 0.637[c]	34.889 ± 0.907[e]
Agitated	20.041 ± 1.100[a]	26.530 ± 0.831[b]	19.594 ± 0.830[a]	32.560 ± 0.829[c]	34.971 ± 1.138[c]
Calm	89.409 ± 1.586[e]	70.544 ± 1.166[b]	80.934 ± 1.165[d]	75.827 ± 1.164[c]	67.871 ± 1.612[a]
Attentive	54.206 ± 1.276[b]	49.844 ± 0.863[a]	48.133 ± 0.864[a]	60.960 ± 0.863[c]	62.342 ± 1.326[c]
Positively occupied	46.450 ± 1.143[d]	15.739 ± 0.772[a]	40.893 ± 0.771[c]	42.587 ± 0.779[d]	23.823 ± 1.204[b]
Curious	38.211 ± 1.306[b]	21.941 ± 0.877[a]	38.215 ± 0.875[b]	39.715 ± 0.875[b]	23.242 ± 1.367[a]
Irritated	18.908 ± 0.976[a]	18.373 ± 0.728[a]	22.992 ± 0.723[a]	22.371 ± 0.728[a]	23.178 ± 1.001[a]
Apathetic	62.071 ± 1.281[d]	17.922 ± 0.880[b]	28.913 ± 0.880[c]	107.370 ± 0.880[e]	9.134 ± 1.344[a]
Happy	32.764 ± 1.228[c]	13.512 ± 0.818[a]	35.204 ± 0.818[c]	55.651 ± 0.818[d]	25.529 ± 1.273[b]
Distressed	12.190 ± 0.855[b]	13.652 ± 0.590[b]	16.124 ± 0.590[c]	9.024 ± 0.590[a]	22.785 ± 0.856[d]
Temperament index[2]	-0.752 ± 0.115[a]	-0.011 ± 0.077[b]	-0.027 ± 0.076[b]	-0.026 ± 0.076[b]	0.740 ± 0.118[c]

[1] QBA refers to qualitative behavior assessment, measured on a 136 mm visual analog scale. A low value (towards zero) indicates no or little expression, where a high value (towards 136) would indicate high or maximum expression.
[2] Temperament index: the first principal component score generated from QBA scores using a Principal Component Analysis in SAS (SAS Institute, Inc., Cary, NC.).
[a,b,c,d,e] Means with different superscripts within rows were different ($P \leq 0.05$).

proving that those evaluators were reading the calves' temperament similarly. For docility score, evaluators 1 and 4 were similar in their scoring and for temperament, evaluators 2 and 6 were similar in their scoring. For QBA, 10 of the 12 attributes had similarities within them; the 2 that did not have any similarities were relaxed and fearful. There were 3 attributes that had 2 different similarities within them including agitated, attentive, and curious. For agitated, evaluators 1 and 5 and evaluators 6 and 8 were similar while for attentive, evaluators 4 and 5 were similar and 6 and 8 were similar again; for curious, evaluators 1, 5, and 6 were similar and evaluators 4 and 8 were similar. For active there were similarities between evaluators 4 and 5, for happy there were similarities between evaluators 1 and 5, and for distressed there were similarities between evaluators 1 and 4. Irritated was the only QBA attribute where evaluator was similar across all 5 evaluators and for temperament index, evaluators 4, 5, and 6 were similar for temperament index. Relaxed, fearful, curious, and distressed were QBA attributes that were consistent with the differences previously discussed with the PCA factor loadings.

Sequence of cattle movement through the chute nested within date of evaluation was also found to be a significant covariate across all traits ($P \leq 0.001$). Sequence in which animals are brought through a working chute system is generally considered to be an important consideration, but very little is reported in literature. It is thought that a more temperamental animal amongst a group of calmer animals will affect the temperament of the calmer animals in a negative way (Petherick et al. 2002), but reports evaluating long-term effects of sequence are lacking. Although sequence nested within date of evaluation was significant for this study, it was left out of the final model for estimation of genetic merit. For more gregarious animals, when they are worked through a chute system they may stick with the group and resist going through the working pens, leading to a later sequence number. Likewise, in the case of more temperamental animals, they may push themselves forward, leading to an earlier sequence number. Due to this, it is likely that sequence could capture genetic variation related to these temperament methods, which may impact genetic merit predictions and is the basis for removing it from the model. Later studies with this dataset may investigate this effect further, including its impact on predictions of genetic merit. The final model for across evaluators, therefore, included fixed effects of date of evaluation, sex, and evaluator that would be used in a traditional animal model in ASReml (Gilmour et al., 2015).

Within Evaluator

Fixed effects of date of evaluation, sex, and fixed covariate of sequence nested within date of evaluation along with a random statement of calf ID were evaluated for significance within evaluator by temperament trait. Least squares means and standard errors were generated for fixed effects, where experiment-wise error rate was controlled using Tukey-Kramer adjustment.

Date of evaluation was significant for 5 of the 6 docility evaluators and 4 of the 5 temperament evaluators (Table 12), where evaluator 7 was the one evaluator for both evaluations that did not have a significant date effect. For temperament, date of evaluation was only significant across years, but was similar within years for evaluator 5. For the 12 QBA attributes, date of evaluation was also significant across most, but not all evaluators (Table 13). There were 7 instances where date of evaluation was significant across both years but not within years and those were evaluator 6 for relaxed, agitated, calm, attentive, and positively occupied, as well as evaluator 4 for curious and apathetic. After that and within Year 1, there were 4 similarities including evaluator 5 for active, evaluator 1 for positively occupied and happy, and evaluator 6 for distressed. In Year 2, there were a few more date of evaluation similarities for evaluators 4, 5, 6, and 8. Date of evaluation was similar for evaluator 4 for QBA attributes of active, fearful, attentive, and irritated, and for evaluator 5 for attributes of relaxed, calm, attentive, and curious. Date of evaluation was also similar for evaluator 6 for curious and for evaluator 8 for relaxed, fearful, agitated, attentive, and curious. When modeling TI, date of evaluation was significant across years but not within years for evaluator 6 and was similar for evaluators 4 and 5 for Year 2 (Table 13). Again, due to there not being literature available for date of evaluation that does not repeat animals, similar findings were reported in Hulsman Hanna et al. (2013) when they looked at season (spring vs. fall) by birth year when evaluating Nellore-Angus cross steers. Hulsman Hanna et al. (2013) found that birth year-season effect was significant ($F < 0.001$) for the 3 cycles of populations used for these evaluations.

Table 12. Least squares means (LSmeans) and standard errors (S.E.) for date of evaluation effect measured within evaluators for docility and temperament.

Evaluator[2] by Temperament trait	**Date of evaluation[1]**			
	1	**2**	**3**	**4**
Docility score[3]				
2	2.100 ± 0.051[a]	1.865 ± 0.050[b]	1.818 ± 0.053[b]	1.528 ± 0.052[c]
4	1.661 ± 0.051[a]	1.428 ± 0.050[b,c]	1.523 ± 0.053[a,b]	1.326 ± 0.052[c]
1	1.907 ± 0.044[a]	1.513 ± 0.043[b]	-	-
3	2.253 ± 0.054[a]	1.982 ± 0.053[b]	-	-
7	-	-	2.462 ± 0.056[a]	2.469 ± 0.055[a]
8	-	-	2.211 ± 0.033[a]	2.023 ± 0.033[b]
Temperament score[4]				
2	1.986 ± 0.065[a,b]	2.041 ± 0.063[b]	2.074 ± 0.068[b]	1.802 ± 0.067[a]
5	1.364 ± 0.047[a]	1.379 ± 0.046[a]	1.638 ± 0.050[b]	1.773 ± 0.049[b]
6	1.863 ± 0.065[a]	2.125 ± 0.064[b]	1.989 ± 0.068[a,b]	1.866 ± 0.067[a]
3	1.725 ± 0.062[a]	1.904 ± 0.060[b]	-	-
7	-	-	2.322 ± 0.065[a]	2.230 ± 0.064[a]

[1] Date of evaluation 1 and 2 refers to October 6 and 7, 2014, respectively, and 3 and 4 refers to September 29 and 30, 2015, respectively.
[2] Evaluators are grouped by year of evaluation instead of numerically.
[3] Docility score: scale of 1 to 6 with 1 = calm and 6 = excited.
[4] Temperament score: scale of 1 to 5, with no intermediate score; 1 = animal walks slowly while allowing close approximation with the observer and 5 = runs the entire time of the observation, jumps against the fence, and tries to attack the observer.
[a,b] Means with different superscripts within rows were different ($P \leq 0.05$).

Table 13. Least squares means (LSmeans) and standard errors (S.E.) for date of evaluation effect measured within evaluators for QBA[1] attributes and temperament index.

	Date of evaluation[2]			
Evaluator[3] **by temperament trait**	**1**	**2**	**3**	**4**
Active				
4	40.595 ± 1.622[a]	32.119 ± 1.595[b]	32.017 ± 1.706[b]	33.844 ± 1.680[b]
5	30.100 ± 1.527[a]	24.876 ± 1.504[a]	31.693 ± 1.607[a]	49.859 ± 1.582[b]
6	80.182 ± 2.060[a]	96.452 ± 2.027[b]	90.694 ± 2.168[b]	79.002 ± 2.134[a]
1	29.223 ± 1.246[a]	13.536 ± 1.236[b]	-	-
8	-	-	79.060 ± 2.586[a]	63.299 ± 2.561[b]
Relaxed				
4	71.743 ± 2.082[a]	60.857 ± 2.050[b,c]	62.558 ± 2.197[b]	53.526 ± 2.157[c]
5	96.198 ± 2.478[a]	85.977 ± 2.437[b]	64.255 ± 2.607[c]	63.640 ± 2.567[c]
6	89.487 ± 2.493[a]	92.915 ± 2.453[a]	52.590 ± 2.623[b]	58.969 ± 2.582[b]
1	86.790 ± 1.656[a]	122.670 ± 1.630[b]	-	-
8	-	-	42.810 ± 2.414[a]	40.954 ± 2.391[a]
Fearful				
4	22.587 ± 1.458[a]	16.256 ± 1.434[b]	26.777 ± 1.530[a]	25.643 ± 1.506[a]
5	19.217 ± 1.114[a]	13.353 ± 1.096[b]	16.397 ± 1.172[a,b]	8.537 ± 1.154[c]
6	14.399 ± 1.118[a]	19.284 ± 1.100[b]	21.387 ± 1.176[b]	17.307 ± 1.158[a,b]
1	11.884 ± 0.719[a]	6.707 ± 0.707[b]	-	-
8	-	-	34.684 ± 1.971[a]	37.839 ± 1.956[a]
Agitated				
4	21.775 ± 1.620[a]	28.249 ± 1.598[b]	29.382 ± 1.695[b]	26.920 ± 1.669[a,b]
5	25.497 ± 1.449[a]	15.605 ± 1.429[b]	22.129 ± 1.528[a]	15.569 ± 1.504[b]
6	21.365 ± 1.927[a]	27.215 ± 1.896[a]	38.200 ± 1.027[b]	45.344 ± 1.996[b]
1	22.384 ± 1.111[a]	12.015 ± 1.091[b]	-	-
8	-	-	38.574 ± 2.073[a]	37.761 ± 2.054[a]
Calm				
4	82.869 ± 2.219[a]	70.544 ± 2.183[b]	67.458 ± 2.329[b,c]	61.408 ± 2.292[c]
5	102.550 ± 2.584[a]	85.568 ± 2.542[b]	65.774 ± 2.718[c]	69.190 ± 2.682[c]
6	99.247 ± 2.085[a]	102.810 ± 2.051[a]	46.294 ± 2.193[b]	50.631 ± 2.159[b]
1	89.325 ± 1.626[a]	118.550 ± 1.601[b]	-	-
8	-	-	47.118 ± 2.853[a]	59.220 ± 2.826[b]
Attentive				
4	61.242 ± 1.540[a]	53.452 ± 1.508[b]	43.843 ± 1.616[c]	43.158 ± 1.591[c]
5	72.252 ± 1.865[a]	57.446 ± 1.840[b]	34.842 ± 1.967[c]	28.388 ± 1.932[c]
6	86.954 ± 1.487[a]	92.036 ± 1.460[a]	30.353 ± 1.561[b]	32.036 ± 1.537[b]
1	77.970 ± 1.373[a]	65.694 ± 1.348[b]	-	-
8	-	-	45.760 ± 2.002[a]	44.348 ± 1.983[a]
Positively occupied				
4	21.758 ± 1.387[a]	8.993 ± 1.362[b]	13.448 ± 1.456[b]	22.260 ± 1.433[a]
5	73.567 ± 1.798[a]	43.714 ± 1.773[b]	31.899 ± 1.902[c]	14.784 ± 1.862[d]
6	70.943 ± 0.682[a]	71.399 ± 0.671[a]	12.861 ± 0.718[b]	12.045 ± 0.708[b]
1	60.163 ± 1.133[a]	62.997 ± 1.111[a]	-	-
8	-	-	4.852 ± 1.099[a]	9.103 ± 1.088[b]

Table 13. Least squares means (LSmeans) and standard errors (S.E.) for date of evaluation effect measured within evaluators for QBA[1] attributes and temperament index (continued).

	Date of evaluation			
Evaluator[3] by temperament trait	**1**	**2**	**3**	**4**
Curious				
4	27.369 ± 1.655[a]	29.074 ± 1.633[a]	16.919 ± 1.738[b]	16.459 ± 1.717[b]
5	66.675 ± 1.942[a]	35.430 ± 1.907[b]	32.426 ± 2.043[b]	19.678 ± 2.012[c]
6	64.951 ± 1.301[a]	75.009 ± 1.277[b]	5.969 ± 1.369[c]	9.940 ± 1.347[c]
1	58.748 ± 1.585[a]	50.589 ± 1.560[b]	-	-
8	-	-	5.541 ± 1.032[a]	7.703 ± 1.022[a]
Irritated				
4	18.679 ± 1.336[a]	14.343 ± 1.317[b]	25.939 ± 1.405[a]	15.874 ± 1.383[a]
5	33.396 ± 1.605[a]	19.025 ± 1.579[b,c]	23.756 ± 1.688[b]	16.095 ± 1.662[c]
6	22.734 ± 1.545[a,b]	25.036 ± 1.520[b]	24.844 ± 1.625[b]	17.064 ± 1.605[a]
1	27.715 ± 1.188[a]	12.128 ± 1.167[b]	-	-
8	-	-	33.472 ± 1.733[a]	11.969 ± 1.716[b]
Apathetic				
4	23.067 ± 1.576[a]	22.183 ± 1.552[a]	14.718 ± 1.658[b]	11.112 ± 1.632[b]
5	47.797 ± 1.778[a]	29.620 ± 1.754[b]	23.119 ± 1.871[b]	13.954 ± 1.842[c]
6	104.070 ± 2.029[a]	114.800 ± 1.991[b]	87.871 ± 2.130[b]	121.400 ± 2.096[c]
1	70.849 ± 1.455[a]	64.549 ± 1.429[b]	-	-
8	-	-	4.266 ± 0.683[a]	2.307 ± 0.678[b]
Happy				
4	29.007 ± 1.193[a]	13.165 ± 1.171[b]	10.173 ± 1.255[b,c]	6.023 ± 1.236[c]
5	63.094 ± 1.682[a]	40.522 ± 1.654[b]	26.458 ± 1.770[c]	13.033 ± 1.742[d]
6	97.292 ± 0.990[a]	104.750 ± 0.974[b]	10.834 ± 1.042[c]	5.357 ± 1.025[d]
1	54.724 ± 1.203[a]	57.008 ± 1.181[a]	-	-
8	-	-	5.250 ± 0.630[a]	1.316 ± 0.624[b]
Distressed				
4	18.031 ± 1.301[a]	7.163 ± 1.277[b]	19.315 ± 1.369[a]	11.452 ± 1.347[b]
5	22.753 ± 1.358[a]	14.866 ± 1.336[b]	18.463 ± 1.429[a,b]	9.072 ± 1.406[c]
6	14.765 ± 0.740[a]	14.562 ± 0.724[a]	4.746 ± 0.774[b]	1.530 ± 0.762[c]
1	20.753 ± 0.820[a]	8.495 ± 0.809[b]	-	-
8	-	-	32.847 ± 1.689[a]	10.084 ± 1.669[b]
Temperament index[4]				
4	-0.236 ± 0.163[a,b]	-0.347 ± 0.161[a]	0.347 ± 0.169[b]	0.316 ± 0.166[b]
5	1.199 ± 0.144[a]	0.485 ± 0.143[b]	-0.753 ± 0.153[c]	-1.108 ± 0.149[c]
6	1.639 ± 0.098[a]	1.696 ± 0.096[a]	-2.038 ± 0.103[b]	-1.658 ± 0.102[c]
1	1.130 ± 0.147[a]	-1.045 ± 0.143[b]	-	-
8	-	-	0.393 ± 0.167[a]	-0.361 ± 0.165[b]

[1] QBA refers to qualitative behavior assessment.
[2] Date of evaluation 1 and 2 refers to October 6 and 7, 2014, respectively, and 3 and 4 refers to September 29 and 30, 2015, respectively.
[3] Evaluators are grouped by year of evaluation instead of numerically.
[4] Temperament index: the first principal component score generated from QBA scores using a Principal Component Analysis in SAS (SAS Institute, Inc., Cary, NC.).
[a,b] Means with different superscripts within rows were different ($P \leq 0.05$).

Within evaluator, docility score and temperament index did not have significant sex effects, yet one evaluator each for temperament score and 2 QBA traits had significant sex effects (Table 14). For temperament score, sex was a significant effect for evaluator 3 and for active and attentive, it was significant for evaluator 1. For all 3 cases, heifers had significantly higher means than steers, consistent with the previous model discussed when evaluator was included in the model and findings of Voisinet et al. (1997) and Riley et al. (2014). The trends discussed when evaluator was included in the model generally remained true in this case, even if differences were significant within an evaluator. This could be partially due to sample size for some evaluators, where additional years of collection may find that within an evaluator sex is a significant effect.

Similar to the previous model that included evaluator, date of evaluation and sex were not significant within all evaluators, yet both have been proven in previous literature to be important factors in evaluating temperament and therefore, regardless of significance, they were both kept as fixed effects for all evaluators. Just like across evaluators, sequence nested within date was also found to be a significant covariate within all evaluators ($P \leq 0.0001$). Again, it was left out of the final model for estimation of genetic merit as it is likely to capture genetic variation related to these temperament evaluations and therefore, possibly impacting genetic merit predictions. The final model for within evaluators included fixed effects of date of evaluation and sex that would be used in a traditional animal model in ASReml (Gilmour et al., 2015).

Table 14. Sample size, least squares means (LSmeans), and standard errors (S.E.) for sex effect measured within evaluators across temperament traits.

	Sex			
	Heifer		Steer	
Evaluator by temperament trait	Number of records	LSmeans ± S.E.	Number of records	LSmeans ± S.E.
Docility score[1]				
1	206	1.723 ± 0.044[a]	212	1.698 ± 0.044[a]
2	371	1.840 ± 0.038[a]	428	1.818 ± 0.035[a]
3	207	2.127 ± 0.053[a]	212	2.108 ± 0.053[a]
4	370	1.516 ± 0.038[a]	429	1.454 ± 0.035[a]
7	164	2.464 ± 0.060[a]	216	2.467 ± 0.052[a]
8	159	2.124 ± 0.035[a]	207	2.110 ± 0.031[a]
Temperament score[2]				
2	371	1.965 ± 0.048[a]	429	1.987 ± 0.045[a]
3	207	1.933 ± 0.061[b]	213	1.696 ± 0.061[a]
5	363	1.555 ± 0.035[a]	425	1.522 ± 0.033[a]
6	369	1.963 ± 0.048[a]	422	1.958 ± 0.045[a]
7	164	2.259 ± 0.069[a]	216	2.282 ± 0.060[a]
QBA[3] attributes				
Active				
1	207	23.311 ± 1.242[b]	213	19.446 ± 1.229[a]
4	371	35.920 ± 1.213[a]	429	33.368 ± 1.127[a]
5	371	34.899 ± 1.142[a]	429	33.365 ± 1.062[a]
6	371	86.709 ± 1.541[a]	429	86.455 ± 1.432[a]
8	164	74.550 ± 2.755[a]	216	67.809 ± 2.404[a]
Relaxed				
1	206	102.690 ± 1.651[a]	213	106.770 ± 1.630[a]
4	371	61.698 ± 1.557[a]	428	62.645 ± 1.449[a]
5	371	77.031 ± 1.853[a]	429	78.004 ± 1.722[a]
6	371	73.279 ± 1.864[a]	429	73.701 ± 1.733[a]
8	164	40.773 ± 2.572[a]	216	42.991 ± 2.244[a]
Fearful				
1	207	9.860 ± 0.717[a]	213	8.730 ± 0.709[a]
4	371	24.086 ± 1.087[a]	427	21.546 ± 1.013[a]
5	371	14.820 ± 0.833[a]	428	13.932 ± 0.774[a]
6	371	18.951 ± 0.836[a]	429	17.238 ± 0.777[a]
8	164	35.507 ± 2.100[a]	215	37.016 ± 1.836[a]
Agitated				
1	207	18.291 ± 1.105[a]	212	16.108 ± 1.097[a]
4	367	26.523 ± 1.212[a]	427	26.640 ± 1.122[a]
5	371	20.690 ± 1.083[a]	425	18.711 ± 1.010[a]
6	371	34.083 ± 1.441[a]	429	31.979 ± 1.339[a]
8	164	37.480 ± 2.209[a]	216	38.855 ± 1.927[a]

Table 14. Sample size, least squares means (LSmeans), and standard errors (S.E.) for sex effect measured within evaluators across temperament traits (continued).

	Sex			
	Heifer		Steer	
Evaluator by temperament trait	**Number of records**	**LSmeans ± S.E.**	**Number of records**	**LSmeans ± S.E.**
Calm				
1	207	101.810 ± 1.622[a]	213	106.070 ± 1.605[a]
4	370	69.555 ± 1.657[a]	428	71.585 ± 1.540[a]
5	371	80.388 ± 1.932[a]	428	81.155 ± 1.797[a]
6	371	73.290 ± 1.559[a]	429	76.202 ± 1.449[a]
8	164	52.223 ± 3.040[a]	216	54.115 ± 2.652[a]
Attentive				
1	207	74.262 ± 1.366[b]	212	69.402 ± 1.355[a]
4	371	51.714 ± 1.149[a]	428	49.133 ± 1.068[a]
5	368	48.474 ± 1.400[a]	429	47.990 ± 1.295[a]
6	371	61.157 ± 1.110[a]	428	59.532 ± 1.032[a]
8	164	46.048 ± 2.133[a]	216	44.060 ± 1.861[a]
Positively occupied				
1	205	61.572 ± 1.130[a]	213	61.588 ± 1.114[a]
4	368	17.847 ± 1.036[a]	427	15.383 ± 0.962[a]
5	369	40.302 ± 1.348[a]	428	41.679 ± 1.251[a]
6	370	41.903 ± 0.511[a]	429	41.721 ± 0.474[a]
8	164	6.739 ± 1.171[a]	126	7.217 ± 1.022[a]
Curious				
1	207	56.139 ± 1.581[a]	213	53.198 ± 1.563[a]
4	369	22.328 ± 1.238[a]	428	22.582 ± 1.149[a]
5	371	37.341 ± 39.764[a]	429	39.764 ± 1.348[a]
6	371	39.579 ± 0.973[a]	429	38.356 ± 0.903[a]
8	164	6.200 ± 1.100[a]	216	7.044 ± 0.960[a]
Irritated				
1	206	20.525 ± 1.185[a]	213	19.318 ± 1.170[a]
4	371	19.768 ± 0.999[a]	428	17.650 ± 0.929[a]
5	371	23.890 ± 1.200[a]	429	22.250 ± 1.115[a]
6	370	22.741 ± 1.157[a]	428	22.098 ± 1.074[a]
8	164	22.028 ± 1.846[a]	216	23.413 ± 1.611[a]
Apathetic				
1	207	67.405 ± 1.451[a]	213	67.993 ± 1.432[a]
4	371	17.090 ± 1.178[a]	429	18.451 ± 1.096[a]
5	370	27.312 ± 1.332[a]	429	29.933 ± 1.236[a]
6	371	105.740 ± 1.514[a]	428	108.330 ± 1.409[a]
8	164	3.754 ± 0.728[a]	216	2.820 ± 0.635[a]
Happy				
1	206	54.504 ± 1.200[a]	213	57.228 ± 1.184[a]
4	371	15.480 ± 0.892[a]	429	13.704 ± 0.828[a]
5	371	35.194 ± 1.258[a]	429	36.359 ± 1.168[a]
6	371	54.667 ± 0.740[a]	429	54.447 ± 0.688[a]
8	164	3.590 ± 0.671[a]	216	2.976 ± 0.586[a]

Table 14. Sample size, least squares means (LSmeans), and standard errors (S.E.) for sex effect measured within evaluators across temperament traits (continued).

	Sex			
	Heifer		Steer	
Evaluator by temperament trait	**Number of records**	**LSmeans ± S.E.**	**Number of records**	**LSmeans ± S.E.**
Distressed				
1	207	15.310 ± 0.818[a]	212	13.938 ± 0.811[a]
4	371	14.881 ± 0.973[a]	429	13.099 ± 0.903[a]
5	371	17.567 ± 1.015[a]	429	15.010 ± 0.944[a]
6	371	9.410 ± 0.550[a]	429	8.392 ± 0.511[a]
8	164	21.835 ± 1.796[a]	216	21.096 ± 1.567[a]
Temperament index[4]				
1	202	0.224 ± 0.146[a]	210	-0.139 ± 0.144[a]
4	361	0.128 ± 0.121[a]	418	-0.088 ± 0.112[a]
5	366	-0.131 ± 0.108[a]	423	0.042 ± 0.100[a]
6	369	-0.122 ± 0.073[a]	426	-0.058 ± 0.068[a]
8	164	0.048 ± 0.177[a]	215	-0.016 ± 0.155[a]

[1] Docility score: scale of 1 to 6 with 1 = calm and 6 = excited.
[2] Temperament score: scale of 1 to 5, with no intermediate score; 1 = animal walks slowly while allowing close approximation with the observer and 5 = runs the entire time of the observation, jumps against the fence, and tries to attack the observer.
[3] QBA refers to qualitative behavior assessment, measured on a 136 mm visual analog scale. A low value (towards zero) indicates no or little expression, where a high value (towards 136) would indicate high or maximum expression.
[4] Temperament index: the first principal component score generated from QBA scores using a Principal Component Analysis in SAS (SAS Institute, Inc., Cary, NC.).
[a,b] Means with different superscripts within rows were different ($P \leq 0.05$).

Genetic Parameters

Across Evaluators

The final model across evaluators was used with known pedigree on the calves, where estimates of additive genetic variance, permanent environmental variance, phenotypic variance, heritability, and the proportion of phenotypic variance due to permanent environmental effects for all 15 traits were generated (Table 15).

Of the 15 traits, QBA attributes of attentive, positively occupied, happy, and temperament index had an estimated additive genetic variance of 0.000 ± 0.000, which means that heritability could not be estimated. It must be noted that our accuracy is relatively low,

Table 15. Genetic parameter estimates ($\hat{\sigma}_a^2$, $\hat{\sigma}_{pe}^2$, and $\hat{\sigma}_p^2$), heritability ($\hat{h}^2$), and proportion of phenotypic variance due to permanent environment effects ($\hat{c}^2$) measured across evaluators[1].

Temperament evaluation	$\hat{\sigma}_a^2$	$\hat{\sigma}_{pe}^2$	$\hat{\sigma}_p^2$	$\hat{h}^2$	$\hat{c}^2$
Docility score[2]	0.802 ± 0.033	0.116 ± 0.031	0.562 ± 0.017	0.143 ± 0.058	0.206 ± 0.055
Temperament score[3]	0.212 ± 0.071	0.211 ± 0.061	0.806 ± 0.029	0.264 ± 0.083	0.262 ± 0.078
QBA[4] attributes					
Active	145.620 ± 52.402	65.072 ± 46.815	991.180 ± 27.479	0.147 ± 0.052	0.066 ± 0.047
Relaxed	199.910 ± 73.350	227.564 ± 66.345	1209.700 ± 36.924	0.165 ± 0.059	0.188 ± 0.055
Fearful	43.377 ± 19.453	67.819 ± 18.280	370.560 ± 10.741	0.117 ± 0.012	0.183 ± 0.049
Agitated	122.710 ± 41.941	120.450 ± 37.176	605.020 ± 19.441	0.203 ± 0.067	0.199 ± 0.062
Calm	281.580 ± 88.055	216.114 ± 76.364	1201.000 ± 39.351	0.234 ± 0.070	0.180 ± 0.065
Attentive	0.000 ± 0.000	53.387 ± 10.985	645.190 ± 16.331	0.000 ± 0.000	0.083 ± 0.017
Positively occupied	0.000 ± 0.000	0.000 ± 0.000	641.410 ± 16.078	Not estimable	Not estimable
Curious	0.185 ± 5.861	0.000 ± 0.000	693.250 ± 18.263	0.000 ± 0.009	0.000 ± 0.000
Irritated	93.792 ± 33.396	112.786 ± 29.838	467.270 ± 15.474	0.201 ± 0.069	0.241 ± 0.065
Apathetic	0.001 ± 0.000	0.000 ± 0.000	1969.000 ± 49.287	Not estimable	Not estimable
Happy	0.000 ± 0.000	0.000 ± 0.000	758.300 ± 18.978	Not estimable	Not estimable
Distressed	32.378 ± 15.616	61.309 ± 15.027	320.46 ± 9.186	0.101 ± 0.048	0.191 ± 0.047
Temperament index[5]	0.000 ± 0.000	0.000 ± 0.000	4.690 ± 0.118	Not estimable	0.000 ± 0.000

[1] $\hat{\sigma}_a^2$ is estimated additive genetic variance, $\hat{\sigma}_{pe}^2$ is estimated permanent environmental variance, $\hat{\sigma}_p^2$ is estimated phenotypic variance, $\hat{h}^2$ is the estimated heritability, the proportion of phenotypic variance due to additive gene effects, and $\hat{c}^2$ is the estimated proportion of phenotypic variance due to permanent environment effects.

[2] Docility score: scale of 1 to 6 with 1 = calm and 6 = excited.

[3] Temperament score: scale of 1 to 5, with no intermediate score; 1 = animal walks slowly while allowing close approximation with the observer and 5 = runs the entire time of the observation, jumps against the fence, and tries to attack the observer.

[4] QBA refers to qualitative behavior assessment, measured on a 136 mm visual analog scale. A low value (towards zero) indicates no or little expression, where a high value (towards 136) would indicate high or maximum expression.

[5] Temperament index: the first principal component score generated from QBA scores using a Principal Component Analysis in SAS (SAS Institute, Inc., Cary, NC.).

which is supported by the moderately large standard errors found for heritability and c^2. Heritability is estimated as the proportion of phenotypic variation due to additive genetic effects. Therefore, non-estimable heritability could be due to 1) dominance and epistatic genetic effects or environmental effects having a larger impact than expected on phenotypic variation, or to 2) the overall phenotypic variation being low in the current population, leading to limitation of properly estimating the proportion due to additive gene effects. There are limitations to our current pedigree as dams born prior to 2012 had unknown parentage and some calves used in the study were bred via natural service in a multi-sire pasture. As parentage testing has not been completed on these calves yet, there is possibility of relationship and inbreeding calculations not reflecting their true levels. This could lead to issues with accurately calculating additive genetic variance or understanding the proportion of phenotypic variance due to additive gene effects. This limitation should be addressed in future studies with this data. In terms of environmental effects, the non-estimable heritability could be due to there being less environmental differences between relatives, especially parents and offspring (Lush, 1994).

In addition, the proportion of phenotypic variance due to permanent environmental effects ($\hat{c}^2$) was not estimable for positively occupied, apathetic, and happy due to their additive variance being equal to zero. For the remaining traits, heritability ranged from low to moderate of 0.000 ± 0.008 (curious) to 0.264 ± 0.083 (TS). Heritability reported for docility score was similar to what Hoppe et al. (2010) reported (0.11 ± 0.080 to 0.33 ± 0.10). As for temperament score, our results found that it was more heritable, with a heritability of 0.264 ± 0.083, compared to that of 0.180 ± 0.020 reported by Barrozo et al. (2012). For $\hat{c}^2$, the proportion ranged from 0.000 ± 0.000 to 0.262 ± 0.078 indicating that there is a low to moderately low correlation of a permanent environment effect to phenotypic variance.

Within Evaluators

Parameters of estimated additive genetic variance, estimated phenotypic variance, and heritability across all 15 traits for each evaluator were generated (Table 16). There were 10 instances that the estimated additive genetic variance was 0.000 ± 0.000 and, therefore, heritability could not be estimated. Of those instances, all but 4 of them were within traits that could not be estimated across evaluators, and the remaining 4 were within QBA traits of curious, apathetic, and distressed. Heritability ranged from low to moderate of 0.000 ± 0.000 (happy – evaluator 5) to 0.566 ± 0.290 (irritated – evaluator 8) for traits that were heritable. Due to some evaluators only being present in one year, sample size limitations were a bigger issue for these estimates of genetic parameters, which are shown with the large standard errors for heritability. As sample size increases in future years of this project, it is likely that the accuracy of this research will increase and the standard errors for heritability will decrease.

Table 16. Genetic parameter estimates ($\hat{\sigma}_a^2$ and $\hat{\sigma}_p^2$) and heritability ($\hat{h}^2$) measured within evaluators across temperament traits[1].

Evaluator by temperament trait	$\hat{\sigma}_a^2$	$\hat{\sigma}_p^2$	$\hat{h}^2$
Docility score[2]			
1	0.151 ± 0.095	0.406 ± 0.030	0.372 ± 0.222
2	0.103 ± 0.054	0.532 ± 0.027	0.193 ± 0.099
3	0.057 ± 0.065	0.589 ± 0.041	0.097 ± 0.110
4	0.009 ± 0.031	0.528 ± 0.026	0.017 ± 0.059
7	0.312 ± 0.164	0.595 ± 0.048	0.524 ± 0.255
8	0.009 ± 0.031	0.528 ± 0.026	0.017 ± 0.059
Temperament score[3]			
2	0.190 ± 0.092	0.868 ± 0.045	0.218 ± 0.103
3	0.231 ± 0.175	0.793 ± 0.058	0.291 ± 0.213
5	0.092 ± 0.049	0.455 ± 0.024	0.202 ± 0.105
6	0.240 ± 0.102	0.874 ± 0.046	0.245 ± 0.112
7	0.189 ± 0.134	0.786 ± 0.059	0.241 ± 0.166
QBA[4] attributes			
Active			
1	136.650 ± 77.987	322.820 ± 24.357	0.423 ± 0.227
4	169.730 ± 67.732	554.290 ± 29.196	0.306 ± 0.116
5	163.660 ± 66.370	521.730 ± 27.618	0.314 ± 0.121
6	282.210 ± 110.86	934.240 ± 49.050	0.302 ± 0.113
8	129.380 ± 202.580	1243.400 ± 91.887	0.104 ± 0.162
Relaxed			
1	119.050 ± 78.839	558.200 ± 39.492	0.213 ± 0.138
4	265.180 ± 106.180	904.630 ± 47.425	0.293 ± 0.112
5	273.540 ± 147.470	1291.800 ± 66.931	0.218 ± 0.111
6	416.390 ± 158.060	1389.400 ± 72.663	0.300 ± 0.108
8	171.710 ± 174.610	1088.400 ± 80.680	0.158 ± 0.158
Fearful			
1	46.780 ± 30.015	109.170 ± 8.396	0.429 ± 0.259
4	109.280 ± 55.243	451.810 ± 23.625	0.242 ± 0.118
5	35.824 ± 23.222	260.710 ± 13.263	0.137 ± 0.880
6	33.078 ± 22.326	262.190 ± 13.301	0.126 ± 0.084
8	219.710 ± 175.700	734.900 ± 56.810	0.299 ± 0.230
Agitated			
1	32.699 ± 31.420	252.230 ± 17.685	0.130 ± 0.130
4	149.190 ± 63.976	542.400 ± 28.460	0.275 ± 0.113
5	127.020 ± 54.716	449.170 ± 23.637	0.283 ± 0.117
6	341.450 ± 104.860	776.150 ± 42.043	0.440 ± 0.124
8	72.963 ± 112.860	801.530 ± 58.879	0.091 ± 0.140

Table 16. Genetic parameter estimates ($\hat{\sigma}_a^2$ and $\hat{\sigma}_p^2$) and heritability ($\hat{h}^2$) measured within evaluators across temperament traits[1] (continued).

Evaluator by temperament trait	$\hat{\sigma}_a^2$	$\hat{\sigma}_p^2$	h^2
Calm			
1	162.310 ± 97.437	542.200 ± 39.104	0.299 ± 0.173
4	432.190 ± 140.350	1029.800 ± 55.770	0.420 ± 0.126
5	392.270 ± 160.840	1417.800 ± 73.995	0.277 ± 0.109
6	368.760 ± 126.100	1083.500 ± 57.018	0.340 ± 0.110
8	361.850 ± 295.130	1527.700 ± 115.130	0.237 ± 0.188
Attentive			
1	0.000 ± 0.000	385.070 ± 26.700	Not estimable
4	0.000 ± 0.000	488.740 ± 24.498	Not estimable
5	0.000 ± 0.000	724.580 ± 36.365	Not estimable
6	70.667 ± 48.196	664.340 ± 33.565	0.106 ± 0.072
8	102.940 ± 112.680	747.960 ± 55.246	0.138 ± 0.149
Positively occupied			
1	35.181 ± 35.364	262.300 ± 18.456	0.134 ± 0.133
4	68.738 ± 45.947	425.840 ± 21.926	0.161 ± 0.106
5	0.000 ± 0.000	683.390 ± 34.298	Not estimable
6	97.411 ± 29.990	264.480 ± 13.935	0.368 ± 0.106
8	16.073 ± 24.549	24.530 ± 16.424	0.072 ± 0.109
Curious			
1	41.899 ± 58.461	516.670 ± 36.010	0.081 ± 0.112
4	0.000 ± 0.000	568.140 ± 28.514	Not estimable
5	0.000 ± 0.000	815.630 ± 40.858	Not estimable
6	146.280 ± 65.119	642.330 ± 33.098	0.228 ± 0.098
8	12.452 ± 24.868	198.270 ± 14.513	0.062 ± 0.125
Irritated			
1	104.600 ± 72.749	294.100 ± 22.072	0.036 ± 0.236
4	102.470 ± 43.682	388.070 ± 20.216	0.264 ± 0.108
5	79.391 ± 50.992	546.800 ± 27.859	0.145 ± 0.092
6	252.650 ± 71.389	506.830 ± 27.861	0.499 ± 0.127
8	329.530 ± 183.290	582.550 ± 47.894	0.566 ± 0.290
Apathetic			
1	44.688 ± 50.974	435.460 ± 30.408	0.103 ± 0.116
4	137.420 ± 72.774	520.720 ± 27.626	0.264 ± 0.135
5	26.965 ± 46.765	658.250 ± 33.107	0.041 ± 0.071
6	254.180 ± 107.99	995.530 ± 51.671	0.255 ± 0.104
8	0.000 ± 0.000	86.664 ± 6.312	Not estimable

Table 16. Genetic parameter estimates ($\hat{\sigma}_a^2$ and $\hat{\sigma}_p^2$) and heritability ($\hat{h}^2$) measured within evaluators across temperament traits[1] (continued).

Evaluator by temperament trait	$\hat{\sigma}_a^2$	$\hat{\sigma}_p^2$	h^2
Happy			
1	21.244 ± 35.134	296.170 ± 20.665	0.072 ± 0.118
4	23.817 ± 24.471	304.290 ± 15.359	0.078 ± 0.080
5	0.000 ± 0.000	587.060 ± 29.408	0.000 ± 0.000
6	202.910 ± 65.831	662.110 ± 34.291	0.307 ± 0.094
8	0.000 ± 0.000	73.745 ± 5.371	Not estimable
Distressed			
1	12.974 ± 19.644	138.170 ± 9.695	0.094 ± 0.141
4	101.540 ± 44.746	376.100 ± 19.680	0.270 ± 0.114
5	70.910 ± 39.254	387.790 ± 19.894	0.183 ± 0.099
6	0.000 ± 0.000	115.000 ± 5.761	Not estimable
8	416.820 ± 180.560	570.020 ± 43.720	0.731 ± 0.295
Temperament index[5]			
1	1.728 ± 1.055	4.338 ± 0.329	0.398 ± 0.230
4	1.964 ± 0.733	5.348 ± 0.290	0.367 ± 0.129
5	0.813 ± 0.440	4.284 ± 0.222	0.190 ± 0.100
6	1.242 ± 0.348	2.752 ± 0.148	0.451 ± 0.116
8	1.881 ± 1.242	5.206 ± 0.406	0.361 ± 0.227

[1] $\hat{\sigma}_a^2$ is estimated additive genetic variance, $\hat{\sigma}_p^2$ is estimated phenotypic variance, and $\hat{h}^2$ is estimated heritability, the proportion of phenotypic variance due to additive gene effects.
[2] Docility score: scale of 1 to 6 with 1 = calm and 6 = excited.
[3] Temperament score: scale of 1 to 5, with no intermediate score; 1 = animal walks slowly while allowing close approximation with the observer and 5 = runs the entire time of the observation, jumps against the fence, and tries to attack the observer.
[4] QBA refers to qualitative behavior assessment, measured on a 136 mm visual analog scale. A low value (towards zero) indicates no or little expression, where a high value (towards 136) would indicate high or maximum expression.
[5] Temperament index: the first principal component score generated from QBA scores using a Principal Component Analysis in SAS (SAS Institute, Inc., Cary, NC.).

Estimated Breeding Value Comparisons

To determine the impact of evaluators on the prediction of genetic merit, estimated breeding value (EBV) comparisons were made based on 1) Pearson and Spearman Rank correlation coefficients, and 2) changes in quartile rankings of animals within a method across evaluators, similar to that found in Hulsman Hanna et al. (2014). Although some evaluators did not have records for both years, the animal model conducted in ASReml generates EBVs for all animals in the pedigree, including those without records.

For docility score (Table 17), there were 4 non-significant Pearson and Spearman Rank correlation coefficients, including evaluator 1 vs. 7 and evaluator 3 vs. 7. All correlation coefficients were significant for temperament score (Table 18) and temperament index had 1 non-significant Pearson correlation coefficient (Table 31). For QBA attributes (Tables 19 to 30), all of the evaluators had non-significant correlation coefficients for at least one trait where active and irritated always had significant Pearson correlation coefficients and active, agitated, calm, irritated, and happy always had significant Spearman Rank correlation coefficients. To serve as a standard comparison, EBV generated when evaluator was included in the model were also used in the correlation coefficient calculations.

If evaluators are interpreting behaviors similarly for a given trait or evaluation method, this should translate to similar EBVs and, therefore, it is expected to have significantly high correlations between evaluators. For the purpose of this experiment, 0.000 to 0.200 is considered a low correlation, 0.200 to 0.400 is moderately low, 0.400 to 0.700 is moderately high, and 0.700 to 1.000 is high. Our model across evaluators does indicate that significantly high correlations is not always true, therefore, some level of differences may exist in their EBV rankings. This is supported by low or non-significant correlations reported for all 15 traits being evaluated.

Table 17. Pearson and Spearman Rank correlation coefficients for docility score estimated breeding values across and within evaluator[1].

	Across[2]	1	2	3	4	7	8
Across[2]		0.641	0.840	0.614	0.761	0.492	0.761
1	0.591		0.428	0.630	0.460	-	0.460
2	0.821	0.371		0.484	0.566	0.327	0.566
3	0.533	0.566	0.388		0.376	-	0.376
4	0.690	0.350	0.483	0.178		0.247	1.000
7	0.464	-	0.286	-	0.261		0.247
8	0.690	0.350	0.483	0.178	1.000	0.261	

[1] Docility score: scale of 1 to 6 with 1 = calm and 6 = excited. Pearson correlations are above and to the right of the diagonal and Spearman correlations are below and to the left of the diagonal. Significant correlations ($P < 0.05$) are displayed in the table. Correlations between two traits with "-" are non-significant ($P > 0.05$).
[2] Estimated breeding values used to calculate correlation coefficients when the model included fixed effects of date of evaluation, sex, evaluator, random effect of calf, and permanent environmental effect to account for repeated records.

Table 18. Pearson and Spearman Rank correlation coefficients for temperament score estimated breeding values across and within evaluator[1].

	Across[2]	2	3	5	6	7
Across[2]		0.873	0.668	0.797	0.862	0.596
2	0.834		0.531	0.656	0.707	0.510
3	0.615	0.531		0.480	0.567	0.088
5	0.773	0.656	0.480		0.618	0.432
6	0.829	0.707	0.567	0.618		0.438
7	0.441	0.510	0.088	0.432	0.438	

[1] Temperament score: scale of 1 to 5, with no intermediate score; 1 = animal walks slowly while allowing close approximation with the observer and 5 = runs the entire time of the observation, jumps against the fence, and tries to attack the observer. Pearson correlations are above and to the right of the diagonal and Spearman correlations are below and to the left of the diagonal. Significant correlations ($P < 0.05$) are displayed in the table.
[2] Estimated breeding values used to calculate correlation coefficients when the model included fixed effects of date of evaluation, sex, evaluator, random effect of calf, and permanent environmental effect to account for repeated records.

Table 19. Pearson and Spearman Rank correlation coefficients for active QBA attribute estimated breeding values across and within evaluator[1].

	Across[2]	**2**	**4**	**5**	**6**	**8**
Across[2]		0.452	0.795	0.697	0.795	0.695
2	0.451		0.437	0.380	0.154	-0.010
4	0.778	0.391		0.568	0.472	0.396
5	0.661	0.357	0.534		0.316	0.309
6	0.783	0.225	0.466	0.299		0.596
8	0.634	0.002	0.373	0.265	0.534	

[1] QBA refers to qualitative behavior assessment, measured on a 136 mm visual analog scale. A low value (towards zero) indicates no or little expression, where a high value (towards 136) would indicate high or maximum expression. Pearson correlations are above and to the right of the diagonal and Spearman correlations are below and to the left of the diagonal. Significant correlations ($P < 0.05$) are displayed in the table.
[2] Estimated breeding values used to calculate correlation coefficients when the model included fixed effects of date of evaluation, sex, evaluator, random effect of calf, and permanent environmental effect to account for repeated records.

Table 20. Pearson and Spearman Rank correlation coefficients for relaxed QBA attribute estimated breeding values across and within evaluator[1].

	Across[2]	**2**	**4**	**5**	**6**	**8**
Across[2]		0.432	0.801	0.818	0.775	0.663
2	0.427		0.530	0.358	0.155	-
4	0.801	0.521		0.447	0.366	0.131
5	0.825	0.362	0.471		0.534	0.457
6	0.737	0.156	0.373	0.501		0.507
8	0.588	-	-	0.444	0.467	

[1] QBA refers to qualitative behavior assessment, measured on a 136 mm visual analog scale. A low value (towards zero) indicates no or little expression, where a high value (towards 136) would indicate high or maximum expression. Pearson correlations are above and to the right of the diagonal and Spearman correlations are below and to the left of the diagonal. Significant correlations ($P < 0.05$) are displayed in the table. Correlations between two traits with "-" are non-significant ($P > 0.05$).
[2] Estimated breeding values used to calculate correlation coefficients when the model included fixed effects of date of evaluation, sex, evaluator, random effect of calf, and permanent environmental effect to account for repeated records.

Table 21. Pearson and Spearman Rank correlation coefficients for fearful QBA attribute estimated breeding values across and within evaluator[1].

	Across[2]	2	4	5	6	8
Across[2]		0.284	0.801	0.735	0.701	0.686
2	0.208		0.119	0.189	0.188	-
4	0.792	0.090		0.575	0.377	0.408
5	0.708	0.175	0.560		0.405	0.269
6	0.692	0.168	0.427	0.335		0.416
8	0.601	-	0.291	0.192	0.428	

[1] QBA refers to qualitative behavior assessment, measured on a 136 mm visual analog scale. A low value (towards zero) indicates no or little expression, where a high value (towards 136) would indicate high or maximum expression. Pearson correlations are above and to the right of the diagonal and Spearman correlations are below and to the left of the diagonal. Significant correlations ($P < 0.05$) are displayed in the table. Correlations between two traits with "-" are non-significant ($P > 0.05$).

[2] Estimated breeding values used to calculate correlation coefficients when the model included fixed effects of date of evaluation, sex, evaluator, random effect of calf, and permanent environmental effect to account for repeated records.

Table 22. Pearson and Spearman Rank correlation coefficients for agitated QBA attribute estimated breeding values across and within evaluator[1].

	Across[2]	2	4	5	6	8
Across[2]		0.437	0.732	0.783	0.820	0.700
2	0.449		0.351	0.376	0.199	-
4	0.745	0.333		0.484	0.383	0.333
5	0.742	0.347	0.475		0.540	0.433
6	0.767	0.222	0.342	0.488		0.608
8	0.611	0.083	0.273	0.369	0.535	

[1] QBA refers to qualitative behavior assessment, measured on a 136 mm visual analog scale. A low value (towards zero) indicates no or little expression, where a high value (towards 136) would indicate high or maximum expression. Pearson correlations are above and to the right of the diagonal and Spearman correlations are below and to the left of the diagonal. Significant correlations ($P < 0.05$) are displayed in the table. Correlations between two traits with "-" are non-significant ($P > 0.05$).

[2] Estimated breeding values used to calculate correlation coefficients when the model included fixed effects of date of evaluation, sex, evaluator, random effect of calf, and permanent environmental effect to account for repeated records.

Table 23. Pearson and Spearman Rank correlation coefficients for calm QBA attribute estimated breeding values across and within evaluator[1].

	Across[2]	**2**	**4**	**5**	**6**	**8**
Across[2]		0.427	0.810	0.820	0.706	0.736
2	0.446		0.365	0.312	0.140	-
4	0.818	0.363		0.544	0.449	0.487
5	0.824	0.337	0.560		0.498	0.538
6	0.639	0.133	0.431	0.442		0.535
8	0.656	0.105	0.452	0.493	0.441	

[1] QBA refers to qualitative behavior assessment, measured on a 136 mm visual analog scale. A low value (towards zero) indicates no or little expression, where a high value (towards 136) would indicate high or maximum expression. Pearson correlations are above and to the right of the diagonal and Spearman correlations are below and to the left of the diagonal. Significant correlations ($P < 0.05$) are displayed in the table. Correlations between two traits with "-" are non-significant ($P > 0.05$).

[2] Estimated breeding values used to calculate correlation coefficients when the model included fixed effects of date of evaluation, sex, evaluator, random effect of calf, and permanent environmental effect to account for repeated records.

Table 24. Pearson and Spearman Rank correlation coefficients for attentive QBA attribute estimated breeding values across and within evaluator[1].

	Across[2]	**2**	**4**	**5**	**6**	**8**
Across[2]		0.232	0.580	0.347	0.358	0.659
2	0.244		0.140	0.131	-0.234	0.087
4	0.540	0.120		-	0.186	0.279
5	0.441	0.132	-		-0.147	-
6	0.342	-0.270	0.190	-0.131		0.218
8	0.521	0.116	0.232	-	0.171	

[1] QBA refers to qualitative behavior assessment, measured on a 136 mm visual analog scale. A low value (towards zero) indicates no or little expression, where a high value (towards 136) would indicate high or maximum expression. Pearson correlations are above and to the right of the diagonal and Spearman correlations are below and to the left of the diagonal. Significant correlations ($P < 0.05$) are displayed in the table. Correlations between two traits with "-" are non-significant ($P > 0.05$).

[2] Estimated breeding values used to calculate correlation coefficients when the model included fixed effects of date of evaluation, sex, evaluator, random effect of calf, and permanent environmental effect to account for repeated records.

Table 25. Pearson and Spearman Rank correlation coefficients for positively occupied QBA attribute estimated breeding values across and within evaluator[1].

	Across[2]	2	4	5	6	8
Across[2]		-0.110	0.495	0.734	0.275	0.369
2	-0.145		-0.248	-	-0.181	-0.120
4	0.550	-0.279		0.285	-	-
5	0.715	-	0.326		-	0.341
6	0.264	-0.140	-	-		-0.093
8	0.364	-0.196	-	0.388	0.070	

[1] QBA refers to qualitative behavior assessment, measured on a 136 mm visual analog scale. A low value (towards zero) indicates no or little expression, where a high value (towards 136) would indicate high or maximum expression. Pearson correlations are above and to the right of the diagonal and Spearman correlations are below and to the left of the diagonal. Significant correlations ($P < 0.05$) are displayed in the table. Correlations between two traits with "-" are non-significant ($P > 0.05$).

[2] Estimated breeding values used to calculate correlation coefficients when the model included fixed effects of date of evaluation, sex, evaluator, random effect of calf, and permanent environmental effect to account for repeated records.

Table 26. Pearson and Spearman Rank correlation coefficients for curious QBA attribute estimated breeding values across and within evaluator[1].

	Across[2]	2	4	5	6	8
Across[2]		0.398	0.753	0.390	0.365	0.525
2	0.397		0.286	-	-	-
4	0.761	0.289		-	0.314	0.270
5	0.333	-	-		-0.183	0.125
6	0.365	-	0.314	-0.183		0.215
8	0.525	-	0.270	0.125	0.215	

[1] QBA refers to qualitative behavior assessment, measured on a 136 mm visual analog scale. A low value (towards zero) indicates no or little expression, where a high value (towards 136) would indicate high or maximum expression. Pearson correlations are above and to the right of the diagonal and Spearman correlations are below and to the left of the diagonal. Significant correlations ($P < 0.05$) are displayed in the table. Correlations between two traits with "-" are non-significant ($P > 0.05$).

[2] Estimated breeding values used to calculate correlation coefficients when the model included fixed effects of date of evaluation, sex, evaluator, random effect of calf, and permanent environmental effect to account for repeated records.

Table 27. Pearson and Spearman Rank correlation coefficients for irritated QBA attribute estimated breeding values across and within evaluator[1].

	Across[2]	2	4	5	6	8
Across[2]		0.480	0.744	0.825	0.759	0.678
2	0.486		0.406	0.378	0.144	0.080
4	0.704	0.397		0.593	0.366	0.398
5	0.805	0.383	0.578		0.520	0.400
6	0.680	0.120	0.285	0.435		0.516
8	0.631	0.114	0.314	0.362	0.478	

[1] QBA refers to qualitative behavior assessment, measured on a 136 mm visual analog scale. A low value (towards zero) indicates no or little expression, where a high value (towards 136) would indicate high or maximum expression. Pearson correlations are above and to the right of the diagonal and Spearman correlations are below and to the left of the diagonal. Significant correlations ($P < 0.05$) are displayed in the table.
[2] Estimated breeding values used to calculate correlation coefficients when the model included fixed effects of date of evaluation, sex, evaluator, random effect of calf, and permanent environmental effect to account for repeated records.

Table 28. Pearson and Spearman Rank correlation coefficients for apathetic QBA attribute estimated breeding values across and within evaluator[1].

	Across[2]	2	4	5	6	8
Across[2]		0.111	0.616	0.634	0.490	0.381
2	-		-	-0.218	0.081	-0.306
4	0.625	-		0.213	0.185	0.288
5	0.644	-0.222	0.323		-	0.300
6	0.505	0.089	0.223	0.137		-
8	0.392	-0.348	0.330	0.324	-	

[1] QBA refers to qualitative behavior assessment, measured on a 136 mm visual analog scale. A low value (towards zero) indicates no or little expression, where a high value (towards 136) would indicate high or maximum expression. Pearson correlations are above and to the right of the diagonal and Spearman correlations are below and to the left of the diagonal. Significant correlations ($P < 0.05$) are displayed in the table. Correlations between two traits with "-" are non-significant ($P > 0.05$).
[2] Estimated breeding values used to calculate correlation coefficients when the model included fixed effects of date of evaluation, sex, evaluator, random effect of calf, and permanent environmental effect to account for repeated records.

Table 29. Pearson and Spearman Rank correlation coefficients for happy QBA attribute estimated breeding values across and within evaluator[1].

	Across[2]	**2**	**4**	**5**	**6**	**8**
Across[2]		0.544	0.650	0.716	0.300	-
2	0.531		0.170	0.364	0.101	-0.093
4	0.661	0.153		0.442	-	0.111
5	0.762	0.316	0.485		-	0.202
6	0.330	0.197	-	0.095		-0.091
8	-0.179	-0.219	-0.069	0.110	-0.146	

[1] QBA refers to qualitative behavior assessment, measured on a 136 mm visual analog scale. A low value (towards zero) indicates no or little expression, where a high value (towards 136) would indicate high or maximum expression. Pearson correlations are above and to the right of the diagonal and Spearman correlations are below and to the left of the diagonal. Significant correlations ($P < 0.05$) are displayed in the table. Correlations between two traits with "-" are non-significant ($P > 0.05$).
[2] Estimated breeding values used to calculate correlation coefficients when the model included fixed effects of date of evaluation, sex, evaluator, random effect of calf, and permanent environmental effect to account for repeated records.

Table 30. Pearson and Spearman Rank correlation coefficients for distressed QBA attribute estimated breeding values across and within evaluator[1].

	Across[2]	**2**	**4**	**5**	**6**	**8**
Across[2]		0.498	0.765	0.861	0.549	0.563
2	0.431		0.347	0.420	0.281	-
4	0.687	0.334		0.607	0.279	0.256
5	0.828	0.347	0.576		0.449	0.325
6	0.415	0.203	0.184	0.291		0.113
8	0.502	-	0.117	0.247	-	

[1] QBA refers to qualitative behavior assessment, measured on a 136 mm visual analog scale. A low value (towards zero) indicates no or little expression, where a high value (towards 136) would indicate high or maximum expression. Pearson correlations are above and to the right of the diagonal and Spearman correlations are below and to the left of the diagonal. Significant correlations ($P < 0.05$) are displayed in the table. Correlations between two traits with "-" are non-significant ($P > 0.05$).
[2] Estimated breeding values used to calculate correlation coefficients when the model included fixed effects of date of evaluation, sex, evaluator, random effect of calf, and permanent environmental effect to account for repeated records.

Table 31. Pearson and Spearman Rank correlation coefficients for temperament index estimated breeding values across and within evaluator.

	Across[2]	2	4	5	6	8
Across[2]		0.510	0.613	-0.254	-	0.202
2	0.364		0.468	-0.457	-0.187	0.057
4	0.602	0.433		-0.650	-0.475	0.524
5	-0.184	-0.444	-0.638		0.518	-0.527
6	0.073	-0.169	-0.444	0.446		-0.530
8	0.201	0.088	0.473	-0.423	-0.427	

[1] Temperament index: the first principal component score generated from QBA scores using a Principal Component Analysis in SAS (SAS Institute, Inc., Cary, NC.). Pearson correlations are above and to the right of the diagonal and Spearman correlations are below and to the left of the diagonal. Significant correlations ($P < 0.05$) are displayed in the table.
[2] Estimated breeding values used to calculate correlation coefficients when the model included fixed effects of date of evaluation, sex, evaluator, random effect of calf, and permanent environmental effect to account for repeated records.

Docility score had a really high correlation coefficient, where there was correlation coefficient of 1.000 between evaluators 4 and 8, indicating that their predictions of genetic merit were identical. Active had a low correlation coefficient between evaluators 2 and 8, indicating a large divergence between their respective observations of cattle scored and how that translated to EBV. Based on the loading plot results generated during PCA, it is known that evaluators 5 and 6 scored differently than the other 3 evaluators. Looking at their correlation coefficients, the majority of their QBA attributes were moderately low. When comparing the correlation coefficients to the results generated for across evaluator, temperament score had very high correlation coefficients for across evaluator vs. evaluators 2 and 6. When looking at the individual evaluator correlation coefficients and comparing them to across evaluator, there is no distinct pattern as they are both higher and lower than the across evaluator model (Tables 17 to 31).

Similar to what was done for Pearson and Spearman Rank correlation coefficients, across evaluator EBV was included in comparing quartiles to serve as a standard to help quantify evaluator bias when considering how animals would be selected (i.e., animals closer to 1 would be more desirable, and, therefore, selected over animals with larger rank values). When assigned to quartiles for a given trait and evaluator and compared to another evaluator for the same trait, it was considered that a change of 3 quartiles was significant re-ranking. Although 2 quartiles may also indicate significant re-ranking, the lower accuracies with the current dataset led us to use a more conservative definition of 3 quartile changes for re-ranking. Less than 3% of individuals had EBVs in different analyses that were 3 quartiles apart when comparing evaluators to the across evaluator model and less than 11% of individuals had EBVs in different analyses that were 3 quartiles apart for docility score (Table 32). Less than 4% of individuals had EBVs in different analyses that were 3 quartiles apart when comparing evaluators to the across evaluator model and less than 9% of individuals had EBVs in different analyses that were 3 quartiles apart for temperament score (Table 32). As for QBA attributes (Appendix Table A6) and temperament index (Appendix Table A7), both had less than 19% of individuals had EBVs in different analyses that were 3 quartiles apart when comparing evaluators to the across evaluator model, while less than 24% of individuals had EBVs in different analyses that were 3 quartiles apart for QBA attributes and less than 30% of individuals had EBVs in different analyses that were 3 quartiles apart for temperament index. Overall, the majority of any given analyses consistently changed 1 quartile and while understanding that there are low accuracies with this experiment due to low numbers, we are more interested in different analyses that changed more than 3 quartiles for the 15 traits being evaluated. Some analyses could naturally be changing more than 3 quartiles for those evaluators that were only available for one year of evaluations versus those evaluators that were available for both years, which is seen in Table 32 for docility score in evaluator 3's comparisons.

Table 32. Comparison of the number and percentage of individuals with estimated breeding values for docility and temperament score that changes *n* quartiles between any two analyses.

	Number of individuals that changed *n* quartiles[1]			
Evaluation	**0**	**1**	**2**	**3**
Docility score[2] for Evaluator				
Across vs. 1	331 (41.272%)	336 (41.895%)	119 (14.838%)	16 (1.995%)
vs. 2	465 (57.980%)	286 (35.661%)	49 (6.110%)	2 (0.249%)
vs. 3	326 (40.648%)	317 (39.526%)	135 (16.833%)	24 (2.993%)
vs. 4	401 (50.000%)	303 (37.781%)	91 (11.347%)	7 (0.873%)
vs. 7	285 (35.536%)	350 (43.641%)	144 (17.955%)	23 (2.868%)
vs. 8	401 (50.000%)	303 (37.781%)	91 (11.347%)	7 (0.873%)
1 *vs.* 2	246 (30.673%)	344 (42.893%)	168 (20.948%)	44 (5.486%)
vs. 3	349 (43.516%)	316 (39.401%)	102 (12.718%)	35 (4.364%)
vs. 4	272 (33.915%)	321 (40.025%)	158 (19.701%)	51 (6.359%)
vs. 7	143 (17.830%)	337 (42.020%)	262 (32.668%)	60 (7.481%)
vs. 8	272 (33.915%)	321 (40.025%)	158 (19.701%)	51 (6.359%)
2 *vs.* 3	266 (33.167%)	320 (39.900%)	179 (22.319%)	37 (4.613%)
vs. 4	296 (36.908%)	337 (42.020%)	140 (17.456%)	29 (3.616%)
vs. 7	284 (35.411%)	298 (37.157%)	161 (20.075%)	59 (7.357%)
vs. 8	296 (36.908%)	337 (42.020%)	140 (17.456%)	29 (3.616%)
3 *vs.* 4	244 (30.424%)	300 (37.406%)	171 (21.322%)	87 (10.848%)
vs. 7	162 (20.200%)	326 (40.648%)	232 (28.928%)	82 (10.224%)
vs. 8	244 (30.424%)	300 (37.406%)	171 (21.322%)	87 (10.848%)
4 *vs.* 7	222 (27.681%)	335 (41.771%)	199 (24.813%)	46 (5.736%)
vs. 8	802 (100.00%)	0 (0.000%)	0 (0.000%)	0 (0.000%)
7 *vs.* 8	222 (27.681%)	335 (41.771%)	199 (24.813%)	46 (5.736%)
Temperament score[3] for Evaluator				
Across vs. 2	501 (62.469%)	252 (31.421%)	47 (5.860%)	2 (0.249%)
vs. 3	368 (45.885%)	321 (40.025%)	100 (12.469%)	13 (1.621%)
vs. 5	410 (51.122%)	344 (42.893%)	44 (5.486%)	4 (0.499%)
vs. 6	500 (62.344%)	254 (31.671%)	46 (5.736%)	2 (0.249%)
vs. 7	317 (39.526%)	291 (36.284%)	168 (20.948%)	26 (3.242%)
2 *vs.* 3	310 (38.653%)	330 (41.147%)	140 (17.456%)	22 (2.743%)
vs. 5	335 (41.771%)	350 (43.641%)	107 (13.342%)	10 (1.247%)
vs. 6	424 (52.868%)	253 (31.546%)	114 (14.214%)	11 (1.372%)
vs. 7	312 (38.903%)	295 (36.783%)	149 (18.579%)	46 (5.736%)
3 *vs.* 5	262 (32.668%)	378 (47.132%)	138 (17.207%)	24 (2.993%)
vs. 6	364 (45.387%)	308 (38.404%)	108 (13.466%)	22 (2.743%)
vs. 7	195 (24.314%)	295 (36.783%)	242 (30.175%)	70 (8.728%)
5 *vs.* 6	324 (40.399%)	366 (45.636%)	100 (12.469%)	12 (1.496%)
vs. 7	236 (29.426%)	321 (40.025%)	193 (24.065%)	52 (6.484%)
6 *vs.* 7	289 (36.035%)	307 (38.279%)	156 (19.451%)	50 (6.234%)

[1] The number of quartile changes was calculated by first assigning an animal's quartile for any given analysis, then finding the difference of each animal's quartile between the two analyses (evaluators in this case) compared. Percentage was calculated by dividing the number of individuals within that category by the total number of animals.

[2] Docility score: scale of 1 to 6 with 1 = calm and 6 = excited.

[3] Temperament score: scale of 1 to 5, with no intermediate score; 1 = animal walks slowly while allowing close approximation with the observer and 5 = runs the entire time of the observation, jumps against the fence, and tries to attack the observer.

In all cases, when comparing quartile changes when evaluator was included in model to each evaluator's score, there are relatively lower percentages of individuals that changed 3 quartiles for both docility and temperament score (Table 32). Furthermore, both attentive and irritated were the most consistent and had the least amount of variation, which is consistent when looking at the evaluator comparisons (Appendix Table A6). Agitated was the most consistent in its percentages of individuals that changed 3 quartiles between analyses with a difference of approximately 5%. This was followed by irritated with an approximated difference of 7%, followed by temperament score, active, and distressed with differences of approximately 8% each, and then relaxed with approximately a 9% difference. The across comparison that was the least consistent and had the largest range of percentages matches what was found with the least consistent evaluator comparisons. The largest difference in percentages to change more than 2 quartiles was for temperament index with approximately 27% and then apathetic with approximately 19%. With there being less variation for agitated, it shows that although there were differences in how evaluators scored calves, how those scores translated to genetic predictions were more similar for agitated than apathetic. The large changes found for TI are expected as it is representing all of the QBA attributes as a whole. We also know there is variation for QBA based the loading plot results as well as the range in correlations amongst QBA attributes.

When looking at docility score, temperament score, and temperament index independently, temperament score had the least amount of variation and, therefore, the least amount of bias across evaluators followed closely by temperament. These results tell us that these two evaluation methods are the ones that are going to reduce evaluator bias and capture the true behavior of the animals that are being evaluated. Although there were 2 QBA attributes that had even less variation than temperament score (agitated and irritated), it has to be taken into account that the QBA attributes are combined to generate an overall temperament index and if they were to be used independently, they would not be capturing all aspects of an animal's behavior. In general, when comparing quartile changes with the across evaluator model, the evaluators that were present in only one year of evaluations had a larger percentage of individuals that changed 3 quartiles. This could be due to these evaluators having evaluated only about half of the number of calves that those evaluators who evaluated in both years did and, therefore, having lower accuracies in their EBVs.

It must be noted that evaluator 4 vs. 8 for docility score was the only comparison in which 100% of individuals changed 0 quartiles indicating that these 2 evaluators scores translated to genetic predictions that were identical, which matches the results from the Pearson and Spearman Rank correlation coefficients. Furthermore, it is interesting to consider the analyses for those evaluators of the second year that replaced those evaluators in the first year that could not return. For docility score, evaluator 3 vs. 7 was the second highest for the percentage of individuals having changed 3 quartiles yet for temperament score, they were the lowest percentage. As for evaluator 1 vs. 8, this analysis fell in the middle of the percentages for docility. When looking at QBA attributes, they were low in the range of percentages for calm and attentive, in the middle for curious and distressed, higher for the remaining attributes, and they were also low for TI. These results are similar to what was found with the statistical analysis, and support that although these evaluators were similar in their cattle experience, they were not similar in these evaluation methods.

Based on these results, we know that evaluator does in fact have an impact on breeding value estimations for docility score, temperament score, and temperament index. Due to evaluators interpreting behavior differently, those differences carry through and are shown in the EBVs. The degree to which evaluators have their own interpretation of temperament has been captured within this study and gives us insight as to which scoring systems could have more evaluation bias compared to others. When looking at temperament studies that are interested in repeatability, such as those that are utilizing the same animal over several evaluations and looking at day of evaluation effect, it is common to use correlations to see how days are correlated. Café et al. (2011) used correlations to look at the different occasions that they evaluated flight speed on the same cattle and evaluated correlation. While comparing only 8 of the 14 occasions, they found that all of the moderate to high correlations were significant ($P < 0.001$) for Brahmans. Considering there has not been research done to estimate the evaluator bias effect in temperament evaluations, correlations have not been used to compare how evaluators score animals. While this is not the first research experiment to utilize quartile comparisons, the current experiment is the first known experiment to utilize them for EBV comparisons to look at evaluator bias when evaluating cattle temperament.

Implications

Although sample size in this experiment was a concerning fact, the general result was that there is evaluation bias across evaluators, which proves that evaluator does impact genetic parameter estimates for temperament evaluations. Due to this, it would naturally impact breeding value estimates if not modeled appropriately. While genetic evaluations include contemporary groups in their models, it may not necessarily account for evaluator differences. Based on our results that evaluators do in fact differ in how they score cattle, it supports that everyone has their own interpretation of an animal’s behavior. Our research does not fully prove that, however, leaving options for future research opportunities. To get a better grasp on this, those future opportunities would need to be sure to use evaluators that could come back for consecutive years of evaluations to ensure that they could capture those evaluator differences.

This research is consistent with other beef cattle temperament in terms of sex comparisons as we found that heifers had significantly more temperamental attributes compared to those of steers. While there is no literature available looking at day of evaluation effect without repeating animals, our research is consistent with previous research looking at season (spring vs. fall) within a birth year, where calves were evaluated at weaning depending on the season in which they were born. Season within a birth year was found to be significant which aligned with our findings of significant date of evaluation effect.

While contemporary group includes ranch location, it does not always factor in if different people were evaluating temperament when working cattle at each location. Based on the results of this study, it could lead to the currently reported EBVs to be skewed due to there being those evaluator differences. With further continuation of this project, the findings could result in breed associations reporting more accurate EBVs with an evaluator effect being factored into temperament evaluations in the future.

CHAPTER 4. GENERAL CONCLUSIONS

Using 3 subjective scoring methods, cattle temperament was evaluated on newly weaned Angus-based calves over the duration of 2 years with the main focus of this research being on the impact of evaluator bias on genetic selection criterion. This is the first report to document the impact of evaluator bias in subjective measures of cattle temperament on genetic selection criteria, therefore, leaving options for future research opportunities.

With sequence nested within date of evaluation always a significant effect and there being little reported in literature about it, it in itself would be an area of interest for further research opportunities to look at the long term temperament effects. With just a few more years of repeating the data collection process that was done in this experiment, sample size would increase greatly as well as accuracy. Ideally it would be nice to have evaluators that could be repeated year after year to be able to get the full impact of their interpretation over a multi-year period. Although this may not always be ideal, a better idea could be made if evaluators could evaluate for at least two consecutive years.

By having the same evaluators for consecutive years, it would also serve as a better model of how evaluations are done in the beef industry. When calves are worked based on management needs, the producer is typically the one that is there for all working events and, therefore, the same person is typically evaluating the animals. For purebred producers, the results from those evaluations are what is being reported to their breed association, which then become EPDs. This experiment has proved that people evaluate cattle differently and if different people are doing evaluations, EPDs could be skewed.

REFERENCES CITED

Apple, J. K., E. B. Kegley, D. L. Galloway, T. J. Wistuba, and L. K. Rakes. 2005. Duration of restraint and isolation stress as a model to study the dark-cutting condition in cattle. J. Anim. Sci. 83:1202-1214.

Barrozo. D., M. E. Buzanskas, J. A. Oliveira. D. P. Munari, H. H. R. Neves, and S. A. Queiroz. 2011. Genetic parameters and environmental effects on temperament score and reproductive traits of Nellore cattle. Animal 6:36-40.

Beef Improvement Federation. 2010. Guidelines for uniform beef improvement programs. 9th ed. http://beefimprovement.org/content/uploads/2015/08/REVISED-MasterEd-BIF-GuidelinesFinal-08-2015.pdf. (Accessed 16 December 2015.)

Boissy, A., A. D. Fisher, J. Bouix, G. N. Hinch, and P. Le Neindre. 2005. Genetics of fear in ruminant livestock. Livest. Prod. Sci. 93:23-32.

Burdick, N. C., R. D. Randel, J. A. Carroll, and T. H. Welsh Jr. 2011. Interactions between temperament, stress, and immune function in cattle. Int. J. Zool. doi:10.1155/2011/373197.

Burrow, H. M. 2001. Variances and covariances between productive and adaptive traits and temperament in a composite breed of tropical beef cattle. Livest. Prod. Sci. 70:213-233.

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

Burrow, H. M., and R. D. Dillon. 1997. Relationships between temperament and growth in a feedlot and commercial carcass traits of *Bos indicus* crossbreds. Aust. J. Exp. Agric. 37:407-411.

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

Cooke, R. F., J. D. Arthington, D. B. Araujo, and G. C. Lamb. 2009a. 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., J. D. Arthington, B. R. Austin, and J. V. Yelich. 2009b. Effects of acclimation to handling on performance, reproductive, and physiological responses of Brahman-crossbred heifers. J. Anim. Sci. 87:3403-3412.

Cooke, R. F. and D. W. Bohnert. 2011. Technical note: Bovine acute-phase response after corticotrophin-release hormone challenge. J. Anim. Sci. 89:252-257.

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

Dobson, H., J. E. Tebble, R. F. Smith, and W. R. Ward. 2001. Is stress really all that important? Theriogenology 55:65-73.

FASS. 2010. Guide for the Care and Use of Agriculture Animals in Research and Teaching. Fed. Anim. Sci. Soc., Savoy, IL.

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

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

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

Gilmour, A. R., B. J. Gogel, B. R. Cullis, S. J. Welham, and R. Thompson. 2015. ASReml User Guide Release 4.1 Functional Specification. VSN International Ltd., Hemel Hempstead, UK.

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

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

Hearnshaw, H., R. Barlow, and G. Want. 1979. Development of a “temperament” or “handling difficulty” score for cattle. Proc. Aust. Assoc. Anim. Breed. Genet. 1:164-166.

Hearnshaw, H. and C. A. Morris. 1984. Genetic and environmental effects on a temperament score in beef cattle. Aust. J. Agric. Res. 35:723-733.

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

Hulsman Hanna, L. L., D. J. Garrick, C. A. Gill, A. D. Herring, J. O. Sanders, and D. G. Riley. 2014. Comparison of breeding value prediction for two traits in a Nellore-Angus crossbred population using different Bayesian modeling methodologies. Genet. Mol. Biol. 37:631-637.

Hulsman Hanna, L. L., D. J. Garrick, C. A. Gill, J. O. Sanders, and D. G. Riley. 2013. Investigation of genomic estimated breeding values and association of methodologies using Bayesian inference in a Nellore-Angus crossbred population for two traits. Ph.D. Dissertation. Texas A&M University, College Station.

Jolliffe, I. 2005. Principal component analysis. Encyclopedia of Statistics in Behavioral Science. University of Aberdeen, Aberdeen, UK.

Kaiser, H. F. 1960. The application of electronic computers to factor analysis. Educ. Psychol. Meas. 20:1.

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

Krohn, C. C., J. G. Jago, and X. Boivin. 2001. The effect of early handling on the socialisation of young calves to humans. Appl. Anim. Behav. Sci. 74:121-133.

Lush, J. L. 1994. Heritability of quantitative characters in farm animals. Heriditas. 35:356-387.

Nkrumah, J. D., D. H. Crews Jr., J. A. Basarab, M. A. Price, E. K. Okine, Z. Wang, C. Li, and S. S. Moore. (2007). Genetic and phenotypic relationships of feeding behavior and temperament with performance, feed efficiency, ultrasound and carcass merit of beef cattle. J. Anim. Sci. 85:2382-2390.

Northcutt, S. and B. Bowman. 2010. Docility genetic evaluation research. ANGUS*Journal* http://www.angus.org/nce/documents/bythenumbersdocility.pdf. (Accessed 16 December 2015).

Oliphint, R. A. 2006. Evaluation of the inter-relationships of temperament, stress responsiveness and immune function in beef calves. M.S. Thesis. Texas A&M University, College Station.

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

Petherick, J. C., V. J. Doogan, B. K. Venus, R. G. Holroyd, and P. Olsson. 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.

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

Ledesma, R. D. and P. Valero-Mora. 2007. Determining the number of factors to retain in EFA: an easy-to-use computer program for carrying out parallel analysis. Pract. Assess. Res. Eval. 12:1-11.

Richardson, M. 2009. Principal component analysis. http://www.sdss.jhu.edu/~szalay/class/2015/SignalProcPCA.pdf. (Accessed 17 March 2016.)

Riley, D. G., C. A. Gill, A. D. Herring, P. K. Riggs, J. E. Sawyer, D. K. Lunt, and J. O. Sanders. 2014. Genetic evaluation of aspects of temperament in Nellore-Angus calves. J. Anim. Sci. 92:3223-3230.

Rolfe, K. M., W. M. Snelling, M. K. Nielsen, H. C. Freetly, C. L. Ferrell, and T. G. Jenkins. 2011. Genetic and phenotypic parameter estimates for feed intake and other traits in growing beef cattle, and opportunities for selection. J. Anim. Sci. 89:3452-3459.

Sant'Anna, A. C., and M. J. R. Paranhos da Costa. 2013. Validity and feasibility of qualitative behavior assessment for the evaluation of Nellore cattle temperament. Livest. Sci. 157:254-262.

Shalizi, C. R. 2016. Advanced Data Analysis from an Elementary Point of View. http://www.stat.cmu.edu/~cshalizi/ADAfaEPoV/ADAfaEPoV.pdf. (Accessed 21 April 2016.)

Stookey, J. M., T. Nickel, J. Hanson, and S. Vandenbosch. 1994. A movement-measuring-device for objectivity measuring temperament in beef cattle and for use in determining factors that influence handling. J. Anim. Sci. 72 (Suppl. 1):207.

Sutherland, M. A., A. R. Rogers, and G. A. Verkerk. 2012. The effect of temperament and responsiveness toward humans on the behavior, physiology and milk production of multi-parous dairy cows in a familiar and novel milking environment. Phy. Behav. 107:329-337.

Waynert, D. F., J. M. Stookey, K. S. Schwartzkoph-Fenswein, J. M. Watts, and C. S. Waltz. 1999. The response of beef cattle to noise during handling. Appl. Anim. Behav. Sci. 62:27-42.

Welfare Quality® Consortium. 2009. Welfare Quality® Assessment Protocol for Cattle. Lelystad, Netherlands.

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

APPENDIX A. SUPPLEMENTARY TABLES

Table A1. Pearson and Spearman rank correlation coefficients for QBA attributes measured within evaluator for evaluator 1[1].

	Active	Relaxed	Fearful	Agitated	Calm	Attentive	Positively occupied	Curious	Irritated	Apathetic	Happy	Distressed
Active		-0.696	0.479	0.594	-0.711	0.297	-0.394	0.122	0.616	-0.253	-0.312	0.512
Relaxed	-0.704		-0.313	-0.630	0.896	-0.299	0.467	-0.215	-0.719	0.219	0.393	-0.614
Fearful	0.353	-0.262		0.509	-0.315	0.253	-0.257	-0.024	0.393	-0.151	-0.200	0.401
Agitated	0.598	-0.638	0.528		-0.651	0.261	-0.392	-	0.630	-0.282	-0.331	0.504
Calm	-0.686	0.875	-0.254	-0.676		-0.289	0.508	-0.196	-0.713	0.295	0.431	-0.579
Attentive	0.278	-0.310	0.173	0.263	-0.286		-0.180	0.278	0.249	-0.123	-0.196	0.282
Positively occupied	-0.422	0.481	-0.216	-0.415	0.525	-0.116		0.037	-0.456	0.404	0.555	-0.329
Curious	-	-0.149	-	-	-0.134	0.292	-		0.166	-0.137	-	-
Irritated	0.553	-0.667	0.340	0.662	-0.648	0.229	-0.442	0.070		-0.405	-0.425	0.676
Apathetic	-0.170	-	-0.137	-0.186	0.149	-	0.372	-0.108	-0.250		0.425	-0.296
Happy	-0.319	0.361	-0.189	-0.361	0.405	-0.155	0.553	-0.060	-0.396	0.457		-0.346
Distressed	0.458	-0.617	0.377	0.525	-0.556	0.217	-0.285	-	-0.396	-0.152	-0.251	

[1] QBA refers to qualitative behavior assessment, measured on a 136 mm visual analog scale. A low value (towards zero) indicates no or little expression, where a high value (towards 136) would indicate high or maximum expression. Pearson correlations are above and to the right of the diagonal and Spearman correlations are below and to the left of the diagonal. Significant correlations ($P < 0.05$) are displayed in the table. Correlations between two traits with "-" are non-significant ($P > 0.05$).

Table A2. Pearson and Spearman rank correlation coefficients for QBA attributes measured within evaluator for evaluator 4[1].

	Active	Relaxed	Fearful	Agitated	Calm	Attentive	Positively occupied	Curious	Irritated	Apathetic	Happy	Distressed
Active		-0.567	0.626	0.562	-0.605	0.365	-0.260	-0.100	0.603	-0.381	-0.072	0.491
Relaxed	-0.585		-0.606	-0.626	0.865	-0.198	-0.167	0.251	-0.565	0.470	0.400	-0.467
Fearful	0.610	-0.675		0.686	-0.641	0.256	0.308	-0.145	0.756	0.306	-0.079	0.695
Agitated	0.544	-0.688	0.683		-0.637	0.268	0.302	-0.064	0.732	-0.309	-0.153	-0.607
Calm	-0.615	0.866	-0.708	-0.696		-0.183	-0.161	0.253	-0.612	0.502	0.389	-0.488
Attentive	0.373	-0.201	0.267	0.270	-0.186		0.386	0.187	0.272	-0.343	0.175	0.190
Positively occupied	0.204	-0.200	0.314	0.279	-0.194	0.307		0.120	0.342	-0.133	0.213	0.348
Curious	-0.098	0.207	-0.196	-0.091	0.212	0.232	0.124		-0.160	0.096	0.415	-0.089
Irritated	0.558	-0.613	0.696	0.684	-0.651	0.270	0.289	-0.232		-0.296	-0.077	0.796
Apathetic	-0.366	0.448	-0.383	-0.317	0.491	-0.236	-0.163	0.078	-0.279		0.396	-0.205
Happy	-	0.344	-0.169	-0.191	0.357	0.179	-	0.330	-0.128	0.425		-
Distressed	0.454	-0.510	0.602	0.514	-0.516	0.208	0.250	-0.185	0.656	-0.225	-	

[1] QBA refers to qualitative behavior assessment, measured on a 136 mm visual analog scale. A low value (towards zero) indicates no or little expression, where a high value (towards 136) would indicate high or maximum expression. Pearson correlations are above and to the right of the diagonal and Spearman correlations are below and to the left of the diagonal. Significant correlations ($P < 0.05$) are displayed in the table. Correlations between two traits with "-" are non-significant ($P > 0.05$).

Table A3. Pearson and Spearman rank correlations for QBA attributes measured within evaluator for evaluator 5[1].

	Active	**Relaxed**	**Fearful**	**Agitated**	**Calm**	**Attentive**	**Positively occupied**	**Curious**	**Irritated**	**Apathetic**	**Happy**	**Distressed**
Active		-0.725	0.449	0.542	-0.670	-0.234	-0.472	-0.328	0.517	-0.264	-0.417	0.361
Relaxed	-0.726		-0.405	-0.447	0.895	0.364	0.614	0.425	-0.419	0.345	0.541	-0.295
Fearful	0.428	-0.411		0.706	-0.425	0.120	-	-	0.699	-	-	0.644
Agitated	0.484	-0.419	0.728		-0.469	0.105	-	-	0.814	-	-	0.734
Calm	-0.659	0.871	-0.435	-0.472		0.347	0.627	0.431	-0.427	0.391	0.539	-0.317
Attentive	0.225	0.349	0.186	0.149	0.343		0.668	0.545	0.178	0.348	0.614	0.178
Positively occupied	-0.485	0.630	-	-	0.646	0.656		0.693	-	0.493	0.796	-
Curious	-0.388	0.436	-	-	0.440	0.545	0.704		-	0.324	0.664	0.070
Irritated	0.437	-0.381	0.702	0.779	-0.406	0.247	-	0.089		-	-	0.764
Apathetic	-0.288	0.348	0.106	0.074	0.383	0.384	0.546	0.401	0.118		0.372	0.114
Happy	-0.405	0.545	0.072	-	0.534	0.609	0.793	0.673	0.099	0.451		-
Distressed	0.318	-0.277	0.669	0.707	-0.332	0.233	0.091	0.118	0.756	0.176	0.139	

[1] QBA refers to qualitative behavior assessment, measured on a 136 mm visual analog scale. A low value (towards zero) indicates no or little expression, where a high value (towards 136) would indicate high or maximum expression. Pearson correlations are above and to the right of the diagonal and Spearman correlations are below and to the left of the diagonal. Significant correlations ($P < 0.05$) are displayed in the table. Correlations between two traits with "-" are non-significant ($P > 0.05$).

Table A4. Pearson and Spearman rank correlations for QBA attributes measured within evaluator for evaluator 6[1].

	Active	**Relaxed**	**Fearful**	**Agitated**	**Calm**	**Attentive**	**Positively occupied**	**Curious**	**Irritated**	**Apathetic**	**Happy**	**Distressed**
Active		-0.606	0.423	0.543	-0.454	0.195	-	-	0.542	-0.263	-	0.074
Relaxed	-0.626		-0.506	-0.742	0.862	0.194	0.393	0.425	-0.477	0.420	0.474	0.121
Fearful	0.528	-0.586		0.663	-0.472	0.100	-	-0.087	0.532	-0.451	-0.140	0.211
Agitated	0.602	-0.730	0.741		-0.729	-	-0.261	-0.289	0.615	-0.447	-0.340	-
Calm	-0.455	0.826	-0.540	-0.686		0.387	0.609	0.605	-0.407	0.391	0.683	0.218
Attentive	0.195	0.109	0.142	0.081	0.327		0.826	0.742	0.236	-0.095	0.776	0.418
Positively occupied	0.101	0.250	-	-	0.483	0.772		0.834	0.080	-	0.929	0.472
Curious	-	0.322	-	-0.143	0.528	0.709	0.755		-	0.073	0.846	0.408
Irritated	0.619	-0.482	0.520	0.613	-0.333	0.401	0.276	0.199		-0.533	-	0.324
Apathetic	-0.381	0.367	-0.373	-0.392	0.278	-0.326	-0.225	-0.200	-0.619		0.117	-0.307
Happy	-	0.353	-0.095	-0.196	0.577	0.688	0.808	0.768	0.154	-0.075		0.420
Distressed	0.100	0.194	0.090	-	0.387	0.615	0.684	0.670	0.375	-0.352	0.661	

[1] QBA refers to qualitative behavior assessment, measured on a 136 mm visual analog scale. A low value (towards zero) indicates no or little expression, where a high value (towards 136) would indicate high or maximum expression. Pearson correlations are above and to the right of the diagonal and Spearman correlations are below and to the left of the diagonal. Significant correlations ($P < 0.05$) are displayed in the table. Correlations between two traits with "-" are non-significant ($P > 0.05$).

Table A5. Pearson and Spearman rank correlations for QBA attributes measured within evaluator for evaluator 8[1].

	Active	Relaxed	Fearful	Agitated	Calm	Attentive	Positively occupied	Curious	Irritated	Apathetic	Happy	Distressed
Active		-0.514	0.562	0.542	-0.529	0.488	-0.373	-0.306	0.494	-0.212	-0.134	0.423
Relaxed	-0.489		-0.507	-0.560	0.778	-0.394	0.397	0.313	-0.473	0.229	0.239	-0.398
Fearful	0.557	-0.549		0.764	-0.522	0.588	-0.212	-0.161	0.561	-0.561	-0.144	0.443
Agitated	0.540	-0.615	0.783		-0.590	0.639	-0.196	-0.137	0.673	-0.149	-0.127	0.497
Calm	-0.520	0.795	-0.539	-0.618		-0.331	0.432	0.328	-0.567	0.189	0.128	-0.486
Attentive	0.470	-0.369	0.563	0.587	-0.319		-	-	0.526	-	-	0.409
Positively occupied	-0.361	0.185	-0.145	-	0.291	0.047		0.743	-0.194	0.602	0.537	-0.227
Curious	-0.348	0.146	-0.148	-	0.217	0.021	0.683		-0.136	0.598	0.507	-0.193
Irritated	0.497	-0.546	0.495	0.586	-0.622	0.427	-	-		-	-	0.800
Apathetic	-0.160	-	-	-	-	-	0.437	0.406	0.148		0.550	-
Happy	-	-	-	-	-0.115	-	0.336	0.398	0.262	0.582		-
Distressed	0.492	-0.510	0.358	0.382	-0.603	0.249	-0.297	-0.242	0.662	-	0.201	

[1] QBA refers to qualitative behavior assessment, measured on a 136 mm visual analog scale. A low value (towards zero) indicates no or little expression, where a high value (towards 136) would indicate high or maximum expression. Pearson correlations are above and to the right of the diagonal and Spearman correlations are below and to the left of the diagonal. Significant correlations ($P < 0.05$) are displayed in the table. Correlations between two traits with "-" are non-significant ($P > 0.05$).

Table A6. Comparison of the number and percentage of individuals with estimated breeding values for QBA[1] attributes and temperament index that changes *n* quartiles between any two analyses.

QBA attributes for Evaluator	Number of individuals that changed *n* quartiles[2]			
	0	**1**	**2**	**3**
Active				
Across vs. 1	296 (36.908%)	354 (44.140%)	120 (14.963%)	32 (3.990%)
vs. 4	442 (55.112%)	306 (38.155%)	50 (6.234%)	4 (0.499%)
vs. 5	373 (46.509%)	322 (40.150%)	97 (12.095%)	10 (1.247%)
vs. 6	450 (56.110%)	291 (36.284%)	58 (7.232%)	3 (0.374%)
vs. 8	375 (46.758%)	301 (37.531%)	113 (14.090%)	13 (1.621%)
1 *vs.* 4	291 (36.284%)	329 (41.022%)	135 (16.833%)	47 (5.860%)
vs. 5	277 (34.539%)	314 (39.152%)	173 (21.571%)	38 (4.738%)
vs. 6	229 (28.554%)	333 (41.521%)	185 (23.067%)	55 (6.858%)
vs. 8	153 (19.077%)	327 (40.773%)	247 (30.798%)	75 (9.352%)
4 *vs.* 5	335 (41.771%)	312 (38.903%)	133 (16.584%)	22 (2.743%)
vs. 6	314 (39.152%)	325 (40.524%)	134 (16.708%)	29 (3.616%)
vs. 8	272 (33.915%)	333 (41.521%)	150 (18.703%)	47 (5.860%)
5 *vs.* 6	262 (32.668%)	309 (38.529%)	184 (22.943%)	47 (5.860%)
vs. 8	247 (30.798%)	307 (38.279%)	193 (24.065%)	55 (6.858%)
6 *vs.* 8	331 (41.272%)	324 (40.399%)	121 (15.087%)	26 (3.242%)
Relaxed				
Across vs. 1	291 (36.284%)	332 (41.397%)	147 (18.329%)	32 (3.990%)
vs. 4	304 (37.905%)	358 (44.638%)	140 (17.456%)	0 (0.000%)
vs. 5	465 (57.980%)	300 (37.406%)	37 (4.613%)	0 (0.000%)
vs. 6	413 (51.496%)	307 (38.279%)	80 (9.975%)	2 (0.249%)
vs. 8	347 (43.267%)	317 (39.525%)	117 (14.589%)	21 (2.618%)
1 *vs.* 4	315 (39.277%)	335 (41.771%)	135 (16.833%)	17 (2.120%)
vs. 5	263 (32.793%)	337 (42.020%)	167 (20.823%)	35 (4.364%)
vs. 6	216 (26.933%)	317 (39.526%)	209 (26.060%)	60 (7.481%)
vs. 8	184 (22.943%)	302 (37.656%)	236 (29.426%)	80 (9.975%)
4 *vs.* 5	241 (30.050%)	386 (48.130%)	161 (20.075%)	14 (1.746%)
vs. 6	208 (25.935%)	377 (47.007%)	207 (25.810%)	10 (1.247%)
vs. 8	161 (20.075%)	382 (47.631%)	239 (29.800%)	20 (2.494%)
5 *vs.* 6	304 (37.905%)	316 (39.401%)	150 (18.703%)	32 (3.990%)
vs. 8	304 (37.905%)	316 (39.401%)	150 (18.703%)	32 (3.990%)
6 *vs.* 8	339 (42.269%)	296 (36.908%)	122 (15.212%)	45 (5.611%)
Fearful				
Across vs. 1	241 (30.050%)	317 (39.526%)	169 (21.072%)	75 (9.352%)
vs. 4	449 (55.985%)	294 (36.658%)	55 (6.858%)	4 (0.499%)
vs. 5	379 (47.257%)	334 (41.646%)	83 (10.349%)	6 (0.748%)
vs. 6	407 (50.748%)	300 (37.406%)	81 (10.100%)	14 (1.746%)
vs. 8	370 (46.135%)	290 (36.160%)	126 (15.711%)	6 (1.995%)
1 *vs.* 4	244 (30.424%)	293 (36.534%)	168 (20.948%)	97 (12.095%)
vs. 5	257 (32.045%)	300 (37.406%)	155 (19.327%)	90 (11.222%)
vs. 6	202 (25.187%)	338 (42.145%)	190 (23.691%)	72 (8.978%)
vs. 8	167 (20.823%)	295 (36.783%)	251 (31.297%)	89 (11.097%)
4 *vs.* 5	357 (44.514%)	315 (39.277%)	113 (14.090%)	17 (2.120%)
vs. 6	301 (37.531%)	317 (39.526%)	141 (17.581%)	43 (5.362%)
vs. 8	275 (34.289%)	290 (36.160%)	181 (22.569%)	56 (6.983%)
5 *vs.* 6	265 (33.042%)	319 (39.776%)	163 (20.324%)	55 (6.858%)

Table A6. Comparison of the number and percentage of individuals with estimated breeding values for QBA[1] attributes and temperament index that changes *n* quartiles between any two analyses (continued).

	Number of individuals that changed *n* quartiles[2]			
QBA attributes for Evaluator	**0**	**1**	**2**	**3**
Fearful cont'd				
5 *vs.* 8	209 (26.060%)	330 (41.147%)	203 (25.312%)	60 (7.481%)
6 *vs.* 8	305 (38.030%)	320 (39.900%)	133 (16.584%)	44 (5.486%)
Agitated				
Across vs. 1	291 (36.284%)	332 (41.397%)	154 (19.202%)	25 (3.117%)
vs. 4	393 (49.002%)	336 (41.895%)	71 (8.853%)	2 (0.249%)
vs. 5	421 (52.494%)	310 (38.653%)	65 (8.105%)	6 (0.748%)
vs. 6	431 (53.741%)	309 (38.529%)	57 (7.107%)	5 (0.623%)
vs. 8	354 (44.140%)	309 (38.529%)	127 (15.835%)	12 (1.496%)
1 *vs.* 4	264 (32.918%)	321 (40.025%)	159 (19.825%)	58 (7.232%)
vs. 5	271 (33.791%)	310 (38.653%)	176 (21.945%)	45 (5.611%)
vs. 6	203 (25.312%)	340 (42.394%)	218 (27.182%)	41 (5.112%)
vs. 8	167 (20.823%)	332 (41.397%)	239 (29.800%)	64 (7.980%)
4 *vs.* 5	309 (38.529%)	318 (39.651%)	141 (17.581%)	34 (4.239%)
vs. 6	268 (33.416%)	321 (40.025%)	160 (19.950%)	53 (6.608%)
vs. 8	260 (32.419%)	293 (36.534%)	201 (25.062%)	48 (5.985%)
5 *vs.* 6	317 (39.526%)	331 (41.272%)	123 (15.337%)	31 (3.865%)
vs. 8	292 (36.409%)	316 (39.401%)	153 (19.077%)	41 (5.112%)
6 *vs.* 8	376 (46.883%)	281 (35.037%)	105 (13.092%)	40 (4.988%)
Calm				
Across vs. 1	292 (36.409%)	341 (42.519%)	138 (17.207%)	31 (3.865%)
vs. 4	474 (59.102%)	288 (35.910%)	40 (4.988%)	0 (0.000%)
vs. 5	463 (57.731%)	291 (36.284%)	45 (5.611%)	3 (0.374%)
vs. 6	355 (44.264%)	327 (40.773%)	105 (13.092%)	15 (1.870%)
vs. 8	360 (44.888%)	321 (40.025%)	111 (13.840%)	10 (1.247%)
1 *vs.* 4	275 (34.289%)	312 (38.903%)	181 (22.569%)	34 (4.239%)
vs. 5	250 (31.172%)	332 (41.397%)	178 (22.195%)	42 (5.237%)
vs. 6	203 (25.312%)	325 (40.524%)	199 (24.813%)	75 (9.352%)
vs. 8	179 (22.319%)	337 (42.020%)	214 (26.683%)	72 (8.978%)
4 *vs.* 5	330 (41.147%)	340 (42.394%)	112 (13.965%)	20 (2.494%)
vs. 6	280 (34.913%)	341 (42.519%)	146 (18.204%)	35 (4.364%)
vs. 8	295 (36.783%)	324 (40.399%)	154 (19.202%)	29 (3.616%)
5 *vs.* 6	340 (42.394%)	288 (35.910%)	128 (15.960%)	46 (5.736%)
vs. 8	333 (41.521%)	307 (38.279%)	126 (15.711%)	36 (4.489%)
6 *vs.* 8	302 (37.656%)	310 (38.653%)	147 (18.329%)	43 (5.362%)
Attentive				
Across vs. 1	236 (29.426%)	336 (41.895%)	174 (21.696%)	56 (6.983%)
vs. 4	331 (41.272%)	318 (39.651%)	131 (16.334%)	22 (2.743%)
vs. 5	313 (39.027%)	329 (41.002%)	125 (15.586%)	35 (4.364%)
vs. 6	244 (30.424%)	345 (43.017%)	172 ()21.446%	41 (5.112%)
vs. 8	328 (40.898%)	316 (39.401%)	136 (16.958%)	22 (2.743%)
1 *vs.* 4	205 (25.561%)	321 (40.025%)	195 (24.314%)	81 (10.100%)
vs. 5	210 (26.185%)	333 (41.521%)	188 (23.441%)	71 (8.853%)
vs. 6	139 (17.332%)	288 (35.910%)	225 (28.055%)	150 (18.703%)
vs. 8	219 (27.307%)	295 (36.783%)	223 (27.805%)	65 (8.105%)

Table A6. Comparison of the number and percentage of individuals with estimated breeding values for QBA[1] attributes and temperament index that changes *n* quartiles between any two analyses (continued).

	Number of individuals that changed *n* quartiles[2]			
QBA attributes for Evaluator	**0**	**1**	**2**	**3**
Attentive cont'd				
4 *vs.* 5	210 (26.185%)	312 (38.903%)	192 (23.940%)	88 (10.973%)
vs. 6	221 (27.556%)	340 (42.394%)	175 (21.820%)	66 (8.229%)
vs. 8	237 (29.551%)	319 (39.776%)	183 (22.818%)	63 (7.855%)
5 *vs.* 6	159 (19.825%)	294 (36.658%)	247 (30.798%)	102 (12.718%)
vs. 8	217 (27.057%)	293 (36.534%)	207 (25.810%)	85 (10.599%)
6 *vs.* 8	239 (29.800%)	283 (35.287%)	209 (26.060%)	71 (8.853%)
Positively occupied				
Across vs. 1	169 (21.072%)	291 (36.284%)	194 (24.190%)	148 (18.454%)
vs. 4	344 (42.893%)	319 (39.776%)	113 (14.090%)	26 (3.242%)
vs. 5	398 (49.626%)	322 (40.150%)	79 (9.850%)	3 (0.374%)
vs. 6	267 (33.292%)	296 (36.908%)	204 (25.436%)	35 (4.364%)
vs. 8	275 (34.289%)	315 (39.277%)	159 (19.825%)	53 (6.608%)
1 *vs.* 4	158 (19.701%)	287 (35.786%)	194 (24.190%)	163 (20.324%)
vs. 5	227 (28.304%)	260 (32.419%)	199 (24.813%)	116 (14.464%)
vs. 6	162 (20.200%)	331 (41.272%)	174 (21.696%)	135 (16.933%)
vs. 8	169 (21.072%)	281 (35.037%)	206 (25.686%)	146 (18.204%)
4 *vs.* 5	286 (35.661%)	332 (40.150%)	140 (17.456%)	54 (6.733%)
vs. 6	179 (22.319%)	310 (38.653%)	229 (28.554%)	84 (10.474%)
vs. 8	286 (35.661%)	322 (40.150%)	140 (17.456%)	54 (6.733%)
5 *vs.* 6	199 (24.813%)	320 (39.900%)	201 (25.062%)	82 (10.244%)
vs. 8	311 (38.778%)	312 (38.903%)	126 (15.711%)	53 (6.608%)
6 *vs.* 8	186 (23.192%)	334 (41.646%)	205 (25.561%)	77 (9.601%)
Curious				
Across vs. 1	272 (33.915%)	344 (42.893%)	144 (17.955%)	42 (5.237%)
vs. 4	458 (57.107%)	281 (35.037%)	56 (6.983%)	7 (0.873%)
vs. 5	254 (31.671%)	321 (40.025%)	190 (23.691%)	37 (4.613%)
vs. 6	298 (37.157%)	309 (38.529%)	156 (19.451%)	39 (4.863%)
vs. 8	321 (40.025%)	305 (38.030%)	143 (17.830%)	33 (4.115%)
1 *vs.* 4	235 (29.302%)	353 (44.015%)	161 (20.075%)	53 (6.608%)
vs. 5	230 (28.678%)	279 (34.788%)	178 (22.195%)	115 (14.399%)
vs. 6	216 (26.933%)	306 (38.155%)	170 (21.197%)	110 (13.716%)
vs. 8	148 (18.454%)	308 (38.404%)	264 (32.918%)	82 (10.244%)
4 *vs.* 5	172 (21.446%)	321 (40.025%)	220 (27.431%)	89 (11.097%)
vs. 6	272 (33.915%)	345 (43.017%)	146 (18.204%)	39 (4.863%)
vs. 8	311 (38.778%)	299 (37.282%)	139 (17.332%)	53 (6.608%)
5 *vs.* 6	197 (24.564%)	271 (33.791%)	199 (24.813%)	135 (16.833%)
vs. 8	223 (27.805%)	285 (35.536%)	207 (25.810%)	87 (10.848%)
6 *vs.* 8	236 (29.426%)	334 (41.646%)	188 (23.441%)	44 (5.486%)
Irritated				
Across vs. 1	308 (38.404%)	341 (42.519%)	130 (16.209%)	23 (2.868%)
vs. 4	426 (53.117%)	285 (35.536%)	82 (10.224%)	9 (1.122%)
vs. 5	436 (54.364%)	319 (39.776%)	44 (5.486%)	3 (0.374%)
vs. 6	398 (49.626%)	309 (38.529%)	86 (10.723%)	9 (1.122%)
vs. 8	367 (45.761%)	319 (39.776%)	105 (13.092%)	11 (1.372%)

Table A6. Comparison of the number and percentage of individuals with estimated breeding values for QBA[1] attributes and temperament index that changes *n* quartiles between any two analyses (continued).

QBA attributes for Evaluator	Number of individuals that changed *n* quartiles[2]			
	0	1	2	3
Irritated cont'd				
1 *vs.* 4	300 (37.406%)	325 (40.524%)	136 (16.958%)	41 (5.112%)
vs. 5	278 (34.663%)	335 (41.771%)	150 (18.703%)	39 (4.863%)
vs. 6	192 (23.940%)	341 (42.519%)	204 (25.436%)	65 (8.105%)
vs. 8	176 (21.945%)	345 (43.017%)	224 (27.930%)	57 (7.107%)
4 *vs.* 5	365 (45.511%)	303 (37.781%)	111 (13.840%)	23 (2.868%)
vs. 6	260 (32.419%)	318 (39.651%)	168 (20.948%)	56 (6.983%)
vs. 8	284 (35.411%)	305 (38.030%)	170 (21.197%)	43 (5.362%)
5 *vs.* 6	275 (34.289%)	339 (42.269%)	159 (19.825%)	29 (3.616%)
vs. 8	268 (33.416%)	335 (41.771%)	154 (19.202%)	45 (5.611%)
6 *vs.* 8	345 (43.017%)	281 (35.037%)	143 (17.830%)	33 (4.115%)
Apathetic				
Across vs. 1	227 (28.304%)	287 (35.786%)	171 (21.322%)	117 (14.589%)
vs. 4	378 (47.132%)	308 (38.404%)	96 (11.970%)	20 (2.494%)
vs. 5	386 (48.130%)	312 (38.903%)	90 (11.222%)	14 (1.746%)
vs. 6	305 (38.030%)	325 (40.524%)	152 (18.953%)	20 (2.494%)
vs. 8	284 (35.411%)	338 (42.145%)	134 (16.708%)	46 (5.736%)
1 *vs.* 4	237 (29.551%)	295 (36.783%)	153 (19.007%)	117 (14.589%)
vs. 5	200 (24.938%)	259 (32.294%)	180 (22.444%)	163 (20.324%)
vs. 6	195 (24.314%)	335 (41.771%)	194 (24.190%)	78 (9.726%)
vs. 8	173 (21.571%)	275 (34.289%)	167 (20.823%)	187 (23.317%)
4 *vs.* 5	276 (34.414%)	311 (38.778%)	162 (20.200%)	53 (6.608%)
vs. 6	233 (29.052%)	321 (40.025%)	186 (23.192%)	62 (7.731%)
vs. 8	282 (35.162%)	327 (40.773%)	136 (16.958%)	57 (7.107%)
5 *vs.* 6	208 (25.935%)	320 (39.900%)	207 (25.810%)	67 (8.354%)
vs. 8	289 (36.035%)	286 (35.661%)	179 (22.319%)	48 (5.985%)
6 *vs.* 8	195 (24.314%)	305 (38.030%)	204 (25.436%)	98 (12.219%)
Happy				
Across vs. 1	311 (38.778%)	357 (44.514%)	107 (13.342%)	27 (3.367%)
vs. 4	377 (47.007%)	314 (39.152%)	103 (12.843%)	8 (0.998%)
vs. 5	407 (50.748%)	348 (43.392%)	45 (5.611%)	2 (0.249%)
vs. 6	267 (33.292%)	333 (41.521%)	159 (19.825%)	43 (5.362%)
vs. 8	137 (17.082%)	308 (38.404%)	213 (26.559%)	144 (17.955%)
1 *vs.* 4	226 (28.180%)	336 (41.895%)	154 (19.202%)	86 (10.723%)
vs. 5	280 (34.913%)	302 (37.656%)	168 (20.948%)	52 (6.484%)
vs. 6	270 (33.666%)	315 (39.277%)	136 (16.958%)	81 (10.100%)
vs. 8	91 (11.347%)	308 (38.404%)	293 (36.534%)	110 (13.716%)
4 *vs.* 5	316 (39.401%)	348 (43.392%)	104 (12.968%)	34 (4.239%)
vs. 6	198 (24.688%)	313 (39.027%)	186 (23.192%)	105 (13.092%)
vs. 8	156 (19.451%)	331 (41.272%)	202 (25.187%)	113 (14.090%)
5 *vs.* 6	258 (32.170%)	292 (36.409%)	158 (19.701%)	94 (11.721%)
vs. 8	197 (24.564%)	335 (41.771%)	187 (23.317%)	83 (10.349%)
6 *vs.* 8	146 (18.204%)	275 (34.289%)	264 (32.918%)	117 (14.589%)
Distressed				
Across vs. 1	324 (40.399%)	276 (34.414%)	161 (20.075%)	41 (5.112%)

Table A6. Comparison of the number and percentage of individuals with estimated breeding values for QBA[1] attributes and temperament index that changes *n* quartiles between any two analyses (continued).

	Number of individuals that changed *n* quartiles[2]			
QBA attributes for Evaluator	**0**	**1**	**2**	**3**
Distressed cont'd				
Across vs. 4	397 (49.501%)	302 (37.656%)	86 (10.723%)	17 (2.120%)
vs. 5	496 (61.845%)	258 (32.170%)	43 (5.362%)	5 (0.623%)
vs. 6	318 (39.651%)	279 (34.788%)	162 (20.200%)	43 (5.362%)
vs. 8	317 (39.526%)	328 (40.898%)	133 (16.584%)	24 (2.993%)
1 *vs.* 4	279 (34.831%)	315 (39.326%)	139 (17.353%)	68 (8.489%)
vs. 5	279 (34.831%)	299 (37.328%)	165 (20.599%)	58 (7.241%)
1 *vs.* 6	260 (32.459%)	277 (34.582%)	183 (22.846%)	81 (10.112%)
vs. 8	148 (18.477%)	311 (38.826%)	271 (33.833%)	71 (8.864%)
4 *vs.* 5	371 (46.317%)	294 (36.704%)	104 (12.984%)	32 (3.995%)
vs. 6	245 (30.587%)	287 (35.830%)	189 (23.596%)	80 (9.988%)
vs. 8	194 (24.220%)	322 (40.200%)	212 (26.467%)	73 (9.114%)
5 *vs.* 6	271 (33.833%)	280 (34.956%)	193 (24.095%)	57 (7.116%)
vs. 8	238 (29.713%)	325 (40.574%)	182 (22.722%)	56 (6.991%)
6 *vs.* 8	211 (26.342%)	292 (36.454%)	212 (26.467%)	86 (10.737%)

[1] QBA refers to qualitative behavior assessment, measured on a 136 mm visual analog scale. A low value (towards zero) indicates no or little expression, where a high value (towards 136) would indicate high or maximum expression.

[2] The number of quartile changes was calculated by first assigning an animal's quartile for any given analysis, then finding the difference of each animal's quartile between the two analyses compared. Percentage was calculated by dividing the number of individuals within that category by the total number of animals.

Table A7. Comparison of the number and percentage of individuals with estimated breeding values for temperament index that changes *n* quartiles between any two analyses.

	Number of individuals that changed *n* quartiles[1]			
Temperament index[2] for Evaluator	**0**	**1**	**2**	**3**
Across vs. 1	300 (37.406%)	292 (36.409%)	153 (19.077%)	57 (7.107%)
vs. 4	364 (45.387%)	306 (38.155%)	115 (14.339%)	17 (2.120%)
vs. 5	193 (24.065%)	298 (37.157%)	162 (20.200%)	149 (18.579%)
vs. 6	206 (25.686%)	323 (40.274%)	187 (23.317%)	86 (24.564%)
vs. 8	230 (28.678%)	311 (38.778%)	197 (24.564%)	64 (7.980%)
1 *vs.* 4	293 (36.534%)	318 (39.651%)	163 (20.324%)	28 (3.491%)
vs. 5	145 (18.080%)	246 (30.673%)	233 (29.052%)	178 (22.195%)
vs. 6	146 (18.204%)	291 (36.284%)	246 (30.673%)	119 (14.838%)
vs. 8	148 (18.454%)	354 (44.140%)	236 (29.426%)	64 (7.980%)
4 *vs.* 5	111 (13.840%)	245 (30.549%)	211 (26.309%)	235 (29.302%)
vs. 6	144 (17.955%)	245 (30.549%)	230 (28.678%)	183 (22.818%)
vs. 8	329 (41.0225)	287 (35.786%)	153 (19.077%)	33 (4.115%)
5 *vs.* 6	308 (38.404%)	319 (39.776%)	138 (17.207%)	37 (4.613%)
vs. 8	139 (17.332%)	262 (32.668%)	205 (25.561%)	196 (24.439%)
6 *vs.* 8	148 (18.454%)	271 (33.791%)	200 (24.938%0	183 (22.818%)

[1] The number of quartile changes was calculated by first assigning an animal's quartile for any given analysis, then finding the difference of each animal's quartile between the two analyses compared. Percentage was calculated by dividing the number of individuals within that category by the total number of animals.

[2] Temperament index: the first principal component score generated from QBA scores using a Principal Component Analysis in SAS (SAS Institute, Inc., Cary, NC.