REFINING AND VALIDATING AN ON-SITE CANINE WELFARE ASSESSMENT TOOL DEVELOPED FOR USE IN COMMERCIAL BREEDING KENNELS

by

Lynda Wambui Mugenda

A Thesis

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Master of Science

Department of Animal Sciences

West Lafayette, Indiana

December 2018

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Candace Croney, Chair

Department of Comparative Pathobiology/Animal Sciences

Dr. Marguerite O'Haire

Department of Comparative Pathobiology

Dr. Marisa Erasmus

Department of Animal Sciences

Approved by:

Ryan A. Cabot

Head of the Graduate Program

To Bella and Kai, my inspiration!

ACKNOWLEDGMENTS

I would like to first and foremost extend my sincerest gratitude to my advisor and mentor, Dr. Candace Croney, for providing me with the opportunity to pursue my degree under her stewardship. Your support, patience and encouragement have been instrumental in making my graduate experience worthwhile, and I will always be grateful to you for that. I would also like to acknowledge the rest of my committee members, Dr. Marisa Erasmus and Dr. Marguerite O'Haire, for their valuable insight and prompt feedback throughout the course of my studies and during the writing process as well. This thesis would not be what it is without their input and contribution. I would also like to sincerely thank Dr. James Ha for his valuable assistance with statistical analysis required for this thesis as well as his moral support. My sincerest thanks also go to Traci Shreyer, Dr. Hannah Flint, Dr. Judith Stella, and Dr. Amy Bauer for their appreciated contribution and feedback throughout the study design, data collection and data analysis process.

I would also like to sincerely thank all the research assistants who assisted me with data collection. Your time and effort will always be appreciated. Additionally, I would like to acknowledge all my lab members for being a source of emotional support and friendship. I will sincerely miss our many laughs, talks and long dog walks. I must also extend my thanks to all the shelters and especially the dog breeders, who opened their doors to us and made this project possible. Without their support and passion to improve the welfare of breeding dogs, this research would not be feasible. Further, I would like to thank my family and friends, both near and far. I would not have had the strength to be where I am today without their support and belief in me. Finally, I would like to acknowledge my heavenly Father who made all this possible, and who continues to give me the courage to achieve all that I set my mind to do.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Author: Mugenda, Lynda, W. MS
Institution: Purdue University
Degree Received: December 2018
Title: Refining and Validating an on-site Canine Welfare Assessment Tool Developed for use in Commercial Breeding Kennels.
Committee Chair: Dr. Candace Croney

Accurate assessments of behavior and welfare are needed to evaluate the state of domestic dogs maintained in commercial breeding (CB) and other types of kennels. Field assessments of dogs' states of being must be reliable, valid and efficient. However, observer subjectivity and situational variation in dogs' responses pose a challenge to incorporating behavioral metrics into welfare assessment tools. The published Field Instantaneous Dog Observation (FIDO) tool, designed to capture the immediately observable physical and behavioral status of dogs in kennels, was thus evaluated on its reliability and validity. Specifically, the main goals were to determine 1) reliability of the behavioral scoring when used by novice raters, 2) whether and to what extent dogs' behavioral responses to stranger-approach changed during a 30-second observation period, and 3) the predictive power of the FIDO scoring on behavioral responses of dogs placed within a standardized arena with a stranger. Behavioral responses to stranger-approach were organized into three categories: red, indicating a fearful response to approach, green, indicating an affiliative or neutral response, and yellow, indicating an ambivalent response. In study one, behavior assessment was conducted by two novice raters with 50 dogs housed at two US shelters. A stranger approached the home pen of each dog in a non-threatening manner, stood quietly, extended a hand to the dog and scored the response while the test was video-recorded. Intra-rater reliability was assessed by comparing each rater's live observation scores with their scores of the same dogs using video recordings. Inter-rater agreement between

scores from video recordings was also calculated. In study two, 81 commercial breeding dogs maintained at four USDA-licensed CB facilities in the US were approached by one observer and scored once every five seconds for 30 seconds. Of the 81 subject dogs, 56 met the criteria for full sampling. In study three, 40 pairs of dogs were scored using the FIDO tool in their home pens at four USDA-licensed CB facilities and assessed on responses to an unfamiliar observer in a field test conducted in an outdoor arena. Behavioral measures from the field test such as latencies to approach the stranger and durations in different areas of the arena were captured from video recordings and subjected to a principal components analysis [PCA] for reduction of variables. A multivariate multiple regression analysis was thereafter used, with principal component scores obtained from the PCA as outcome variables, and FIDO scores as predictor variables, accounting for other factors such as breed and group composition. Results from study one indicated that raters showed almost perfect agreement between their own scores of live and video-recorded shelter dog responses (kappa = 0.83, 0.89) and between each other’s video-recorded scores (kappa = 0.87), indicating high intra- and inter-rater reliability. Results from study two indicated that over a 30-second time frame with five-second increments, 91% of the dogs showed no change in their behavioral response to approach. This suggests that the first five seconds of scoring provide a reliable time point for assessing behavior using the FIDO tool and indicates no benefit to extending the FIDO scoring period to gauge dogs’ immediate responses to stranger-approach. Results from study three showed that dogs scored as red, as determined by the FIDO scoring, also showed higher scores associated with avoidance, indicated by greater time spent away and more time taken to approach the unfamiliar person in the field test ($p = 0.039$). FIDO scoring was, however, not significantly predictive of other behavioral responses such as interaction with the unfamiliar person. How dogs were housed was also significantly associated with incidences of escape attempts from the arena; dogs paired as a mixed sex (i.e. male-female) showed lower scores on incidences of escape attempts than dogs paired as a same sex (i.e. female-female) ($p = 0.003$). Taken together, these results suggest that the FIDO tool can be used by individuals without much expertise in canine behavior, and also attest to the practicality of the tool via a reliable five-second approach. Further, results also indicate that the FIDO scoring can help to identify fearful dogs in need of greater socialization towards unfamiliar people. Future research would, however, be valuable in validating behavioral scores obtained using the tool against long-term indicators of overall welfare.

CHAPTER 1. INTRODUCTION

Commercial dog breeding remains a polemical debate in the United States today despite the efforts implemented in improving this industry. In the public domain, commercial breeding facilities are often labelled as “puppy mills” or “puppy farms” and are frequently portrayed as filthy, substandard kennels where dogs are neglected or abused. A puppy mill describes a large, deficient breeding operation ran by individuals with little to no concern for the welfare of puppies or breeding dogs (National Animal Interest Alliance [NAIA], n.d.). Dogs in these kinds of facilities are often described as malnourished, fearful, abused, dirty and having many medical issues that go unchecked (Humane Society Veterinary Medical Association [HSVMA], 2013). While some substandard commercial breeding facilities may fit this description, some sources report that there has been an overall decline in the existence of puppy mills especially with the increasing oversight of the Animal Welfare Act and the efforts by the United States Department of Agriculture [USDA] to be transparent in their inspections (The Inquirer, 2009; USDA, 2017). However, public perception, most often perpetuated by animal welfare activists and organizations, still continues to paint the entire commercial breeding industry as puppy mills, undesirably slowing down the progress being made by good and legitimate commercial breeders in improving kennel standards (NAIA, n.d.). The outcome of these perceptions can be exemplified by the new ban imposed in California to stop pet stores from selling commercially bred animals, making it the first state to carry this law (National Public Radio, 2017). Furthermore, counties such as DuPage and Will of Illinois have recently proposed a bill to ban the retail sale of commercially bred pets in pet stores (Chicago Tribune, 2017), and there seems to be a continuing rise in public welfare concerns surrounding commercial breeding in general.

The demand for purebred dogs drives the commercial breeding industry. Human beings have bred and continue to breed dogs for various purposes, ranging from functionality (e.g. hunting, service and police dogs) to research, dog showing and companionship (McGreevy and Nicholas, 1999). Within the United states, anyone with five or more breeding females is considered a commercial breeder, is required to obtain a USDA license and is subject to regular inspections if selling to pet stores or via online (USDA, 2016). However, different definitions exist among different states. In Indiana, for example, a commercial breeder is defined as a person who keeps more than 20 "unaltered" female dogs that are at minimum 12 months old (Indiana State Board of Animal Health [BOAH], 2010). Such discrepancies in what constitutes a commercial breeder potentially create areas of concern especially regarding inspection requirements and regulation of facilities (American Society for the Prevention of Cruelty to Animals [ASPCA], 2015). Currently, only 26 states have laws that regulate commercial kennels in the United States (ASPCA, 2015). Likewise, the commercial breeding industry has been criticized by some groups, such as the Humane Society of the United States [HSUS], for not being sufficiently stringent in their inspections and regulations (HSUS, 2018). There have indeed been efforts made by the USDA to make inspections of licensed breeders available online for the public, although recently, ease of access to such reports has been reduced (Journal of the American Veterinary Medical Association [JAVMA] news, 2017). All of these factors aid in propagating the distrust and suspicion surrounding the commercial breeding industry today.

Aside from the concern surrounding the regulation and welfare of dogs kept in breeding facilities, there is also increasing debate on the industry's effects on pet population in the United

States. Some stakeholders hold the view that commercial breeding exacerbates the issue of pet overpopulation by contributing to shelter euthanasia rates (Fennell, 1999; Patronek and Glickman, 1994). Indeed, pet overpopulation has often been measured or characterized by the percentage of companion animals being euthanized in animal shelters (Fennell, 1999; Olson and Moulton, 1993; Zawistowski et al., 1998). Over the years, however, euthanasia has decreased, with the percentage estimated to be roughly 23% in 2018, down from 36% in 2011 (ASPCA, 2018), and mainly due to current initiatives set in place such as spay and neuter programs, education programs and higher numbers of adoptions from shelters and rescues (Protopopova, 2016). Likewise, research findings have shown that the number of purebred dogs found in shelters is very low (~5%), strongly contradicting the argument that commercial breeders are significantly contributing to overpopulation (NAIA, 2015). Some experts have also speculated that the blame placed on the production of puppies in contributing to the overpopulation of companion animals is unfounded as dog breeders are only responding to “consumer demand” (Fennell, 1999). Such blame on producers has been equated to "blaming the waiters in restaurants for obesity" (Stigler, 1964, p. 29). Finally, others have argued that the statistics associated with pet overpopulation in shelters are unclear and have been deemed unreliable, overestimated and a mere reflection of unsubstantiated assumptions (Clancy and Rowan, 2003; Rowan, 1992). Accurate and objective assessment of what causes pet overpopulation thus remains a contentious issue.

While there has been a plethora of studies conducted on dogs in shelters and other kenneled environments, there is currently little research that has been done in the commercial breeding population of dogs. Of the few studies that have been published, findings from owner reports suggest that there may be a significant difference in behavioral and health outcomes between dogs believed to be acquired from commercial breeding facilities and pet dogs acquired from other sources, with the former group exhibiting higher incidences of fear, house soiling and contextually inappropriate behaviors (McMillan et al., 2011). Another study based on surveys also reported more adverse behaviors in puppies purchased from pet stores compared to those purchased from non-commercial breeders (McMillan et al., 2013). However, conclusions about welfare in commercial dog breeding are to date primarily being drawn from owner reports, necessitating more objective research. A recent study on commercial breeding facilities that investigated the impact of kennel flooring on dog foot health, showed that the majority of dogs assessed in southern Indiana facilities did not have major foot issues as a function of their flooring and cleanliness, and most dogs had few dental and ear health concerns (Hurt, 2016; Stella et al., 2018). Limitations notwithstanding, this study provides a first glance at the welfare of commercial breeding dogs, as well as valuable information that is contrary to public perceptions that all commercial breeding facilities are filthy, substandard, and teeming with dog welfare issues.

The importance of studying canine welfare in commercial breeding facilities has implications not just for the dogs themselves, but for other stakeholders such as pet purchasers. Several cases exist of puppy owners having complained of purchasing puppies with health or behavioral problems from pet stores (e.g. ABC Action news, 2018). As many pet stores acquire puppies from commercial breeding facilities, understanding the underlying causes of some of these concerns is vital and would necessitate tracing back to the environmental health of breeding dogs kept in such facilities. Likewise, knowing the welfare status of this population of dogs would also help improve the chances that they are appropriately suited for environments in which they are likely to reside in once rehomed (Bateson, 2010).

While understanding the welfare of dogs kept in commercial breeding facilities is important, one of the challenges facing objective assessment of welfare is the lack of adequate assessment tools to be used on-site in these establishments. Currently, a newly developed assessment tool exists that aims to objectively identify the welfare state of dogs in commercial breeding operations. However, to enhance its robustness, further investigation of its reliability and validity is necessary. The present study will therefore seek to evaluate the reliability and validity of the assessment tool, and in the process investigate indicators of behavioral well-being in commercial breeding dogs.

CHAPTER 2. LITERATURE REVIEW

2.1 What Is Animal Welfare?

Animal welfare concerns how an animal is coping with the conditions in which it lives (Word Organization for Animal Health [OIE], 2017). Over the past several years, however, the term has stirred up debate in the scientific community with regard to the definition of animal welfare. Broom (1986), for instance, highlighted that an animal's welfare involves its ability to cope with its environment and emphasized that an animal's difficulty in coping or failure to cope could occur independently of suffering or pain. Carpenter (1980), on the other hand, defined animal welfare as the degree to which an animal could adapt in its environment *without* suffering (as reviewed in Fraser et al., 1997), thus implying that welfare would only be considered poor after suffering occurs. Others have also defined animal welfare in relation to harmony (both physical and psychological) between an animal and its environment (e.g. Hurnik, 1988; Van Putten, 2000). While no formal or scientific definition of animal welfare exists, there is indeed consensus that the study of animal welfare is fundamental for the improvement of animals' lives (Millman et al., 2004).

It is generally believed that inspiration for the study of agricultural animal welfare first stemmed from the publication of the book, *Animal Machines* by Ruth Harrison in 1964 (Van de Weerd and Sandilands, 2008). Harrison criticized the use of intensive animal farming practices and use of drugs in animal production in the UK, which instigated a public outcry on the welfare of farm animals and led to the appointment of a committee to investigate these concerns (Van de Weerd and Sandilands, 2008). The Brambell committee was thus established to address the welfare concerns of intensely farmed animals, leading to the creation of the "Five Freedoms": 1) freedom from hunger and thirst, 2) freedom from discomfort, 3) freedom from pain, injury and disease, 4) freedom to express normal behavior, and 5) freedom from fear and distress – all of which emphasize physical, mental and psychological well-being (McCulloch, 2013). Extending from this framework, various scientists and researchers have over time refined their conceptions of animal welfare and the associated implications for welfare assessment, leading to different "schools" of thought about what is important for animals (Duncan, 2005; Fraser and Weary, 2003; Fraser, 2009). Some scholars have approached animal welfare solely in terms of the animal's ability to freely express natural behaviors according to their adaptations (e.g. Kiley-Worthington, 1989; Rollin, 2007), while others have characterized the term in relation to physical health and biological functioning (e.g. Barnett and Hemsworth, 2003); still, others have approached animal welfare in terms of the animal's subjective experiences such as what the animal wants or feels (e.g. Dawkins, 1990, 2004; Duncan, 1996, 2004). While it can be argued that each of these approaches are valid and important by themselves, having a narrow and one-dimensional view of what animal welfare entails can lead to oversights in major areas of welfare, such as the experience of pleasure and its impact on animals' states of being (Mench, 1998). Hence, a more comprehensive perception of animal welfare that has come to be accepted in the scientific community encompasses all three views; that an animal is physically healthy (free from disease, injury, hunger), lives a reasonably natural life (performs behavioral urges and needs), and is free from suffering (pain, fear, boredom) while also enjoying some pleasures (play, companionship, exploration) (Fraser, 2003; Fraser et al., 1997; Green and Mellor, 2011; Hemsworth et al., 2015).

2.2 Challenges Facing Animal Welfare Assessment

Beyond the discourse surrounding the definition or conception of animal welfare lies the dispute on what constitutes good welfare. Animal welfare exists on a continuum from very poor to very good (Broom, 1991) and changes with time, creating a potential for disagreement on what is acceptable or not. Further, a farmer, a veterinarian, a scientific researcher, an animal welfare advocate and a member of the general public may all come to different conclusions concerning the acceptability of poor welfare in an individual animal (Fraser, 2008). There is also increasing agreement that these different conclusions stem from ethical concerns that are modified by the values people hold regarding what is better or worse for animals (Lassen et al., 2006; Miele et al., 2011), and which differ not only among people, but also across different regions and cultures (Serpell, 2004). In an attempt to determine an agreed level of poor welfare, Barnett and Hemsworth (1990) proposed a cut-off point whereby animals that sustained an elevation of greater than 40% in free corticosteroid concentrations were at more risk than controls in incurring fitness reduction, decrease in growth rate or immunosuppression. While the idea of a cut-off point in assessing animal welfare risk and poor welfare may seem plausible, it carries with it numerous problems which relate back to the conceptions of animal welfare and the underlying ethical concerns. Broom (1991), for instance, considers the extent of difficulty an animal displays in coping. Even though there may be no reduced fitness, suppressed growth or disease susceptibility in animals below the cut-off point, these animals may still have poorer welfare relative to others that are not finding it difficult to cope with the environment (Mendl, 1991).

Also emphasized in the literature is the fact that animal welfare is multidimensional, involving aspects of physical and biological functioning, behavior and mental health (Botreau et al., 2007; CAST, 2018; Miele et al., 2011; Scott et al., 2001). This in turn translates into a multidimensional approach to assessing animal welfare. It has been widely accepted that there is no singular way of assessing welfare, but that different measures ought to be used and validated against each other (Mason and Mendl, 1993; Swaisgood, 2007). Issues of welfare assessment, however, still arise due to the different values underlying the measures used (Fraser, 2003) and the interpretation of these measures (Rushen, 1991). For instance, a main problem often associated with the misinterpretation of measures is the inadequate understanding of the functional and biological significance of the measures being used (Dawkins, 1998; Terlouw et al., 1991). Despite these challenges, however, animal welfare as a scientific discipline continues to grow and improve for the benefit of animals.

2.3 Scientific Assessment of Animal Welfare

Scientific assessment of welfare often reflects back to the concept of stress (Morgan and Tromborg, 2007; Swaisgood, 2007). Stress responses, both behavioral and physiological, are triggered when homeostasis of an animal is threatened (Moberg, 2000). While moderate stress is considered necessary, for instance to improve vigilance (Wiepkema and Koolhaas, 1993), responses that affect normal functioning subject the animal to “distress”, which is what distinguishes good stress (eustress) from bad stress (Moberg, 2000). The success or failure of adaptive stress responses often depend on the ability of one’s body to adequately “compensate” in the presence of a challenge (Mitchell and Kettlewell, 1998; Veissier and Boissy, 2007). Thus, inadequate compensation can often lead to difficulty in coping, while decompensation can lead to failure in coping and detrimental effects on the body (Wiepkema and Koolhaas, 1993). For the simple purpose of clarification, however, the term “stress” and “distress” will be used interchangeably throughout the remainder of the text.

As mentioned before, challenges exist in the assessment of welfare. However, an understanding of the impact of stress remains a priority for animal welfare scientists. Indicators of welfare that have thus become accepted as valid assessments of welfare include physical, behavioral and cognitive measures of welfare.

2.3.1 Physical Measures

Physical metrics of welfare are highly relevant in providing information about the health and welfare status of animals and can refer to both the internal and external changes that occur in an animal's body when attempting to cope with its environment. External physical measures are often times easily measurable, and can include incidences of injury and disease, measures of body condition, as well as production parameters such as reproductive performance (i.e. numbers of stillborn or reduced litter sizes), growth rates, and mortality rates among others (Broom and Fraser, 2007). Such measures can also provide insight into the impact of housing and/or management practices on the welfare of animals as a group. For example, poor housing conditions have been associated with increased foot lesions and broken bones in birds (Appleby and Hughes, 1991). Internal physical measures, on the other hand, include physiological metrics such as cortisol levels, catecholamines and measures of heart rates which despite not being easily or directly measurable, still provide valuable information on the welfare state of an animal. The most commonly used physiological measures of stress include those associated with activation of the hypothalamic-pituitary-adrenal axis (HPA axis), the sympathetic nervous system (SNS), and the immune system, all of which will be briefly discussed.

Activation of the HPA axis has been considered a useful measure of stress and has traditionally often been a primary means by which to assess stress responses in animals. Briefly, HPA axis activation due to a stressor results in the release of corticotrophin-releasing factor

(CRF) from the hypothalamus, which in turn leads to the release of Adrenocorticotrophic Hormone (ACTH) from the anterior pituitary, and finally the release of glucocorticoids from the adrenal cortex (as reviewed in Otovic and Hutchinson, 2015). Glucocorticoids vary between species, with mammals and fish releasing cortisol, and birds and rodents releasing corticosterone (Broom and Fraser, 2007). Likewise, glucocorticoids can be monitored via different samples. Hiby et al. (2006) assessed the physiological response of dogs entering a rehoming kennel via urine samples while Schatz and Palme (2001) assessed the stress responses via fecal cortisol metabolites in cats and dogs. Other common approaches that are used in animal welfare research include salivary cortisol sampling e.g. in assessing stress responses associated with transport in calves (Fell and Shutt, 1986), hair cortisol sampling e.g. in assessing chronic stress in cats and dogs (Accorsi et al., 2008), and plasma cortisol sampling e.g. in assessing stress responses in intensively housed pigs (Parrott et al., 1989). Though glucocorticoids can be a useful metric of stress, challenges with regard to interpretation still exist, as stressors may not always activate the system, and glucocorticoids may increase even in situations that do not constitute *distress or negative stress*, such as in sexual behavior, exercise, social interaction or courtship (CAST, 2018).

The sympathetic nervous system (SNS), on the other hand, is one of the two main divisions of the autonomic nervous system (ANS) and involves the release of catecholamines epinephrine (adrenaline) and norepinephrine (noradrenaline) from the adrenal medulla. These hormones ultimately activate the fight-or-flight response during a perceived threat by increasing heart rate and blood pressure, contracting muscles, converting glycogen to glucose for energy and suppressing all body processes that are not critical in order to prepare an animal to react (Hydbring-Sandberg et al., 2004; Reeder and Kramer, 2005). The common measures of the sympathetic nervous system that have been widely assessed in animals are heart rate and heart rate variability, and core body temperature (Lewis et al., 2008; Marchant et al., 1997; von Borell et al., 2007). Heart rate variability has, however, come to be considered a more robust measure of cardiac activity than heart rate because it takes into consideration the interactions of the antagonistic sympathetic and parasympathetic nervous systems (von Borell et al., 2007), thereby accounting for the challenges in interpreting heart rate responses (i.e. distinguishing between arousal and stress).

Stress may also mediate the immune system through the endocrine and sympathetic systems (Wiepkema and Koolhaas, 1993). According to Khansari et al. (1990), stress signals disrupt the homeostasis of various hormones such as glucocorticoids which can significantly impact the immune response in general. Glucocorticoids and catecholamines have, for instance, been found to inhibit the synthesis of proinflammatory cytokines that typically activate the immune system, thereby inducing immunosuppression (Wiegers et al., 2005) and increasing susceptibility to disease. However, the type, intensity and duration of the stressor matters; acute and chronic stressors affect the immune responses differently, with acute stress being more likely to enhance the immune system (Dhabhar, 2000) and chronic stress being more likely to suppress the immune system (Wiegers et al., 2005). This has important implications for animals kept in environments in which chronic stress is experienced.

While physical measures of welfare can sufficiently indicate poor welfare and suffering in animals through measures such as lesions or reduced production, other measures of welfare such as behavior are important to assess, particularly in situations whereby physical well-being may not have been compromised or may not yet be evident.

2.3.2 Behavioral Measures

Behavior can be very valuable in assessing an animal's state as it provides an easily observed indicator of an animal's perception and experience of its environment (Swaisgood, 2007). However, behaviors are susceptible to misinterpretations because of 1) deciding whether a behavior truly reflects poor welfare (e.g. stereotypies), and 2) determining how much performance of a behavior indicates poor welfare (Dawkins, 1998; Duncan and Dawkins, 1983). Therefore, it may be necessary to monitor a variety of behaviors in concert and compare across other metrics (i.e. physiological and physical) in order to accurately characterize the welfare of an animal (Swaisgood, 2007). Common behavioral measures that are often used to assess animal welfare include measures of fear (e.g. avoidance responses, Hemsworth and Barnett, 1991; Tanida et al., 1995), pain (e.g. abnormal gait associated with lameness in dairy cows, Rushen et al., 2007), sickness behaviors (e.g. decreased feeding and lethargy, Fogsgaard et al., 2012; Stella et al., 2011), abnormal behaviors (e.g. feather pecking in hens and tail-biting in pigs, El-Lethey et al., 2000; Taylor et al, 2010) and stereotypies. Stereotypies are generally defined as repetitive unvarying behavior patterns with no apparent function (Mason, 1991; Mason et al., 2007). Examples include circling and tail-chasing in dogs, head-shaking in domestic fowl, bar-biting in stall-housed sows and stall-kicking in horses (Broom and Fraser, 2007). These behaviors have been linked to poor welfare, especially when animals are kept in environments or under conditions that lead to frustration such as during food restriction (Lawrence and Terlouw, 1993), or when animals experience fear or unavoidable stress, as well as physical restraint and lack of stimulation (Mason et al., 2007). However, the relationship between stereotypic behaviors and welfare is complex; stereotypic performance may also be a coping mechanism and hence may actually serve a purpose**.** Hansen and Jeppesen (2006), for instance, showed that high stereotyping minks also exhibited a less fearful temperament than low stereotyping minks. Other studies have also revealed an association between stereotypies and increased reproductive success (Jeppesen et al., 2004) as well as lower baseline cortisol levels (Bildsøe et al., 1991). The link between stereotypies and poor welfare is further complicated with the revelation that improving welfare does not automatically reduce the performance of stereotypies (Cooper et al., 1996) and that stereotypies do not always correlate with other signs of poor welfare such as corticosteroid levels (Mason and Latham, 2004). Regardless, stereotypies should be taken as a cautionary sign of potential suffering and stress, especially when they result in harmful behaviors such as self-injury or aggression.

2.3.3 Cognitive Measures

Cognitive measures of welfare are becoming increasingly accepted as valuable indicators of animal welfare. Research on humans has consistently shown that emotional states influence cognitive processes such as memory, attention and judgment, suggesting a reliable interaction between emotion and cognition (Haselton and Nettle, 2015; Mineka and Sutton, 1992). For example, individuals with a negative affective state are more likely to remember negative events than positive ones, focus more of their attention on negative events, and interpret ambiguous stimuli or situations more negatively, thereby possessing a negative judgment bias (Mendl et al., 2009). Currently, animal studies have also begun incorporating assessment of cognitive measures, particularly judgment biases, in order to gain insight into the affective or emotional states of animals (Baciadonna and McElligott, 2015; Harding et al., 2004; Mendl and Paul, 2004; Roelofs et al., 2016). For instance, pigs that had been previously trained to discriminate between positive and negative auditory cues showed differential responses to an ambiguous cue as a function of their housing quality (Douglas et al., 2012); specifically, pigs housed in an enriched environment were more likely to respond to an ambiguous sound more optimistically than pigs housed in a barren environment (Douglas et al., 2012). Studies on cognitive biases have also been carried out in birds (Wichman et al., 2012), sheep (Doyle et al., 2010), and dogs (Mendl et al., 2010). While physical and behavioral measures remain worthwhile in objectively assessing animal welfare, difficulties still exist in the interpretations of results (i.e. cortisol levels and stereotypic behaviors) particularly with regard to the valence (Mason and Rushen, 2006; Mendl et al., 2009; Paul et al., 2005). Therefore, combining cognitive assessments with existing measures could provide an opportunity to strengthen assessments of animal welfare, as cognitive measures provide insight into the affective states of animals.

2.4 Assessment of Welfare in Domestic Dogs

As discussed in the previous section, various measures of welfare can be utilized across different species so long as there is correct interpretation and knowledge of the species of interest. While there is little known about the welfare status of commercial breeding dogs, much can be inferred from what is currently known in other domestic dogs such as those used for research and those maintained in shelters. Similar to other species, current approaches to welfare assessment in dogs often rely on behavioral measures such as stereotypies and abnormal behaviors, physical and physiological measures such as health status, cortisol levels and heart rate, and more recently, cognitive measures such as judgement biases.

Behavioral indicators of welfare in dogs have been extensively studied in those kept in laboratory (e.g. Hall et al., 2017; Hetts et al., 1992) and in shelter (e.g. Hennessy et al., 2001) environments. Behavior, however, varies within and even between dogs within similar situations, suggesting that dogs either perceive stressors differently or just cope differently (Rooney et al., 2009; Stephen and Ledger, 2005). Such differences in coping styles may reflect differences in morphology, genetics, early life experiences (e.g. Foyer et al., 2014), breed (e.g. Morrow et al., 2015) and other factors that may all impact the temperament and behavior of individual dogs (Stephen and Ledger, 2005). Protopopova et al. (2014) found that dogs of medium and large breed and dogs with long hair were more likely to keep further away and express more panting than small-sized dogs, and dogs with a history of abuse were rated as showing more aggression and avoidance towards unfamiliar humans and dogs, more hyperactivity and more persistent barking compared to a control group (McMillan et al., 2015). All these various aspects ultimately influence how dogs behave, which in turn influence how behavioral measures are captured. Thus, to accurately assess canine welfare, consideration of time spent displaying certain behaviors and the contexts in which such behaviors are exhibited become necessary.

Abnormal behavior, both in the form of stereotypies and in the altered frequency of normal behaviors, have been used as indicators of poor welfare or stress in dogs. To date, presence of stereotypic or repetitive behavior has strongly been linked to poor welfare in captive animals (Mason and Latham, 2004). Evaluation of stereotypic or repetitive behaviors in shelter and other kenneled dogs has shown that particular behaviors are exhibited by dogs experiencing distress or mild stress (Stephen and Ledger, 2005). Expressions of circling, repetitive pacing, wall-bouncing and tail-chasing provide some examples of such stereotypic behaviors (Denham et al., 2014). For instance, laboratory dogs housed in small cages spent 14% of their time engaging in behaviors such as whirling, circling and pacing, while dogs housed in outdoor pens showed none of these behaviors (Hetts et al, 1992), suggesting the effects of different housing conditions on canine welfare. Furthermore, dogs that were singly housed showed a higher percentage of repetitive behaviors such as circling and pacing compared to dogs that were group housed (Beerda et al., 1999a; Hubrecht et al., 1992; Mertens and Unshelm, 1996) suggesting that social restriction is a source of stress in dogs. Stereotypies may therefore provide valuable information on how dogs fare within their environment, but interpretation of the context is also important. For example, Overall and Dunham (2002) found that dogs can still engage in repetitive behaviors independently of their social or physical environments, suggesting an underlying brain abnormality in some dogs. Other stress-related behaviors that have been studied and found in dogs include responses such as body shaking and paw-lifting, and oral behaviors such as lip-licking and yawning (Beerda et al., 1997, 1998, 1999a; Rooney et al., 2007). For instance, Schilder and Van Der Borg (2004) assessed the behavioral impacts of using the shock collar during training and found that shocked dogs showed significantly more lip-licking than controls, suggesting that shock collar training is a stressful experience.

Variation in frequency of normal behaviors as a function of an animal's environment is also an important indicator of poor welfare. In dogs, this may extend from increased activity to unresponsiveness, or through changes in species-typical behaviors such as elimination, grooming, barking, panting or feeding (Protopopova, 2016). For example, dogs that were subjected to social and spatial restriction showed increased frequencies of auto-grooming, paw-lifting, lower body posture, circling and vocalizations as well as reduced locomotor activity (Beerda et al., 1999a). Interpretation of these behaviors is, however, not always forthright. Measures of activity can become difficult to interpret as both inactivity (e.g. Hubrecht et al., 1992) and a higher level of activity (e.g. Hetts et al., 1992) have been associated with chronic stress as a function of social isolation (Beerda et al., 2000). Likewise, excessive barking has been suggested to be indicative of elevated stress and anxiety (Hetts et al., 1992; Rooney et al., 2009; Tod et al., 2005) and has been included as a behavioral indicator of poor welfare in many studies. Yet, barking may also accompany social interactions such as during play or in soliciting attention

(Bradshaw and Nott, 1995; Pongrácz et al., 2010), or it may be an evolutionary mechanism for warning and guarding (Adams and Johnson, 1994), and differences in barking may also reflect diversity across breeds. Careful assessment of behaviors under normal conditions in dogs is therefore imperative in order to determine any deviation from the norm.

Physiological measures provide valuable information about the internal state of an animal in response to potential stressors. In dogs, there has been considerable research done on assessing physiological measures of stress, including measures of catecholamines (Beerda et al., 1996), immune function (Beerda et al., 1999b), and heart rate (Vincent and Leahy, 1997). However, HPA-activation indicators, and thus, cortisol levels, have been among the most investigated (e.g. Blackwell et al., 2010; Hennessy et al., 1997; Kobelt et al., 2003). Hennessy et al. (1997), for instance, found that dogs entering a shelter had higher plasma cortisol levels during their first three days compared to dogs that had stayed for a longer period of time (longer than 10 days) and dogs that were living in homes as pets. As highlighted before, however, interpretation of physiological measures can be difficult, and must also account for context (Titulaer et al., 2013). For instance, in military dogs that were subjected to auditory and visually fear-inducing stimuli, salivary cortisol level changes were not adequately reflected during these experiences (Foyer et al., 2016). While some studies have shown a strong correlation between physiological and behavioral measures (e.g. high urinary cortisol/creatinine ratios with higher frequency of behavioral indicators of chronic stress in Beerda 1999a,b), many studies have revealed inconsistent or weak correlations of physiological indicators with behavioral indicators of stress (e.g. Hennessy et al., 2001; Hiby et al., 2006; Part et al., 2014; Rooney et al., 2007). This weak relationship has been suggested to result from the selective breeding of specific behaviors such as docility, which may influence how dogs respond to stressors behaviorally (Blackwell et al., 2010; Hewson et al., 2007). Additionally, as previously discussed, cortisol levels do not always provide information on the emotional valence of an experience. An increase in cortisol levels may reflect an increase in arousal but not necessarily distress, which may also explain why the correlation between cortisol levels and behavioral indicators of stress are weak (Protopopova, 2016).

Physical health metrics that have been utilized in other species such as farm animals have also been used in dogs (e.g. body condition in Part et al., 2014). However, measures such as reproductive performance may not be applicable in domestic dogs because many are spayed for a variety of reasons (King et al., 2012). In contrast, reproductive performance, for instance number of young produced, provides a vital metric of welfare in breeding dogs given the purpose for which they are raised.

Cognitive measures of welfare have also been used to assess the affective states of dogs. Though not as common as the traditional metrics of welfare, cognitive measures are slowly becoming more integrated and accepted as viable indicators of welfare, with many recent studies focusing on judgment biases (Bateson and Nettle, 2015; Mendl et al., 2009; Roelofs et al., 2016). Mendl et al. (2010) found that dogs exhibiting more separation-related behaviors (i.e. vocalizing, scratching the door, destructive behavior and toileting) also showed a more negative judgment bias characterized by higher latencies to approach a food bowl placed in an ambiguous location, thus suggesting a negative affective state. However, despite the benefits of incorporating cognitive measures into the emotional assessment of canine welfare, feasibility and application remain limiting factors in many studies, as cognitive bias tests involve substantial training beforehand which may be time-consuming or impractical in particular contexts such as shelters or other large-scale facilities (Wells et al., 2017).

In conclusion, different measures have been used to evaluate the welfare of domestic dogs in various environments such as shelters, laboratories, and training facilities among others. Such measures therefore provide a framework from which to develop assessment tools that can be used to evaluate the welfare of dogs kept in commercial breeding facilities.

2.5 Welfare Concerns of Breeding Dogs

Assessment of welfare in shelter, laboratory and working dogs provides an important avenue into understanding the welfare status of commercial breeding dogs, as both populations of dogs experience challenges associated with the impact of kenneling. Breeding dogs, however, also face additional challenges that distinguish them from other populations of dogs. Factors impacting the welfare of breeding dogs are therefore discussed below.

2.5.1 Housing and Management

Welfare concerns surrounding the housing and management conditions of shelter and other kenneled dogs also extend to dogs kept in commercial breeding facilities. Previous studies have revealed several factors that impact the welfare of kenneled dogs, and which may also affect the welfare of dogs used for commercial breeding. These include; noise levels, limited human interaction and barren environments.

Dogs kept in kennels often bark or vocalize for a number of reasons e.g. due to the presence of humans during husbandry procedures or due to external noises (Sales et al., 1997). However, loud barking (e.g. over 100dB) has been suggested to be stressful for other dogs and may even cause hearing damage that could potentially compromise the welfare of dogs (Scheifele et al., 2012). In breeding dogs, this may be of concern because dogs are kept in kenneled environments for a very long time and may therefore suffer negative consequences associated with prolonged exposure to high noise levels. Noise levels may also cause additional stress to pregnant or nursing bitches which may impact the welfare of their puppies as well. Breeding dogs may also face welfare issues as a result of having limited access to human interaction. Some studies done on other kenneled dogs have shown that at least some dogs prefer human contact to conspecifics, suggesting that human interaction may be of great importance to their overall well-being (Shiverdecker et al., 2013; Tuber et al., 1996, 1999; Wells, 2004). However, due to the higher dog to caretaker ratio in commercial breeding kennels, some individual dogs may not receive adequate social interactions with humans. It is possible that limited human interaction may lead to frustration in breeding dogs in commercial kennels due to inability to receive solicited attention and/or stimulation, which could potentially impact their welfare in the long-term. Aside from social contact, another key factor that is of concern in commercial breeding and other kenneled dogs is the level of mental stimulation provided. Many kennel environments provide dogs with only limited opportunity to engage in species-typical behaviors such as play or activity. Yet, previous studies have shown that there can be positive effects associated with providing enrichments in the form of exercise (e.g. Menor-Campos et al., 2011) and toys (e.g. Wells, 2004), or via physical enhancements such as increased kennel complexity (Hubrecht, 1993, 2002). As not much is known about the level of mental stimulation provided in commercial breeding facilities, investigating and addressing these concerns would be important in order to ascertain ways of improving the welfare of these dogs. Besides housing environments, breeding dogs also face additional concerns associated with genetics and socialization.

2.5.2 Genetics

Central to the debate surrounding commercial breeding are concerns about the health and behavioral welfare of dogs bred and raised in these establishments. Across the human and animal literature, it is well known that development of behavior is influenced by the interaction of both genetic and environmental factors (Grandin and Deesing, 2014; Hirsch, 1963). Many studies on canine behavior have indeed shown evidence of a genetic component for several behavioral traits such as aggression (e.g. van der Borg et al., 2003), fearfulness (e.g. Goddard and Beilharz, 1984), human aversion (e.g. Brown et al., 1978) and human-directed contact seeking (e.g. Persson et al., 2015) among others. Furthermore, the propensity for some behavioral outcomes such as activity, nervousness, trainability and friendliness across different breeds (Ruefenacht et al., 2002) has been shown to be partly heritable (Hradecká et al., 2015; Wilsson, 2016). For this reason, it is now commonly accepted that selective breeding of dogs based on temperament provides an effective way of influencing behavior (Wauthier and Williams, 2018). In a study carried out to assess the efficacy of selective breeding in Dutch Rottweilers, exclusion of fearful and aggressive Rottweilers from breeding (as determined by the Socially Acceptable Behavior [SAB] test (Planta and De Meester, 2007), resulted in a reduction in reported fear and aggression within the Dutch Rottweiler pedigrees compared to those not included in the breeding policy (van der Borg et al., 2017), thus suggesting the importance and benefit of carefully selecting breeding pairs. Likewise, according to Wauthier and Williams (2018), careful screening and genetic testing can also significantly reduce the possibility of inherited illnesses, health problems or defects typically found in many purebred dogs (e.g. Asher et al., 2009; McGreevy and Nicholas, 1999), by preventing the reproduction of unhealthy or predisposed individuals. In commercial breeding facilities within the United States, such industry standards for genetic and behavioral testing are not officially required, creating an area of concern about the unaddressed behavioral and medical issues likely present in dogs still residing in these facilities (McMillan, 2017) and in dogs acquired from these facilities (e.g. McMillan et al., 2011, 2013). Researching and addressing these concerns would be important because such issues may persist beyond the breeding facility and into the homes of new owners, increasing the likelihood of such dogs being relinquished to shelters or eventually euthanized.

2.5.3 Socialization

Beyond the effects of genetics, early environment and experiences in dogs can influence how dogs behave and respond to the world around them. Socialization refers to the process of gradually exposing a dog or puppy to its environment, i.e. to new experiences, objects, sounds, textures, humans and other animals, in a safe and positive manner (Boxall et al., 2004; Horwitz, 1999). In breeding dogs as in any other population of domestic dogs, this process should begin early in a dog's life, particularly during the sensitive periods of a dog's development (Scott, 1962; Serpell and Jagoe, 1995). Briefly, early development of the dog can be divided into three major periods; the primary period, the socialization period and the juvenile period (Battaglia, 2009; Howell et al., 2015; Scott, 1958, 1962). During the primary period, beginning from birth until roughly three weeks of age, puppies are helpless and dependent on maternal care due to limited motor and sensory capabilities, relying mostly on their sense of touch (Fox, 1971). The period following after, termed the socialization period, (three - 12 weeks), has been described as a "critical" time for forming attachments or "primary social relationships" (Scott, 1958). During this phase, mothers begin the process of separating from their pups, encouraging stronger littermate interaction. This is also the period in which puppies are most vulnerable to social influences such as human interaction and traumatic events that may impact their later development (Scott, 1963). Pierantoni et al. (2011), for instance, found that puppies separated from their mother and littermates too early (i.e. 30-40 days) reportedly showed increased prevalence for undesirable behaviors such as fearfulness during walks, attention-seeking, reactivity to noises and excessive barking. Finally, the juvenile period, which lasts until the onset of sexual maturity, is associated with further growth and development of pups, as well as increasing independence from maternal care (Scott, 1963).

Lack of appropriate socialization particularly during the socialization period has been linked with the development of behavioral concerns such as fearfulness, separation anxiety, destructive behavior and stranger-directed aggression in dogs (Denenberg and Landsberg, 2008). This has important implications for both puppies adopted out of commercial breeding kennels, and breeding dogs intended to be rehomed after their breeding careers. Problem behaviors such as excessive fear identified in dogs acquired from pet stores and in former breeding dogs may be attributed to inadequate socialization during puppyhood (McMillan et al., 2011, 2013); rehomed dogs that have never been exposed to experiences and situations typical of urban or suburban life (i.e. meeting strangers, walking on a busy street, being petted and hugged) may have trouble adjusting into such environments due to lack of adequate stimuli exposure and human contact (McMillan, 2017). Challenges, however, exist that may limit opportunities for socialization in commercial breeding kennels. Exposure to humans, though necessary, may be restricted due to biosecurity measures intended to protect the health of the dogs, as would be the case in many biomedical kennels. Likewise, depending on the type or location of the facility, availability of different people, objects or environments may be scarce (Wauthier and Williams, 2018). Lastly, because of the high number of dogs kept in commercial breeding facilities, time taken to adequately assure proper socialization of each dog may be a limiting factor, along with the knowledge level of caretakers concerning proper socialization practices (Howell et al., 2015). Such factors necessitate more research to be carried out in this population in order to ensure well-adjusted puppies and dogs.

In summary, investigation of welfare in commercial breeding dogs is important in order to address dog quality of life and the public concerns surrounding the breeding industry with respect to genetics, health and behavior, as well as to determine areas that may require improvement based on science. Central to this goal is the need for reliable and valid assessment tools that can be objectively used to assess the welfare of dogs on-site in such environments.

CHAPTER 3. REFINING ON-SITE CANINE WELFARE ASSESSMENT TOOLS

As accentuated throughout the literature, welfare is a multifaceted paradigm that integrates both subjective (e.g. behavior/emotional state) and objective (e.g. physical health) aspects of an animal's quality of life (Smulder et al., 2006). Accurately measuring and monitoring animal welfare is therefore a key factor in the development and refinement of welfare assessment tools. One of the challenges facing assessment of welfare in the field is deciding which measures of welfare are both practical and efficient. For instance, while physiological measures may provide useful information regarding an animal's response to stressful situations, collecting saliva or urine samples from every animal may not be feasible in the field due to limited time, sampling expertise and financial resources. Choosing measures that provide important insights into an animal's welfare within a realistic and appropriate time frame is thus important in the development of on-site assessment tools. Besides assessing indicators of well-being, on-site welfare assessment is also useful in examining the influences of resources such as diet, management practices and housing as well as identifying welfare risk factors – all of which can play an active role in improving the welfare of animals (Main et al., 2003). Currently, different assessment tools have been established for farm species. The Welfare Quality protocol, which has been considered highly reliable and valid, was developed in Europe with the aim of objectively assessing on-farm welfare of cattle, pigs and poultry based on four main principles of animal welfare; good feeding, good housing, good health and appropriate behavior (Canali and Keeling, 2009; Temple et al., 2011). Other assessment systems have been developed for use on farm animals, for instance to assess lameness (e.g. Manson and Leaver, 1988; Sprecher et al., 1997), pain (e.g. Viscardi et al., 2017), and body condition (e.g. Edmonson et al., 1989). While animal-based physical measures such as lameness, cleanliness and body condition can generally be objectively scored with strong agreement between different observers (Fraser, 2003; De Rosa et al., 2015; German et al., 2006), assessment of behavior still faces criticism from many in the scientific field, particularly because it can be anthropomorphic and thus, more subjective (Burghardt, 2003; Knierim and Winckler, 2009). Nonetheless, behavioral assessment is important and should be factored within any welfare assessment tool to adequately capture how an animal is coping.

Several canine assessment tools currently exist. The Shelter Quality protocol, for example, was established to provide a valid, reliable and practical means for assessing dog welfare in long-term shelters (Barnard et al., 2016a). This protocol shares the same welfare principles of appropriate behavior, good health and physical condition as the Welfare Quality protocol, but with a focus on canine welfare. A key advantage to this tool for use in kenneled dogs is that it focuses on measures that can be assessed on an individual level as well as on a group and facility level. The individual assessment category, for instance, is comprised of measures of body condition, cleanliness, skin condition, lameness, coughing and reactions towards humans, all of which can provide insight into an individual dog's welfare state at a given time (Barnard et al., 2016a). Likewise, behavioral assessment tests have also been developed to assess dogs' suitability for different purposes such as various types of work (e.g. Goddard and Beilharz, 1986; Knol et al.,1988; Serpell and Hsu, 2001; Sinn et al., 2010; Slabbert and Odendaal, 1999) or to identify potential behavioral concerns such as aggression and fear (e.g. Duffy et al., 2014; King, 2003; Netto and Planta, 1997; van der Borg et al., 1991, 2010). For example, the behavioral assessment for rehoming K9's (B.A.R.K) protocol was developed in Australia to accurately evaluate the behavioral appropriateness of shelter dogs as companion animals (Mornement et al., 2014). While many of these behavioral assessment tools have been developed for shelter and working dogs, the purposes for which they were designed greatly limit their use in the field. For instance, many behavioral assessment tools were developed to predict future behavior by evaluating the consistency of dog behavior across different contexts, thus serving as screening tools and not *welfare* assessment tools. Likewise, the Shelter Quality protocol, which was designed to assess welfare in shelter dogs, is limited in its use in the field due to its lengthiness and level of detail, making it difficult to be used by lay people on-site. The newly developed Field Instantaneous Dog Observation [FIDO] tool was thus established to practically and easily assess the behavioral and physical welfare status of commercial breeding dogs living in large-scale establishments (for more details, see Bauer et al., 2017). Specifically, it was developed with the aim of providing a relatively quick and accurate assessment of dog welfare at a given time through capturing immediately observable measures (Bauer et al., 2017). However, as noted by the authors, additional refinements and validation of the tool are necessary.

Within the literature on behavioral testing in dogs (Diederich and Giffroy, 2006), a common theme that plagues a majority of tests is insufficient reliability and validity (Taylor and Mills, 2006), which creates variation in how they are conducted and applied (King et al., 2012). For any assessment tool to be considered objective and successful, it must be both consistent (reliable) and accurate (valid) in assessing the measures for which it was designed (King et al., 2012). Therefore, the first step in developing and refining an objective tool is examining its reliability, particularly when it pertains to behavior (Knierim and Winckler, 2009).

3.1 Reliability

Reliability concerns the degree to which measurements are both repeatable and consistent (Carmines and Zeller, 1979; Martin and Bateson, 1993), and is typically a prerequisite for validity (Jones and Gosling, 2005). There are different ways to assess reliability depending on the type of measurement and purpose of the tool. Intra-rater reliability is one type of reliability that assesses the consistency of reports by a single observer (Martin and Bateson, 1993). In a review carried out by Taylor and Mills (2006), it was established that only a few behavioral assessment studies on dogs explicitly reported on intra-rater reliability. For instance, Murphy (1995) reported reasonable intra-rater reliability (61.5%) after evaluating for consistency within 11 trainers reassessing the suitability of a potential guide dog; time interval between ratings ranged from one week to six months. Likewise, Valsecchi et al. (2011) reported moderate to high levels of intra-rater reliability correlations (0.42 - 0.97) of two raters when assessing the reliability of a temperament test after one month for shelter dogs. Lack of or poor intra-rater reliability is particularly concerning because it not only affects interpretations of the variables being measured, but also influences the repeatability of the test especially if one observer is used (Diesel et al., 2008; Taylor and Mills, 2006). For example, Mornement et al. (2014) reported a generally weak test-retest reliability of the behavioral assessment (B.A.R.K) protocol after only 24 hours, a finding that can be partially explained by their failure to first control for changes in the observer via assessment of intra-rater reliability.

Extending from the evaluation of intra-rater reliability is the critical assessment of inter-rater reliability. Inter-rater reliability measures how well different observers will assess the same individual or animal (Taylor and Mills, 2006). This is particularly important as welfare assessment tools are usually developed to be used interchangeably by different people or entities;

therefore, agreement and consistency in reports between individuals are fundamental. A few behavioral assessment studies on dogs (e.g. Diesel et al., 2008; Mornement et al., 2014; Sherman et al., 2015; Sinn et al., 2010; Valsecchi et al., 2011), albeit not many, have reported inter-rater reliability ranging from moderate to high. Though encouraging, more studies ought to be explicit in the assessment and reporting of reliability given its importance in the development of an objective assessment tool (Jones and Gosling, 2005).

Lastly, another important form of assessing reliability is test-retest reliability. This assessment entails repeating the same test with the same subject after a period of time has passed, after which a comparison of the measures being tested is made (Diederich and Giffroy, 2006; Harvey et al., 2016). Because test-retest reliability describes the extent to which a dog's behavioral scores remain stable over time, consistency of behavior is vital if any predictions are to be made about behavioral outcomes in similar situations (De Meester et al., 2011; Taylor and Mills, 2006). For instance, after assessing 112 dogs on their level of aggression to different stimuli, Netto and Planta (1997) retested 37 dogs after six months and found a substantial mean correlation (0.77) of behavioral measures (i.e. snapping and biting/attack) between test and re-test, suggesting that the test was reliable. Klausz et al. (2014) also retested 19 out of 73 dogs after one year to assess the consistency of aggression and fear and found no significant differences between variables measured in the test and retest, though no actual reliability index was reported to sufficiently say that the test was reliable. As concluded in a review by Brady et al. (2018), many studies on dogs still fail to report on the reliability of tests reflecting a need for more standardization and objectivity.

3.2 Validity

Another crucial factor that is important to consider, and a necessary step when developing and refining behavioral assessment tools, is evaluation of their validity. Validity describes the extent to which a test is accurate and meaningful (Martin and Bateson, 1993). Different forms of validity exist. Construct validity, for instance, describes the extent to which items within a test measure what they were designed to measure (Taylor and Mills, 2006). In many canine behavioral assessment studies, this form of validity is often understood best in relation to both convergent validity, i.e. when related measures within a test show high correlation, and discriminant validity, i.e. when unrelated measures do not correlate with each other (Taylor and Mills, 2006). Criterion validity, in contrast, describes an association between results of a test and an external and independent measure of behavior (Barnard et al., 2012). For many of the behavioral assessment tools developed for shelter and working dogs, criterion validity has often been assessed by comparing test results with i) owner reports of dogs' behavior (e.g. Christensen et al., 2007; Marder et al., 2013; Stephen and Ledger, 2007; van der Borg et al., 1991), ii) training or adoption outcomes i.e. successful or not (e.g. Asher et al., 2013; Duffy and Serpell, 2012; Duffy et al., 2014; Harvey et al., 2016; Slabbert and Odendaal, 1999; Sinn et al., 2010), iii) expert ratings of dogs' behavior (e.g. Valsecchi et al., 2011), iv) previous behavioral history (e.g. Bollen and Horowitz, 2008; Kroll et al., 2004; Netto and Planta, 1997; van der Borg et al., 2010), and v) dogs' behavior in a different context/environment (e.g. Barnard et al., 2016b; Döring et al., 2017a,b) as a means of predicting behavior. While some studies have shown some success in predicting behavior (e.g. Asher et al., 2013; Marder et al., 2013; Valsecchi et al., 2011), others have reported poor predictive validity (e.g. Christensen et al., 2007; Mornement et al., 2014, 2015). For instance, Christensen et al. (2007) reported that 40.9% of dogs that had successfully passed a modified version of the Assess-A-Pet (Sternberg, 2004) test in the shelter exhibited lunging, growling, snapping and/or biting after adoption, suggesting that the test failed to accurately identify certain types of aggression in dogs while in the shelter. Likewise, amongst the dogs that had initially passed the behavioral assessment for the rehoming of K9's (B.A.R.K) protocol in the shelter and subsequently rehomed, the predictive validity for all measures except fear and friendliness was poor (Mornement et al., 2014). Such studies highlight the need to adequately and rigorously assess the validity of tests, especially if such tests are to be used as the basis for making decisions such as whether to euthanize or not based on behavioral outcomes. This need is further evidenced by the fact that many shelters do not use standardized protocols to evaluate dogs, but instead use invalidated behavioral tests with varying levels of accuracy to determine the suitability of dogs for adoption (Mornement et al., 2010). Extending from this, validation of the FIDO tool is a necessary step in order to determine how useful and accurate it is in the field.

3.3 Thesis Objectives

As indicated by the review of the literature, challenges and concerns exist relating to the reliability and validity of behavioral assessment tools currently in use for dogs. There also remains a need for reliable, valid, objective and practical tools to be utilized in commercial breeding facilities. The main goal of this project, therefore, was to assess the reliability and validity of the newly developed FIDO tool. Physical measures contained within the FIDO tool include measures of body condition, body cleanliness, and physical health ranging from nasal and ocular discharge to lameness and lesions (for further details, see Hurt, 2016). The behavioral portion assesses a dog's response to approach by a stranger as a proxy measure of socialization towards unfamiliar people (Bauer et al., 2017). In a recent study conducted to evaluate the reliability of the FIDO tool, results revealed near perfect agreement between experts and between novice raters on the physical metrics of the dogs evaluated (Bauer et al., 2017), further supporting the idea that physical measures can be reliably and objectively evaluated. Behavior, on the other hand, which has shown to represent an important indicator of welfare, particularly in dogs, and which forms a critical part of many assessment tools (Wemelsfelder, 2007), faces many challenges with regard to objective assessment (Meagher, 2009). In light of these concerns, the present project sought to exclusively investigate the reliability and validity of the behavioral scoring system of the FIDO tool. With regard to reliability, the first study aimed to assess the intra-rater reliability and inter-rater reliability of the behavior scoring when used by novices with training. Furthermore, in line with the aim of developing a practical tool that is field-ready and efficient, the second study aimed to assess a minimal reliable duration for which observations could be made and behaviors objectively evaluated. With respect to the third study on validity, the main goal was to investigate whether the behavioral scoring system of the FIDO tool had predictive power in a standardized field test evaluating dogs' responses to a stranger.

CHAPTER 4. EVALUATING THE RELIABILITY OF A NEWLY DEVELOPED BEHAVIORAL SCORING SYSTEM (FIDO SCORING) FOR USE IN COMMERCIAL BREEDING FACILITIES

Introduction

Despite the high level of public demand for purebred dogs driving the commercial breeding industry, little is known about the welfare of dogs kept in such large-scale facilities. Furthermore, until recently no tools that could aid in the assessment of welfare in this population of dogs had been established to be used on-site. The Field Instantaneous Dog Observation [FIDO] tool was thus developed to provide a practical yet efficient means of assessing welfare in commercial breeding dogs using overt behavioral and physical measures (Bauer et al., 2017). Behavioral and welfare assessment tools have also been developed for shelter and working dogs (e.g. Barnard et al., 2016a; Diederich and Giffroy, 2006; Jones and Gosling, 2005; Taylor and Mills, 2006), however, constraints of time and applicability greatly limit their use in commercial breeding facilities. For instance, a majority of behavioral tests developed often aim to predict future behavior by assessing temperament, i.e. consistency of behavior in dogs across different situations. While genetic influences, often captured via temperament and personality assessments, play a key role in the development of individual characteristics of behavior, effects of early experience and environment are equally important (Scott and Fuller, 2012). The behavioral measures incorporated in the FIDO tool were therefore developed to specifically capture aspects of environmental influences such as socialization towards people via responses of dogs to a stranger's approach using a scoring system. Behavioral responses of dogs not only provide insight into the reactivity of dogs at a given point in time (i.e. responses to a potentially mild stressor), but also enable quick identification of areas of concern such as fear of strangers that may need further evaluation and intervention, all of which provide a reference point for understanding important facets of welfare in dogs kept in commercial breeding facilities. Nonetheless, as with many other behavioral assessment tools already developed, concerns surrounding the objective assessment of behavior still persist due to effects of observer subjectivity (Brady et al., 2018; Diederich and Giffroy, 2006; King et al., 2012). The purpose of this study was to therefore evaluate the reliability of the behavioral portion of the FIDO tool. Reliability concerns the extent to which measurements are consistent and repeatable and involves assessing consistency of reports within (intra-rater reliability) and between (inter-rater reliability) assessors (Martin and Bateson, 1993). Evaluation of intra- and inter-rater reliability is a key aspect in observational studies of animal behavior as well as in assessment tools (Kaufman and Rosenthal, 2009). In a pilot study conducted by the developers of the new FIDO tool, results of the inter-rater reliability between two trained novice raters on behavioral responses showed substantial agreement (kappa = 0.74), reflecting good reliability and a step in the right direction with regard to the refinement of a reliable behavioral assessment tool (Bauer et al., 2017). Extending from the pilot study, the first study of this project aimed to further investigate the intra- and inter-rater reliability of the scoring system of the FIDO tool when used by two novice raters with more extensive training on the tool. Because the focus of this study was on how well two raters could agree with themselves and with each other while scoring dogs, rather than on how well the tool applied in commercial breeding kennels, shelter dogs, which were more conveniently located at a local humane society than participating breeding kennels were used to evaluate the reliability of the scoring. This also provided an opportunity to assess the tool's applicability in different types of kennels despite the fact that it was developed explicitly to meet the needs of commercial breeding kennels.

Study one: investigating reliability of the FIDO scoring

4.1 Ethics Statement

The procedures described were reviewed and approved by the Purdue University Institutional Animal Care and Use Committee, and permission to visit shelters was granted by shelter management prior to the commencement of the study.

4.2 Materials and Methods

4.2.1 The FIDO tool - Scoring system

Within the FIDO tool, behavioral responses of a dog to stranger-approach are classified into three categories: red, yellow or green (RYG). Responses classified as red include fearful behaviors, characterized by fleeing or freezing; stereotypic behaviors; and offensively aggressive behaviors (that may also be rooted in fear) characterized by hard or forward body language. Responses classified as green include affiliative behaviors such as approaching or soliciting attention from the unfamiliar person or continuing to engage in a behavior despite the presence of the stranger (neutral response). Finally, responses classified as yellow include those that are neither clearly red nor clearly green, or which show ambivalent body language such as approach and avoid (Bauer et al., 2017).

4.2.2 Subjects

Behavioral assessment was conducted on 50 adult dogs (ages 1-9 years old) of mixed breed, sex and reproductive status housed singly at two US shelters in Indiana; one in Lafayette (shelter A: n = 21) and one in Indianapolis (shelter B: n = 29). Responses to approach were independently collected by two novice raters; rater one: female (25 years old) and rater two:

male (22 years old) trained by an expert on FIDO prior to the beginning of the study. Rater two was used as the gold standard for this study due to being most similar to expert scoring in practice tests.

4.2.3 Study Procedures

Data collection at shelter A was carried out on May 24th, 2017 between 10:30am and 11:30am before the shelter officially opened to the public in order to minimize potential interruptions during scoring. Each rater was similarly dressed in neutral colored clothes (black top/t-shirt and dark blue/ black pants) and refrained from wearing any perfume or jewelry to minimize overt variation in appearance. Prior to the actual data collection, a practice run was carried out on eight puppies housed in a separate room, whose scores were not included in the final sample. Each rater approached the kennel of each dog in a non-threatening manner, stood with their side roughly 0.5 m away from the front of the kennel, and oriented towards the dog while avoiding direct eye contact that could be considered threatening (van der Borg et al., 1991). Each rater then bent slightly at the knees and extended a hand toward the dog for five seconds before scoring the dog's response as either red, yellow or green. Every approach was also video-recorded using a GoPro camera attached to the rater's forehead during the entire process. The order of approach and scoring was alternated for each rater, such that one rater did not immediately score a row of dogs that had just been evaluated by the other. This was carried out to give each row of dogs time (approximately five to seven minutes) to settle down before being observed by the next rater, and to minimize order effects. At the end of the trip, a total of 31 dogs had been scored, but it was later determined that there was inconsistent approach by the raters for 16 dogs. Thus, data from those dogs were discarded and a second trip was made on June 2nd, 2017 to carry out scoring on another 16 dogs. As is typical of shelters, however, dogs that were to be scored by both raters were shifted in between raters' scoring. Therefore, only six dogs were consistently scored by both raters and utilized for the study, reducing the total sample size for shelter A to 21 dogs. Data collection at shelter B was carried out a week later on June 9th, 2017 between 11:15am and 12:30pm before the shelter officially opened to the public. Raters dressed similarly (dark colored clothes) and followed the same procedures as were used in shelter A. A total of 29 dogs were scored from shelter B.

4.2.4 Video Recordings

Responses to approach were evaluated again from video recordings a month later to allow sufficient time to pass before reassessment. Each rater rescored the behavioral responses (red, yellow or green) of the dogs from both shelters off their own videos first, followed by each other's videos. One video for rater one was discarded because of a missing live score, reducing the total number of videos for rater one to 49 instead of 50. All 50 videos for rater two were, however, utilized for analysis.

4.2.5 Data Analysis

Intra-rater reliability

As highlighted in a review by Taylor and Mills (2006), evaluation of intra-rater reliability is a crucial aspect of behavior assessment in dogs. Following a recommendation by Martin and Bateson (1993), intra-rater reliability was assessed using video recordings in order to control for behavioral changes of the dogs rather than the raters. Each rater's scores (red, yellow or green) from their live observations were compared to their scores of the same dogs from video recordings, and a Cohen's weighted kappa was calculated (Landis and Koch, 1977; Mielke et al., 2009). Cohen's weighted kappa was utilized over percent agreement as a measure of intra-rater reliability because it assesses the extent to which agreement within or between raters would be better than chance (Hoehler, 2000), making it a more stringent statistic. The weighted kappa statistic was calculated using Statistical Analysis System, Version 9.4 (SAS Institute Inc., Cary, NC). Based on Landis and Koch ratings (1977), kappa statistics of “less than 0.00 indicate poor agreement, 0.00-0.20 indicate slight agreement, 0.21-0.40 indicate fair agreement, 0.41-0.60 indicate moderate agreement, 0.61-0.80 indicate substantial agreement, and 0.81-1.00 indicate almost perfect agreement”. A minimum kappa value of 0.7 has, however, been suggested to be appropriate (Kiddie and Collins, 2014).

Inter-rater reliability

Similar to Valsecchi et al. (2011), inter-rater reliability between the two novice raters was also assessed using video recordings. Each rater scored the behavioral responses of dogs across both shelters from the other rater’s videos, and these scores were then compared to each rater’s own individual video scores for the same dogs. Cohen’s weighted kappa was also calculated using SAS and interpreted in the same way as outlined in the previous section. For both forms of reliability, prevalence indices [PI] (a measure of homogeneity of response) were not calculated because the present study utilized ordinal scores, of which PIs for weighted kappa have not been developed (Burn and Weir, 2011).

4.3 Results

As intra-rater reliability provides insight on the tool itself rather than the subjects being assessed, raters’ scores were aggregated across the two shelters. Results indicated that there was almost perfect agreement (rater one: kappa = 0.83; rater two: kappa = 0.89) between each rater’s live scores and their scores of the same dogs’ responses from video recordings (Table 1). Inter-rater reliability, on the other hand, was calculated based on how well each rater's own scores from video agreed with the other rater's scores for the same video. Results indicated almost perfect agreement between raters based on videos (rater one videos: kappa = 0.86; rater two videos: kappa = 0.88), and an overall almost perfect agreement (kappa = 0.87) across all videos. Table 2 displays percent agreement and kappa values for inter-rater reliability for both raters, followed by an average inter-rater reliability weighted kappa statistic (Table 3). When compared against each other, raters' scores of each other's videos were relatively consistent, suggesting that discrepancies were few (Figure 1).

Table 1. Intra-rater reliability: Percent agreement and weighted kappa with 95% confidence intervals for the behavioral assessment of shelter dogs based on live and video recordings.

Observer	% Agreement	Kappa (95% C.I)
Rater one (female)	90%	0.83 (.68, .97)
Rater two (male)	94%	0.89 (.77, 1.00)

Table 2. Inter-rater reliability: Percent agreement and weighted kappa with 95% confidence intervals for the behavioral assessment of shelter dogs based on video recordings.

	% Agreement	Kappa (95 % C.I)
Rater 1 videos	92%	0.86 (0.72, .99)
Rater 2 videos	94%	0.88 (0.75, 1.00)

Table 3. Averaged inter-rater reliability: Percent agreement and weighted kappa with 95% confidence intervals for the behavioral assessment of shelter dogs based on video recordings.

Observer	% Agreement	Kappa (95 % C.I)
Rater one and Rater two	93%	0.87 (0.78, 0.96)

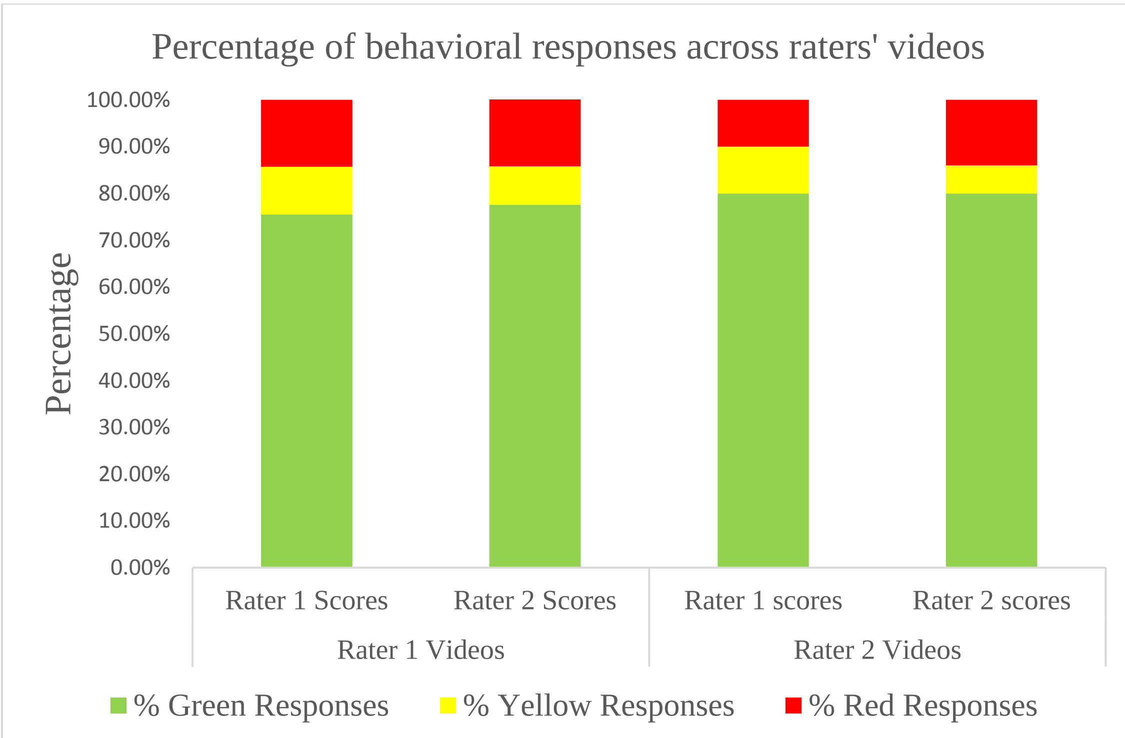

Figure 1. Percentage of red, yellow and green responses across raters based on video recordings.

Study two: investigating reliability of the assessment period

Previous research assessing the behavioral reactions of dogs to stranger-approach have utilized different assessment periods to capture dogs' behavioral responses. Some studies have suggested the immediate response (e.g. Slabbert and Odendaal, 1999) to be most informative of a dog's reaction, while others have suggested that observation periods lasting five seconds (e.g. Bollen and Horowitz, 2008), 10 seconds (e.g. van der Borg et al., 1991), 30 seconds (e.g. Barnard et al., 2016a), or even two minutes (e.g. Kiddie and Collins, 2014) can be sufficient to gain insight into dogs' responses to a stranger's approach. Until now, however, no studies have explicitly been carried out to determine whether changes in a dog's behavioral reaction to a stranger occur as a function of assessment duration. Because dog behavior can be variable between and particularly within dogs, it is important that behavioral assessment tools factor in the reliability of the duration of the test (De Meester et al., 2011). Given that one of the intentions of the FIDO tool is to provide an objective, but practical, and field-ready means of assessing the welfare of commercial breeding dogs, the second study aimed to investigate whether, and to what extent, dogs' behavioral responses to stranger-approach within their home pens changed during a 30-second observation period. Commercial breeding kennels were utilized for this portion of the study in order to gather insight into this population of dogs, and because it allowed for the assessment of multiple facilities.

4.4 Ethics Statement

Owners of commercial breeding facilities who participated in this study provided informed consent to the testing procedures and were informed of their freedom to withdraw from the study at any point without penalty. All experimental procedures were approved by the Purdue University Institutional Review Board and Institutional Animal Care and Use Committee.

4.5 Materials and Methods

4.5.1 The FIDO Tool - Scoring system

The behavioral component of the FIDO tool was also utilized for this portion of the study as previously outlined.

4.5.2 Subjects

Four USDA-licensed commercial breeding facilities in Arthur, Illinois were recruited for participation in the study carried out in October 2017. After factoring in dynamics such as availability and location of facilities of interest, total number of dogs per facility, and management practices per facility (i.e. provision of free access to outdoor runs), it was determined that 20 dogs per facility would be most feasible. Prior to actual data collection, pilot testing was carried out in July 2017 on a separate group of 26 dogs at two facilities in Odon, Indiana to establish the practicality of the study, but were otherwise not included in the final analysis due to changes made in the methodology of the study. All dogs that met the study criteria were randomly sampled at all four facilities. These included dogs that were over one year of age and generally healthy. To protect the welfare of pregnant bitches and puppies, bitches in the last two weeks of pregnancy and those nursing puppies were not included in the study. A total of 81 (59 females, 22 males) adult dogs of various breeds across all four facilities were therefore assessed on-site (Table 4). Dogs were group-housed (i.e. 3-4 dogs per pen) in facility one and pair-housed in facilities two, three and four. In all four facilities, dogs had continuous access to the indoor and outdoor portions of their home pens. Kennel sizes also varied by facility to meet or exceed the USDA requirements for dogs (USDA, 2017). Table 4 summarizes the demographics of dogs at each facility.

Table 4. Summary of number of dogs assessed, sex, age (mean ±SD) and number of dogs per breed for each facility.

Facility	Total dogs assessed	Sex	Mean Age ±SD (months)	Breed
1	23	F (18) M (5)	41.62±20.27	Poodle (1) French bulldog (4) Shiba Inu (13) Golden Retriever (5)
2	20	F (14) M(6)	34.11±21.83	Yorkshire Terrier (9) Pembroke Welsh Corgi (6) Shih Tzu (5)
3	19	F (14) M(5)	31.42±16.48	French Bulldog (5) Siberian Husky (11) Labrador Retriever (1) Pomsky (1) German Shepherd (1)
4	19	F (12) M(7)	26.63±15.65	French Bulldog (5) Boston Terrier (14)

4.5.3 Study Procedure

Data collection for all facilities was carried out within one day, beginning at 7:45am in facility one and ending at 4:00pm in facility four, with total testing time for each facility running 45 minutes. All four facilities were roughly 15-20 minutes apart from each other. Prior to the commencement of the study at each facility, the facility owner was given a briefing on what the testing procedures would entail, and thereafter given time to read and sign the consent form. A copy remained with them for their records. A practice run was also carried out at facility one on a group of puppies housed in a separate room before the actual data collection. At each facility, the primary observer (female) stood perpendicular to the front of the pen containing the selected dog, bent slightly at the knees and extended her hand to the dog before immediately recording the dog’s response as either red, yellow or green. A second person (female) standing adjacent to the primary observer and video-recording the approach then signaled to the primary observer each time five seconds had passed, after which the primary observer oriented toward the dog (without making direct eye contact) and recorded the immediate response (red, yellow or green) again. This process continued every five seconds until 30 seconds had passed. Because previous findings have shown a conflict arising between food motivation and fear (e.g. Jones and Boissy, 2011), a treat was offered at the end of the observation period (if the focal dog was still in the kennel) either directly to the dog through the pen bars, or placed directly in the feed dish, and a note was made on whether the dog took the treat or not. Lastly, as randomization was done on the dogs and not the pens, some pens contained more than one focal dog, and therefore scoring of each dog was done simultaneously in those particular pens.

4.5.4 Data analysis

Descriptive statistics were utilized to determine the percentage of dogs that changed in their behavioral response during the 30-second live observation period. A total of seven data points were established for each dog, with the first data point representing the immediate response (zero seconds) and the remaining data points representing each of the remaining consecutive five second observation periods. Due to the fact that some of the dogs were unavailable to be scored during the whole observation period because of moving in and out of their pens, a criteria was set such that only dogs with four data points or more (representing at least half the observation time) were included for analysis. Thus, out of a total of 81 dogs, only 56 dogs met the criteria for analysis.

4.6 Results

Results indicated that of the 56 dogs that met the criteria for sampling, 91% (n = 51) showed no change in their behavioral responses during the 30-second observation period. Across facilities, 72% (n = 40) consistently showed green responses throughout the 30-second assessment period, characterized mostly by solicitation of attention and friendly approaches, and occasionally by neutral responses (i.e. undisturbed behavior). The remaining dogs consistently showed red responses characteristic of fear (i.e. staying at the back of the kennel) [14% (n = 8)] and yellow responses characterized by approach and avoid behavior [5% (n = 3)] (Figure 2). When evaluated by facility, differences were also seen. In facility one, for instance, 53% (n = 9/17) consistently showed green responses while 41% (n = 7/17) consistently showed red responses. In facility two, 80% (n = 8/10) consistently showed green responses while 20% (n = 2/10) were divided equally between red and yellow responses. In facility three, 80% (n = 12/15) consistently showed green responses, and in facility four, 79% (n = 11/14) consistently showed green responses while 14% (n = 2/14) showed yellow responses (Figure 3). For the few dogs that showed changes in their behavioral responses [9% (n = 5)], changes occurred after five seconds in one dog (green to yellow - facility four), after 10 seconds in two dogs (red to yellow - facility one; green to yellow - facility three), after 20 seconds in one dog (red to yellow - facility three) and after 25 seconds in another dog (yellow to red - facility three). Lastly, all but one dog that showed fearful responses did not consume the treat offered at the end of test.

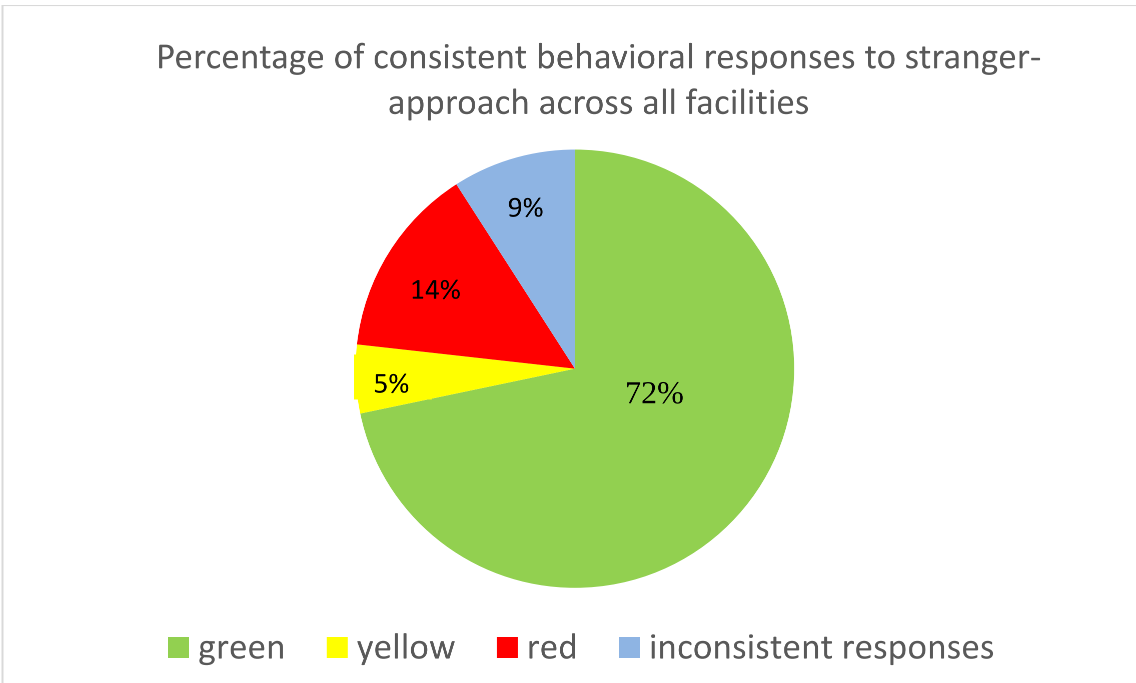

Figure 2. Percentage of dogs' behavioral responses (red, yellow and green) during a 30-second observation period across all facilities.

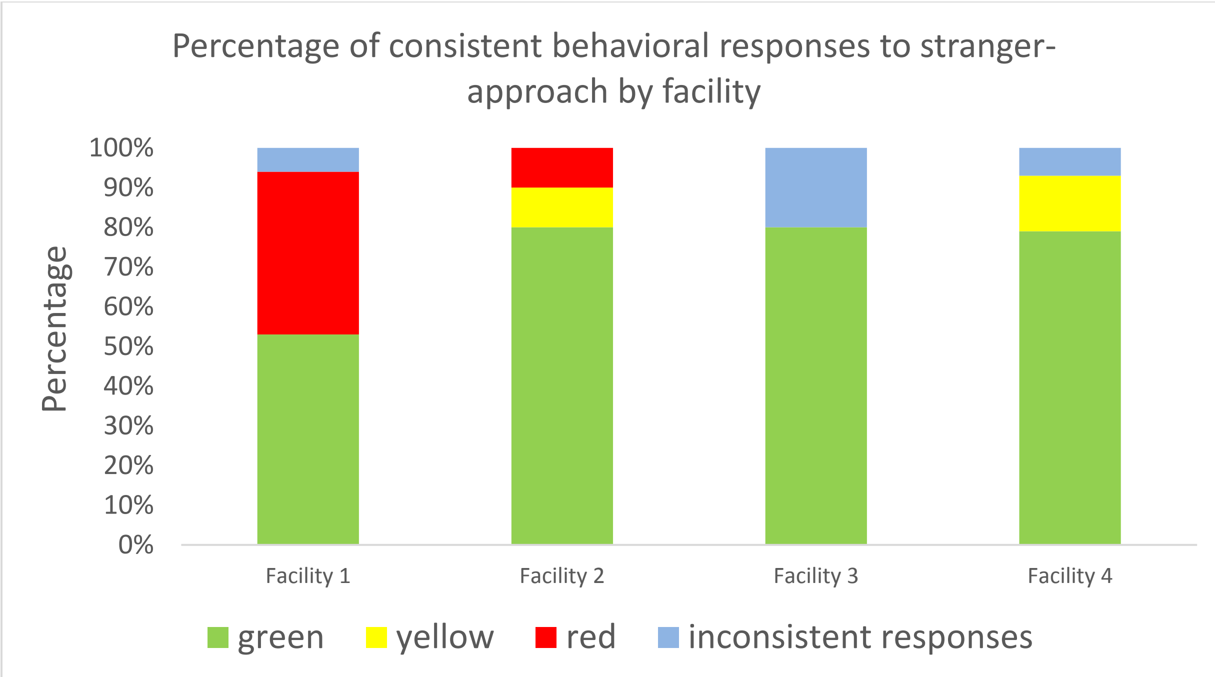

Figure 3. Percentage of dogs' behavioral responses (red, yellow and green) during a 30-second observation period by facility.

4.7 Discussion

The first goal of this study was to evaluate the intra- and inter-rater reliability of the newly developed FIDO tool when used by novice raters. Use of videos to assess both forms of reliability was employed for two reasons. Firstly, dogs were housed in shelters. Therefore, the likelihood of repeating the test on the same dogs on different occasions was unlikely to be feasible due to the possibility of the dogs being moved or rehomed in the interim between test and retest. Secondly, previous studies have indicated that a dog may react differently to different assessors. Wells and Hepper (1999), for instance, found that dogs showed more defensive-aggressive reactions (i.e. barking and maintaining eye contact) towards men than women, and Lore and Eisenberg (1986) showed that dogs more readily approached a woman sitting at the end of a cage than a man sitting in a similar manner. Video recordings thus provided a better avenue to assess agreement within and between raters as it eliminated the potential confound of effects of gender on the behavioral responses of dogs.

As indicated by the findings of the first study, novice raters showed almost perfect agreement both within themselves and with each other when assessing the behavioral responses of shelter dogs to a stranger's approach, further indicating the reliability of the FIDO scoring. This has important implications particularly with regard to the practicality of the FIDO tool for individuals without expertise in canine behavior. A high level of intra-rater reliability, for example, reflects more confidence in the assessment of behavior by an individual (Diesel et al., 2008), which is important considering that many on-site evaluations are undertaken by individual inspectors or investigators. Inter-rater reliability was also almost perfect between raters, further providing evidence that the tool can be reliably used by different observers, and supporting previous findings reported from the FIDO tool pilot testing that showed substantial agreement (kappa = 0.74) between novice raters (Bauer et al., 2017). Disagreement within and between raters occurred between the green and yellow categories, and the red and yellow categories, but not between the green and red categories. This could be attributed to the fact that the yellow category describes behaviors that are contradictory or ambivalent (i.e. barking while tail wagging), and hence interpretation of the term "ambivalent" may have been deemed as more difficult, reflecting a need for more training on this category. Valsecchi et al. (2011) also reported disagreement between raters on the sociability of dogs to humans due to the differential interpretations of "friendly" and "exuberant". Additionally, as the yellow category also captures behavioral responses that are not clearly green or red, raters may have been more inclined to score a yellow when in doubt. Such disagreement, however, occurred minimally and did not impact the overall reliability of the test, suggesting that the scoring system was effective in identifying reliable measures of behavioral responses to approach. For the second objective of the study, findings indicated that for a majority of the commercial breeding dogs, there was no change in the behavioral response to stranger-approach over a 30-second observation period. This suggests that there is no benefit to extending the observation time beyond five seconds and provides support that this is a reliable and practical time point from which to evaluate a dog's response to stranger-approach. For the few dogs that showed a change in their behavioral responses during different time points, individual factors such as genetics (Svartberg, 2006) may have influenced their responses. The finding that dogs exhibiting fearful responses also declined to take the treat further supports the notion that fearful animals are less likely to eat (Boissy, 1995; Rushen et al., 1999), and provides some validity of the tool in identifying fearful dogs. Comparable results of the conflict between food motivation and fear have also been found in species such as pigs (e.g. Dalmau et al., 2009), sheep (e.g. Romeyer and Bouissou, 1992) and hens (e.g. Jones et al., 1981). Interestingly, red responses were predominantly found in two of the four facilities. Variations between facilities can be attributed to the effect of breed or may reflect differences in the level of exposure or socialization of dogs to unfamiliar people, as well as other dog management factors. Further research is therefore needed to determine which factors may influence the behavioral responses of dogs to approach across different types of facilities. The majority of the behavioral responses were, however, classified as green across facilities which provides valuable insight into the general behavioral responses of these dogs in relation to unfamiliar people. Akin to the previous findings and comments noted by the developers of the FIDO tool (Bauer et al., 2017), this finding contradicts the notion that dogs believed to be from commercial breeding facilities are uniformly fearful of unfamiliar humans, as reflected in owner report studies (e.g. McMillan et al., 2011). However, discrepancies could have also been found because the test was conducted on-site within the dogs' pens rather than in a different environment such as in a home. Therefore, many dogs may not have shown (extremely) fearful responses simply because they were in a relatively familiar environment (Bauer et al., 2017). It should also be noted that the commercial breeding facilities that chose to participate in the study may not be representative of all such facilities in the United States, and that their practices may be skewed toward more positive outcomes for dogs as demonstrated by dogs' behavior. More research across multiple facilities in different states is therefore needed to validate these findings.

In both objectives, emphasis was placed on the tool and not on the characteristics of the dogs themselves. Therefore breed, sex and age differences were not accounted for. While many studies have reported an effect of gender and reproductive status (e.g. De Meester et al., 2008; Starling et al., 2013), age (e.g. Döring et al., 2016), and breed (e.g. King et al., 2003) on the behavioral responses of dogs to different social stimuli including humans, the current study did not prioritize these factors because the goals were to: 1) investigate how well two raters agreed on the general reactions of dogs to human approach irrespective of dogs' demographics, and 2) determine whether these responses changed as a function of observation time. The present study also only utilized an approach to the outside of the pen. While the behavioral responses of dogs (both shelter and breeding) to a stranger's approach to the pen provides some insight into their behavior at a particular point in time, little can be said about how such dogs would respond if the pen door were opened or if the dog was placed in a different environment such as their exercise yard. More research is therefore needed to verify how dogs might respond in different situations, such as in an opened pen or an outdoor area. Likewise, it would also be valuable to study how dogs might respond to different stimuli, for instance unfamiliar objects. This could provide insight into whether responses of fear are generalizable or specific and may also provide more understanding of the interactions between nature and nurture. Lastly, because the present study only utilized two raters for the first objective, and one rater for the second objective, more research may be needed to adequately test the reliability of behavioral responses across multiple raters.

Limitations notwithstanding, findings of the present study provide support that behavioral responses of dogs to stranger-approach can be reliably scored by different raters, and within a practical time frame. This has important implications with regard to its use by different evaluators in a field setting. More research is, however, needed to further refine and evaluate the validity of the tool in order to assess how accurately the tool captures the behavioral status of dogs.

CHAPTER 5. EVALUATING THE VALIDITY OF A NEWLY DEVELOPED BEHAVIORAL SCORING SYSTEM (FIDO SCORING) FOR USE IN COMMERCIAL BREEDING FACILITIES

Introduction

Behavioral testing in domestic dogs continues to be a growing area of focus within the canine literature. As such, the need to develop reliable, valid and practical tools remains an important quest for many researchers. While many behavioral assessment tools have been developed for use in shelter and working dog programs, no field-ready assessment tool had been previously developed to be used in commercial breeding facilities despite the level of public concern surrounding this population of dogs. Stemming from this paucity, the Field Instantaneous Dog Observation [FIDO] tool was established to quickly identify areas of concern via evaluation of behavioral and physical measures (Bauer et al., 2017).

Central to the behavioral measures incorporated in the FIDO tool is the assessment of dogs' behavioral responses to a stranger's approach. Tests assessing responses to strangers have been widely utilized in many behavioral studies as a means of gauging the fearfulness or sociability of dogs to unfamiliar humans (e.g. Barrera et al., 2010; Conley et al., 2014; Döring et al., 2014, 2016, 2017a; Lore and Eisenberg, 1986; Pullen et al., 2012), and are often included as subtests within temperament tests (e.g. Ethotest (Lucidi et al., 2005), Dog Mentality Assessment [DMA] (Svartberg, 2002), Emotional Reactivity Test [ERT] (Sherman et al., 2015) designed to determine the appropriateness of dogs for different types of work or as rehoming candidates. Fear towards people is often considered to be an undesirable emotional state that not only impacts the welfare of dogs via potential long-term activation of stress responses (Foyer et al., 2016; Jones and Boissy, 2011) or reduced lifespans (Dreschel, 2010), but also compromises the chances of successful adaptability in instances such as rehoming or in working programs (Goddard and Beilharz, 1984). Shin and Shin (2017), for instance, found that dogs with low levels of sociability during an exposure test to an unfamiliar human showed significantly higher levels of cortisol concentrations compared to dogs showing high sociability; comparable results have also been found in other species such as cattle (e.g. Breuer et al., 2003). Dogs that do not exhibit a fearful response or those that recover fairly quickly from a startle response to a novel stimulus are thus deemed as more stable, more stress resistant and suitable for potential roles such as companionship or work than those who respond with prolonged fear (Foyer et al., 2016).

Fear describes a response to a specific stimulus perceived as threatening (Forkman et al., 2007). This is somewhat distinguished from anxiety, which refers to the anticipation of a potential danger that is more non-specific (Adolphs, 2013). Fear may be triggered by environmental stimuli which are novel, intense (e.g. loud or big), evolutionary (i.e. predators), related to social interactions with conspecifics, or which have been previously paired with aversive experiences, i.e. conditional stimuli (Gray, 1987). Development of fear in dogs may also be strongly influenced by environmental factors such as exposure or lack thereof to stimuli particularly during the socialization period in puppies (Kutsumi et al., 2013), which may potentially lead to fearfulness, described as the general tendency to react in a fearful manner to different novel situations (Boissy, 1995). For commercial breeding dogs, fear may be reflected in their behavioral responses to unfamiliar people and novel situations due to inadequate socialization and exposure, creating an area of concern that is necessary to explore, particularly if such dogs are to be rehomed thereafter.

Various methods have been employed to assess fear in domestic dogs within the literature. Many behavioral tests, for instance, involve presenting dogs with a range of novel stimuli/situations (e.g. people, objects, sounds, places) in a standardized manner, after which behavioral coding is carried out via measurements of frequency, latency and duration of behaviors of interest (e.g. Barnard et al., 2016b; Flint et al., 2018) or via ratings of behaviors presented on a predefined scale (e.g. 0 – 5) reflecting the intensity of the behavior being measured (e.g. Döring et al., 2017a; Goold and Newberry, 2017). Because of the difficulty in measuring the *conscious* emotion of fear in animals (Adolphs, 2013), fear is often inferred from the behavioral responses characterized in terms of the four F's: fight (attack/aggression), flight (avoidance), freeze (immobility) or flirt (submission) (Marks and Nesse, 1994). For instance, fear in pigs is typically interpreted from avoidance behavior directed towards a stationary or approaching observer (Powell et al., 2016), and in chickens, via the duration of tonic immobility (i.e. freezing), a validated fear assessment in birds (Jones and Waddington, 1992; Pichova et al., 2016). Reliable behavioral measures of fear have also been identified in dogs, particularly measures capturing latencies to approach a stimulus, responses to a stimulus (i.e. approaching or running away from the stimulus), and time spent near or away from a stimulus (King et al., 2003), and which have been partially validated via assessment of the positive effects of anxiolytic drugs on these response measures to novel stimuli (e.g. Ley et al., 2007).

As reiterated within the canine literature, issues surrounding the reliability and validity of many tests intended to predict behavior continue to be a concern. The main aim of the present study was to therefore evaluate the validity of FIDO scoring by assessing the extent to which dogs' responses to a stranger's approach, as captured by the tool, predicted their behavioral responses in the presence of a stranger within an outdoor arena test. Similar methods of evaluating validity have been utilized in many animal studies within the literature (for a full review, see Gosling, 2001). Döring et al. (2017a), for instance, evaluated the predictive power of a simple human encounter test on a behavior test performed on laboratory dogs. As previously mentioned, the FIDO scoring classifies dogs' responses to stranger-approach into three categories; red - describing fearful, stereotypic or aggressive behaviors; yellow - describing ambivalent behaviors such as approach and avoid, and green - describing affiliative or attention-seeking behaviors or neutral behaviors. The main hypothesis was that dogs exhibiting behaviors categorized as red during a brief approach test by an unfamiliar person would also show responses characteristic of fear such as longer latencies to approach or greater avoidance of an unfamiliar person when inside an experimental test arena. Likewise, it was hypothesized that responses characteristic of affiliation during the approach test in the home pen would also be expressed during the standardized arena test, as indicated by shorter latencies to approach and more time spent in proximity and interacting with the unfamiliar person. For dogs characterized as yellow, it was hypothesized that they would exhibit intermediate responses during the experimental arena test. Other behaviors of interest that were hypothesized to corroborate the main hypotheses above were also collected, including elimination behaviors (urination and defecation), escape attempts and stereotypic behaviors.

Because some commercial breeding facilities provide dogs with indoor/outdoor access in the home pens, an issue that often arises when scoring dogs is the abundant loss of data due to dogs being absent from their home pens during scoring. To address this concern, a secondary question was embedded in the study in order to determine if there was any relationship between how dogs scored when locked indoors (allowing for a full set of data) versus when provided free access to their outdoor runs during scoring. Despite the potential benefit of scoring dogs while locked in with regard to provision of a complete dataset, it was considered important to determine whether scoring dogs during lock-in would indirectly induce stress due to their inability to escape, thereby influencing their responses to a stranger's approach.

5.1 Ethics Statement

All experimental procedures were approved by the Purdue University Institutional Review Board and Institutional Animal Care and Use Committee. Commercial breeding facility owners that participated in the study were provided with consent forms prior to the study and were informed that they could withdraw from the study at any point. Each facility owner was also compensated for their participation in the study with a $25 visa gift card.

5.2. Materials and Methods

5.2.1 Subjects

All dogs that met the study criteria were randomly sampled across four USDA-licensed commercial breeding facilities in Odon, Indiana. These included small breed dogs that were over one year of age, generally healthy, and pair-housed. To protect the welfare of pregnant bitches and puppies, bitches in the last two weeks of pregnancy and those nursing puppies were not included in the study. Since dogs were evaluated as a pair within a pen, ten randomly selected pens per facility were utilized. Prior to data collection, pilot testing was undertaken in one facility, but was not included in the final group of facilities because it did not meet certain criteria for the study (i.e. housing of pair-grouped dogs). Thus, a total of 80 small breed adult dogs (20 males and 60 females) were assessed in this study (Table 5).

5.2.2 Facilities

Facilities were chosen so as to be comparable to each other based on the type of breeds kept, and the number of dogs per pen. Thus, facilities with small breeds and pair housing were recruited for this study. At all four facilities, dogs were allowed free access to the outdoor portions of their pens. In facilities one, three and four, dogs had continuous access to the outdoor run between 4:30am and 6:30pm before being locked indoors overnight, while dogs in facility two had 24-hour access to the outdoor run. Because one of the objectives of the study was to test the behavioral responses of dogs to stranger-approach while they were routinely locked in, facility two agreed to lock their dogs indoors for the brief period of data collection (roughly two hours) needed for the purpose of this study. However, testing for the remaining three facilities occurred during normal lock-in hours. Dogs additionally had access to grass exercise yards for roughly half an hour six days per week in facility one, several hours two to three times per week in facility two, one hour five times per week in facility three, and 10-15 minutes three to four times per week in facility four. In facility one, indoor pens measured 126cm × 82cm and outdoor portions measured 242cm × 80cm. With regard to exercise yards, facility one had four separate yards, two at the front of the facility (10.15m × 15.24m and 10.15m × 20.30m) and two at the back (each 15.24m × 17.77m), and dogs were allowed access to particular yards based on their location in the facility. In facility two, indoor pens measured 88cm × 60cm and outdoor portions measured 227cm × 64cm, while the exercise yard measured 16.03m ×10.21m. In facility three, indoor pens measured 144cm × 87cm and outdoor portions measured 289cm × 117cm, while the exercise yard measured 9.75m × 20.73m. In facility four, indoor pens measured 140cm × 112cm while the outdoor portions measured 224cm × 91cm, and the exercise yard measured 10.97m × 29.26m.

Table 5. Summary of number of dogs assessed, sex, age (mean ±SD), number of dogs per breed, group composition and reproductive status.

Facility	Dogs assessed	Sex	Age (mean±SD)	Breed	Group Composition	Reproductive status (females)
1	20	4 M, 16F	34.2±14.81	Miniature Pinscher (6) Maltese (7) Yorkshire (5) Morkie (2)	Female Pair (n = 8) Male Pair (n = 2) Mixed Pair (none)	Pregnant (none)
2	20	6M, 14F	32.6±14.70	West highland white Terrier (12) Cairn Terrier (8)	Female Pair (n = 4) Male Pair (none) Mixed Pair (n = 6)	Pregnant (n = 3)
3	20	5M, 15F	44.05±20.03	Shiba Inu (7) Shetland Sheepdog (13)	Female Pair (n = 5) Male Pair (none) Mixed Pair (n = 5)	Pregnant (n = 1)
4	20	5M, 15F	35.3±16.21	Mini Schnauzer (10) Coton de Tulear (6) Havanese (4)	Female Pair (n = 5) Male Pair (none) Mixed Pair (n = 5)	Pregnant (n = 2)

5.2.3 Study Procedures

The study was conducted over a three-day period during the month of May 2018. For the first two days, assessment of dogs was carried out using the FIDO scoring with the dogs having free access to their outdoor runs (day one) and locked indoors (day two) during scoring. On day three, arena testing in a field was done.

On the eve of the first day of data collection, each dog in each selected pen at all facilities was marked for identification by the caretaker using either a green paint (PetPaint Pet Hair Spray, 5-Ounce, Greyhound Green) or a pink paint (PetPaint Pet Hair Spray, 5-Ounce, Poodle Pink) provided by the researchers. This was done to differentiate the pair of dogs during scoring and during video analysis. On day one, data collection began at 8:00am in facility one and ended at around 5:00pm in facility four. For this test day, dogs in each of the ten pens across each facility were assessed by a primary observer (female) while having free indoor/outdoor access. Each pen was approached and a behavioral response (either red, yellow or green) to stranger-approach was recorded within five seconds, all while being videotaped using a head-mounted GoPro camera. Because the dogs in each pen had free access to the outdoor portions of their pens (i.e. outdoor runs), a second observer (female) was stationed on the outside portion of the pen to score the behavioral responses of those dogs that ran out or were already on the outside. Inter-rater reliability between the two observers was carried out prior to testing day on 20 dogs and an achieved 85% percent agreement was considered adequate. At each facility, the testing period took roughly ten minutes to complete, after which set-up of the test arena for the field approach took place (discussed in detail later). This was done on day one to allow the dogs roughly 48 hours to acclimate to the test arena during their exercise periods prior to the actual field approach test (e.g. Bollen and Horowitz, 2008). On day two, data collection commenced in the evening at

6:30pm in facility two and ended at 8:00pm in facility one. Testing of dogs was carried out at least 20 minutes after dogs were locked indoors in order to allow them time to settle down. Testing procedures as well as the primary observer remained the same as day one, with the only difference being that dogs had no access to the outdoor run, hence a second observer was not needed. On the final third day, the field approach test was carried out in the exercise yard with the primary observer, commencing at 8:00am in facility four and ending at 9:00pm in facility one. In each facility, testing took roughly two to three hours to complete, including camera and visual barrier set-up.

5.2.4 Outside Test Arena

Set-up for the outdoor testing took place in each facility's outside exercise yard. Because facility one had four exercise yards (two at the back of the facility and two at the front), four test arenas were built in this facility in order to utilize environments most familiar to the dogs in the study. Set-up for the outside test arena began by marking circles on the ground with chalk spray (Krylon 5894 Marking Chalk White) around a focal point, with one-meter, two-meter and three-meter radii from the center. A circular test arena was then created using orange colored snow fencing (roughly 66ft per arena) attached to eight plastic stakes placed along the perimeter of the 3-meter line. This was done in order to standardize all test arenas (Taylor and Mills, 2006). A portable start box measuring 4ft × 6ft was then constructed using the remaining snow fence affixed to two plastic stakes and attached to the arena fence during testing. A schematic representation of the experimental set-up is provided in Figure 4.

5.2.5 Field Approach Test

Before commencement of the test in each facility, a visual barrier (a long surgical drape) was clipped either on a metal fence running in front of the outdoor runs, or metal stakes placed between the outdoor runs and the exercise yard. This was done to prevent the dogs that had not yet been tested and those not in the study from having visual access to those being tested. Furthermore, two video cameras (Sony Handycam HDR-CX405 1080p HD Video Camera Camcorders) set up on tripod stands were strategically placed on the outside of the test arena (Figure 4). At the start of the test, the primary observer sat on a five-gallon bucket placed at the center of the test arena. Thereafter, the primary caretaker and a helper (one that was most familiar to the dogs) each carried a dog from a pen and placed them into a marked area within the attached start box. A second observer (female) acted as the timer and signaled to the caretaker and helper when it was time to release the dogs into the start box and vacate. Once the two dogs were placed on the ground, the test officially began and lasted for five minutes (e.g. Wilsson and Sundgren, 1998). During this time, the female primary observer inside the arena remained passive, but could briefly pet the dog if interaction was solicited (e.g. Barnard et al., 2016b). Procedures remained consistent across all four facilities, and behaviors collected were analyzed from video recordings (Table 6).

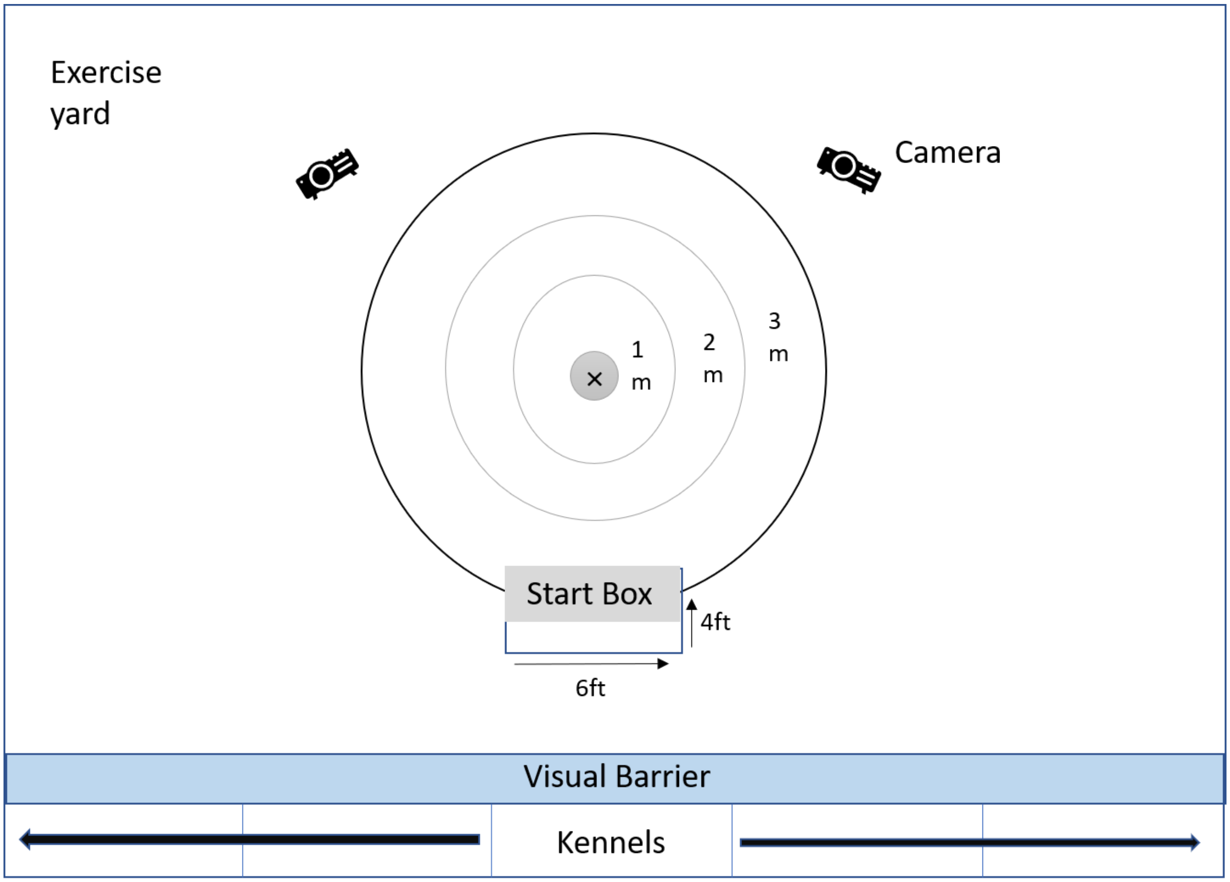

Figure 4. A diagrammatic representation of the experimental set-up.

× - Location of unfamiliar person

5.2.6 Video Coding

Video coding took place using the Behavioral Observation Research Interactive Software [BORIS] (Friard and Gamba, 2016) during the month of June 2018. Table 6 provides an ethogram of the behaviors coded. Two observers were used in the coding process, with one observer coding frequency and duration of behaviors of interest (as per the ethogram) exhibited in the test arena, and the second observer coding latencies to approach, durations within and frequencies of entry into different locations in the test arena. Video data from four dogs was excluded from analysis due to poor video quality.

Table 6. A list of behaviors included in the ethogram.

Behavior	Description	Measure
Approach	Both front legs cross over the three-meter, two-meter or one-meter line	Latency to 1st approach
Location (start box, 3m, 2m, 1m)	Dog has a majority of body within the zone. If a dog is halfway on a line and orienting towards the observer, choose the zone in which the front two legs are placed. If facing the opposite direction, choose the zone in which the two hind legs are placed as this is the closest point to the observer	Duration, frequency of entries
Interaction with observer	Dog solicits interaction with the observer through making physical contact either in the form of licking, jumping, nose contact (sniffing or nudging) or mouthing ***Jumping:*** The dog stands on its hind legs while placing its front legs against the observer's lap or side (Harvey et al., 2016) ***Licking:*** The dog licks the observer's hands (only exposed area) or clothes ***Nose contact*:** • Sniffing – Dog's nose to any area of the observer (Hubrecht et al., 1992) • Nudging – Dog pushes its nose against the observer ***Mouthing:*** Dog's mouth is in contact with observer or observer's clothes or shoes; this is accompanied by soft and loose body language or body language indicative of play such as a play bow	Latency, duration, frequency
Defecation	Dog's hind end lowered, back arched and tail held out to discharge faeces from the body (Overall, 2014)	Duration, frequency
Urination	Dog lifts or raises leg or lowers hind end relatively fast to release urine	Duration, frequency
Aggression towards observer	Dog barks, growls, lunges or snaps/bite at the observer during the period within the testing arena ***Barking*:** Dog produces a hard bark towards the observer accompanied by a tense, rigid and forward body language; head and lips forward, mouth opening, and shutting repeatedly; dog should be oriented at observer and can be backing up, moving towards or stationary while barking (Overall, 2014) ***Lunging:*** Dog moves toward the observer quickly and aggressively with a sudden forward thrust	Duration, frequency

	Growling: Dog produces a low buzzing sound towards the observer (Overall, 2014) ***Biting/snapping*:** Dog shows snapping movement (opening and closing of the mouth accompanied by showing the teeth), or seizes observer with teeth (Overall, 2014)	
Escape attempts	Dog attempts to leave the testing arena under or through the fence either by biting onto the fence or sticking its nose through the fence (Kiddie and Collins, 2014)	Duration, frequency
Escape	Dog leaves test arena	Yes/no
Repetitive movements	Dog shows repetitive circling or pacing for more than 3 times ***Repetitive pace:*** Dog repeatedly paces in a fixed route (>3 times) (Stephen and Ledger, 2005) and/or ***Circling:*** Dog walks around in a circle repeatedly (> 3 times) (Stephen and Ledger, 2005)	Duration, frequency

Table 7. Descriptive statistics of behaviors displayed during the experimental arena test.

	Proportion of dogs (%)	Total frequencies	Average duration ± SD (s)
Latency to approach one-meter line	-	-	84.42 ± 113.56
Latency to approach two-meter line	-	-	27.78 ± 70.97
Latency to approach three-meter line	-	-	1.82 ± 2.52
Latency to interact	-	-	166 ± 122.88
Within one meter of unfamiliar person	73.68	395	18.02 ± 26.92
Within two meters of unfamiliar person	94.74	1337	34.39 ± 29.43
Within three meters of unfamiliar person	100	1763	114.71 ± 66.05
Within start box	100	839	93.86 ± 67.42
Jumping at unfamiliar person	25	55	2.25 ± 8.58
Licking unfamiliar person	6.58	17	0.59 ± 2.84
Sniffing unfamiliar person	30.26	52	1.55 ± 4.11
Nudging unfamiliar person	0	0	-
Mouthing unfamiliar person	5.26	4	0.20 ± 1.00
Aggressive behaviors (barking, lunging, growling, biting) towards unfamiliar person	3.95	8	1.55 ± 8.08
Defecation	19.74	15	2.18 ± 5.35
Urination	72.37	207	5.07 ± 6.26
Escape attempts	67.11	242	17.04 ± 24.96
Repetitive movements	5.26	20	2.33 ± 14.29

Number of all dogs across facilities (N=76)
Multiple categories apply

5.2.7 Data analysis

All statistical analysis was carried out in STATA (version 15).

Intra-rater Reliability

Upon conclusion of video coding, intra-rater reliability was assessed on a random selection of 20% of the observations (16 dogs) (e.g. Barnard et al., 2012). Intra-class correlation coefficient [ICC] (a measure of reliability of measurements) was used to assess intra-rater reliability for durations, latencies and frequencies of all the behavioral measures incorporated in the ethogram. ICC’s and their confidence intervals were calculated for each rater using a single rating, 2-way mixed-effects model with absolute agreement (Koo and Li, 2016). Guidelines for interpretation of ICC coefficients suggest that below 0.40 is poor, between 0.40 and 0.59 is fair, between 0.60 and 0.74 is good, and above 0.75 is excellent (Cicchetti, 1994). For both raters, all ICC estimates were above 0.80 and hence deemed excellent.

Chi-square test

To determine whether dogs' responses to a stranger when locked indoors differed from when allowed free access, a Pearson's chi-square test of independence (McHugh, 2013) was performed. Firstly, a chi-square test was done to compare RYG [i.e. red, yellow, green] scoring when dogs were locked in, with RYG scoring when dogs were *inside* their pen during free access. A second chi-square test was performed to compare RYG scoring when dogs were locked in, with RYG scoring when dogs were either inside their pen or in the outdoor run (thus scored by the second observer) during free access, allowing for a full dataset of scores. The second chi-square test was carried out to determine if there was any change in the relationship of the two scoring conditions with the inclusion of dogs scored on the outside portion of the pens.

Principal Components Analysis (PCA)

Behaviors that occurred in less than 10 dogs during the field test, i.e. nudging, licking, mouthing, human-directed aggression and repetitive movements (Table 7), were excluded from further analysis following recommendations by Peduzzi et al. (1995). The remaining behavioral variables were first subjected to a principal components analysis (PCA) with orthogonal varimax rotation in order to reduce the number of variables (Tabachnick and Fidell, 2013). As suggested by Budaev (2010), only variables with eigenvalues of greater than one and loadings of 0.30 and above were interpreted. Sampling adequacy was also assessed via the Bartlett sphericity test (to test if correlations are zero) and the Kaiser-Meyer-Olkin (KMO) measure, and only considered appropriate with a significant Bartlett test and a minimum KMO of 0.50 (Budaev, 2010).

Multivariate Multiple Regression Analysis

Based on the PCA, component scores were generated for each subject on each of the identified components, and thereafter included in a multivariate multiple regression analysis as outcome variables. Building of the regression model was carried out in a backward stepwise process as recommended by Dohoo et al. (2010). Each potential predictor (i.e. RYG score [both for when dogs were locked indoors and when allowed free access], sex, breed, age, group composition and reproductive status [pregnant/not pregnant]) was first assessed individually against all outcome variables in a multivariate regression analysis and included in the full model if they met a liberal p-value of 0.20 (Dohoo et al., 2010). Backward stepwise regression was then used in order to further reduce the number of independent variables, with variables being removed if they failed to meet a p-value of 0.10. As RYG score was the primary variable of interest, this variable was kept in the final model regardless of significance. Random effects of facility and pen were tested. Assumptions of normality and homoscedasticity for all outcome variables were also examined both visually and statistically as recommended by Ghasemi and Zahediasl (2012).

5.3 Results

Chi-square Test

The chi-square test of independence performed to examine the relationship between scoring condition (free access and locked indoors) and RYG scoring showed a significant association between RYG scoring when dogs were locked in and when they were scored on the interior portion of their pen during free access, χ^2 (4, N = 34) = 15.05, p = 0.005. Furthermore, results showed a significant association between RYG scoring when dogs were locked in and when allowed free access (i.e. scored either inside or outside the pen), χ^2 (4, N = 80) = 19.96, p = 0.001, reflecting no change in the relationship with the inclusion of dogs scored on the outside portion of their pens. A follow-up test with a spearman's rank correlation further supported this result, revealing a moderate correlation between scoring while locked in and scoring when allowed free access, r_s (78) = 0.46, n = 80, $p < .001$, as per interpretation guidelines outlined in Taylor (1990). Upon evaluation of the trend of responses from free access to lock-in, more than half of the dogs (69%) remained consistent in their responses to a stranger's approach between conditions. For the dogs that showed a change in their responses, only ten dogs changed from green during free access, to red during lock-in, and one dog from yellow during free access, to red during lock-in (Table 8). Further descriptive statistics are summarized in Table 8.

Table 8. Descriptive statistics for RYG scoring (free access) by RYG scoring (locked in).

	RYG scoring (free access)			
RYG scoring (locked in)	Red	Yellow	Green	Total
Red	21	1	10	32
Yellow	2	1	2	5
Green	8	2	33	43
Total	31	4	45	N = 80

Principal Components Analysis (PCA)

The PCA revealed seven principal components with eigenvalues of greater than one that together explained 80.74% of the total variance. Sampling adequacy of the correlation matrix was deemed appropriate (Bartlett's sphericity test $\chi^2 = 1784.47$, df = 231, $p < 0.001$; KMO = 0.57). Based on the output (Table 9), the first principal component (PC1) [avoidance activity] accounted for 17.65% of the total variance and showed high positive loadings for frequencies of dogs within the start box and within two and three meters of the unfamiliar person. A second component (PC2) [approach with affiliation] accounted for 12.63% of the total variance and showed high positive loadings for frequency and duration of jumping as well as time spent within one meter of the unfamiliar person. A third component (PC3) [investigation of observer] explaining 11.85% of the total variance had high positive loadings for frequency and duration of sniffing, urination and time spent within one meter of the unfamiliar person. A fourth component (PC4) [avoidance] accounting for 10.75% of the total variance had high positive loadings for time spent in the start box and latency to approach the three-meter and two-meter line. Component five (PC5) [escape and avoid] accounted for 9.97% of the total variance and had high positive loadings for duration and frequency of escape attempts, and a negative loading for latency to approach the three-meter line. Component six (PC6) [exploration/marking] explained 9.50% of the total variance and had high positive loadings for duration and frequency of urination, time spent within three meters of the unfamiliar person, latency to approach the one-meter line, and latency to interact. Lastly, component seven (PC7) [defecation] explained 8.40% of the total variance and had strong positive loadings for duration and frequency of defecation.

Table 9. PCA output showing the loadings of each variable on each of the seven components. Loadings of 0.30 and above are shown in bold.

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Time jump.	0.0019	**0.5811**	-0.0337	-0.0032	-0.0198	0.0453	-0.0206
Freq. jump.	0.0027	**0.5759**	-0.0174	-0.0023	-0.0154	0.0219	-0.0226
Time sniff.	-0.1074	0.0004	**0.5458**	0.0120	0.0957	-0.0318	-0.0206
Freq. sniff.	-0.0534	-0.0168	**0.5801**	0.0278	0.0553	-0.0534	0.0289
Time urinate	0.1744	-0.2111	0.2439	-0.0168	-0.1491	**0.3312**	-0.0470
Freq. urinate	0.1537	-0.1982	**0.3001**	-0.0065	-0.1472	**0.3337**	-0.0278
Time escape att.	-0.0039	-0.0120	0.1103	0.0717	**0.5886**	0.0118	0.0156
Freq. escape att.	0.1166	-0.0812	-0.0111	0.0343	**0.5700**	-0.0204	-0.0538
Time defecate	-0.0047	-0.0043	0.0050	0.0142	-0.0125	-0.0097	**0.7037**
Freq. defecate	0.0039	-0.0168	-0.0021	-0.0020	0.0034	0.0122	**0.6972**
Time in area one	0.0536	**0.3605**	**0.3375**	0.0537	-0.0518	-0.0234	-0.0011
Time in area two	0.2868	0.0138	-0.0105	-0.1082	-0.1794	0.1136	0.0425
Time in area three	-0.0299	0.0772	-0.0051	-0.1939	0.0529	**0.5899**	0.0184
Freq. area one	0.2811	0.2295	0.1662	0.0152	-0.0804	-0.0696	0.0340
Freq. area two	**0.4321**	0.1178	0.0167	-0.0394	-0.0473	0.0195	0.0143
Freq. area three	**0.5145**	-0.0076	-0.0966	0.0161	0.0712	0.0132	-0.0060
Freq. start box	**0.4931**	-0.0584	-0.1064	0.1145	0.1526	-0.0784	-0.0159
Time start box	0.1067	0.0096	0.0600	**0.6017**	0.1095	-0.0077	0.0529
Latency one-meter	-0.1964	0.1243	-0.1277	0.2841	0.0800	**0.3244**	0.0210
Latency two-meter	-0.0798	0.0264	-0.0057	**0.5341**	-0.0554	0.0025	-0.0313
Latency three-meter	0.0360	-0.1310	0.0080	**0.3782**	**-0.4182**	-0.1600	-0.0484
Latency to interact	-0.0257	0.0080	-0.1404	0.2318	0.0245	**0.5175**	-0.0090

PC1 [Avoidance activity]; **PC2** [Approach with affiliation]; **PC3** [Investigation of observer]; **PC4** [Avoidance]; **PC5** [Escape and avoid]; **PC6** [Exploration/marking]; **PC7** [Defecation].

Multivariate Multiple Regression Analysis

A multivariate multiple regression was conducted to predict all seven PC scores based on potential predictors of interest. Several predictor variables were tested (i.e. RYG score (for both scoring conditions), sex, breed, age, group composition and reproductive status e.g. pregnant/not pregnant), but only RYG score and group composition were included in the final model, accounting for breed and length of time within the test arena. Because RYG scores for the two scoring conditions (locked in and free access) were moderately correlated ($r_s = 0.46$), only RYG scores for dogs during lock-in were included and interpreted because of i) the consistency of one observer and one location (i.e. inside the pen), and ii) the underlying conceptual similarity to the field arena test (i.e. dogs were spatially restricted). Assumptions of normality and homoscedasticity were met for PC1 [avoidance activity], PC3 [investigation of observer], PC4 [avoidance], PC5 [escape and avoid] and PC6 [exploration/marking] scores via the Shapiro-Wilk test, kernel density graphs and a scatterplot of residuals, however visual inspection of PC2 [approach with affiliation] and PC7 [defecation] scores also deemed them acceptable for analysis. Random effects of facility and pen tested were found to be non-significant, and therefore excluded from further analysis.

Results showed that the multivariate multiple regression explained a significant amount of variation for PC1 [avoidance activity] ($R^2 = 0.49$, $p < 0.001$), PC3 [investigation of observer] ($R^2 = 0.34$, $p = 0.023$), PC4 [avoidance] ($R^2 = 0.41$, $p = 0.003$), PC5 [escape and avoid] ($R^2 = 0$ .42, $p = 0.002$), and PC6 [exploration/marking] ($R^2 = 0.56$; $p < 0.001$) based on a significant F-test when compared to an intercept-only model. The models for PC2 [approach with affiliation] and PC7 [defecation] did not explain a significant amount of variation ($R^2 = 0.24$, $p = 0.249$ and $R^2 = 0.19$, $p = 0.541$, respectively). Further examination of individual predictors within significant models indicated that RYG scoring was only significantly associated with PC4 [avoidance], comprised of time spent in the start box and latency to approach the three-meter and two-meter line. Compared to dogs that scored green, dogs that scored red had significantly higher PC4 scores (beta:0.82, 95% CI [0.04, 1.61], $p = 0.039$), indicating that they spent more time in the start box, and had a longer latency to approach the three-meter and two-meter line. No significant differences were found between dogs that scored green and dogs that scored yellow ($p = 0.480$), and between dogs that scored red and dogs that scored yellow ($p = 0.659$). RYG scoring was, however, not predictive of the remaining significant PC scores (Figure 5).

Group composition was also found to be significantly associated with PC5 [escape and avoid], comprised of frequency and duration of escape attempts. Dogs housed as a female pair had significantly higher PC5 scores than dogs housed as a mixed pair [i.e. male and female] (beta: 1.19, 95% C.I [0.43, 1.94], $p = 0.003$), indicating that they spent more time engaging in and had higher frequencies of escape attempts from the test arena. No significant differences were found between the other group compositions.

Breed was included in the multivariate model to control for potential effects, and was found to be significantly associated with PC1 [avoidance activity] ($p = 0.004$), PC4 [avoidance] ($p = 0.020$), PC5 [escape and avoid] ($p = 0.002$) and PC6 [exploration/marking] ($p = 0.003$). Time spent in the arena was also accounted for in all models and found to be significant in PC1 [avoidance activity] ($p < 0.001$), PC4 [avoidance] ($p = 0.040$), PC5 [escape and avoid] ($p = 0.009$) and PC6 [exploration/marking] ($p < 0.001$).

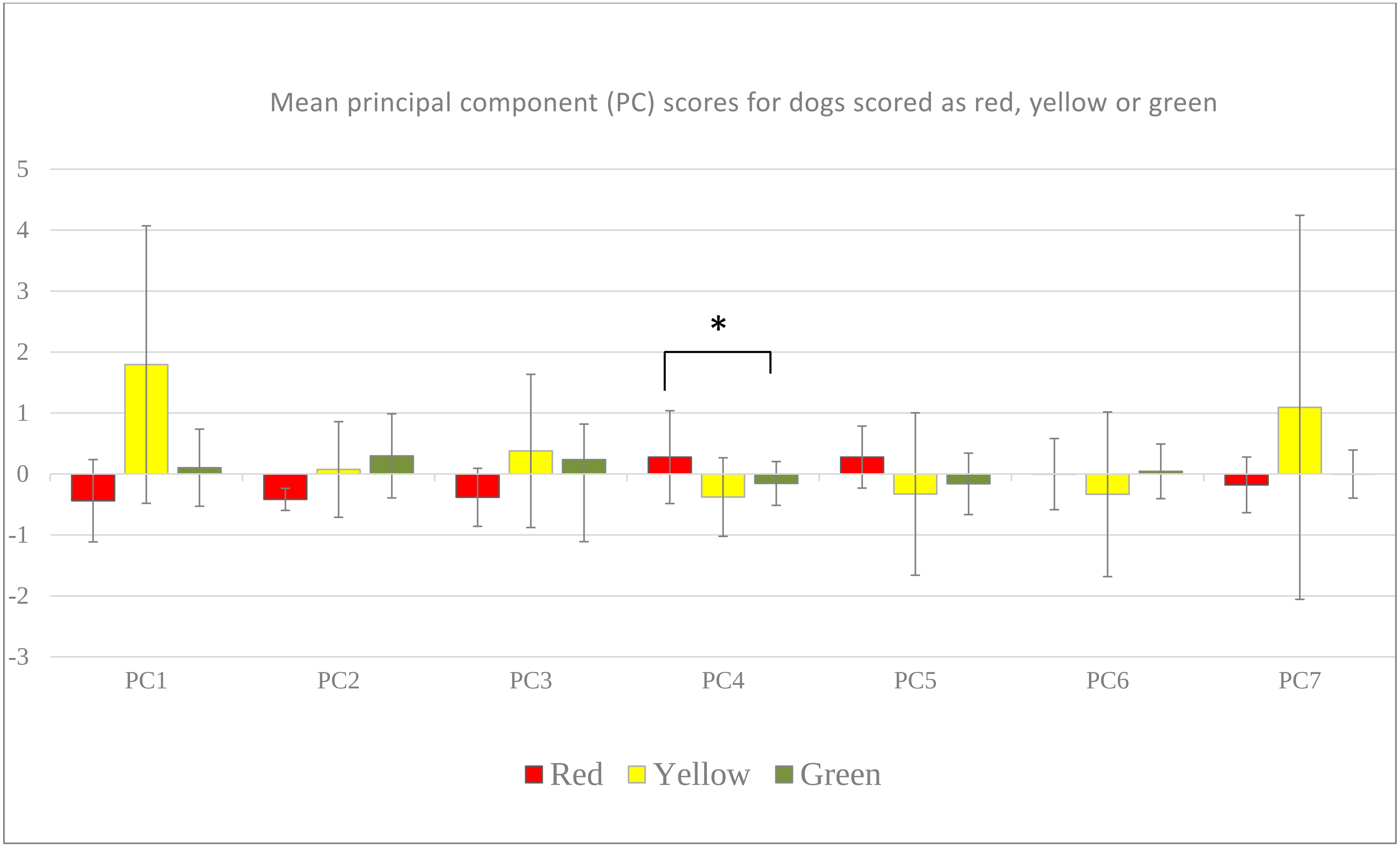

Figure 5: Mean scores for each of the principal components (1-7) for dogs scored as red, yellow and green. **PC1** [Avoidance Activity]; **PC2** [Approach with affiliation]; **PC3** [Investigation of observer]; **PC4** [Avoidance]; **PC5** [Escape and avoid]; **PC6** [Exploration/marking]; **PC7** [Defecation]

5.4 Discussion

Findings of the relationship between the RYG scoring conditions revealed a significant association between RYG scores when dogs were locked indoors and when allowed free access. A majority of the dogs (69%) remained consistent in their responses to stranger-approach between the two conditions, providing some valuable insight into the applicability of the differing scoring situations. Further evaluation of the trend of scores between free access and locked in also showed that only a few dogs (14%) changed from green or yellow during free access to red during lock-in, suggesting that the majority may not have necessarily found the process of being scored while locked in stressful. Taken together, these findings suggest that scoring dogs while locked in can be useful in situations that require a full dataset of dogs to be scored. The lack of a strong correlation between scoring conditions could, however, be attributed to several factors. First, the use of different observers between scoring conditions may have played a role in influencing how dogs responded, for instance due to overt differences in appearance or subtle nuances in approach. While both female observers were unfamiliar to the dogs for this study, it is possible that some dogs may have responded differently between free access and lock-in due to physical differences between the observers. Wells and Hepper (1999), for instance, assert that a dog's breed, age, and previous experience with different people as well as the observer's physical appearance, behavior, age, race and odor all play a role in influencing a dog's behavioral response towards the person. Differences in scores between conditions could also be attributed to variations in scoring between the two observers. While both observers reached 85% percent agreement prior to actual scoring, it is possible that observers may have differed in their interpretation of certain behavioral responses, leading to some differences in scoring between conditions.

Findings of the multivariate regression analysis indicated that dogs scored as red showed significantly higher PC4 scores than dogs scored as green. Because PC4 had high loadings for time spent in the start box, and latency to approach the three-meter and two-meter line, this finding suggests that "red" dogs spent more time away from the unfamiliar person (i.e. in the start box) and showed longer latencies to move towards the unfamiliar person than "green" dogs. As previously mentioned, avoidance/approach behavior has been consistently evaluated in many fear assessment studies particularly on farm species such as pigs, sheep, cattle, and birds (for a full review, see Forkman et al., 2007), via measurements of latencies to approach the stimulus, time spent near the stimulus, and frequency and duration of interactions with the stimulus (e.g. Breuer et al., 2000; Hemsworth et al., 1996). Tanida et al. (1995), for instance, found that non-handled pigs showed greater avoidance of a human during a behavior test by keeping further away, and taking longer to approach and/or interact with an unfamiliar human. Likewise, studies within the canine literature have also associated reduced fear towards a novel stimulus with more approach behaviors, and increased fear of a stimulus with more avoidance (Conley et al., 2014; King et al., 2003; Ley et al., 2007; Stellato et al., 2017). Findings from the present study therefore provide further evidence of the aforementioned suppositions, and support the main hypothesis that dogs scored as red would also exhibit greater avoidance of an unfamiliar person when inside the experimental arena compared to dogs scored as green. Findings for dogs scored as yellow relative to those scored as red or green was, however, difficult to infer due to the very small sample size ($n = 5$) and should thus be interpreted with caution.

Interestingly, how dogs were housed and tested (i.e. mixed vs same sex pair) was significantly predictive of PC5 scores comprising escape attempts. Specifically, dogs housed as a female pair spent more time engaging in escape behavior than dogs housed as a mixed pair. A possible explanation for this is may be related to the interaction between males and females on behavioral responses exhibited in the arena. For instance, previous research findings have suggested male dogs to be bolder and less prone to displaying fearful behavior than females (e.g. Döring et al., 2009; Lund et al., 1996; Starling et al., 2013), hence females tested with males may have been less likely to try and escape as a result of "social buffering" associated with being in the presence of unafraid males (e.g. Hennessy et al., 2008).

Contrary to our expectations, RYG scoring was not significantly predictive of the PC scores comprising behaviors indicative of proximity and interaction such as jumping (PC2) and sniffing (PC3), or escape attempts (PC5). Graphs in Figure 5, however, show a tendency for "green" dogs to exhibit higher PC2 and PC3 scores, and lower PC5 scores than "red" dogs, suggesting that dogs scoring green exhibited more affiliative and investigatory behaviors, and less escape behaviors than dogs scoring red. Non-significance of results could have therefore been due to limitations of sample size and insufficient power. Interpretation of findings for dogs scored as yellow should again be made with caution because of the very small sample size. Findings for PC1 [avoidance activity] (comprising frequencies within the start box and the different areas), PC6 [exploration/marking] (comprising urination, latency to interact, time spent within three meters and latency to cross the one-meter line), and PC7 [defecation] (comprising defecation) were also non-significant, suggesting that RYG scoring was not predictive of these scores. This could be partly explained by the nature of the test arena itself, as it provided an opportunity for dogs to engage in other motivated behaviors such as exploration and marking (Boissy, 1995; Waiblinger et al., 2006), independent of how they were scored during the approach test. Factors such as age, sex and reproductive status (i.e. pregnant or not) were also not significant predictors of behaviors within the test arena, likely due to little variation within these variables.

In relation to study design, behavioral testing was conducted on pairs of dogs throughout. This was mainly because the majority of such facilities socially housed their dogs (in pairs or groups). To therefore reduce the likelihood of indirectly inducing stress by isolating conspecifics, testing dogs in pairs was found most appropriate. Previous studies have, for example, shown that isolating conspecifics during testing can impact how animals respond to stressful situations (Boissy, 1995; Donald et al., 2011), and that social partners can help mitigate the responses to stressors by providing a buffer during such situations (Kiyokawa and Hennessy, 2017; Nakamura et al., 2016). Furthermore, as suggested by Döring et al. (2017a), consistency of context between testing was important. Döring et al. (2017a), for instance, attribute their lack of correlation and predictive power between a human encounter test and a behavior test to the fact that dogs were tested as a group in the former, and individually in the latter. Also, because the human approach test was conducted in a familiar environment (i.e. in the home pen), it was imperative that the test arena be constructed in a familiar area such as the exercise yard, in order to minimize the confounding effects of environmental novelty (Waiblinger et al., 2006).

With regard to breed effects, interpretations of specific breed differences were not of primary interest in this study. This was because of the small number of dogs within each breed category, making it difficult to compare or generalize across breed groups. Furthermore, because of the limitations surrounding behavioral research on breed differences (i.e. variation in methodology utilized [e.g. survey-based vs experimental studies], overrepresentation of some breeds over others [e.g. those easily recruitable] and variation in the categorization of behaviors of interest [e.g. fear vs dominance-related aggression]) (Bradley et al., 2011; Mehrkam and

Wynne, 2014), accurate and meaningful conclusions on breed differences would be difficult to draw given this study's constraints.

Limitations:

One limitation of the present study is the sample size used. Due to the nature of the study, factors such as availability of facilities to recruit, time invested by caretakers in the study, criteria for recruitment, and number of total dogs per facility limited the number of facilities available for participation in the study. Thus, the number of dogs accessible may not have been sufficient to achieve even more robust results. Future research using a larger sample size and a greater number of facilities may therefore be necessary to validate and expand on the present results.

Another possible limitation for the study may be the novelty of the test arena itself. While efforts were made to habituate the dogs to the test arena prior to actual testing, it is possible that the novelty of the arena might have influenced some of the behavioral responses displayed by the dogs such as exploration, marking, defecation, and escape attempts. If so, this could have interfered with the measurement of responses towards the unfamiliar person such as interaction. Future studies could therefore benefit from ensuring that dogs are fully habituated to the test arena, such that any responses exhibited would be in response to the unfamiliar person. Likewise, because of the constraints of time and resources, arena testing was carried out in one day. Time of day could have therefore influenced the behavioral responses of dogs within the test arena. For instance, dogs tested in the morning and afternoon hours could have been influenced by the effects of increasing temperatures (which may have impacted elimination and activity rates) than dogs tested in the evening hours. Furthermore, because of the repeated testing of different pens (and different group compositions) within the same arena, eliminative behaviors such as defecation and urination/marking could have also been socially triggered for assorted reasons (Lisberg and Snowdon, 2009, 2011), and thus not reflective of responses associated with the presence of an unfamiliar person. Careful deciphering of such behaviors within the context of the aforementioned limitations would be necessary for future studies before any definite conclusions can be drawn.

Overall, the present study provides preliminary evidence that dogs classified as fearful towards an unfamiliar person, as determined by the FIDO tool, also show greater avoidance of the unfamiliar person during a standardized arena test. This finding provides important insight, as one of the main purposes of the tool is to identify dogs that would benefit from adequate and/or improved socialization towards strangers, a crucial element particularly for dogs likely to be rehomed afterward. Findings of the present study thus suggest that using a five-second approach test can practically and accurately inform an observer about the immediate state of a dog, providing an avenue for quick identification of welfare concerns associated with fear towards people. Limitations notwithstanding, the predictive validity of the FIDO scoring is generally supported in this study.

CHAPTER 6. CONCLUSION AND IMPLICATIONS

The need to understand the welfare of dogs kept in commercial breeding facilities within the United States inspired the execution of this project. Despite commercial breeding being an important source for companion dogs to date, a review of the literature revealed a scarcity of research conducted on this population of dogs. For the few studies that had been published, findings largely drew from owner reports and surveys (e.g. McMillan et al., 2011, 2013; McMillan, 2017), and thus, were potentially biased or constrained by verification of source of origination and lack of direct observation of on-site at commercial kennels. Objective assessment of welfare with an eye toward finding ways to support and improve it was therefore the goal of this project. In a previous study published by Bauer et al. (2017), a newly developed assessment tool (i.e. the Field Instantaneous Dog Observation [FIDO] tool), was established to objectively identify potential welfare concerns in dogs kept in breeding operations. The behavioral portion of the tool, which was the focus of this project, consists of three categories characterizing the response of a dog to a stranger's approach: red, indicating fear-related behaviors such as freezing, aggression or flight, green-indicating affiliative or neutral responses, and yellow-indicating ambivalent behaviors or behaviors neither green nor red. Extending from that study, the present project sought to explore further the robustness of the tool by evaluating its reliability and validity, while in the process investigating behavioral indicators of well-being in breeding dogs. Findings showed high intra- and inter-rater reliability between novice raters, as well as a reliable and practical time point from which to evaluate dogs' responses. Findings also showed evidence of the predictive validity of the FIDO tool when compared to behavioral responses of dogs to a stranger within a standardized test arena.

Implications and future directions

On-site assessment requires the use of reliable, valid and feasible tools that are both practical and objective. The first part of the project therefore provides support on the reliability of the FIDO tool. Results indicate that with training, individuals with relatively little knowledge on canine behavior can show strong agreement on behavioral assessment, suggesting the tool's potential use with kennel evaluators or caretakers, for instance, during routine site inspections. Furthermore, as indicated in the second study, the FIDO tool provides a practical and efficient way of assessing behavior, as no benefit of extending observation time of the dog is supported. The implication of this finding is especially applicable in high volume facilities, for which time taken to assess dogs is limited.

The second part of the project also provides support on the validity of the FIDO tool. Results indicate that dogs scored as red also demonstrate fearful responses characterized by avoidance behavior when placed in a standardized outside arena. This finding has important implications. First, it provides evidence of the accurate identification of fearful dogs, therefore contributing to the validity of the scoring system of the tool. Secondly, because the FIDO tool offers a proxy measure of how well-socialized a dog is to strangers, this finding also implies that the FIDO tool can quickly but accurately identify dogs that may need interventions, for instance, via improved socialization towards people. Fearful behavior in breeding dogs is of particular concern because of the negative physiological impacts associated with prolonged fear (Foyer et al., 2016), and the resultant implications for puppy behavior. Identification and improvement of such behaviors is therefore essential for both dogs likely to be rehomed, and for puppies that come from commercial facilities. Beyond breeding establishments, the FIDO tool can also be potentially useful for other kenneled settings such as shelters and research facilities that are intent on identifying problem behaviors in a quick and objective way.

For this project, reliability between raters was assessed via video recordings. While this was necessary to avoid the potential confound of differential responses by dogs to a male and female assessor, future research investigating how dogs respond to a male and female human would be worthwhile. This would be especially important for situations in which either a male or female inspector or caretaker is needed to assess kennels.

Likewise, while the FIDO tool was originally established to assess a dog's response to a human approach, it would be interesting and useful to assess how dogs respond to other stimuli such as novel objects or situations, for instance via battery-style testing, in order to decipher whether fearful behavior is specific or global. This would provide critical information on the plausibility and suitability of interventions that would be necessary to improve behavior.

Findings of the validity study also presented insightful results on the ability of the FIDO scoring to predict fearful dogs. An avenue for future research would be to further investigate the validity of the tool by assessing it against long-term metrics of stress and welfare, such as hair cortisol or reproductive performance. Determining whether a short, quick test can indeed be predictive of long-term welfare would be highly beneficial for caretakers, as it would provide a practical way of continuously monitoring and improving the welfare of dogs.

In summary, this thesis sets an important foundation into the objective assessment of welfare in commercial breeding dogs. Despite limitations, findings for this project offer a useful segue into upcoming work in this area that may provide a foundation for future research critical for the welfare of this under-investigated population of dogs.

REFERENCES

ABC Action News. (2018). Puppy owners paid thousands for pets that were supposed to be healthy - but the pets nearly died**.** Retrieved from https://www.abcactionnews.com/news/region-north-pinellas/largo/puppy-owners-paid-thousands-for-pets-that-were-supposed-to-be-healthy-but-the-pets-nearly-died

Accorsi, P. A., Carloni, E., Valsecchi, P., Viggiani, R., Gamberoni, M., Tamanini, C., & Seren, E. (2008). Cortisol determination in hair and faeces from domestic cats and dogs. *General and Comparative Endocrinology*, *155*(2), 398–402. https://doi.org/10.1016/j.ygcen.2007.07.002

Adams, G. J., & Johnson, K. G. (1994). Behavioural responses to barking and other auditory stimuli during night-time sleeping and waking in the domestic dog (Canis familiaris). *Applied Animal Behaviour Science*, *39*(2), 151–162. https://doi.org/10.1016/0168-1591(94)90135-X

Adolphs, R. (2013). The biology of fear. *Current Biology*, *23*(2), R79-R93.

American Society for the Prevention of Cruelty to Animals [ASPCA]. (2015). State Puppy Mill Chart. Retrieved from https://www.aspca.org/sites/default/files/puppy-mill-chart.pdf

American Society for the Prevention of Cruelty to Animals [ASPCA]. (2018). Shelter Intake and Surrender. Retrieved from https://www.aspca.org/animal-homelessness/shelter-intake-and-surrender/pet-statistics

Appleby, M. C., & Hughes, B. O. (1991). Welfare of laying hens in cages and alternative systems: environmental, physical and behavioural aspects. *World's Poultry Science Journal*, *47*(July), 109. https://doi.org/10.1079/WPS19910013

Asher, L., Diesel, G., Summers, J. F., McGreevy, P. D., & Collins, L. M. (2009). Inherited defects in pedigree dogs. Part 1: disorders related to breed standards. *The Veterinary Journal*, *182*(3), 402-411.

Asher, L., Blythe, S., Roberts, R., Toothill, L., Craigon, P. J., Evans, K. M., ... & England, G. C. (2013). A standardized behavior test for potential guide dog puppies: Methods and association with subsequent success in guide dog training. *Journal of Veterinary Behavior: Clinical Applications and Research*, *8*(6), 431-438.

Baciadonna, L., & McElligott, A. G. (2015). The use of judgement bias to assess welfare in farm livestock. *Animal Welfare*, *24*(1), 81–91. https://doi.org/10.7120/09627286.24.1.081

Barnard, S., Siracusa, C., Reisner, I., Valsecchi, P., & Serpell, J. A. (2012). Validity of model devices used to assess canine temperament in behavioral tests. *Applied Animal Behaviour Science*, *138*(1-2), 79-87.

Barnard, S., Pedernera, C., Candeloro, L., Ferri, N., Velarde, A., & Villa, P. D. (2016a). Development of a new welfare assessment protocol for practical application in long-term dog shelters. *Veterinary Record*, *178*(1), 18. https://doi.org/10.1136/vr.103336

Barnard, S., Marshall-Pescini, S., Passalacqua, C., Beghelli, V., Capra, A., Normando, S., ... & Valsecchi, P. (2016b). Does subjective rating reflect behavioural coding? Personality in 2 month-old dog puppies: an open-field test and adjective-based questionnaire. *PloS one*, *11*(3), e0149831.

Barnett, J. L., & Hemsworth, P. H. (1990). The validity of physiological and behavioural measures of animal welfare. *Applied Animal Behaviour Science*, *25*(1–2), 177–187. https://doi.org/10.1016/0168-1591(90)90079-S

Barnett, J. L., & Hemsworth, P. H. (2003). Science and its application in assessing the welfare of laying hens in the egg industry. *Australian Veterinary Journal*, *81*(10), 615–624. https://doi.org/10.1002/sce.10098

Barrera, G., Jakovcevic, A., Elgier, A. M., Mustaca, A., & Bentosela, M. (2010). Responses of shelter and pet dogs to an unknown human. *Journal of Veterinary Behavior: Clinical Applications and Research*, *5*(6), 339-344.

Bateson, M., & Nettle, D. (2015). Development of a cognitive bias methodology for measuring low mood in chimpanzees. *PeerJ, 3*, e998. https://doi.org/10.7717/peerj.998

Bateson, P. (2010). Independent Inquiry into Dog Breeding, (January), 1–65.

Battaglia, C. L. (2009). Periods of early development and the effects of stimulation and social experiences in the canine. *Journal of Veterinary Behavior: Clinical Applications and Research*, *4*(5), 203-210.

Bauer, A., Beck, A., Stella, J., & Croney, C. (2016). Overpopulation or Too Many Unwanted Pets? Perspective on Concepts and Management Approaches.

Bauer, A. E., Jordan, M., Colon, M., Shreyer, T., & Croney, C. C. (2017). Evaluating FIDO: Developing and pilot testing the Field Instantaneous Dog Observation tool. *Pet Behaviour Science*, (4), 1-16.

Beach, F. A., & Gilmore, R. W. (1949). Response of male dogs to urine from females in heat. *Journal of Mammalogy*, *30*(4), 391-392.

Beerda, B., Schilder, M. B. H., Janssen, N. S. C. R. M., & Mol, J. A. (1996). The use of saliva cortisol, urinary cortisol, and cateoholamine measurements for a noninvasive assessment of stress responses in dogs. *Hormones and Behavior*, *30*(3), 272–279. https://doi.org/10.1006/hbeh.1996.0033

Beerda, B., Schilder, M. B. H., Van Hooff, J. A. R. A. M., & De Vries, H. W. (1997). Manifestations of chronic and acute stress in dogs. *Applied Animal Behaviour Science*, *52*(3–4), 307–319. https://doi.org/10.1016/S0168-1591(96)01131-8

Beerda, B., Schilder, M. B. H., Van Hooff, J. A. R. A. M., De Vries, H. W., & Mol, J. A. (1998). Behavioural, saliva cortisol and heart rate responses to different types of stimuli in dogs. *Applied Animal Behaviour Science*, *58*(3–4), 365–381. https://doi.org/10.1016/S0168-1591(97)00145-7

Beerda, B., Schilder, M. B., Van Hooff, J. A., De Vries, H. W., & Mol, J. A. (1999a). Chronic stress in dogs subjected to social and spatial restriction. I. Behavioral responses. *Physiology & Behavior*, *66*(2), 233-242.

Beerda, B., Schilder, M. B., Bernadina, W., Van Hooff, J. A., De Vries, H. W., & Mol, J. A. (1999b). Chronic stress in dogs subjected to social and spatial restriction. II. Hormonal and immunological responses. *Physiology & Behavior*, *66*(2), 243-254.

Bildsøe, M., Heller, K. E., & Jeppesen, L. L. (1991). Effects of immobility stress and food restriction on stereotypies in low and high stereotyping female ranch mink. *Behavioural Processes*, *25*(2–3), 179–189. https://doi.org/10.1016/0376-6357(91)90020-Z

Blackwell, E. J., Bodnariu, A., Tyson, J., Bradshaw, J. W. S., & Casey, R. A. (2010). Rapid shaping of behaviour associated with high urinary cortisol in domestic dogs. *Applied Animal Behaviour Science*, *124*(3–4), 113–120. https://doi.org/10.1016/j.applanim.2010.02.011

Boleij, H., Klooster, J. van t., Lavrijsen, M., Kirchhoff, S., Arndt, S. S., & Ohl, F. (2012). A test to identify judgement bias in mice. *Behavioural Brain Research*, *233*(1), 45–54. https://doi.org/10.1016/j.bbr.2012.04.039

Bollen, K. S., & Horowitz, J. (2008). Behavioral evaluation and demographic information in the assessment of aggressiveness in shelter dogs. *Applied Animal Behaviour Science*, *112*(1-2), 120-135.

Boissy, A. (1995). Fear and fearfulness in animals. *The quarterly review of biology*, *70*(2), 165-191.

Botreau, R., Veissier, I., Butterworth, A., Bracke, M. B. M., & Keeling, L. J. (2007). Definition of criteria for overall assessment of animal welfare. In *Animal Welfare* (Vol. 16, pp. 225–228).

Boxall, J., Heath, S., Bate, S., & Brautigam, J. (2004). Modern concepts of socialisation for dogs: Implications for their behaviour, welfare and use in scientific procedures. *Alternatives to laboratory animals: ATLA*, *32*, 81-93.

Bradley, J. (2011). The relevance of breed in selecting a companion dog. *The relevance of breed in selecting a companion dog.*

Bradshaw, J. W. S., & Nott, H. M. R. (1995). Social and Communication Behaviour of Companion Dogs. *The Domestic Dog: Its Evolution, Behaviour and Interactions with People*, 115–130.

Brady, K., Cracknell, N., Zulch, H., & Mills, D. S. (2018). A Systematic review of the reliability and validity of behavioural tests used to assess behavioural characteristics important in working dogs. *Frontiers in Veterinary Science*, *5*, 103.

Breuer, K., Hemsworth, P. H., Barnett, J. L., Matthews, L. R., & Coleman, G. J. (2000). Behavioural response to humans and the productivity of commercial dairy cows. *Applied animal behaviour science*, *66*(4), 273-288.

Breuer, K., Hemsworth, P. H., & Coleman, G. J. (2003). The effect of positive or negative handling on the behavioural and physiological responses of nonlactating heifers. *Applied Animal Behaviour Science*, *84*(1), 3-22.

Broom, D. M. (1986). Indicators of poor welfare. *British veterinary journal*, *142*(6), 524-526.

Broom, D. M. (1991). Animal welfare: concepts and measurement. *Journal of animal science*, *69*(10), 4167-4175.

Broom, D. M. (2001). Coping, stress and welfare. *Coping with challenge: Welfare in animals including humans*, 1-9.

Broom, D. M., & Fraser, A. F. (2007). *Domestic animal behaviour and welfare* (Vol. 1). https://doi.org/10.1079/9781845932879.0000

Brown, C. J., Murphree, O. D., & Newton, J. E. O. (1978). The effect of inbreeding on human aversion in pointer dogs. *Journal of Heredity*, *69*(6), 362-365.

Budaev, S. V. (2010). Using principal components and factor analysis in animal behaviour research: caveats and guidelines. *Ethology*, *116*(5), 472-480.

Burghardt, W. F. (2003). Behavioral considerations in the management of working dogs. *Veterinary Clinics of North America - Small Animal Practice*, *33*(2), 417–446. https://doi.org/10.1016/S0195-5616(02)00133-X

Burn, C. C., & Weir, A. A. (2011). Using prevalence indices to aid interpretation and comparison of agreement ratings between two or more observers. *The Veterinary Journal*, *188*(2), 166-170.

Canali, E., & Keeling, L. (2009). Welfare Quality® project: from scientific research to on farm assessment of animal welfare. *Italian Journal of Animal Science*, *8*(sup2), 900-903.

Carmines, E. G., & Zeller, R. A. (1979). Reliability and validity assessment. *Sage university papers series. Quantitative applications in the social sciences. 17*.

Chicago Tribune (2017). Proposed bill would ban sale of commercially bred pets in Will, DuPage counties. Retrieved from http://www.chicagotribune.com/suburbs/naperville-sun/news/ct-nvs-commercial-pet-sale-ban-bill-st-1224-20171222-story.html

Christensen, E. L., Scarlett, J., Campagna, M., & Houpt, K. A. (2007). Aggressive behavior in adopted dogs that passed a temperament test. *Applied Animal Behaviour Science*, *106*(1-3), 85-95.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological assessment*, *6*(4), 284.

Clancy, E., & Rowan, A. (2003). Companion Animal Demographics in the United States: A Historical Perspective. *The State of the Animals II*, *2003*, 9–26.

Conley, M. J., Fisher, A. D., & Hemsworth, P. H. (2014). Effects of human contact and toys on the fear responses to humans of shelter-housed dogs. *Applied Animal Behaviour Science*, *156*, 62-69.

Cooper, J. J., Ödberg, F., & Nicol, C. J. (1996). Limitations on the effectiveness of enviornmental improvement in reducing steretypic behaviour in bank voles (Clethrionomys glareolus). *Applied Animal Behaviour Science*, *48*, 237–248.

Council for Agricultural Science and Technology [CAST]. (2018). *Scientific, Ethical, and Economic Aspects of Farm Animal Welfare.* Task Force Report 143. CAST, Ames, Iowa.

Dalmau, A., Fabrega, E., & Velarde, A. (2009). Fear assessment in pigs exposed to a novel object test. *Applied Animal Behaviour Science*, *117*(3-4), 173-180.

Dawkins, M. S. (1990). From an animal's point of view: motivation, fitness, and animal welfare. *Behavioral and brain sciences*, *13*(1), 1-9.

Dawkins, M. S. (1998). Evolution and Animal Welfare. *The Quarterly Review of Biology*, *73*(3), 305–328. https://doi.org/10.1086/420307

Dawkins, M. S. (2004). Using behaviour to assess animal welfare. *Animal welfare*, *13*(1), 3-7.

De Meester, R. H., De Bacquer, D., Peremans, K., Vermeire, S., Planta, D. J., Coopman, F., & Audenaert, K. (2008). A preliminary study on the use of the Socially Acceptable Behavior test as a test for shyness/confidence in the temperament of dogs. *Journal of Veterinary Behavior: Clinical Applications and Research*, *3*(4), 161-170.

De Meester, R. H., Pluijmakers, J., Vermeire, S., & Laevens, H. (2011). The use of the socially acceptable behavior test in the study of temperament of dogs. *Journal of Veterinary Behavior: Clinical Applications and Research*, *6*(4), 211–224. https://doi.org/10.1016/j.jveb.2011.01.003

Denenberg, S., & Landsberg, G. M. (2008). Effects of dog-appeasing pheromones on anxiety and fear in puppies during training and on long-term socialization. *Journal of the American Veterinary Medical Association*, *233*(12), 1874-1882.

Denham, H. D., Bradshaw, J. W., & Rooney, N. J. (2014). Repetitive behaviour in kennelled domestic dog: Stereotypical or not?. *Physiology & behavior*, *128*, 288-294.

De Rosa, G., Grasso, F., Winckler, C., Bilancione, A., Pacelli, C., Masucci, F., & Napolitano, F. (2015). Application of the Welfare Quality protocol to dairy buffalo farms: Prevalence and reliability of selected measures. *Journal of dairy science*, *98*(10), 6886-6896.

Dhabhar, F. S. (2000). Acute stress enhances while chronic stress suppresses skin immunity The role of stress hormones and leucocyte trafficking. *Annals of the New York Acadamy of Sciences*, *917.1*, 876–893.

Diederich, C., & Giffroy, J. M. (2006). Behavioural testing in dogs: A review of methodology in search for standardisation. *Applied Animal Behaviour Science*, *97*(1), 51–72. https://doi.org/10.1016/j.applanim.2005.11.018

Diesel, G., Brodbelt, D., & Pfeiffer, D. U. (2008). Reliability of assessment of dogs' behavioural responses by staff working at a welfare charity in the UK. *Applied Animal Behaviour Science*, *115*(3–4), 171–181. https://doi.org/10.1016/j.applanim.2008.05.005

Dohoo, I., Martin, W., & Stryhn, H. (2010). *Veterinary Epidemiological Research* (2nd ed.). VER.

Donald, R. D., Healy, S. D., Lawrence, A. B., & Rutherford, K. M. (2011). Emotionality in growing pigs: is the open field a valid test?. *Physiology & behavior*, *104*(5), 906-913.

Döring, D., Haberland, B. E., Ossig, A., Küchenhoff, H., Dobenecker, B., Hack, R., ... & Erhard, M. H. (2014). Behavior of laboratory beagles towards humans: Assessment in an encounter test and a simulation of experimental situations. *Journal of Veterinary Behavior: Clinical Applications and Research*, *9*(6), 295-303.

Döring, D., Haberland, B. E., Ossig, A., Küchenhoff, H., Dobenecker, B., Hack, R., ... & Erhard, M. H. (2016). Behavior of laboratory beagles: Assessment in a standardized behavior test using novel stimuli and situations. *Journal of Veterinary Behavior: Clinical Applications and Research*, *11*, 18-25.

Döring, D., Haberland, B. E., Bauer, A., Dobenecker, B., Hack, R., Schmidt, J., & Erhard, M. H. (2017a). Consistency in behavior: Evaluation of behavior tests in laboratory beagles. *Journal of Veterinary Behavior: Clinical Applications and Research*, *21*, 59-63.

Döring, D., Nick, O., Bauer, A., Küchenhoff, H., & Erhard, M. H. (2017b). Behavior of laboratory dogs before and after rehoming in private homes. *ALTEX-Alternatives to animal experimentation*, *34*(1), 133-147.

Döring, D., Roscher, A., Scheipl, F., Küchenhoff, H., & Erhard, M. H. (2009). Fear-related behaviour of dogs in veterinary practice. *The Veterinary Journal*, *182*(1), 38-43.

Douglas, C., Bateson, M., Walsh, C., Bédué, A., & Edwards, S. A. (2012). Environmental enrichment induces optimistic cognitive biases in pigs. *Applied Animal Behaviour Science*, *139*(1–2), 65–73. https://doi.org/10.1016/j.applanim.2012.02.018

Doyle, R. E., Fisher, A. D., Hinch, G. N., Boissy, A., & Lee, C. (2010). Release from restraint generates a positive judgement bias in sheep. *Applied Animal Behaviour Science*, *122*(1), 28–34. https://doi.org/10.1016/j.applanim.2009.11.003

Dreschel, N. A. (2010). The effects of fear and anxiety on health and lifespan in pet dogs. *Applied Animal Behaviour Science*, *125*(3-4), 157-162.

Duffy, D. L., & Serpell, J. A. (2012). Predictive validity of a method for evaluating temperament in young guide and service dogs. *Applied Animal Behaviour Science*, *138*(1-2), 99-109.

Duffy, D. L., Kruger, K. A., & Serpell, J. A. (2014). Evaluation of a behavioral assessment tool for dogs relinquished to shelters. *Preventive Veterinary Medicine*, *117*(3–4), 601–609. https://doi.org/10.1016/j.prevetmed.2014.10.003

Dunbar, I. F. (1977). Olfactory preferences in dogs: the response of male and female beagles to conspecific odors. *Behavioral Biology*, *20*(4), 471-481.

Duncan, I. J. (1996). Animal welfare defined in terms of feelings. *Acta Agriculturae Scandinavica. Section A. Animal Science. Supplementum (Denmark)*.

Duncan, I. J. (2004). A concept of welfare based on feelings. *The well-being of farm animals: Challenges and solutions*, 85-101.

Duncan, I. J. H. (2005). Science-based assessment of animal welfare: farm animals. *Revue Scientifique Et Technique-Office International Des Epizooties*, *24*(2), 483–492. Retrieved from https://www.researchgate.net/profile/Ian_Duncan6/publication/7410806_Science-based_assessment_of_animal_welfare_Farm_animals/links/549aee930cf2d6581ab2df00.pdf

Duncan, I. J. H., & Dawkins, M. S. (1983). The Problem of Assessing "Well-Being" and "Suffering" in Farm Animals. In *Indicators Relevant to Farm Animal Welfare* (pp. 13–24). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-009-6738-0_2

Edmonson, A. J., Lean, I. J., Weaver, L. D., Farver, T., & Webster, G. (1989). A Body Condition Scoring Chart for Holstein Dairy Cows. *Journal of Dairy Science*, *72*(1), 68–78. https://doi.org/10.3168/jds.S0022-0302(89)79081-0

El-Lethey, H., Aerni, V., Jungi, T. W., & Wechsler, B. (2000). Stress and feather pecking in laying hens in relation to housing conditions. *British Poultry Science*, *41*(1), 22–28. https://doi.org/10.1080/00071660086358

Fell, L. R., & Shutt, D. A. (1986). Adrenocortical Response of Calves To Transport Stress As Measured By Salivary Cortisol. *Canadian Journal of Animal Science*, *66*(3), 637–641. https://doi.org/10.4141/cjas86-070

Fennell, L. (1999). Beyond Overpopulation: A Comment on Zawistowski et al. and Salman et al. *Journal of Applied Animal Welfare Science*. Retrieved from http://www.tandfonline.com/doi/abs/10.1207/s15327604jaws0203_4

Flint, H. E., Coe, J. B., Serpell, J. A., Pearl, D. L., & Niel, L. (2018). Identification of fear behaviors shown by puppies in response to nonsocial stimuli. *Journal of Veterinary Behavior*, *28*, 17-24.

Fogsgaard, K. K., Røntved, C. M., Sørensen, P., & Herskin, M. S. (2012). Sickness behavior in dairy cows during Escherichia coli mastitis. *Journal of Dairy Science*, *95*(2), 630–638. https://doi.org/10.3168/jds.2011-4350

Foyer, P., Bjällerhag, N., Wilsson, E., & Jensen, P. (2014). Behaviour and experiences of dogs during the first year of life predict the outcome in a later temperament test. *Applied Animal Behaviour Science*, *155*, 93–100. https://doi.org/10.1016/j.applanim.2014.03.006

Foyer, P., Svedberg, A. M., Nilsson, E., Wilsson, E., Faresjö, Å., & Jensen, P. (2016). Behavior and cortisol responses of dogs evaluated in a standardized temperament test for military working dogs. *Journal of Veterinary Behavior: Clinical Applications and Research*, *11*, 7–12. https://doi.org/10.1016/j.jveb.2015.09.006

Fraser, D. (2003). Assessing animal welfare at the farm and group level: The interplay of science and values. *Animal Welfare*, *12*(4), 433–443.

Fraser, D. (2008). Understanding animal welfare. *Acta Veterinaria Scandinavica*. https://doi.org/10.1186/1751-0147-50-S1-S1

Fraser, D. (2009). Assessing animal welfare: different philosophies, different scientific approaches. *Zoo Biology: Published in affiliation with the American Zoo and Aquarium Association*, *28*(6), 507-518.

Fraser, D., & Weary, D. M. (2003). Quality of life for farm animals: linking science, ethics and animal welfare. *The Wellbeing of Farm Animals Challenges and Solutions*, 39–60. https://doi.org/10.1002/9780470344859

Fraser, D., Weary, D., Pajor, E. A., & Milligan, B. N. (1997). A scientific concept of animal welfare that reflects ethical concerns. *Animal Welfare*, *6*(2), 187–205. Retrieved from http://animalstudiesrepository.org/ethawel/1/

Friard, O., & Gamba, M. (2016). BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. *Methods in Ecology and Evolution*, *7*(11), 1325-1330.

Fox, M. W. (1971). Overview and critique of stages and periods in canine development. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, *4*(1), 37-54.

German, A. J., Holden, S. L., Moxham, G. L., Holmes, K. L., Hackett, R. M., & Rawlings, J. M. (2006). A simple, reliable tool for owners to assess the body condition of their dog or cat. *The Journal of nutrition*, *136*(7), 2031S-2033S.

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: a guide for non-statisticians. *International journal of endocrinology and metabolism*, *10*(2), 486.

Goddard, M. E., & Beilharz, R. G. (1984). A factor analysis of fearfulness in potential guide dogs. *Applied Animal Behaviour Science*, *12*(3), 253-265.

Goddard, M. E., & Beilharz, R. G. (1986). Early prediction of adult behaviour in potential guide dogs. *Applied Animal Behaviour Science*, *15*(3), 247–260. https://doi.org/10.1016/0168-1591(86)90095-X

Goold, C., & Newberry, R. C. (2017). Modelling personality, plasticity and predictability in shelter dogs. *Royal Society open science*, *4*(9), 170618.

Gosling, S. D. (2001). From mice to men: what can we learn about personality from animal research?. *Psychological bulletin*, *127*(1), 45.

Grandin, T., & Deesing, M. J. (2014). Behavioral genetics and animal science. *Genetics and the Behavior of Domestic Animals (Second Edition)*, 1-40.

Gray, J. A. (1987). The psychology of fear and stress. Cambridge: Camridge University Press.

Green, T. C., & Mellor, D. J. (2011). Extending ideas about animal welfare assessment to include "quality of life" and related concepts. *New Zealand Veterinary Journal*, *59*(6), 263–271. https://doi.org/10.1080/00480169.2011.610283

Hall, L. E. S., Robinson, S., Finch, J., & Buchanan-Smith, H. M. (2017). The influence of facility and home pen design on the welfare of the laboratory-housed dog. *Journal of pharmacological and toxicological methods*, *83*, 21-29.

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in quantitative methods for psychology*, *8*(1), 23.

Hansen, S. W., & Jeppesen, L. L. (2006). Temperament, stereotypies and anticipatory behaviour as measures of welfare in mink. *Applied Animal Behaviour Science*, *99*(1–2), 172–182. https://doi.org/10.1016/j.applanim.2005.10.005

Harding, E. J., Paul, E. S., & Mendl, M. (2004). Animal behaviour: cognitive bias and affective state. *Nature*, *427*(6972), 312.

Harvey, N. D., Craigon, P. J., Sommerville, R., McMillan, C., Green, M., England, G. C., & Asher, L. (2016). Test-retest reliability and predictive validity of a juvenile guide dog behavior test. *Journal of Veterinary Behavior: Clinical Applications and Research*, *11*, 65-76.

Haselton, M. G., Nettle, D., & Murray, D. R. (2015). The evolution of cognitive bias. *The handbook of evolutionary psychology*, 1-20.

Hemsworth, P. H., & Barnett, J. L. (1991). The effects of aversively handling pigs, either individually or in groups, on their behaviour, growth and corticosteroids. *Applied Animal Behaviour Science*, *30*(1–2), 61–72. https://doi.org/10.1016/0168-1591(91)90085-C

Hemsworth, P. H., Price, E. O., & Borgwardt, R. (1996). Behavioural responses of domestic pigs and cattle to humans and novel stimuli. *Applied animal behaviour science*, *50*(1), 43-56.

Hennessy, M. B., Davis, H. N., Williams, M. T., Mellott, C., & Douglas, C. W. (1997). Plasma cortisol levels of dogs at a county animal shelter. *Physiology and Behavior*, *62*(3), 485–490. https://doi.org/10.1016/S0031-9384(97)80328-9

Hennessy, M. B., Voith, V. L., Mazzei, S. J., Buttram, J., Miller, D. D., & Linden, F. (2001). Behavior and cortisol levels of dogs in a public animal shelter, and an exploration of the ability of these measures to predict problem behavior after adoption. *Applied Animal Behaviour Science*, *73*(3), 217–233. https://doi.org/10.1016/S0168-1591(01)00139-3

Hennessy, M. B., Zate, R., & Maken, D. S. (2008). Social buffering of the cortisol response of adult female guinea pigs. *Physiology & behavior*, *93*(4-5), 883-888.

Hetts, S., Derrell Clark, J., Calpin, J. P., Arnold, C. E., & Mateo, J. M. (1992). Influence of housing conditions on beagle behaviour. *Applied Animal Behaviour Science*, *34*(1–2), 137–155. https://doi.org/10.1016/S0168-1591(05)80063-2

Hewson, C. J., Hiby, E. F., & Bradshaw, J. W. S. (2007). Assessing quality of life in companion and kenelled dogs: a critical review. *Animal Welfare*, *16*(S), 89–95.

Hiby, E. F., Rooney, N. J., & Bradshaw, J. W. S. (2006). Behavioural and physiological responses of dogs entering re-homing kennels. *Physiology and Behavior*, *89*(3), 385–391. https://doi.org/10.1016/j.physbeh.2006.07.012

Hirsch, J. (1963). Behavior genetics and individuality understood. *Science*, *142*(3598), 1436-1442.

Hoehler, F. K. (2000). Bias and prevalence effects on kappa viewed in terms of sensitivity and specificity. *Journal of clinical epidemiology*, *53*(5), 499-503.

Horwitz, D. (1999). Counseling pet owners on puppy socialization and establishing leadership. *Veterinary Medicine-Bonner Springs Then Edwardsville-*, *94*, 149-156.

Howell, T. J., King, T., & Bennett, P. C. (2015). Puppy parties and beyond: the role of early age socialization practices on adult dog behavior. *Veterinary Medicine: Research and Reports*, *6*, 143-153.

Hradecká, L., Bartoš, L., Svobodová, I., & Sales, J. (2015). Heritability of behavioural traits in domestic dogs: A meta-analysis. *Applied Animal Behaviour Science*, *170*, 1-13.

Hubrecht, R. C. (1993). A comparison of social and environmental enrichment methods for laboratory housed dogs. *Applied Animal Behaviour Science*, *37*(4), 345-361.

Hubrecht, R. C. (2002). Comfortable quarters for dogs in research institutions. *Comfortable quarters for laboratory animals*, 56-64.

Hubrecht, R. C., Serpell, J. A., & Poole, T. B. (1992). Correlates of pen size and housing conditions on the behavior of kenneled dogs. *Applied Animal Behaviour Science*, *34*(4), 365–383.

Humane Society of the United States [HSUS]. (2018). The Horrible Hundred 2018. Retrieved from https://blog.humanesociety.org/wp-content/uploads/2018/05/Horrible-Hundred-May-9-2018-FULL-FINAL.pdf

Humane Society Veterinary Medical Association [HSVMA]. (2013). Veterinary Report of Puppy Mills. Retrieved from http://www.hsvma.org/assets/pdfs/hsvma_veterinary_report_puppy_mills.pdf

Hurnik, J. F. (1988). Welfare of farm animals. *Applied Animal Behaviour Science*, *20*(1–2), 105–117. https://doi.org/10.1016/0168-1591(88)90130-X

Hurt, M. J. (2016). Evaluating the physical welfare of dogs in commercial breeding facilities in the United States. M.S thesis, Purdue University, USA. https://docs.lib.purdue.edu/open_access_theses/967/

Hydbring-Sandberg, E., von Walter, L. W., Höglund, K., Svartberg, K., Swenson, L., & Forkman, B. (2004). Physiological reactions to fear provocation in dogs. *Journal of Endocrinology*, *180*(3), 439–448. https://doi.org/10.1677/joe.0.1800439

Indiana State Board of Animal Health [BOAH]. (2010). Commercial Dog Breeder and Broker. Retrieved from https://www.in.gov/boah/2549.htm

Jeppesen, L. L., Heller, K. E., & Bildsøe, M. (2004). Stereotypies in female farm mink (Mustela vison) may be genetically transmitted and associated with higher fertility due to effects on body weight. *Applied Animal Behaviour Science*, *86*(1–2), 137–143. https://doi.org/10.1016/j.applanim.2003.11.011

Jones, A. C., & Gosling, S. D. (2005). Temperament and personality in dogs (Canis familiaris): a review and evaluation of past research. *Applied Animal Behaviour Science*, *95*(1-2), 1-53.

Jones, B., & Boissy, A. (2011). Fear and other negative emotions. *Animal welfare*, 78-97.

Jones, R. B., Duncan, I. J. H., & Hughes, B. O. (1981). The assessment of fear in domestic hens exposed to a looming human stimulus. *Behavioural Processes*, *6*(2), 121-133.

Journal of the American Veterinary Medical Association [JAVMA] news. (2017). APHIS pulls welfare information from website. Retrieved from https://www.avma.org/News/JAVMANews/Pages/170401e.aspx

Kaufman, A. B., & Rosenthal, R. (2009). Can you believe my eyes? The importance of interobserver reliability statistics in observations of animal behaviour. *Animal Behaviour*.

Khansari, D. N., Murgo, A. J., & Faith, R. E. (1990). Effects of stress on the immune system (review). *Immunol Today*, *11*(5), 170–175. Retrieved from https://www.sciencedirect.com/science/article/pii/016756999090069L

Kiddie, J. L. ., & Collins, L. M. . (2014). Development and validation of a quality of life assessment tool for use in kennelled dogs (Canis familiaris). *Applied Animal Behaviour Science*, *158*, 57–68. https://doi.org/10.1016/j.applanim.2014.05.008

Kiley-Worthington, M. (1989). Ecological, ethological, and ethically sound environments for animals: Toward symbiosis. *Journal of Agricultural Ethics*, *2*(4), 323–347. https://doi.org/10.1007/BF01826810

King, T., Hemsworth, P. H., & Coleman, G. J. (2003). Fear of novel and startling stimuli in domestic dogs. *Applied Animal Behaviour Science*, *82*(1), 45-64.

King, T., Marston, L. C., & Bennett, P. C. (2012). Breeding dogs for beauty and behaviour: Why scientists need to do more to develop valid and reliable behaviour assessments for dogs kept as companions. *Applied Animal Behaviour Science*. https://doi.org/10.1016/j.applanim.2011.11.016

Kiyokawa, Y., & Hennessy, M. B. (2017). Comparative studies of social buffering: A consideration of approaches, terminology, and pitfalls. *Neuroscience & Biobehavioral Reviews*.

Klausz, B., Kis, A., Persa, E., Miklósi, Á., & Gácsi, M. (2014). A quick assessment tool for human-directed aggression in pet dogs. *Aggressive Behavior*, *40*(2), 178–188. https://doi.org/10.1002/ab.21501

Knierim, U., & Winckler, C. (2009). On-farm welfare assessment in cattle: Validity, reliability and feasibility issues and future perspectives with special regard to the Welfare Quality?? approach. *Animal Welfare*, *18*(4), 451–458. https://doi.org/10.4081/ijas.2005.223

Knol, B. W., Roozendaal, C., van den Bogaard, L., & Bouw, J. (1988). The suitability of dogs as guide dogs for the blind: criteria and testing procedures. *The Veterinary Quarterly*, *10*(3), 198–204. https://doi.org/10.1080/01652176.1988.9694171

Kobelt, A. J., Hemsworth, P. H., Barnett, J. L., & Butler, K. L. (2003). Sources of sampling variation in saliva cortisol in dogs. *Research in Veterinary Science*, *75*(2), 157–161. https://doi.org/10.1016/S0034-5288(03)00080-8

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine*, *15*(2), 155-163.

Kroll, T. L., Houpt, K. A., & Erb, H. N. (2004). The use of novel stimuli as indicators of aggressive behavior in dogs. *Journal of the American Animal Hospital Association*, *40*(1), 13-19.

Kutsumi, A., Nagasawa, M., Ohta, M., & Ohtani, N. (2013). Importance of puppy training for future behavior of the dog. *Journal of veterinary medical science*, *75*(2), 141-149.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, 159-174.

Lassen, J., Sandøe, P., & Forkman, B. (2006). Happy pigs are dirty! - conflicting perspectives on animal welfare. *Livestock Science*, *103*(3), 221–230. https://doi.org/10.1016/j.livsci.2006.05.008

Lawrence, A. B., & Terlouw, E. M. (1993). A review of behavioral factors involved in the development and continued performance of stereotypic behaviors in pigs. *Journal of Animal Science*, *71*(10), 2815–2825. https://doi.org/10.2527/1993.71102815x

Lewis, C. R. G., Hulbert, L. E., & McGlone, J. J. (2008). Novelty causes elevated heart rate and immune changes in pigs exposed to handling, alleys, and ramps. *Livestock Science*, *116*(1–3), 338–341. https://doi.org/10.1016/j.livsci.2008.02.014

Ley, J., Coleman, G. J., Holmes, R., & Hemsworth, P. H. (2007). Assessing fear of novel and startling stimuli in domestic dogs. *Applied animal behaviour science*, *104*(1-2), 71-84.

Lisberg, A. E., & Snowdon, C. T. (2009). The effects of sex, gonadectomy and status on investigation patterns of unfamiliar conspecific urine in domestic dogs, Canis familiaris. *Animal Behaviour*, *77*(5), 1147-1154.

Lisberg, A. E., & Snowdon, C. T. (2011). Effects of sex, social status and gonadectomy on countermarking by domestic dogs, Canis familiaris. *Animal behaviour*, *81*(4), 757-764.

Lore, R. K., & Eisenberg, F. B. (1986). Avoidance reactions of domestic dogs to unfamiliar male and female humans in a kennel setting. *Applied Animal Behaviour Science*, *15*(3), 261-266.

Lucidi, P., Bernabò, N., Panunzi, M., Dalla Villa, P., & Mattioli, M. (2005). Ethotest: A new model to identify (shelter) dogs' skills as service animals or adoptable pets. *Applied Animal Behaviour Science*, *95*(1-2), 103-122.

Lund, J. D., Agger, J. F., & Vestergaard, K. S. (1996). Reported behaviour problems in pet dogs in Denmark: age distribution and influence of breed and gender. *Preventive Veterinary Medicine*, *28*(1), 33-48.

Main, D. C. J., Kent, J. P., Wemelsfelder, F., Ofner, E., & Tuyttens, F. A. M. (2003). Applications for methods of on-farm welfare assessment. *Animal Welfare*, *12*(4), 523–528.

Manson, F. J., & Leaver, J. (1988). Incidence of concentrate amount on locomotion and lameness in dairy cattle. *Animal Production*, *47*(2), 185–190.

Marchant, J. N., Rudd, A. R., & Broom, D. M. (1997). The effects of housing on heart rate of gestating sows during specific behaviours. *Applied Animal Behaviour Science*, *55*(1–2), 67–78. https://doi.org/10.1016/S0168-1591(97)00022-1

Marder, A. R., Shabelansky, A., Patronek, G. J., Dowling-Guyer, S., & D'Arpino, S. S. (2013). Food-related aggression in shelter dogs: A comparison of behavior identified by a behavior evaluation in the shelter and owner reports after adoption. *Applied animal behaviour science*, *148*(1-2), 150-156.

Marks, I., & Nesse, R. M. (1994). Fear and fitness: An evolutionary analysis of anxiety disorders. *Ethology and sociobiology*, *15*(5-6), 247-261.

Martin, P., & Bateson, P. P. G. (1993). Measuring behaviour: An introductory guide.

Mason, G. J. (1991). Stereotypies: a critical review. *Animal behaviour*, *41*(6), 1015-1037.

Mason, G., Clubb, R., Latham, N., & Vickery, S. (2007). Why and how should we use environmental enrichment to tackle stereotypic behaviour?. *Applied Animal Behaviour Science*, *102*(3-4), 163-188.

Mason, G., & Mendl, M. (1993). Why is there no simple way of measuring animal welfare? *Animal Welfare*, *2*(4), 301–319. Retrieved from papers://031fc22f-ec67-46dc-8154-9b1109ad3ac4/Paper/p94

Mason, G. J., & Rushen, J. (2006). A decade or more's progress in the understanding of stereotypic behaviour. Stereotypies in Captive Animals.

McCulloch, S. P. (2013). A Critique of FAWC's Five Freedom as a Framework for the Analysis of Animal Welfare. *Journal of Agricultural and Environmental Ethics*, *26*(5), 959–975. Retrieved from https://idp.springer.com/authorize/casa?redirect_uri=https://link.springer.com/article/10.1007/s10806-012-9434-7&casa_token=CvkBBFx8CigAAAAA:sDOh8INcFbluNR-T-0zRWeXCjYX5L06yaVZNFB4wdPtcX-fummaM_cDuihkVrEL0GJ_W5Nxjtxcxvif6

McGreevy, P. D., & Nicholas, F. W. (1999). Some Practical Solutions to Welfare Problems in Dog Breeding. *Animal Welfare*, *8*(4), 329-341.

McHugh, M. L. (2013). The chi-square test of independence. *Biochemia medica: Biochemia medica*, *23*(2), 143-149.

McMillan, F. D. (2017). Behavioral and psychological outcomes for dogs sold as puppies through pet stores and/or born in commercial breeding establishments: Current knowledge and putative causes. *Journal of Veterinary Behavior: Clinical Applications and Research*, *19*, 14-26.

McMillan, F. D., Duffy, D. L., & Serpell, J. A. (2011). Mental health of dogs formerly used as "breeding stock" in commercial breeding establishments. *Applied Animal Behaviour Science*, *135*(1–2), 86–94. https://doi.org/10.1016/j.applanim.2011.09.006

McMillan, F. D., Duffy, D. L., Zawistowski, S. L., & Serpell, J. A. (2015). Behavioral and Psychological Characteristics of Canine Victims of Abuse. *Journal of Applied Animal Welfare Science*, *18*(1), 92–111. https://doi.org/10.1080/10888705.2014.962230

McMillan, F. D., Serpell, J. A., Duffy, D. L., Masaoud, E., & Dohoo, I. R. (2013). Differences in behavioral characteristics between dogs obtained as puppies from pet stores and those obtained from noncommercial breeders. *Journal of the American Veterinary Medical Association*, *242*(10), 1359–1363. https://doi.org/10.2460/javma.242.10.1359

Meagher, R. K. (2009). Observer ratings: Validity and value as a tool for animal welfare research. *Applied Animal Behaviour Science*, *119*(1–2), 1–14. https://doi.org/10.1016/j.applanim.2009.02.026

Mehrkam, L. R., & Wynne, C. D. (2014). Behavioral differences among breeds of domestic dogs (Canis lupus familiaris): Current status of the science. *Applied Animal Behaviour Science*, *155*, 12-27.

Mench, J. (1998). Why it is important to understand animal behavior. *ILAR Journal, 39*(1), 20–26. https://doi.org/10.1093/ilar.39.1.20

Mendl, M. (1991). Some problems with the concept of a cut-off point for determining when an animal's welfare is at risk. *Applied Animal Behaviour Science, 31*(1–2), 139–146. https://doi.org/10.1016/0168-1591(91)90161-P

Mendl, M., Burman, O. H. P., Parker, R. M. A., & Paul, E. S. (2009). Cognitive bias as an indicator of animal emotion and welfare: Emerging evidence and underlying mechanisms. *Applied Animal Behaviour Science*, *118*(3–4), 161–181. https://doi.org/10.1016/j.applanim.2009.02.023

Mendl, M., Brooks, J., Basse, C., Burman, O., Paul, E., Blackwell, E., & Casey, R. (2010). Dogs showing separation-related behaviour exhibit a 'pessimistic'cognitive bias. *Current Biology*, *20*(19), R839-R840.

Mendl, M., & Paul, E. S. (2004). Consciousness, emotion and animal welfare: insights from cognitive science. *Animal Welfare*, *13*(1), 17-25.

Menor-Campos, D. J., Molleda-Carbonell, J. M., & López-Rodríguez, R. (2011). Effects of exercise and human contact on animal welfare in a dog shelter. *Veterinary Record-English Edition*, *169*(15), 388.

Mertens, P. A., & Unshelm, J. (1996). Effects of group and individual housing on the behavior of kennelled dogs in animal shelters. *Anthrozoos*, *9*(1), 40–51. https://doi.org/10.2752/089279396787001662

Miele, M., Veissier, I., Evans, A., & Botreau, R. (2011). Animal welfare: establishing a dialogue between science and society. *Animal Welfare*, *20*, 103–117. Retrieved from https://www.researchgate.net/profile/Isabelle_Veissier/publication/230851710_Animal_welfare_Establishing_a_dialogue_between_science_and_society/links/0fcfd5054786e04696000000/Animal-welfare-Establishing-a-dialogue-between-science-and-society.pdf

Mielke Jr, P. W., Berry, K. J., & Johnston, J. E. (2009). Unweighted and weighted kappa as measures of agreement for multiple judges. *International Journal of Management*, *26*(2), 213.

Millman, S. T., Duncan, I. J. H., Stauffacher, M., & Stookey, J. M. (2004). The impact of applied ethologists and the International Society for Applied Ethology in improving animal welfare. In *Applied Animal Behaviour Science* (Vol. 86, pp. 299–311). https://doi.org/10.1016/j.applanim.2004.02.008

Mineka, S., & Sutton, S. K. (1992). Cognitive biases and the emotional disorders. *Psychological Science*, *3*(1), 65-69.

Mitchell, M. A., & Kettlewell, P. J. (1998). Physiological Stress and Welfare of Broiler Chickens in Transit: Solutions Not Problems! *Poultry Science*, *77*(12), 1803–1814. https://doi.org/10.1093/ps/77.12.1803

Moberg, G. P. (2000). Biological response to stress: impli- cation for animal welfare. In *The biology of animal stress.* (pp. 123–146).

Morgan, K. N., & Tromborg, C. T. (2007). Sources of stress in captivity. *Applied Animal Behaviour Science*, *102*(3–4), 262–302. https://doi.org/10.1016/j.applanim.2006.05.032

Mornement, K. M., Coleman, G. J., Toukhsati, S., & Bennett, P. C. (2010). A review of behavioral assessment protocols used by australian animal shelters to determine the adoption suitability of dogs. *Journal of Applied Animal Welfare Science*, *13*(4), 314–329. https://doi.org/10.1080/10888705.2010.483856

Mornement, K. M., Coleman, G. J., Toukhsati, S., & Bennett, P. C. (2014). Development of the behavioural assessment for re-homing K9’s (B.A.R.K.) protocol. *Applied Animal Behaviour Science*, *151*, 75–83. https://doi.org/10.1016/j.applanim.2013.11.008

Mornement, K. M., Coleman, G. J., Toukhsati, S. R., & Bennett, P. C. (2015). Evaluation of the predictive validity of the Behavioural Assessment for Re-homing K9's (BARK) protocol and owner satisfaction with adopted dogs. *Applied Animal Behaviour Science*, *167*, 35-42.

Morrow, M., Ottobre, J., Ottobre, A., Neville, P., St-Pierre, N., Dreschel, N., & Pate, J. L. (2015). Breed-dependent differences in the onset of fear-related avoidance behavior in puppies. *Journal of Veterinary Behavior: Clinical Applications and Research*, *10*(4), 286–294. https://doi.org/10.1016/j.jveb.2015.03.002

Murphy, J. a. (1995). Assessment of the temperament of potential guide dogs. *Anthrozoos*, *8*(4), 224–228 ST–Assessment of the temperament of pot. https://doi.org/10.2752/089279395787156536

Nakamura, K., Ishii, A., Kiyokawa, Y., Takeuchi, Y., & Mori, Y. (2016). The strain of an accompanying conspecific affects the efficacy of social buffering in male rats. *Hormones and behavior*, *82*, 72-77.

National Animal Interest Alliance [NAIA]. (n.d.) Policy Statement: Dogs. Retrieved from http://www.naiaonline.org/about-us/position-statements/dogs/

National Animal Interest Alliance [NAIA]. (2015). Survey of Shelter Dog Composition: Mutts vs. Purebreds. Retrieved from http://www.shelterproject.naiaonline.org/purebred/

National Public Radio (2017). Fears of a Black Market after California Bans Some Commercial Breeding. Retrieved from https://www.npr.org/2017/11/01/560933215/fears-of-a-black-market-after-calif-bans-some-commercial-breeding

Netto, W. J., & Planta, D. J. U. (1997). Behavioural testing for aggression in the domestic dog. *Applied Animal Behaviour Science*, *52*(3–4), 243–263. https://doi.org/10.1016/S0168-1591(96)01126-4

Olson, P. N., & Moulton, C. (1993). Pet (Dog and Cat) Overpopulation in the United-States. *Journal of Reproduction and Fertility,* 433–438. Retrieved from isi:A1993MB82100054

Otovic, P., & Hutchinson, E. (2015). Limits to using HPA axis activity as an indication of animal welfare. *Altex*. https://doi.org/10.14573/altex.1406161

Overall, K. L. (2014). The ethogram project. *Journal of Veterinary Behavior: Clinical Applications and Research*, *9*(1), 1-5.

Overall, K. L., & Dunham, A. E. (2002). Clinical features and outcome in dogs and cats with obsessive-compulsive disorder: 126 cases (1989–2000). *Journal of the American Veterinary Medical Association*, *221*(10), 1445-1452.

Pal, S. K. (2003). Urine marking by free-ranging dogs (Canis familiaris) in relation to sex, season, place and posture. *Applied Animal Behaviour Science*, *80*(1), 45-59.

Parrott, R. F., Misson, B. H., & Baldwin, B. A. (1989). Salivary cortisol in pigs following adrenocorticotrophic hormone stimulation: Comparison with plasma levels. *British Veterinary Journal*, *145*(4), 362–366. https://doi.org/10.1016/0007-1935(89)90034-1

Part, C. E., Kiddie, J. L., Hayes, W. A., Mills, D. S., Neville, R. F., Morton, D. B., & Collins, L. M. (2014). Physiological, physical and behavioural changes in dogs (Canis familiaris) when kennelled: Testing the validity of stress parameters. *Physiology and Behavior*, *133*, 260–271. https://doi.org/10.1016/j.physbeh.2014.05.018

Paul, E. S., Harding, E. J., & Mendl, M. (2005). Measuring emotional processes in animals: The utility of a cognitive approach. *Neuroscience and Biobehavioral Reviews*. https://doi.org/10.1016/j.neubiorev.2005.01.002

Patronek, G. J., & Glickman, L. T. (1994). Development of a model for estimating the size and dynamics of the pet dog population. *Anthrozoös*, *7*(1), 25-42.

Peduzzi, P., Concato, J., Feinstein, A. R., & Holford, T. R. (1995). Importance of events per independent variable in proportional hazards regression analysis II. Accuracy and precision of regression estimates. *Journal of clinical epidemiology*, *48*(12), 1503-1510.

Persson, M. E., Roth, L. S. V., Johnsson, M., Wright, D., & Jensen, P. (2015). Human-directed social behaviour in dogs shows significant heritability. *Genes, Brain and Behavior*, *14*(4), 337-344.

Pierantoni, L., Albertini, M., & Pirrone, F. (2011). Prevalence of owner-reported behaviours in dogs separated from the litter at two different ages. *The Veterinary Record*, *169*(18), 468.

Pichova, K., Nordgreen, J., Leterrier, C., Kostal, L., & Moe, R. O. (2016). The effects of food-related environmental complexity on litter directed behaviour, fear and exploration of novel stimuli in young broiler chickens. *Applied Animal Behaviour Science*, *174*, 83-89.

Planta, J. U. D., & De Meester, R. (2007). Validity of the Socially Acceptable Behavior (SAB) test as a measure of aggression in dogs towards non-familiar humans. *Vlaams Diergeneeskundig Tijdschrift*, *76*(5), 359-368.

Pongrácz, P., Molnár, C., & Miklósi, Á. (2010). Barking in family dogs: An ethological approach. *Veterinary Journal*, *183*(2), 141–147. https://doi.org/10.1016/j.tvjl.2008.12.010

Powell, C., Hemsworth, L. M., Rice, M., & Hemsworth, P. H. (2016). Comparison of methods to assess fear of humans in commercial breeding gilts and sows. *Applied Animal Behaviour Science*, *181*, 70-75.

Protopopova, A. (2016). Effects of sheltering on physiology, immune function, behavior, and the welfare of dogs. *Physiology and Behavior*. https://doi.org/10.1016/j.physbeh.2016.03.020

Protopopova, A., Mehrkam, L. R., Boggess, M. M., & Wynne, C. D. L. (2014). In-kennel behavior predicts length of stay in shelter dogs. *PLoS ONE*, *9*(12), e114319. https://doi.org/10.1371/journal.pone.0114319

Pullen, A. J., Merrill, R. J., & Bradshaw, J. W. (2012). The effect of familiarity on behaviour of kennel housed dogs during interactions with humans. *Applied animal behaviour science*, *137*(1-2), 66-73.

Reeder, D. M., & Kramer, K. M. (2005). Stress in Free-Ranging Mammals: Integrating Physiology, Ecology, and Natural History. *Journal of Mammalogy*, *86*(2), 225–235. https://doi.org/10.1644/BHE-003.1

Roelofs, S., Boleij, H., Nordquist, R. E., & van der Staay, F. J. (2016). Making Decisions under Ambiguity: Judgment Bias Tasks for Assessing Emotional State in Animals. *Frontiers in Behavioral Neuroscience*, *10*. https://doi.org/10.3389/fnbeh.2016.00119

Rollin, B. E. (2007). Cultural variation, animal welfare and telos. *Animal Welfare*, *16*(SUPPL.), 129–133.

Romeyer, A., & Bouissou, M. F. (1992). Assessment of fear reactions in domestic sheep, and influence of breed and rearing conditions. *Applied Animal Behaviour Science*, *34*(1), 93-119.

Rooney, N. J., Gaines, S. A., & Bradshaw, J. W. S. (2007). Behavioural and glucocorticoid responses of dogs (Canis familiaris) to kennelling: Investigating mitigation of stress by prior habituation. *Physiology and Behavior, 92*(5), 847–854. https://doi.org/10.1016/j.physbeh.2007.06.011

Rooney, N., Gaines, S., & Hiby, E. (2009). A practitioner's guide to working dog welfare. *Journal of Veterinary Behavior: Clinical Applications and Research*, *4*(3), 127–134. https://doi.org/10.1016/j.jveb.2008.10.037

Rowan, A. N. (1992). Shelters and pet overpopulation: A statistical black hole. *Anthrozoös*, *5*(3), 140-143.

Ruefenacht, S., Gebhardt-Henrich, S., Miyake, T., & Gaillard, C. (2002). A behaviour test on German Shepherd dogs: heritability of seven different traits. *Applied Animal Behaviour Science*, *79*(2), 113-132.

Rushen, J. (1991). Problems Associated with the Interpretation of Physiological Data in the Assessment of Animal-Welfare. *Applied Animal Behaviour Science*, *28*(4), 381–386.

Rushen, J., Pombourcq, E., & Passillé, A. M. de. (2007). Validation of two measures of lameness in dairy cows. *Applied Animal Behaviour Science*, *106*(1–3), 173–177. https://doi.org/10.1016/j.applanim.2006.07.001

Rushen, J., Taylor, A. A., & de Passillé, A. M. (1999). Domestic animals' fear of humans and its effect on their welfare. *Applied Animal Behaviour Science*, *65*(3), 285-303.

Sales, G., Hubrecht, R., Peyvandi, A., Milligan, S., & Shield, B. (1997). Noise in dog kennelling: Is barking a welfare problem for dogs? *Applied Animal Behaviour Science*, *52*(3–4), 321–329. https://doi.org/10.1016/S0168-1591(96)01132-X

Salman, M. D., New, Jr, J. G., Scarlett, J. M., Kass, P. H., Ruch-Gallie, R., & Hetts, S. (1998). Human and animal factors related to relinquishment of dogs and cats in 12 selected animal shelters in the United States. *Journal of Applied Animal Welfare Science*, *1*(3), 207-226.

Schatz, S., & Palme, R. (2001). Measurementof faecal cortisol metabolits in cats and dogs: A non-invasive method for evaluating adrenocortical function. *Vet Res Com*, *25*, 271–287. Retrieved from https://link.springer.com/article/10.1023/A:1010626608498

Scheifele, P., Martin, D., Clark, J. G., Kemper, D., & Wells, J. (2012). Effect of kennel noise on hearing in dogs. *American journal of veterinary research*, *73*(4), 482-489.

Schilder, M. B. H., & Van Der Borg, J. A. M. (2004). Training dogs with help of the shock collar: Short and long term behavioural effects. *Applied Animal Behaviour Science*, *85*(3–4), 319–334. https://doi.org/10.1016/j.applanim.2003.10.004

Shin, Y. J., & Shin, N. S. (2017). Relationship between sociability toward humans and physiological stress in dogs. *Journal of Veterinary Medical Science*, *79*(7), 1278-1283.

Scott, J. P. (1958). Critical periods in the development of social behavior in puppies. *Psychosomatic medicine*, *20*(1), 42-54.

Scott, J. P. (1962). Critical periods in behavioral development. *Science*, *138*(3544), 949-958.

Scott, J. P. (1963). The process of primary socialization in canine and human infants. *Monographs of the Society for Research in Child Development*, 1-47.

Scott, E. M., Nolan, A. M., & Fitzpatrick, J. L. (2001). Conceptual and Methodological Issues Related to Welfare Assessment: A Framework for Measurement. *Acta Agriculturae Scandinavica, Section A — Animal Science*, *51*(sup030), 5–10. https://doi.org/10.1080/090647001316922983

Scott, J. P., & Fuller, J. L. (2012). Genetics and the Social Behavior of the Dog.

Serpell, J. A. (2004). Factors influencing human attitudes to animal and their welfare. *Animal Welfare*, *13*, 145–151. Retrieved from issn 0962-7286

Serpell, J. A., & Hsu, Y. (2001). Development and validation of a novel method for evaluating behavior and temperament in guide dogs. *Applied Animal Behaviour Science, 72*(4), 347–364. https://doi.org/10.1016/S0168-1591(00)00210-0

Serpell, J., & Jagoe, J. A. (1995). Early experience and the development of behaviour. *The Domestic Dog: its evolution, behaviour and interactions with people. Cambridge University Press*, *Cambridge*, 79-102.

Sherman, B. L., Gruen, M. E., Case, B. C., Foster, M. L., Fish, R. E., Lazarowski, L., … Dorman, D. C. (2015). A test for the evaluation of emotional reactivity in Labrador retrievers used for explosives detection. *Journal of Veterinary Behavior: Clinical Applications and Research*, *10*(2), 94–102. https://doi.org/10.1016/j.jveb.2014.12.007

Shiverdecker, M. D., Schiml, P. A., & Hennessy, M. B. (2013). Human interaction moderates plasma cortisol and behavioral responses of dogs to shelter housing. *Physiology & behavior*, *109*, 75-79.

Sinn, D. L., Gosling, S. D., & Hilliard, S. (2010). Personality and performance in military working dogs: Reliability and predictive validity of behavioral tests. *Applied Animal Behaviour Science*, *127*(1–2), 51–65. https://doi.org/10.1016/j.applanim.2010.08.007

Slabbert, J. M., & Odendaal, J. S. J. (1999). Early prediction of adult police dog efficiency - A longitudinal study. *Applied Animal Behaviour Science*, *64*(4), 269–288. https://doi.org/10.1016/S0168-1591(99)00038-6

Smulder, D., Verbeke, G., Mormède, P., & Geers, R. (2006). Validation of a behavioural observation tool to assess pig welfare. *Physiology & Behavior*, *89*(3), 438–447. https://doi.org/http://dx.doi.org/10.1016/j.physbeh.2006.07.002

Sprecher, D. J., Hostetler, D. E., & Kaneene, J. B. (1997). A lameness scoring system that uses posture and gait to predict dairy cattle reproductive performance. *Theriogenology*, *47*(6), 1179–1187. https://doi.org/10.1016/S0093-691X(97)00098-8

Starling, M. J., Branson, N., Thomson, P. C., & McGreevy, P. D. (2013). Age, sex and reproductive status affect boldness in dogs. *The Veterinary Journal*, *197*(3), 868-872.

Stella, J. L., Lord, L. K., & Buffington, C. T. (2011). Sickness behaviors in response to unusual external events in healthy cats and cats with feline interstitial cystitis. *Journal of the American Veterinary Medical Association*, *238*(1), 67-73.

Stella, J., Hurt, M., Bauer, A., Gomes, P., Ruple, A., Beck, A., & Croney, C. (2018). Does Flooring Substrate Impact Kennel and Dog Cleanliness in Commercial Breeding Facilities?. *Animals, 8*(4), 59.

Stellato, A. C., Flint, H. E., Widowski, T. M., Serpell, J. A., & Niel, L. (2017). Assessment of fear-related behaviours displayed by companion dogs (Canis familiaris) in response to social and non-social stimuli. *Applied Animal Behaviour Science*, *188*, 84-90.

Stephen, J. M., & Ledger, R. A. (2005). An audit of behavioural indicators of poor welfare in kennelled dogs in the United Kingdom. *Journal of Applied Animal Welfare Science*, *8*(2), 79–95.

Stephen, J., & Ledger, R. (2007). Relinquishing dog owners' ability to predict behavioural problems in shelter dogs post adoption. *Applied Animal Behaviour Science*, *107*(1-2), 88-99.

Sternberg, S. (2004). Temperament evaluation of shelter dogs. Illinois, Tawzer Dog Videos.

Stigler, G. J. (1964). The intellectual and the marketplace. *Trans-Action*, *1*(5), 11–12. https://doi.org/10.1007/BF02813609

Svartberg, K. (2006). Breed-typical behaviour in dogs—historical remnants or recent constructs?. *Applied Animal Behaviour Science*, *96*(3-4), 293-313.

Swaisgood, R. R. (2007). Current status and future directions of applied behavioral research for animal welfare and conservation. *Applied Animal Behaviour Science*, *102*(3–4), 139–162. https://doi.org/10.1016/j.applanim.2006.05.027

Tabachnick, B. G., & Fidell, L. S. (2013). *Using Multivariate Statistics* (6th ed.). Pearson Education.

Tanida, H., Miura, A., Tanaka, T., & Yoshimoto, T. (1995). Behavioral response to humans in individually handled weanling pigs. *Applied Animal Behaviour Science*, *42*(4), 249–259. https://doi.org/10.1016/0168-1591(94)00545-P

Taylor, N. R., Main, D. C., Mendl, M., & Edwards, S. A. (2010). Tail-biting: a new perspective. *The Veterinary Journal*, *186*(2), 137-147.

Taylor, K. D., & Mills, D. S. (2006). The development and assessment of temperament tests for adult companion dogs. *Journal of Veterinary Behavior: Clinical Applications and Research*, *1*(3), 94–108. https://doi.org/10.1016/j.jveb.2006.09.002

Taylor, R. (1990). Interpretation of the correlation coefficient: a basic review. *Journal of diagnostic medical sonography*, *6*(1), 35-39.

Temple, D., Manteca, X., Velarde, A., & Dalmau, A. (2011). Assessment of animal welfare through behavioural parameters in Iberian pigs in intensive and extensive conditions. *Applied Animal Behaviour Science*, *131*(1–2), 29–39. https://doi.org/10.1016/j.applanim.2011.01.013

Terlouw, E. M. C., Lawrence, A. B., & Illius, A. W. (1991). Influences of feeding level and physical restriction on development of stereotypies in sows. *Animal Behaviour*, *42*(6), 981–991. https://doi.org/10.1016/S0003-3472(05)80151-4

The Inquirer (2009). USDA dog kennel inspection reports to be posted online. Retrieved from http://www.philly.com/philly/blogs/pets/USDA_dog_kennel_inspection_reports_now_online.html

Titulaer, M., Blackwell, E. J., Mendl, M., & Casey, R. A. (2013). Cross sectional study comparing behavioural, cognitive and physiological indicators of welfare between short and long term kennelled domestic dogs. *Applied Animal Behaviour Science*, *147*(1–2), 149–158. https://doi.org/10.1016/j.applanim.2013.05.001

Tod, E., Brander, D., & Waran, N. (2005). Efficacy of dog appeasing pheromone in reducing stress and fear related behaviour in shelter dogs. *Applied Animal Behaviour Science*, *93*(3–4), 295–308. https://doi.org/10.1016/j.applanim.2005.01.007

Tuber, D. S., Hennessy, M. B., Sanders, S., & Miller, J. A. (1996). Behavioral and glucocorticoid responses of adult domestic dogs (Canis familiaris) to companionship and social separation. *Journal of Comparative Psychology*, *110*(1), 103.

Tuber, D. S., Miller, D. D., Caris, K. A., Halter, R., Linden, F., & Hennessy, M. B. (1999). Dogs in animal shelters: problems, suggestions, and needed expertise. *Psychological Science*, *10*(5), 379-386.

United States Department of Agriculture [USDA]. (2016). Regulatory Compliance for Commercial Dog Breeders: Introductory Course. Retrieved from https://www.aphis.usda.gov/aphis/ourfocus/animalwelfare/caw/education+and+training/ct_commercial_dog_breeders

United States Department of Agriculture [USDA] Animal Care (2017). Dog Breeder Resource Guide. Retrieved from https://www.aphis.usda.gov/aphis/ourfocus/animalwel fare/caw.

United States Department of Agriculture-Animal and Plant Health Inspection Services [USDA-APHIS]. (2017). Animal Welfare Act Inspections. Retrieved from https://www.aphis.usda.gov/aphis/ourfocus/animalwelfare/sa_awa/ct_awa_inspections

Valsecchi, P., Barnard, S., Stefanini, C., & Normando, S. (2011). Temperament test for re-homed dogs validated through direct behavioral observation in shelter and home environment. *Journal of Veterinary Behavior: Clinical Applications and Research*, *6*(3), 161–177. https://doi.org/10.1016/j.jveb.2011.01.002

Van de Weerd, H., & Sandilands, V. (2008). Bringing the issue of animal welfare to the public: A biography of Ruth Harrison (1920–2000). *Applied Animal Behaviour Science*, *113*(4), 404-410.

Van den Berg, L., Schilder, M. B. H., & Knol, B. W. (2003). Behavior genetics of canine aggression: behavioral phenotyping of golden retrievers by means of an aggression test. *Behavior genetics*, *33*(5), 469-483.

van der Borg, J. A. M., Beerda, B., Ooms, M., de Souza, A. S., van Hagen, M., & Kemp, B. (2010). Evaluation of behaviour testing for human directed aggression in dogs. *Applied Animal Behaviour Science*, *128*(1–4), 78–90. https://doi.org/10.1016/j.applanim.2010.09.016

van der Borg, J. A., Graat, E. A., & Beerda, B. (2017). Behavioural testing based breeding policy reduces the prevalence of fear and aggression related behaviour in Rottweilers. *Applied Animal Behaviour Science*, *195*, 80-86.

van der Borg, J. A. M., Netto, W. J., & Planta, D. J. U. (1991). Behavioural testing of dogs in animal shelters to predict problem behaviour. *Applied Animal Behaviour Science*, *32*(2–3), 237–251. https://doi.org/10.1016/S0168-1591(05)80047-4

Van Putten, G. (2000). An ethological definition of animal welfare with special emphasis on pig behaviour. *Veeru.Reading.Ac.Uk*, 89–103. Retrieved from http://www.veeru.rdg.ac.uk/organic/proc/vanP.htm

Veissier, I., & Boissy, A. (2007). Stress and welfare: Two complementary concepts that are intrinsically related to the animal's point of view. *Physiology and Behavior*, *92*(3), 429–433. https://doi.org/10.1016/j.physbeh.2006.11.008

Vincent, I. C., & Leahy, R. A. (1997). Real-time non-invasive measurement of heart rate inworking dogs: a technique with potential applications in the objective assessment of welfare problems. *The Veterinary Journal*, *153*(2), 179–183. https://doi.org/10.1016/S1090-0233(97)80038-X

Viscardi, A. V., Hunniford, M., Lawlis, P., Leach, M., & Turner, P. V. (2017). Development of a Piglet Grimace Scale to Evaluate Piglet Pain Using Facial Expressions Following Castration and Tail Docking: A Pilot Study. *Frontiers in Veterinary Science*, *4*. https://doi.org/10.3389/fvets.2017.00051

von Borell, E., Langbein, J., Després, G., Hansen, S., Leterrier, C., Marchant-Forde, J., … Veissier, I. (2007). Heart rate variability as a measure of autonomic regulation of cardiac activity for assessing stress and welfare in farm animals - A review. *Physiology and Behavior*, *92*(3), 293–316. https://doi.org/10.1016/j.physbeh.2007.01.007

Waiblinger, S., Boivin, X., Pedersen, V., Tosi, M. V., Janczak, A. M., Visser, E. K., & Jones, R. B. (2006). Assessing the human–animal relationship in farmed species: a critical review. *Applied Animal Behaviour Science*, *101*(3-4), 185-242.

Wauthier, L. M., & Williams, J. M. (2018). Using the Mini C-BARQ to Investigate the Effects of Puppy Farming on Dog Behaviour. *Applied Animal Behaviour Science*.

Wells, D. L. (2004). A review of environmental enrichment for kennelled dogs, Canis familiaris. *Applied Animal Behaviour Science*, *85*(3-4), 307-317.

Wells, D. L., & Hepper, P. G. (1999). Male and female dogs respond differently to men and women. *Applied Animal Behaviour Science*, *61*(4), 341-349.

Wells, D. L., Hepper, P. G., Milligan, A. D. S., & Barnard, S. (2017). Cognitive bias and paw preference in the domestic dog (Canis familiaris). *Journal of Comparative Psychology*, *131*(4), 317–325. https://doi.org/10.1037/com0000080

Wemelsfelder, F. (2007). How Animals Communicate Quality of Life : The Qualitative Assessment of Behaviour. *Animal Welfare*, *16*(S), 25–31. https://doi.org/Article

Wichman, A., Keeling, L. J., & Forkman, B. (2012). Cognitive bias and anticipatory behaviour of laying hens housed in basic and enriched pens. *Applied Animal Behaviour Science*, *140*(1–2), 62–69. https://doi.org/10.1016/j.applanim.2012.05.006

Wiegers, G. J., Stec, I. E. M., Sterzer, P., & Reul, J. M. H. M. (2005). Glucocorticoids and the immune response. *Techniques in the Behavioral and Neural Sciences*, *15*(PART 2), 175–191. https://doi.org/10.1016/S0921-0709(05)80055-0

Wiepkema, P. R., & Koolhaas, J. M. (1993). Stress and animal welfare. *Animal welfare*, *2*(3), 195-218.

Wilsson, E. (2016). Nature and nurture—how different conditions affect the behavior of dogs. *Journal of Veterinary Behavior: Clinical Applications and Research*, *16*, 45-52.

Wilsson, E., & Sundgren, P. E. (1998). Behaviour test for eight-week old puppies—heritabilities of tested behaviour traits and its correspondence to later behaviour. *Applied Animal Behaviour Science*, *58*(1-2), 151-162.

World Organization for Animal Health [OIE]. (2017). Animal welfare at a glance. Retrieved from http://www.oie.int/en/animal-welfare/animal-welfare-at-a-glance/

Zawistowski, S., Morris, J., Salman, M. D., & Ruch-Gallie, R. (1998). Population dynamics, overpopulation, and the welfare of companion animals: new insights on old and new data. *Journal of Applied Animal Welfare Science : JAAWS*, *1*(3), 193–206. https://doi.org/10.1207/s15327604jaws0103_1