



University of Pennsylvania
ScholarlyCommons

Master of Science in Animal Welfare and
Behavior Capstone Projects

School of Veterinary Medicine

11-30-2022

Assessment of Welfare of Shelter Cats and Dogs: Inter-Observer Reliability Using the Penn Vet Quality of Life Tool App

Edienna Perez
perezep8@gmail.com

Follow this and additional works at: https://repository.upenn.edu/awbcapstones_vet

Recommended Citation

Perez, Edienna, "Assessment of Welfare of Shelter Cats and Dogs: Inter-Observer Reliability Using the Penn Vet Quality of Life Tool App" (2022). *Master of Science in Animal Welfare and Behavior Capstone Projects*. 9.

https://repository.upenn.edu/awbcapstones_vet/9

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/awbcapstones_vet/9
For more information, please contact repository@pobox.upenn.edu.

Assessment of Welfare of Shelter Cats and Dogs: Inter-Observer Reliability Using the Penn Vet Quality of Life Tool App

Abstract

To increase positive shelter animal welfare, we must find the causes of poor animal welfare on an individual level by using a reliable and valid assessment. In order to analyze inter-observer reliability, two raters of different backgrounds simultaneously assessed the welfare of 80 randomly selected animals (40 cats and 40 dogs) in a US shelter using the Penn Vet Quality of Life Tool App. Data collection occurred one to three times a week between the hours of 8 am and 2 pm from December 2021 to March 2022. Environmental variables showed good agreement, most physical health variables showed moderate agreement (e.g., fecal scoring and dog coat condition) and some behavioral variables (e.g., fear and aggression) frequently showed poor agreement. The poor level of reliability may be driven by factors such as the rater's experience and understanding of animal behavior, their familiarity with the shelter animals, familiarity with the app, or differences in the rater's gender. To address these challenges and better understand the factors that impact the reliability of animal welfare assessments, we recommend future studies: (1) train raters before data collection, (2) recruit more raters, and (3) record animals during live assessments for later review.

Degree Type

Thesis

Degree Name

Master of Science in Animal Welfare and Behavior (MSc AWB)

First Advisor

Lauren Powell, PhD

Second Advisor

Brittany Watson, MS, VMD, PhD, DACVPM

Keywords

inter-observer reliability, shelter animals, animal welfare assessment, Penn Vet, quality of life

**ASSESSMENT OF WELFARE OF SHELTER CATS AND DOGS:
INTER-OBSERVER RELIABILITY USING THE
PENN VET QUALITY OF LIFE TOOL APP**

Edienna Perez

Capstone Thesis

Master of Science (MSc) in Animal Welfare and Behavior Program

School of Veterinary Medicine, University of Pennsylvania

Mentors: Dr. Lauren Powell, PhD¹ and Dr. Brittany Watson, MS/VMD/PhD/DACVPM¹

Date: November 30, 2022.

¹ Shelter Medicine Program, School of Veterinary Medicine, University of Pennsylvania, Philadelphia, Pennsylvania

Abstract

To increase positive shelter animal welfare, we must find the causes of poor animal welfare on an individual level by using a reliable and valid assessment. In order to analyze inter-observer reliability, two raters of different backgrounds simultaneously assessed the welfare of 80 randomly selected animals (40 cats and 40 dogs) in a US shelter using the Penn Vet Quality of Life Tool App. Data collection occurred one to three times a week between the hours of 8 am and 2 pm from December 2021 to March 2022. Environmental variables showed good agreement, most physical health variables showed moderate agreement (e.g., fecal scoring and dog coat condition) and some behavioral variables (e.g., fear and aggression) frequently showed poor agreement. The poor level of reliability may be driven by factors such as the rater's experience and understanding of animal behavior, their familiarity with the shelter animals, familiarity with the app, or differences in the rater's gender. To address these challenges and better understand the factors that impact the reliability of animal welfare assessments, we recommend future studies: (1) train raters before data collection, (2) recruit more raters, and (3) record animals during live assessments for later review.

Keywords: inter-observer reliability, shelter animals, animal welfare assessment, Penn Vet, quality of life

Dedication

I dedicate this thesis to my mother, Ana I. Caraballo, and my best friend Bella.

Mami: I believe that this Master's degree is yours as much as it is mine. Thank you for staying awake until 2 am to make sure I didn't fall asleep while writing my papers. I appreciate you bringing me food and coffee during class. I will always be grateful for those "believe in yourself beba" texts. Because of you, I have learned to be kind, thorough, and resilient.

Bella: I owe my whole career to you. I have learned so much about dog behavior because of you. I hope that I will help people and animals replicate the bond that we have. From middle school to the present day, you lift my spirits and never leave my side. Thank you for being the sassiest companion.

Acknowledgments

Throughout my data collection and writing of this thesis, I have received support that is unparalleled.

I would like to extend a heartfelt thanks to our Associate Dean for One Health, Dr. Jennifer Punt, who truly gave me the best welcome to this graduate program and continued to make me feel empowered throughout my academic career at Penn.

Thank you to my professor, Dr. Thomas Parsons, for hosting thought-provoking discussions and helping me grow as a critical thinker. You have shown me that solving one problem is enough to leave a foundation for others to learn more.

Special thanks to my mentors Dr. Lauren Powell and Dr. Brittany Watson, the Co-Head of the Behavior, Welfare, and Human-Animal Interaction Research Division, for giving me access to the Penn Vet Quality of Life in Shelters App. I truly appreciate your guidance during data collection and analysis, as well as in writing this thesis. You both are trailblazers and I am honored to have worked with you.

Shelter A, especially the Behavior Team, thank you for going above and beyond to find raters that could assist me. Huge thank you to the raters/staff who dedicate their time to our common goal which is to save and rehome animals. I also want to thank the Foster Team for giving me the opportunity to foster animals (Puffy, Ashley and Pepe) who needed help overcoming their behavioral and medical issues. I gained a new appreciation for shelter animal welfare and medicine.

Thank you to Ryan Miller, the Director for Academic Support at the Weingarten Center, for promptly sharing feedback on my thesis and providing a safe space for me to discuss my obstacles with research.

I could not have completed this thesis without my dear friends, Asia T. Weeks, Kevin Cataldo, Kiara L. Graham, and Brielle Nieves who reminded me of my purpose and encouraged breaks when I needed rest. I would also like to thank my sisterhood within Chi Upsilon Sigma National Latin Sorority, Inc for continuously educating, elevating, and empowering me so that I can take on any new endeavor.

Lastly, I want to thank all Latina researchers who came before me, especially because they did not allow their circumstance to outweigh their passion. I am now a part of the 4% of Latinas in the US who have a Master's degree and hope to help other women navigate graduate school. Thank you all for your substantial contribution to science and positive impact on younger women.

Introduction

Animal Welfare in Shelters

It is estimated that 6.3 million animals enter the U.S. shelter system every year (ASCPA, 2019). Out of the 6.3 million animals, approximately 3.2 million are cats and 3.1 million are dogs (ASCPA, 2019). Animals of different temperaments, ages, breeds, etc, may be surrendered, found, or transferred from other facilities, and end up in the U.S. shelter system.

The primary aim of a shelter is to provide temporary housing to companion animals before they can be placed in a new home. Shelters are known to be unpredictable, have limited resources and may receive negative publicity about the quality of life of their animals. Kennels usually lack space, mental stimulation and limit interaction between dogs, humans, and other environments (Taylor et al., 2007a; Barrera et al., 2010). Some animals deteriorate to a point where they may be deemed unadoptable either due to medical or behavioral issues. In some cases, the result is euthanasia. The shelter environment can be stressful for animals and can negatively impact their quality of life (QoL). Wiseman-Orr et al. (2006) defined quality of life (QoL) as

the subjective and dynamic evaluation by the individual of its circumstances (internal and external) and the extent to which these meet its expectations (which may be innate or learned and which may or may not include anticipation of future events), which results in, or includes, an affective (emotional) response to those circumstances. (p.578)

To decrease the negative effects on welfare, increase adoptions, and decrease mortality rates in shelters, we must find the causes of poor animal welfare on an individual level. According to Broom (2011), "animal welfare is a potentially

measurable quality of a living animal at a particular time" (p. 122). Identifying causes of negative animal welfare can be time-consuming and require a lot of assistance from the shelter staff. Assessing welfare is complex, therefore, Botreau et al (2007) stated that "an overall assessment requires a multicriteria evaluation" (p. 225). Behavioral, physical, and mental health assessments give us insight into not only what is occurring within the animal, but also how their environment affects them holistically (Barnard et al., 2016).

Assessing Quality of Life in Animal Shelters

To properly assess QoL, the goal is to develop a practical tool with measures that are easy to record, can be applied objectively, and assessed in a timely manner (Barnard et al., 2016). QoL assessments began in humans (Maille et al., 1997) and were adapted to animals. The QoL of farm animals was the first to be assessed in the Welfare Quality Consortium. More recently, Kiddie & Collins (2014) and Barnard (2016) also developed animal welfare assessments for use in shelter dogs to give better understanding of the shortcomings within the shelter system and highlight how we could improve the recording of animal-based measures.

Kiddie & Collins (2014) investigated the reliability of their animal welfare assessment and found poor internal consistency. Animal-based measures (e.g., behavior) were observed four times in one day. Two raters were allowed to observe the dogs simultaneously during the first three observations (lasting 2 minutes) yet for the last one (lasting 4 minutes), they had to assess the dogs separately. During the last session, raters also recorded observations about the dogs' physical health. The study included 74 assessors, who were staff members with varying experience, and 202 dogs, who were chosen after alphabetical sorting. This sorting method was used to

avoid choosing animals that had a specific appearance or temperament. Yet, randomizing a list of all animals across the 13 centers could have provided a more diverse list of subjects. During the live observation, assessors filled out a questionnaire which helped increase the accuracy of their answers. Yet the observers were unable to revisit questions that were included earlier in the assessment to add to their observation of animal-based measures. The average amount of time to complete the assessment was 75.32 minutes. This is a long time to engage an observer during a single session but also very time-consuming for shelter staff on a long-term basis. Although training was provided, it occurred two months prior to the observation sessions which could affect the assessors' memory of the protocol. The protocol may have been easier to apply in a shelter setting if it had originally been reviewed by a large group of experts. Assessors were offered training at each center but due to the shelter's inability to spare staff and/or time, up to two staff members at each center attended the training. The staff members who did not attend the training were given a written research procedure and ethogram.

A few years later, dogs in long-term shelters were observed to test the reliability, validity, and feasibility of a new animal welfare assessment called the Shelter Quality Protocol. The assessment, which took approximately 172.41 minutes, evaluated environmental, animal-based, and management-based measures. Barnard et al., (2016) believed that their scoring system was simple, objective, and included ordinal scales. Four experts were chosen to set the 'gold standard' by reaching a consensus after reviewing video/images and prior to developing a training program. Inter-observer reliability was assessed when four female veterinarians participated in the training program by sharing their ratings for the videos/images. The study also included an analysis of the level of agreement between the 'gold standard' and the veterinarians' responses during training. The results may be biased by a lack of a

larger and more diverse group during training (i.e., male veterinarians, behavior experts, etc). Animals were randomly selected from the shelter (mean population = 89.27 dogs) to reduce bias. 29 assistants helped 29 assessors in 29 shelters which provided the opportunity to assess different animals in similar environments.

Before arriving at each shelter, both groups of observers had to pass an exam after classroom and field training. The small lapse of time between training and observations allowed better recall than the training protocol set by Kiddie and Collins (2014). On the other hand, there were a few shortcomings in this study. The methodology of observing videos/images in training sessions, compared to observing live animals during data collection may present different values of inter-observer reliability. The inter-observer reliability was substantial for the Body Condition Score (BCS), reaction to human test and barking level. Moderate agreement was seen when assessors were asked about shivering. A change in the scoring system (from ordinal to binary) was necessary for some variables because the level of agreement was fair (e.g., skin condition and panting). The new scoring system for skin condition also asked for the presence or absence of lesions rather than the number of lesions. This would help increase the level of agreement because, in the previous version, the inter-observer reliability was affected by the visibility and variability of the lesion. Recommendations for future studies include repeating this protocol during the winter to see if the inter-observer reliability changes due to seasons. Another shortcoming occurred during live observation of the health parameters (e.g., the presence or absence of diarrhea or coughing). Some assessors and/or assistants were recording animal-based measures while considering all animals in the pen and not animals individually. Lastly, the authors found it difficult to analyze the frequency of repetitive behaviors due to the small sample size of animals showing stereotypies.

The Penn Vet QoL tool addresses some of the shortcomings from previous studies, such as those from Kiddie and Collins (2014) and Barnard et al (2016). In contrast to these studies, the new tool cut the assessment time by more than half. The Penn Vet QoL tool also includes an assessment for individual animals similar to the UK Assessment for QoL (Kiddie and Collins, 2014) but in contrast to the Shelter Quality Protocol (Barnard et al., 2016). It also provides the opportunity for the environmental, behavioral, and physical health of each animal to be evaluated. Barnard et al (2016) included ordinal scales in their assessment yet this QoL tool includes ordinal scales with more details within the behavioral and physical health categories which could improve the accuracy of the data. The Penn Vet QoL tool may allow future raters to record extensive information about individual animals in a short amount of time and with a smaller number of raters, which could help shelters increase positive animal welfare and number of adoptions.

Interobserver Reliability

The design of an animal welfare assessment can have a direct effect on the outcome of an assessment, therefore affecting the future welfare of the animal (Taylor and Mills, 2006). An unreliable and invalid behavioral assessment can lead to serious consequences before adoption, such as staff injury and wrongful euthanasia, and after adoption, such as owner injury, relinquishment, and euthanasia (Scarlett et. al, 2007; Patronek et. al, 2019). It is therefore important to assess the reliability and validity of any new tool that is designed to measure the individual welfare of shelter animals.

Interobserver reliability refers to reliability between two or more raters. High inter-observer reliability means that the assessment can provide us with consistent

information about an animal's quality of life over time (Carmines and Zeller, 1979). The aim of this study was to test if 2 raters could simultaneously assess the welfare of each animal in a similar way, regardless of their educational levels and backgrounds.

Methods

The Penn Vet Quality of Life Tool

The Penn Vet Quality of Life Tool was developed with the intention that the QoL of shelter animals could be evaluated both at a certain moment and over time and if their QoL was negative, shelter staff could prioritize getting them adopted or come up with an individual intervention plan. The continuous assessments should also allow shelter staff to evaluate if the intervention plans were improving the animals' QoL.

The Penn Vet Quality of Life in Shelters Tool is available in Apple Store and Google Play making it accessible to all shelters. It includes an assessment that is more applicable to shelters in the United States. The assessments can be completed by several raters, in under 20 minutes, and by any staff member or volunteer (regardless of their educational background) in the shelter setting. Each rater can access the app via their cell phone. The app includes a questionnaire for each animal regarding welfare categories which include environmental, behavioral, and physical health. These three categories will be used to assess the quality of life (QoL) of the animals in the shelter. For every welfare category, there are several binary, ordinal, categorical, multiple-choice, and open-ended questions within each category.

Background

To begin the analysis of the QoL of each shelter animal, we recorded the animals' characteristics (sex, altered status, breed, age, weight, length of stay, intake type). Length of stay (LOS) is the number of days that the animal has been at Shelter A since their most recent arrival and was calculated as the number of days between intake and the assessment.

Environmental Health

Guidelines of husbandry (ex. enclosure design) were derived from the Association of the Shelter Veterinarians' Guidelines for Standards of Care in Animal Shelters and Five Freedoms from the Farm Animal Welfare Council.

To analyze the environmental health of a shelter dog, the QoL tool records details about the feeding schedule, food type, water availability, availability of fresh water, type of housing, enrichment schedule, enrichment types, presence of a compartment layout, floor type, space availability, bedding condition, cleanliness, exercise frequency, exercise duration and noise level in a kennel.

To analyze the environmental health of a shelter cat, the QoL tool records details about the feeding schedule, food type, water availability, availability of fresh water, type of housing, presence of a compartment layout, hiding places, elevated spaces, size of the cattery, number of cats in the cattery, litter availability, and food availability.

Behavioral Health

The Penn Vet QoL tool was developed using previous research that included tools that assessed animal welfare such as the Canine Behavioral Assessment and

Research Questionnaire (C-BARQ), ethograms of poor welfare indicators (Beerda et al., 1997; Beerda et al., 1998; Beerda et al., 2000; Stephen & Ledger, 2005), and the FAS and frustration scale (see introduction). The C-BARQ was developed by researchers within the Center for the Interaction of Animals and Society of the University of Pennsylvania. Dr. James Serpell was one of those researchers and recently helped develop this QoL tool at Penn Vet.

To analyze the behavioral health of a shelter dog, the QoL tool records details about behavior medications, behavior modification plans, food consumption, activity prior to feeding, overnight habits, in-kennel behavior towards handlers, in-kennel behavior towards other dogs, out-of-kennel behavior towards handlers, and reaction towards toys. Specific behaviors were assessed such as fear, anxiety, arousal, etc. This category included instructions (e.g., approaching the kennel without making eye contact) before tests were performed. Tests included standing in front of the kennel without giving attention, crouching near the kennel for 30 seconds, walking a social dog by the test dog's kennel, taking a dog out of the kennel, and into a separate area, etc. The tool included 5-point scale questions with descriptions at points 1, 3 and 5. Each description included information about body language signs.

To analyze the behavioral health of a shelter cat, the QoL tool records details about litter habits, food consumption, activity prior to feeding, overnight habits, posture, in-cage behavior towards other cats, in-cage behavior towards handlers, and reaction to toys. This category included instructions (e.g., extend one hand towards the cat and let the cat sniff your hand) before petting the cat. Tests included standing in front of the cage door while greeting the cat in a high-pitched tone, petting the cat, and using a string toy to engage them.

Physical Health

The body condition chart assessment by The American Animal Hospital Association (AAHA) will be used in the physical health category. In addition, the fecal scoring system used in this assessment is the WALTHAM fecal scoring system.

To analyze physical health, the QoL tool included mostly all ordinal questions (e.g., presence of a URI) with the exception of a binary question (e.g., whether or not a cat urinated). The raters were asked to rate the cats' and dogs' body condition scores (BCS), coat conditions, and feces. Raters were also asked if an upper respiratory infection (URI), lameness, or other medical condition was present.

Procedure

Before arriving at Shelter A on data collection days, the shelter team sent over an animal inventory list to the researcher the night before. Animals (both cats and dogs) were excluded if under the age of 6 months. Any animals that were listed as seized were excluded. All the animals, 6 months and above and legally under the care of the shelter, were placed into an excel spreadsheet and numbered.

Randomizer.org was used to randomize the list of animals into a list of 15 animals. A new excel spreadsheet was created with each animal's information (e.g., breed, source, length of stay) and distributed to raters the next morning.

Data collection occurred one to three times a week between the hours of 8 am and 2 pm from December 2021 to March 2022. The data collection typically began between the hours of 8 am and 11 am which is the time frame for cleaning in the shelter. It was important to arrive early enough to assess the animals because parts of the questionnaire asked about the kennel's condition (e.g., ripped sheets) and the

animal's behavior (e.g., amount of food/breakfast consumed). The sessions typically lasted between 3 to 4 hours. Within those 3 to 4 hours, 6 to 8 animals were assessed. Seven animals (6 in one session and 1 in another session) were assessed using Qualtrics instead of the QoL app due to technical difficulties.

Study Site

The shelter that was involved in this study will be referred to as Shelter A. Shelter A is an open-admission shelter located in New Jersey, USA, described further in Table 1. The shelter took in approximately 3,190 dogs and 2,081 cats within the last calendar year (January 1st - December 31st).

At Shelter A, 160 welfare assessments were completed by 13 raters in groups of two: Observer A remained the same for all assessments, unlike Observer B. There were 13 data collection days in total, which occurred between December 2021 and March 2022. During each assessment, both raters observed the same animal at the same time using a mobile application.

Table 1. Shelter's characteristics

Shelter Characteristics	
Shelter Type	Mixed
Admission Type	Open
Total animal intake in the last calendar year (Jan 1st - Dec 31st)	5994
Total dog intake in the last calendar year (Jan 1st - Dec 31st)	3190
Average dog length of stay (days) in the last calendar year (Jan 1st - Dec 31st)	9.2
Other animal intakes in the last calendar year (Jan 1st - Dec 31st)	723
Total cat intake in the last calendar year (Jan 1st - Dec 31st)	2081
Average cat length of stay (days) in the last calendar year (Jan 1st - Dec 31st)	36.7
Other animal length of stay (days) in the last calendar year (Jan 1st - Dec 31st)	19.1

Animal Characteristics

Date of birth, sex, breed, intake source, intake date, and outcome date were retrieved from PetPoint which is where Shelter A maintains its animal records (PetPoint Data Management System, Version 5, Pethealth Software Solutions Inc., USA).

Statistical Analysis

The descriptive characteristics of each animal (i.e., sex, breed, age, fixed status, weight, length of stay, intake type, outcome type) are provided in Table 2. To calculate average values, all values for each characteristic were added and divided by the total number of animals in that species (i.e., 40 dogs or 40 cats).

To analyze inter-observer reliability, Intraclass Correlation Coefficients (ICC) were calculated using two-way random-effects models based on absolute agreement from a single rater. The following ICC categories describe the level of agreement between 2 raters: <0.5 (Poor), $0.5 < 0.75$ (Moderate), $0.75 < 0.9$ Good), 0.9 (Excellent) (Koo & Li, 2016). Statistical analyses were conducted in IBM SPSS Statistics Windows version 27.0. P values < 0.05 were considered statistically significant.

Results

Animal characteristics

Table 2. Characteristics of all animals assessed at Shelter A (n = 80).

Characteristics			
Dogs		Cats	
Sex (% , n)			
Male	55.0%, 22	52.5%, 21	
Female	45.0%, 18	47.5%, 19	
Breed (% , n)			
Herding Group	10.0%, 4	American Shorthair	2.5%, 1
Hound Group	10.0%, 4	Domestic Shorthair	90.0%, 36
Toy Group	5.0%, 2	Domestic Longhair	0
Non-Sporting Group	7.5%, 3	Russian Blue	2.5%, 1
Sporting Group	15.0%, 6	Siamese	5.0%, 2

Terrier Group	40.0%, 16	
Working Group	7.5%, 3	
Misc. Class	5.0%, 2	
Age (% , n)		
Puppy/Kitten (up to 6 months)	n/a	n/a
Young Adult 6 months - 2 years)	35.0%, 14	35.0%, 14
Adult (2-8 years)	55.0%, 22	55.0%, 22
Senior (8+ years)	10.0%, 4	10.0%, 4
Spay/Neuter (% , n)	50.0% neutered, 20 22.5% spayed, 9 27.5% not fixed, 11	45.0% neutered, 18 45.0% spayed, 18 10.0% not fixed, 4
Weight (lbs, std dev)	48.38 pounds, 17.45	8.9 pounds, 2.42
Length of stay¹ (days, std dev)	173.79 days, 143.43	121.85 days, 125.03

Intake Type (% , n)		
Owner relinquishment	27.5%, 11	72.5%, 29
Stray	10.0%, 4	22.5%, 9
Seized	n/a	n/a
Returned Adoption	2.5%, 1	2.5%, 1
Transfer from another animal shelter	60.0%, 24	2.5%, 1
Outcome Type (% , n)		
Adoption	80.0%, 32	60.0%, 24
Euthanasia	17.5%, 7	0
Return to owner	2.5%, 1	7.5%, 3
Transferred to another animal shelter	0	5.0%, 2
Died	0	0
Available for adoption	0	27.5%, 11

¹*Length of stay (LOS) accounts for animals who have had an outcome and does not include the 11 cats that remain in the shelter.*

The average age of dogs was 44.4 months (SD: 30.38). The approximate average weight was 48.3 pounds (SD: 17.44) since one animal's weight was not recorded in the pet's record or during the assessment. The average age of cats was 41.85 (SD: 33.77). The average weight was 8.95 pounds (SD: 2.42). Of the 80 shelter animals, 70% (n = 56) were adopted. 8.75% (n = 7) were euthanized. 5% (n = 4) were returned to the owner, 2.5% (n = 2) were transferred to another facility, and 13.75% (n = 11) remained in the shelter system.

Rater Characteristics

Of the 13 raters who completed quality of life assessments, there were 11 women and 2 men. The raters were between the ages of 20 and 50. Six raters were on the behavior team. Two raters were on the Doggy Day Care Team. One rater was on the foster team. One rater was a volunteer. One rater was on the veterinary clinic team. One rater was on the transportation team.

Inter-observer Reliability: Species-Specific Observations

Dogs

When assessing dogs, consistent agreement (85-100%) between raters was present in the environmental health section (see Table 3). The strongest agreements (100%) were seen when raters were asked for information about food frequency, food

match, water availability, water replacement, housing, and enrichment for dogs. The Intraclass Correlation and p-values for the dog assessments can be found in Table 5.

Within the dog assessments, 'good' agreement was seen in 1 behavioral health section which asked raters about how much food the dog consumed in the morning (ICC=0.813). Moderate agreement was seen in 8 behavioral health sections such as arousal around other dogs (ICC=0.544). Raters had 'poor' agreement in 9 behavioral health sections such as fear while being in the kennel (ICC=0.086).

In the physical health sections in dog assessments, raters had 'good' agreement when asked about 2 variables: lameness (ICC=0.901) or another medical issue (ICC=0.772). Moderate agreement was seen in 3 physical health sections such as arousal around other dogs (ICC=0.544). Raters had 'poor' agreement when asked to observe 1 physical health variable which was the dogs' coat condition (ICC=0.347).

Cats

When assessing cats, consistent agreement (85-100%) between raters was present in the environmental health section (see Table 4). The strongest agreements (100%) were seen in ratings about food frequency, food match, water availability, water replacement, group hiding, and group stretch. The Intraclass Correlation and p-values for the cat assessments can be found in Table 6.

Within the cat assessments, 'moderate' agreement between raters was present when they were asked to discuss 8 behavioral health variables including body posture (ICC=0.543). Raters had 'poor' agreement in 8 behavioral health sections such as levels of arousal towards handlers (ICC=0).

In the physical health sections in cat assessments, 'good' agreement was seen for 2 variables: signs of an upper respiratory infection (ICC=1) or another medical issue

(ICC=0.922). Raters had 'poor' agreement when they were asked about 1 physical health variable, the fecal scores (ICC=0.31).

Table 3. Percentage of agreement between 2 raters during dog environmental health assessments.

Categories	% agreement
Environmental Health	
Frequency of meals	100
Appropriate diet	100
Access to water	100
Water replacement	100
Housing	100
Enrichment	100
Compartmentalization	67.50
Floor space	90
Ability to stretch	90

Presence of bedding	85
Cleanliness	97.50

Table 4. Percentage of agreement between 2 raters during the cat environmental health assessments.

Categories	% agreement
Environmental Health	
Frequency of meals	100
Appropriate diet	100
Access to water	100
Water replacement	97
Presence of compartments	54
Access to hiding area (single)	92
Access to elevated space (single)	77

Access to separate areas of living (single)	92
Ability to stretch (single)	85
Access to hiding area (group)	100
Access to elevated space (group)	96
Separate areas of living spaces	92
Ability to stretch (group)	85
Spot cleaning	95
Deep cleaning	90

Table 5. Values for Intraclass Correlation and p-value for the dog assessments.

Categories	Single Measures	
	<i>ICC</i>	<i>P</i>
Environmental Health		
Frequency of exercise	0.424	0.003
Duration of exercise	0.469	0.001
Noise	0.220	0.064
Behavioral Health		
Amount of food consumption	0.813	<0.001
Activity level during feeding	0.569	<0.001
Level of overnight activity	0.473	0.001
Animal's reaction to being approached while in the kennel	0.586	<0.001
Animal's fear response to being approached while in the kennel	0.086	0.300
Animal's arousal response while being approached while in the kennel	0.687	<0.001

Animal's threat response while being approach while in the kennel	0.695	<0.001
Reaction to crouch approach by human	0.707	<0.001
Fear shown during crouch approach	0.259	0.062
Presence or absence during crouch approach	0.647	<0.001
Positive interaction with another dog	0.343	0.013
Presence/absence of fear around another dog	-0.009	0.521
Level of arousal around another dog	0.544	<0.001
Presence or absence of threatening behavior around another dog	0.444	0.003
Out-of-kennel reaction to human approach	0.389	0.015
Out-of-kennel fear response to human approach	0.484	0.003

Out-of-kennel arousal response to human approach	0.316	0.042
Reaction to toy(s)	0.689	<0.001
Physical Health		
Body Condition Score (BCS)	0.601	0.001
Coat Condition	0.347	0.013
Upper Respiratory Infection (URI)	0.661	<0.001
Fecal Score	0.547	<0.001
Presence or absence of lameness	0.901	<0.001
Medical conditions (other)	0.772	<0.001

Table 6. Values for Intraclass Correlation and p-value for the cat assessments.

Categories	Single Measures	
	<i>ICC</i>	<i>P</i>
Environmental Health		
Size of enclosure (single)	0.462	0.040
Number of cats in enclosure	0.992	<0.001
Size of enclosure (group)	0.998	<0.001
Group access to litter	0.867	<0.001
Group access to food	0.900	<0.001
Group access to water	0.885	<0.001
Noise level in cattery	0.113	0.142
Behavioral Health		
Amount of food consumed	0.634	<0.001
Level of activity	0.681	<0.001
Level of overnight activity	0.634	<0.001

Cat's body posture	0.543	0.001
Presence or absence of fearful behavior around another cat	0.335	0.027
Level of arousal around another cat	-0.045	0.596
Presence or absence of threatening behavior around another cat	-0.054	0.615
Cat's response to handler approach	0.623	<0.001
Presence or absence of fearful behavior around handler	0.272	0.044
Level of arousal around handler	0	0.5
Presence or absence of threatening behavior around handler	0.657	<0.001
Presence of positive response to handler	0.547	0.001
Presence or absence of fearful behavior during petting	0.269	0.066
Level of arousal during petting	0.047	0.361

Presence or absence of threatening behavior during petting	0.468	0.004
Reaction to toy(s)	0.69	<0.001
Physical Health		
Body Condition Score (BCS)	0.721	<0.001
Coat Condition	0.59	<0.001
Upper Respiratory Infection (URI)	1.00	1.00
Fecal Score	0.31	0.057
Medical conditions (other)	0.922	<0.001

Discussion

The aim of the study was to test if 2 raters could simultaneously assess the welfare of shelter cats and dogs in a similar way, regardless of their educational levels and backgrounds. The data suggests that the most consistent and strongest agreement (up to 100%) between raters was seen in the environmental health section for both dogs and cats. This section mainly contains binary questions, which may be easier to answer than scale questions. Raters had moderate agreement on 3 physical health variables for dogs which were body condition score, signs of upper respiratory

infection, and fecal score. Moderate agreement was also seen in 2 physical health variables for cats: body condition score and signs of an upper respiratory infection. Additionally, canine coat condition and feline fecal score showed poor agreement. There was also poor agreement across 8 behavioral variables, including the cat's level of fear, threat, and arousal around other cats and handlers. Poor agreement was also seen across half of the behavioral variables for the dogs including level of fear and threat outside of the kennel.

The interobserver reliability between all 13 raters may have been affected by the following: their animal-related experience, familiarity with the QoL app, familiarity with the animals in Shelter A, gender, and the window of time during the assessment. In addition, the raters experienced limitations during data collection such as limited staff availability, a low number of raters per assessment, and limited knowledge of cat behavior.

Difference in Rater's Experience

Differences in raters' experience may have affected their assessment of behavior and physical health in shelter dogs and cats. Similar to Travnik et al. (2022), our study included raters who have different experiences with animals, specifically cats, and knowledge of animal behavior. Six of these raters, who have different years of species-related experience, are on the dog and cat behavior teams at Shelter A. The behavior team consists of employees who have worked at the shelter for years and have species-specific certifications (e.g., CDT-KA).

Behavioral Health in Dogs

Overall, all raters had both moderate and poor agreement within the behavioral assessments in dogs which may be due to differences in their ability to identify and report various behaviors. Mariti et al. (2012) found that the educational level of dog owners had no impact on their ability to correctly identify behavioral indicators of stress, however other studies have found that more experienced dog behavior professionals had more success at identifying fearful behaviors in dogs than those who were less experienced (Wan et al., 2012).

Behavioral Health in Cats

Cat behavioral variables, such as the cat's level of arousal around other cats, fear towards the handler, and the cat's arousal level around the handler, also showed poor agreement. Finka et al (2022) stated that higher human-cat interaction (HCI) scores were weakly correlated with raters having animal-related work experience. In this study, the primary rater expressed that their knowledge of cat behavior was limited which made it difficult to determine whether some cats were experiencing discomfort and pain. Conversely, there were 2 cat behavior experts who served as secondary raters in separate sessions who had different levels of agreement for the questions about cats' fear of other cats, cats' reaction toward the handler approaches, cats' fear towards the handler, cats' arousal around the handler, and the cats' assessment of the threat. When the raters were asked to approach and attempt to pet the cats, data shows that raters also had different levels of agreement related to the cats' level of fear and aggressive or threatening behaviors. Da Graça Pereira et al (2014) found that veterinarians, veterinary nurses, and owners all had a similar understanding of general cat behavior. Yet, there was a difference between individuals within these respective groups who were and were not cat owners. Higher

values, indicating a better understanding of cat behavior, were seen when cat owners completed the questionnaire rather than the professionals who did not own a cat.

Physical Health in Dogs

Good levels of agreement were seen when raters were asked about signs of lameness or a medical issue were present in dogs. For 3 behavioral variables (body condition score, signs of an upper respiratory infection, and fecal scoring), the level of agreement was moderate.

Poor inter-observer reliability was seen when raters observed coat condition and it may be due to the differences in the rater's experiences. Previous research has found intraobserver reliability between dermatologists and surgeons was moderate to substantial (Devriendt et al, 2021), although intraobserver agreement was higher among dermatologists in comparison to surgeons and interns (Devriendt et al, 2021).

Physical Health in Cats

Excellent levels of agreement were present for 3 physical variables: whether or not a cat urinated, signs of an upper respiratory infection, and medical issues. Two physical variables, body condition score and coat condition, showed moderate agreement.

Differences in the rater's experiences could be a factor in why poor inter-observer reliability was seen in fecal scoring. The Waltham Fecal Scoring System has been reliable and validated. This scoring system continues to be used as a gold standard in the animal welfare field. Yet there have been inconsistencies in agreement noted between individuals with varying degrees of experience (Cavett et al, 2021). There

was a fair agreement ($k=0.34$) between the scores of the veterinarian and the lay public (Cavett et al, 2021). In comparison, the agreement between veterinarians was higher ($k= 0.54$ to 0.61). Similarly, in this study, the raters' differences in experience may have affected their ability to report fecal scores. Two raters, including the primary rater, had experience as veterinary assistants. These raters assessed five animals and out of those five animals, the raters gave the same fecal score for only one animal. All raters had 'poor' agreement ($ICC=0.31$) for fecal scores for cats at Shelter A.

Familiarity of Rater with Animals

It is possible that increased familiarity between humans and companion animals leads to a better understanding of animal behavior (Travnik et al., 2022). Perhaps some dogs and cats in previous studies showed more signs of anxiety and fear with some raters than others as a result of their familiarity with the rater which may have led to poor inter-rater reliability.

Dogs' Familiarity with Humans

The presence of an observer may affect the behavior and performance of a dog (Horn et al., 2013a). The familiarity of the rater with the animal can affect the level of agreement between raters (Gosling, 1998) as dogs can behave differently around familiar people and unfamiliar people (Kerepesi et al, 2015). In contrast to the study by Bollen and Horowitz (2008), the unfamiliar raters in this study were diverse and did not: (1) belong to a specific gender and (2) wear the same outfit. In this study, there was moderate agreement between raters in regard to the dog's reactions to them approaching their kennels. Diesel (2008) also found moderate agreement

between staff when assessing dogs' behavioral responses to a person approaching their kennel. Some raters, such as the primary rater, were not familiar with the shelter dogs yet studies such as Gácsi et al., 2001 showed that a 30-minute-long interaction could be enough to the dog to initiate social interaction towards the rater. In one study, dogs who were housed alone or housed with other dogs had different reactions to experimenters. Positive experiences with humans may also increase the chance of shelter dogs approaching unfamiliar people, which may explain the small difference in the dogs' reactions towards familiar and unfamiliar people (Kerepesi et al, 2015). Additionally, shelter dogs may display more fear-appeasement behaviors towards an unfamiliar person but also show more social behaviors than owned dogs (Barrera et al., 2010).

Cat's Familiarity with Humans

Jacobson et al., (2022) found substantial to perfect agreement between raters when asked about cats' socialization yet in this study, poor to moderate agreement was found between raters. In some studies, cats have interacted with humans who are attentive (Ito et al., 2016) regardless if they are familiar or not (Mertens and Turner, 1988; Vitale and Udell, 2019). Similar to Ellis et. al (2015), some cats demonstrated negative behaviors (e.g., fear, threat, arousal) when familiar handlers approached and attempted to pet them. Other cats showed positive behaviors (e.g., rubbing) when around an unfamiliar human in a laboratory and shelter setting (Podberscek et al., 1991; Tuozi et al., 2021). Therefore, it appears that the differences in the level of familiarity between raters and the shelter cats may not explain the poor reliability found in the current study.

Familiarity with the Quality of Life Application

Rater's familiarity with the QoL app may have affected the way they interpreted the questions, therefore affecting their answers and ultimately the interobserver reliability. Raters were not given training on how to use the application. A brief summary of the aim of the study was given and then the data collection began. Each rater (besides the primary rater) saw the application for the first time either the night before or the morning of the data collection day. To reduce bias, raters were told to stay focused on their own assessment and avoid discussion. The questions have a question icon next to them which provides the context in case a rater has difficulty understanding or answering a question. After the assessments were complete, some raters admitted that some questions were unclear, yet they did not use the information icon to help clarify the question's purpose. In other studies, the lack of familiarity with an assessment has had a negative effect on inter-observer reliability, so it is possible that training provided before the assessment may help to improve reliability. The primary rater was the only one who was familiar with the app, so having this advantage may have affected the way they responded to prompts. Yet similar to the findings of Kiddie and Collins (2014), consensual drift may have been reduced by preventing raters from discussing observations during the assessment and submitting answers that were influenced by the other rater.

Raters' Gender

The difference in raters' gender could have an effect on the animal's interaction with the handler and the level of agreement between raters. Schöberl et al. (2009) stated that owner interaction style and owner gender affect situational stress levels in dogs. The animal welfare assessments were not completed by owners but in this study, it

is possible that the gender of the rater may have impacted the animal's response. Although one male rater was more familiar with the dogs in Shelter A, some dogs were less aroused and showed fewer signs of fear around an unfamiliar female. Previous research has found women appeared to be more compassionate, informed, and likely to form strong emotional attachments to individual animals than males (Kellert & Berry, 1987). This may show a difference that relates to a dog's past experience with a rater of a specific gender. Dogs showed a stronger decrease in their barking and their tendency to look towards the human whenever the subject was a woman than a man, suggesting that dogs may be more defensively-aggressive towards men than women (Wells & Hepper, 1999). This may have affected the outcome of the arousal and crouch scores and potentially decreased the inter-observer reliability.

Limitations

Effect of Staff Availability on Time of Assessments

The shelter is open to staff at 8 am and once kennel staff arrives, the cleaning process begins. Not all animal assessments began at 8 am due to staff availability which may have compromised the reliability of some variables. For factors such as fecal score, the raters had to rely on the kennel staff's interpretation of each animal's stool sample, if the cage/kennel was already cleaned once the raters arrived. The staff, who also have varying degrees of experience, may have provided different descriptions of the animal's stool or the raters may have interpreted the staff member's descriptions differently. In some cases, raters also found it difficult to select a fecal score due to the animal smearing their feces throughout the kennel which could hinder reliability.

To prevent poor agreement in the future while using this tool, raters could undergo a training period. Training can include a session where they rate different samples of stool within the shelter environment. In cases where raters do not see feces, perhaps adding a “not applicable” option would help prevent misinterpretation of the staff’s description of the feces.

Effect of Number of Raters on Inter-observer Reliability

Raters in pairs, where one rater remained consistent, assessed the animals in each session. The values were distributed among several categories: poor, moderate, and good reliability. There may have been a limitation where one individual’s ratings may have affected the inter-observer reliability values altogether if their assessments were consistently inaccurate. The values may increase in a future study if more than two raters or a single pair of raters assess at least 80 animals (40 cats and 40 dogs), giving us more information to ascertain the reliability of the questions. It may be best to record animals during live observations so that multiple people can review the recordings and come to an agreement on the conditions of the environment and the animals’ interactions with their environment.

Conclusion

The aim of this study was to analyze the inter-observer reliability between two raters, of different backgrounds, during animal welfare assessments via the Penn Vet Quality of Life in Shelters App. Although environmental variables showed good agreement, most physical health variables showed moderate agreement and behavioral variables (e.g., fear, anxiety, arousal, and aggression) frequently showed poor agreement. This may be due to the rater’s experience, familiarity with the

animals, familiarity with the app, or differences in the rater's gender. In future studies, three approaches may help address these challenges: (1) train raters before data collection, (2) recruit more raters, and (3) record animals during live assessments for later review. With these considerations, we hope that future assessments completed via the Penn Vet QoL app will not only help us perform reliable testing but also increase our understanding of shelter animal welfare overall.

References

- Barnard, S., Pedernera, C., Candeloro, L., Ferri, N., Velarde, A., & Dalla Villa, P. (2016). Development of a new welfare assessment protocol for practical application in long-term dog shelters. *Veterinary Record*, *178*(1), 18-18.
- 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*, *5*(6), 339-344.
- 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](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](https://doi.org/10.1016/s0168-1591(97)00145-7)
- Beerda, B., Schilder, M. B. H., van Hooff, J. A. R. A. M., de Vries, H. W., & Mol, J. A. (2000). Behavioural and hormonal indicators of enduring environmental stress in dogs. *Animal Welfare*, *9*, 49-62.
- 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.
- Botreau, R., Veissier, I., Butterworth, A., Bracke, M. B. M., & Keeling, L. J. (2007). Definition of criteria for overall assessment of animal welfare. *Animal Welfare*, *16*(2), 225-228.

<http://www.ingentaconnect.com/content/ufaw/aw/2007/00000016/00000002/art00030>

Broom, D. M. (2011). A history of animal welfare science. *Acta Biotheoretica*, 59(2), 121–137. <https://doi.org/10.1007/s10441-011-9123-3>

Carmines, E. G., Zeller, R. A. (1979). *Reliability and Validity Assessment*. Sage Publications, Inc., pp. 11.

Cavett, C. L., Toner, M., Marks, S. L., Winston, J. A., Gilor, C., & Rudinsky, A. J. (2021). Consistency of faecal scoring using two canine faecal scoring systems. *Journal of Small Animal Practice*, 62(3), 167-173.

Devriendt, N., Rodrigues, T. C. N., Vandenabeele, S., Favril, S., Biscop, A., Marynissen, S., ... & De Rooster, H. (2021). Validation of a skin and coat scoring protocol in dogs. *Vlaams Diergeneeskundig Tijdschrift*, 90(5).

Da Graça Pereira, G., Fragoso, S., Morais, D., Villa de Brito, M. T., & de Sousa, L. (2014). Comparison of interpretation of Cat's behavioral needs between veterinarians, veterinary nurses, and cat owners. *Journal of Veterinary Behavior*, 9(6), 324–328. <https://doi.org/10.1016/j.jveb.2014.08.006>

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), 171-181.

Ellis, S. L. H., Thompson, H., Guijarro, C., & Zulch, H. E. (2015). The influence of body region, handler familiarity and order of region handled on the domestic cat's response to being stroked. *Applied Animal Behaviour Science*, 173, 60-67.

- Finka, L. R., Ripari, L., Quinlan, L., Haywood, C., Puzzo, J., Jordan, A., Tsui, J., Foreman-Worsley, R., Dixon, L., & Brennan, M. L. (2022). Investigation of humans individual differences as predictors of their animal interaction styles, focused on the domestic cat. *Scientific Reports*, *12*(1).
<https://doi.org/10.1038/s41598-022-15194-7>
- Gácsi, M., Topál, J., Miklósi, Á., Dóka, A., & Csányi, V. (2001). Attachment behavior of adult dogs (*Canis familiaris*) living at rescue centers: forming new bonds. *Journal of Comparative Psychology*, *115*(4), 423.
- Gosling, S. D. (1998). Personality dimensions in spotted hyenas (*Crocuta Crocuta*). *Journal of Comparative Psychology*, *112*(2), 107–118.
<https://doi.org/10.1037/0735-7036.112.2.107>
- Kellert, S. R., & Berry, J. K. (1987). Attitudes, knowledge, and behaviors toward wildlife as affected by gender. *Wildlife Society Bulletin (1973-2006)*, *15*(3), 363-371.
- Kerepesi, A., Dóka, A., & Miklósi, Á. (2015). Dogs and their human companions: the effect of familiarity on dog–human interactions. *Behavioural processes*, *110*, 27-36.
- Horn, L., Huber, L., & Range, F. (2013a). The importance of the secure base effect for domestic dogs – evidence from a manipulative problem-solving task. *PLoS ONE*, *8*(5). <https://doi.org/10.1371/journal.pone.0065296>
- Ito, Y., Watanabe, A., Takagi, S., Arahori, M., & Saito, A. (2016). Cats beg for food from the human who looks at and calls to them: ability to understand humans' attentional states. *Psychologia*, *59*(2-3), 112-120.

- Jacobson, L. S., Ellis, J. J., Janke, K. J., Giacinti, J. A., & Robertson, J. V. (2022). Behavior and adoptability of hoarded cats admitted to an animal shelter. *Journal of feline medicine and surgery*, 24(8), e232-e243.
- 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.
- Koo & Li (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine*, 15(2), 155-163.
- List of breeds by Group. American Kennel Club. (n.d.). Retrieved April 28, 2022, from <https://www.akc.org/public-education/resources/general-tips-information/dog-breeds-sorted-groups/>
- Maille, A. R., Koning, C. J. M., Zwinderman, A. H., Willems, L. N. A., Dijkman, J. H., & Kaptein, A. A. (1997). The development of the 'Quality-of-life for Respiratory Illness Questionnaire (QOL-RIQ)': a disease-specific quality-of-life questionnaire for patients with mild to moderate chronic non-specific lung disease. *Respiratory medicine*, 91(5), 297-309.
- Mariti, C., Gazzano, A., Moore, J., Baragli, P., Chelli, L., & Sighieri, C. (2012). Perception of dogs' stress by their owners. *Journal of Veterinary Behavior*, 7(4), 213-219. <https://doi.org/10.1016/j.jveb.2011.09.004>
- Mertens, C., & Turner, D. C. (1988). Experimental analysis of human-cat interactions during first encounters. *Anthrozoös*, 2(2), 83-97.

Patronek, G. J., Bradley, J., & Arps, E. (2019). What is the evidence for reliability and validity of behavior evaluations for shelter dogs? A prequel to "No better than flipping a coin". *Journal of Veterinary Behavior*, 31, 43-58.

PetHealth Software Solutions, Inc. (2022). PetPoint Data Management System Version (5).

Pet statistics. ASPCA. (n.d.). Retrieved November 20, 2022, from <https://www.asPCA.org/helping-people-pets/shelter-intake-and-surrender/pet-statistics>

Podberscek, A. L., Blackshaw, J. K., & Beattie, A. W. (1991). The behaviour of laboratory colony cats and their reactions to a familiar and unfamiliar person. *Applied Animal Behaviour Science*, 31(1-2), 119-130.
[https://doi.org/10.1016/0168-1591\(91\)90159-u](https://doi.org/10.1016/0168-1591(91)90159-u)

Research randomizer. Pair of dice. (n.d.). Retrieved from <https://www.randomizer.org/>

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.

Schöberl, G. I. (2009). *Effects of interaction style, attachment, relationship and personality on stress coping in human-dog dyads*. na.

Stephen, J.M., Ledger, R.A., 2005. An audit of behavioral indicators of poor welfare in kennelled dogs in the United Kingdom. *J. Appl. Anim. Welf. Sci.* 8, 79-95.

Taylor, K.D., Mills, D.S., 2006. The development and assessment of temperament tests for adult companion dogs. *J. Vet. Behav.* 1, 94-108.

- Taylor, K.D., Mills, D.S., 2007a. The effect of the kennel environment on canine welfare: a critical review of experimental studies. *Anim. Welf.* 16, 435-447.
- Taylor, K.D., Mills, D.S., 2007b. Is quality of life a useful concept for companion animals? *Anim. Welf.* 16, 55-65.
- Travnik, I. C., Machado, D. S., & Sant'Anna, A. C. (2022). Do you see the same cat that I see? Inter-and intra-observer reliability for Qualitative Behaviour Assessment as temperament indicator in domestic cats. *Animal Welfare*, 31(3), 319-327.
- Tuozi, A., Arhant, C., Anderle, K., Backes, J., Cords, C., Magierski, V., ... & Windschnurer, I. (2021). Effects of Human Presence and Voice on the Behaviour of Shelter Dogs and Cats: A Preliminary Study. *Animals*, 11(2), 406.
- Vitale, K. R., & Udell, M. A. R. (2019). The quality of being sociable: The influence of human attentional state, population, and human familiarity on domestic cat sociability. *Behavioural Processes*, 158, 11-17.
<https://doi.org/10.1016/j.beproc.2018.10.026>
- Wan, M., Bolger, N., & Champagne, F. A. (2012). Human perception of fear in dogs varies according to experience with dogs. *PLoS one*, 7(12), e51775.
- Wells, D.L., Hepper, P.G., Male and female dogs respond differently to men and women, *Applied Animal Behaviour Science*, Volume 61, Issue 4, 1999, Pages 341-349, ISSN 0168-1591, [https://doi.org/10.1016/S0168-1591\(98\)00202-0](https://doi.org/10.1016/S0168-1591(98)00202-0). (<https://www.sciencedirect.com/science/article/pii/S0168159198002020>)

Wiseman-Orr, M. L., Scott, E. M., Reid, J., & Nolan, A. M. (2006). Validation of a structured questionnaire as an instrument to measure chronic pain in dogs on the basis of effects on health-related quality of life. *American Journal of Veterinary Research*, 67(11), 1826–1836.

<https://doi.org/10.2460/ajvr.67.11.1826>