



Pet Behaviour Science
2026, Vol. 20, 18 - 36
doi:10.21071/pbs.vi20.18281



UNIVERSIDAD DE CORDOBA

Amy E. Bauer¹

Monica Colon²

Mary Jordan³

Traci Shreyer⁴

Candace Croney^{5,*}

*Email: ccroney@purdue.edu

¹ College of Veterinary Medicine,
Department of Comparative Pathobiology,
Purdue University, USA

² University of Detroit Mercy, Detroit,
Michigan

³ Center for Animal Welfare Science,
Department of Comparative Pathobiology
and Animal Science, Purdue University,
USA

Working with FIDO: Pilot testing the field instantaneous dog observation (fido) tool in the animal shelter environment

Amy E. Bauer¹, Monica Colon², Mary Jordan³,
Traci Shreyer⁴, Candace Croney^{5,*}

Abstract: In the United States, approximately three to four million dogs enter animal shelters each year. These dogs must cope with unfamiliar caretakers and conspecifics in a novel, often loud, environment which can negatively impact their welfare. It is therefore important to evaluate the welfare of dogs housed in animal shelters and assist those that are having difficulty coping with their new environments. Here we aimed to pilot test the behavioral component of the Field Instantaneous Dog Observation (FIDO) tool, a non-invasive tool designed for the assessment of kennel dog welfare, in a shelter environment. Upon the approach of a stranger, dogs were scored as either Green (affiliative or neutral response), Yellow (ambivalent response), or Red (fearful response). Inter-rater reliability (IRR) between four novice raters using the FIDO tool, and agreement between novice raters and an expert in canine behavior were evaluated. In addition, changes in dog response based upon caretaker presence or absence were explored. Overall, most dogs were scored as Green in response to approach (>79% of sample across studies). Cohen's kappa test revealed that novice raters had moderate levels of agreement ($k=0.41$, 86% percent agreement), while the agreement between novice raters and the expert was substantial ($k=0.77$, 91% percent agreement). Finally, Wilcoxon signed-rank test revealed no statistically significant difference in dogs' responses to approach whether the caretaker was present or absent. The FIDO tool has the potential to provide rapid insight into a dog's behavioral status in a shelter environment and can be quickly and reliably taught to novice users as a practical evaluative tool.

Keywords: animal shelter; behavior; canine; dogs; welfare assessment

HIGHLIGHTS

- The behavioral component of the Field Instantaneous Dog Observation (FIDO) tool was pilot tested for inter-rater reliability between novice raters and between expert and novice raters in the animal shelter environment.
- Most dogs assessed were scored Green (affiliative, neutral) in response to the approach of an unfamiliar person.
- The presence of a caretaker had no statistically significant effect on the response to approach.

INTRODUCTION

In the United States, an estimated 6.5 million animals enter shelter facilities each year. Of these, approximately 3.3 million are dogs (ASPCA 2021). Dogs housed in shelters lack control over their surroundings (Beerda et al. 1999), which may negatively impact their welfare as a result of high levels of stress and anxiety. Stressors include exposure to unfamiliar environments, caretakers, and dogs, as well as loud noises (Coppola et al. 2010; Part et al. 2014). Additionally, transport of dogs between rescue organizations and animal shelters has become commonplace in the United States (Scarlett 2013; Simmons and Hoffman 2016), which may also prevent dogs from habituating to the shelter environment. Stress associated with all of these factors can affect the welfare of dogs housed in shelters, and adoption and retention in new homes (Tuber et al. 1999; Part et al. 2014).

It is important for caretakers in shelters to monitor the dogs in their care. Use of animal-based measures to assess dog welfare may lead to improved outcomes through early identification of dogs who are struggling to cope in the shelter environment. Identifying these dogs can allow for implementation of supportive measures before chronic behavioral indicators of distress occur and interventions are more difficult to implement (Belshaw et al. 2015; Protopopova 2016). This may be especially important for dogs that remain in the shelter environment for prolonged periods of time. In a recent study of dogs in a commercial breeding (CB) kennel, investigators reported that dogs characterized prior to re-homing as being sociable towards people, also showed less fear of both social and non-social stimuli after re-homing (Barnard et al. 2023). While CB kennels and shelters differ in the environment and population characteristics of the dogs they house, they can share stressors such as noise levels. Barnard and colleagues (2023) also demonstrated that daily management and handling practices, such as performing low stress handling, can help promote more positive social interactions between dogs and caretakers, which in turn may translate to more positive interactions towards strangers.

Welfare assessment tools intended for use in a field setting, such as an animal shelter or CB kennel must be applicable in those settings, in addition to being practical, reliable, and valid (Meagher 2009; Mugenda et al. 2019, Barnard et al. 2021). Practicality, reliability, and validity are particularly important when a tool is based upon observer ratings, as they can be subjective. The use of observer ratings to evaluate behavior and welfare has been studied in several species, including dogs (Sinn et al. 2010; Muri et al. 2013; Petelle and Blumstein 2014; Dalla Costa et al. 2015; Fratkin et al. 2015; Barnard et al. 2016; Mugenda et al. 2019). The application of observer



Pet Behaviour Science
2026, Vol. 20, 18 - 36
doi:10.21071/pbs.vi20.18281



UNIVERSIDAD D CORDOBA

Amy E. Bauer¹

Monica Colon²

Mary Jordan³

Traci Shreyer⁴

Candace Croney^{5,*}



Pet Behaviour Science
2026, Vol. 20, 18 - 36
doi:10.21071/pbs.vi20.18281



UNIVERSIDAD DE CORDOBA

Amy E. Bauer¹

Monica Colon²

Mary Jordan³

Traci Shreyer⁴

Candace Croney^{5,*}

ratings to assess the welfare of dogs housed in shelters has been of particular interest (Diesel et al. 2008; Kiddie and Collins 2014; Barnard et al. 2015), as they often require little observer training, are easy to perform, and are cost effective since they can be conducted without specialized equipment. Overall, behavioral observations are less invasive, minimizing additional handling stress, which can confound physical and physiological measures of interest and negatively impact dog welfare. In general, observer ratings appear to be a viable means of assessing behavior and welfare (Meagher 2009). Nevertheless, tools intended for use by novice raters without extensive training in behavior and welfare science must be evaluated to determine how well they are able to produce ratings that agree with each other and with experts in canine behavior and welfare.

The purpose of the present study was to evaluate the applicability of the behavioral component of the Field Instantaneous Dog Observation (FIDO) tool to dogs housed in shelter environments. This tool was designed to be rapid, non-invasive, easy for novice raters to use, and field-ready (Bauer et al. 2017; Mugenda et al. 2019). The FIDO tool includes ten measures of physical health and cleanliness (Body Condition Score, Body Cleanliness, Presence of Nasal Discharge, Ocular Discharge, Tear Staining, Observable Sneezing, Coughing, Missing Fur or Poor Coat, Presence of Wounds, Sores or Lesions, and Observable Lameness) as well as the Red, Yellow, Green behavioral component which is the focus of this study (Bauer et al. 2017). The behavioral aspect of the FIDO tool was intended to capture fearful responses to approach by an unfamiliar person, since fear of unfamiliar people may be an indicator of poor human-directed socialization (Serpell and Jagoe 1995; Lindsay 2000). In the context of the shelter environment, displays of fear may also affect adoptability (Weiss et al. 2012), as fearful dogs tend to hide from prospective adopters, while dogs that are more immediately pro-social and move forward in their pens are more easily seen and more likely to be adopted.

Pilot-testing of the behavioral component of the FIDO tool in commercial breeding facilities has reported high levels of inter-rater reliability between a pair of expert raters and a pair of novice raters (Bauer et al. 2017). The first objective of this study was to test the IRR (inter-rater reliability) of the behavioral component of the FIDO tool, as used by novice raters in a shelter environment. The second objective of this study was to evaluate agreement in categorizing dogs' responses to approach between novice raters and an expert in canine behavior and welfare. The final objective of this study was to determine if dogs' responses to approach changed when a caretaker was present, compared to the caretaker being absent. Although the presence of a

caretaker did not significantly influence the overall responses of dogs at commercial breeding facilities to approach by an unfamiliar person (Bauer et al. 2017), the presence of a known individual was found to reduce aggression in a shelter setting (Kis et al. 2014). Moreover, in a study exploring the relationship between the attitudes of animal shelter staff and the approach behavior of dogs, shelters where staff reported more positive attitudes toward dogs and utilized positive handling were found to have a greater proportion of dogs that approached the unfamiliar experimenter (Arhant and Troxler 2014).



Pet Behaviour Science
2026, Vol. 20, 18 - 36
doi:10.21071/pbs.vi20.18281



UNIVERSIDAD DE CORDOBA

Amy E. Bauer¹

Monica Colon²

Mary Jordan³

Traci Shreyer⁴

Candace Croney^{5,*}

METHODS

Statement of approval

The procedures described in this manuscript were reviewed and approved by the Purdue Institutional Animal Care and Use Committee.

Facilities and subjects

Observations were conducted in two non-profit animal shelters located in Indiana, USA. Facility 1 was a limited-intake shelter which had municipality contracts to care for stray animals. Facility 1 admitted 1,177 dogs in 2016 (year the study was performed) and reported an annual live release rate of 92% for all admitted animals in that same year. Dogs at Facility 1 were housed in 1.5m by 1.8m pens. Facility 2 was also a limited intake shelter but did not have any municipality contracts. Facility 2 admitted 1,492 dogs in 2016 and reported an annual live release rate of 95% for all admitted animals that year. [Table 1](#) details the number of dogs observed at each facility by the novice raters. For the novice and expert raters, 20 dogs were observed at Facility 1 and 38 dogs at Facility 2 for a total of 58 dogs.



Pet Behaviour Science
2026, Vol. 20, 18 - 36
doi:10.21071/pbs.vi20.18281



UNIVERSIDAD DE CORDOBA

Amy E. Bauer¹

Monica Colon²

Mary Jordan³

Traci Shreyer⁴

Candace Croney^{5,*}

Facility	Number observed both with and without caretaker	Number observed without caretaker only	Total
1	13	1	14
2	19	1	20
Total	32	2	34

Table 1. Number of dogs observed at each facility by novice raters. Two dogs were not present for the approach trial with the caretaker present. These dogs are included in the second column and were not included in the statistical analysis for Objective 3, the effect of caretaker presence on response to approach. All dogs observed with the caretaker present were also observed without the caretaker present.

Approach test protocol

An important goal of the FIDO Tool is to allow raters with limited training and experience to quickly utilize response to approach as a surrogate for welfare. However, raters did need to be introduced to the tool and to approach techniques. The two experts in canine behavior and animal welfare science who developed and refined the FIDO tool (Bauer et al. 2017) trained the novice raters in both the approach technique and the identification of behaviors that would fit into the Red, Yellow, and Green categories. Approach training focused on implementing a non-threatening approach to the kennel. Raters practiced stepping toward the front of the pen with the body turned sideways. The rater closest to the door would then bend and extended a hand to the front of the pen. All raters were instructed to avoid making eye contact with the dogs on the initial approach (Bauer et al. 2017).

The dogs tested were all singly housed and remained in their home pens for the duration of the test. Two raters stood next to each other, approximately 0.5 m away from the pen door. The raters turned their bodies so that their sides faced the door to the kennel and approached the door as described above. Each dog's immediate response to the closest rater's approach was recorded as either: 1) Red, indicating a fearful response or stereotypic behaviors; 2) Green, indicating an affiliative or neutral/undisturbed response; or 3) Yellow, indicating an ambivalent response that could not clearly be classified as either Red or Green (Bauer et al. 2017).

Affiliative responses classified as Green included behaviors such as approaching the observer or soliciting the observer's attention. Neutral/undisturbed responses categorized as Green included maintenance behaviors such as resting, eating, drinking, or grooming. Responses classified as Yellow included alternately approaching and retreating from the observer or growling while wagging the tail in a neutral position. All potential manifestations of fear (flight, fight, or freeze; Lindsay 2001) were recorded as Red. (Bauer et al. 2017).

Objective 1: inter-rater reliability of novices

Four female novice raters (raters: A: a veterinary student; B: an undergraduate student; C: a graduate student in Animal Welfare Science; and D: a veterinarian with a Ph.D. in Comparative Epidemiology and Public Health) evaluated the response to approach of 34 dogs from Facilities 1 and 2 in July 2016. Breed information was available for 26 of the 34 dogs observed (76.5%) with pitbull or pitbull mixes (23.5%, n=8) and chihuahua or chihuahua mixes (14.7%, n=5) predominating. Pointers or pointer mixes were represented by 2 dogs (5.9%). All other breeds were represented by a single individual as listed in [Table 2](#).

Prior to evaluating the dogs, the raters were trained as described above by an expert in canine behavior and welfare on how to approach each dog and evaluate their response to the approach. Every third pen was selected for the approach test. Raters A and B approached and rated all dogs, while raters C or D stood a few steps behind and scored the dog's response from a distance. Due to space constraints, raters C and D alternated, so that they each were involved in the assessment of 17 dogs ([Table 3](#)). The selected dog's response to approach was categorized independently by each of the raters as Red, Yellow, or Green (RYG), as defined above.

Percent agreement and weighted kappa values using Cohen's Kappa were calculated for the responses to approach. Weighted kappa was calculated using proc frequency, with 0.0001 added to 0 values to facilitate the analysis (SAS, Version 9.4, SAS Institute Inc., Cary, NC). Agreement was defined based on the scale: kappa < 0.00 poor agreement, kappa = 0 - 0.20 slight agreement, 0.21 - 0.40 fair agreement, 0.41 - 0.60 moderate agreement, 0.61 - 0.80 substantial agreement, 0.81 - 1.00 almost perfect agreement, as suggested by Landis & Koch (1977). The weighted kappa values were averaged to provide an overall kappa score following Light's solution for fully crossed designs with three or more raters (Hallgren 2012). As kappa statistics can be affected by homogeneous responses, a prevalence index (PI) was also calculated for each kappa value using the formula $PI = |g - o| / N$. In



Pet Behaviour Science
2026, Vol. 20, 18 - 36
doi:10.21071/pbs.vi20.18281



UNIVERSIDAD DE CORDOBA

Amy E. Bauer¹

Monica Colon²

Mary Jordan³

Traci Shreyer⁴

Candace Croney^{5,*}



Pet Behaviour Science
2026, Vol. 20, 18 - 36
doi:10.21071/pbs.vi20.18281



UNIVERSIDAD DE CORDOBA

Amy E. Bauer¹

Monica Colon²

Mary Jordan³

Traci Shreyer⁴

Candace Croney^{5,*}

this formula, g is the number of agreed upon rankings (for example where both raters coded a dog's response to approach as Green), o is the number of instances where the particular coding (Green in this case) was not present, and N is the total number of observations. A PI close to 1 indicates that the sample is homogeneous in a way that would influence the kappa statistic. A PI close to 0 indicates that the sample is more heterogeneous and less likely to have influenced the kappa statistic (Burn and Weir 2011).

Objective 2: agreement between novices and an expert rater

Three additional novice raters (E: a female undergraduate student; F: a male undergraduate student; and G: a female graduate student in Animal Welfare Science) evaluated the response to approach of 58 previously untested dogs from Facilities 1 and 2 and their observations were compared with those of an additional female expert in canine behavior and welfare who holds a Ph.D. in Applied Ethology (rater H). Breeds of dogs were not recorded for this portion of the study.

This group of novice raters was introduced to the approach test in the same fashion as the previous group of novice raters. Additionally, they were provided with training on interpreting dog responses to approach using video clips. This group of novices also had a practice session at Facility 1 prior to the evaluation session. On the day of the approach tests, each of the novice raters approached the front of the dog's pen with the expert as described in the approach test protocol. Next, each novice-expert pair recorded each dog's immediate response to approach as Red, Yellow, or Green. Raters E and F each observed the response to approach of 19 dogs with the expert rater, while rater G observed the response to approach of 20 dogs with the expert rater. Percent agreement and weighted kappa values using Cohen's Kappa were calculated for each novice-expert pair using SAS, Version 9.4 (SAS Institute Inc., Cary, NC). All weighted kappa values were averaged to provide an overall kappa score (Hallgren 2012). Prevalence indices were also calculated for each pair. Agreement was defined based on the scale used in Objective 1.

Objective 3: effect of caretaker presence

After each selected dog at the facility was evaluated by the novice raters (Objective 1), raters waited 30 minutes before repeating the approach test (as described above) with a caretaker (i.e., shelter staff employee) to gauge whether the dog's responses to approach changed with the presence of a familiar person. The evaluations for this study were performed by raters A and B. The caretaker stood at the front of the run where they were visible to the dog being approached. Thirty-four dogs were observed by the novice



Pet Behaviour Science
2026, Vol. 20, 18 - 36
doi:10.21071/pbs.vi20.18281



UNIVERSIDAD DE CORDOBA

Amy E. Bauer¹

Monica Colon²

Mary Jordan³

Traci Shreyer⁴

Candace Croney^{5,*}

raters in Objective 1 ([Table 1](#)). Due to the wait period between the test of inter-rater reliability for novice raters and the test of effect on caretaker presence, two of the dogs that were included in Objective 1 were not available for repeat assessment in Objective 3. The reason for the dogs' absence was not recorded. They may have been taken outside for exercise with staff members or been visiting with prospective adopters. Thus, thirty-two of the dogs originally observed by the novice raters were also observed in the presence of a caretaker. Breed information was available for 24 of the 32 dogs observed (75.0%). As in Objective 1, pitbull or pitbull mixes (25.0%, n=8) and chihuahua or chihuahua mixes (12.5%, n=4) predominated ([Table 2](#)).

Percent agreement, weighted kappa values, and PIs were calculated for the responses to approach, as described in Objective 1. Where the raters agreed upon the dogs' responses to approach, a Wilcoxon signed-rank test was performed to determine if there was a statistically significant difference between the proportions of Red, Yellow, and Green responses to approach when a caretaker was present, as compared to the caretaker not being present, using SPSS Statistics for Windows, Version 22.0 (IBM Corp. Released 2013. Armonk, NY: IBM Corp.).



**Pet
Behaviour
Science**
open access journal

Pet Behaviour Science
2026, Vol. 20, 18 - 36
doi:10.21071/pbs.vi20.18281



UNIVERSIDAD DE CORDOBA

Amy E. Bauer¹

Monica Colon²

Mary Jordan³

Traci Shreyer⁴

Candace Croney^{5,*}

Identified breed or mix	Number observed in Objective 1	Number observed in Objective 3
Akita mix	1	1
Australian Cattle Dog	1	1
Beagle	1	0
Chihuahua or Chihuahua mix	5	4
Coonhound mix	1	1
Dachshund mix	1	1
Foxhound	1	1
Labrador mix	1	1
Pitbull or Pitbull mix	8	8
Pointer or Pointer mix	2	2
Rat Terrier mix	1	1
Retriever	1	1
Shepherd mix	1	1
Undefined mix	1	1
Unidentified breed	8	8
Total dogs observed	34	32

Table 2. Identified breeds of dogs observed in Objective 1 and Objective 3. All identified breeds were represented by a single individual with the exceptions of pitbulls and pitbull mixes, chihuahuas and chihuahua mixes, and pointers and pointer mixes. Eight of the observed dogs did not have an identified breed or breed mix.

RESULTS

Objective 1: inter-rater reliability of novices

The proportion of responses to approach falling into each RYG category, percent agreement, and Kappa's Cohen for each pair is presented in [Table 3](#).



Pet Behaviour Science
2026, Vol. 20, 18 - 36
doi:10.21071/pbs.vi20.18281



UNIVERSIDAD DE CORDOBA

Amy E. Bauer¹

Monica Colon²

Mary Jordan³

Traci Shreyer⁴

Candace Croney^{5,*}

		Red (%)	Yellow (%)	Green (%)	Percent Agreement	Kappa (95% CI)	PI
Pair 1	Rater A	0 (0.00%)	7 (20.59%)	27 (79.41%)	82.35%	0.315 (-0.082, 0.713)	0.706
	N=34 Rater B	0 (0.00%)	3 (8.82%)	31 (91.18%)			
Pair 2	Rater A	0 (0.00%)	0 (0.00%)	17 (100%)	94.12%	0.0001 (-0.098, 0.100)	0.941
	N=17 Rater C	0 (0.00%)	1 (5.88%)	16 (94.12%)			
Pair 3	Rater A	0 (0.00%)	7 (41.18%)	10 (58.82%)	64.71%	0.164 (-0.131, 0.446)	0.588
	N=17 Rater D	0 (0.00%)	1 (5.88%)	16 (94.12%)			
Pair 4	Rater B	0 (0.00%)	1 (5.88%)	16 (94.12%)	100%	0.999 (0.963, 1.000)	0.882
	N=17 Rater C	0 (0.00%)	1 (5.88%)	16 (94.12%)			
Pair 5	Rater B	0 (0.00%)	2 (11.76%)	15 (88.23%)	94.12%	0.638 (-0.003, 1.000)	0.823
	N=17 Rater D	0 (0.00%)	1 (5.88%)	16 (94.12%)			
Overall					86.27%	0.405 (0.336, 0.474)	

Table 3. Proportions of response to approach, percent agreement, weighted kappa, and prevalence indices (PI) when response to approach was evaluated by novice raters A, B, C and D. The overall values were calculated as the average of the percent agreements and kappa values for each novice pair.

There was a total of 102 pairs of assessments. When all approach tests were analyzed, the raters agreed on 86.27% of the responses to approach (N=88), with an average weighted kappa score of 0.41, indicating moderate agreement. PI values ranged from 0.588 to 0.941 with a mean value of 0.788. The high level of agreement and high mean PI may have biased the kappa statistic. The predominance of a single type of response can negatively influence the kappa estimate (Hallgren 2012). While the confidence interval for the overall kappa statistic does not contain zero, the confidence intervals

for kappa for each novice pair do contain zero. This indicates that a portion of the raters' agreement could be due to chance. A larger sample size for each pair could help to resolve this problem (Sim and Wright 2005).

Objective 2: agreement between novices and an expert rater

There was a total of 58 pairs of assessments. The proportion of responses to approach, percent agreement, and weighted kappa for each of the novice raters and expert rater are listed in [Table 4](#). When all approach tests were analyzed, the raters agreed on 91.38% of the responses to approach (N=53). The average weighted kappa score was 0.77, indicating substantial agreement. PI values ranged from 0.550 to 0.737, with a mean PI of 0.603. Although there was a higher level percent agreement between the raters than in Objective 1, the mean PI is lower, indicating more diversity in the responses to approach. This increased diversity may account, at least in part, for the higher overall weighted kappa score.



Pet Behaviour Science
2026, Vol. 20, 18 - 36
doi:10.21071/pbs.vi20.18281



UNIVERSIDAD DE CORDOBA

Amy E. Bauer¹

Monica Colon²

Mary Jordan³

Traci Shreyer⁴

Candace Croney^{5,*}

		Red (%)	Yellow (%)	Green (%)	Percent Agreement	Kappa (95% CI)
Pair 6	Expert	0 (0.00%)	2 (10.53%)	17 (89.47%)	84.21%	0.496 (0.043, 0.948)
N=19	Rater E	0 (0.00%)	5 (26.32%)	14 (73.68%)		
Pair 7	Expert	2 (10.53%)	2 (10.53%)	15 (78.94%)	97.74%	0.893 (0.676, 1.000)
N=19	Rater F	2 (10.53%)	1 (5.26%)	16 (84.21%)		
Pair 8	Expert	4 (20.00%)	0 (0.00%)	16 (80.00%)	95.00%	0.925 (0.777, 1.000)
N=20	Rater G	4 (20.00%)	1 (5.00%)	15 (75.00%)		
Overall					91.38%	0.774 (0.667, 0.881)

Table 4. Percent agreement and weighted kappa between raters without expertise in canine behavior and an expert rater when scoring dogs' responses to approach. The overall values were calculated as the average of the percent agreements and kappa values for each expert-novice pair.

Objective 3: effect of caretaker presence

The proportion of responses to approach identified by raters A and B based on caretaker presence is shown in [Table 5](#).



Pet Behaviour Science
2026, Vol. 20, 18 - 36
doi:10.21071/pbs.vi20.18281



UNIVERSIDAD DE CORDOBA

Amy E. Bauer¹

Monica Colon²

Mary Jordan³

Traci Shreyer⁴

Candace Croney^{5,*}

	Caretaker not present (N=34)			Caretaker present (N=32)		
	Red	Yellow	Green	Red	Yellow	Green
	(%)	(%)	(%)	(%)	(%)	(%)
Rater A	0 (0.00%)	7 (20.59%)	27 (79.41%)	0 (0.00%)	6 (18.75%)	26 (76.47%)
Rater B	0 (0.00%)	3 (8.82%)	31 (91.18%)	0 (0.00%)	4 (12.50%)	28 (87.50%)
Percent agreement		82.35%			87.50%	
Kappa (95% CI)		0.315 (-0.082, 0.713)			0.529 (0.130, 0.928)	
PI		0.706			0.687	

Table 5. Distribution of responses to approach and rater agreement with and without a caretaker present

High proportions (>76%) of Green responses to approach were identified both with, and without, a caretaker present and no Red responses to approach were recorded. The raters agreed on the responses to approach, with and without a caretaker present, for 24 dogs (75.00%). The PI without a caretaker present was 0.706. The PI with a caretaker present was 0.687. Upon analysis of the agreed-upon paired observations from both shelters (N=24), a Wilcoxon signed rank test determined that the presence of a caretaker did not produce a statistically significant change in response to approach ($T=-1.00$, $P=0.32$).

DISCUSSION

In this study, the behavioral portion of the FIDO tool was tested in the context of animal shelters. The objectives of the study were to measure IRR (inter-rater reliability) between novice raters, determine agreement between novice raters and an expert in canine behavior in interpreting dogs' responses to approach, and evaluate the effect of caretaker presence on the response of a dog to approach by a stranger. Most dogs assessed were scored as Green, indicating affiliative or neutral responses to approach. The predominance of Green responses to approach is not surprising, as both facilities are classified as limited intake shelters although Facility 1 is partially funded by municipal contracts for housing of stray animals.



Pet Behaviour Science
2026, Vol. 20, 18 - 36
doi:10.21071/pbs.vi20.18281



UNIVERSIDAD D CORDOBA

Amy E. Bauer¹

Monica Colon²

Mary Jordan³

Traci Shreyer⁴

Candace Croney^{5,*}

The percentage of dogs displaying social/less fearful behavior in this study (>70%) is comparable with that found in previous studies. In a no-kill, open-admission shelter in the US, 89.9% of tested dogs showed no concerning behavior during a stranger approach test (McGuire & Song 2023). Similarly, in testing a welfare assessment tool in 29 European animal shelters, Barnard et al. (2015) found that approximately 80% of the dogs tested reacted to the presence of a person in a sociable manner, as opposed to displaying fear, or defensive or offensive aggression.

The high proportion of Green responses reported in this study likely influenced the IRR observed between novice raters (raters A, B, C, and D). Although the overall percent agreement was 86.27%, the IRR in this portion of the study was moderate ($\kappa = 0.41$) (Landis & Koch 1977). Generally, a homogeneous sample, such as the one seen in this objective, biases kappa values in a negative direction (Hallgren 2012). This negative bias may explain the lower-than-expected kappa value obtained for the novice raters. Increasing the variability in dogs' response to approach through increasing the number of animal shelters participating in the study and including those that do not have limited intake may help to address the issue of homogeneous sampling.

Space constraints resulted in modification of the number of raters for each approach test. Light's solution for using the kappa statistic to compare more than two raters is most applicable for a fully crossed design where all the raters assess all of the dogs (Hallgren 2012). Deviation from this may have affected the average weighted kappa value. Rating only in pairs or including space limitations in the experimental design can help to ensure a fully crossed design in future studies.

The majority of the dogs evaluated in Objective 2 by the novice raters E, F, and G and the expert rater (rater H) also displayed Green responses to approach (>74%). Overall agreement between these raters and the expert rater was substantial. Since the expert's evaluations were used as a reference standard in the behavioral portion of the FIDO, the high level of agreement obtained indicates that, with a modest amount of training, novice raters can classify the response of dogs to the approach of a stranger accurately. This is consistent with findings that non-experts can reliably utilize ratings of dog behavior in a similar fashion to experts (Fratkin et al. 2015; Bauer et al. 2017) and that the IRR of shelter staff members' evaluation of dogs' behavior increases with formal training or experience with dog behavior (Diesel et al. 2008).



Pet Behaviour Science
2026, Vol. 20, 18 - 36
doi:10.21071/pbs.vi20.18281



UNIVERSIDAD DE CORDOBA

Amy E. Bauer¹

Monica Colon²

Mary Jordan³

Traci Shreyer⁴

Candace Croney^{5,*}

In comparison with the novice raters in Objective 1, those in Objective 2 had additional training in interpreting dog behavior. These raters also had a practice session. This additional training may have increased the novices' confidence in using the FIDO tool, which may have led to less homogeneity in coding responses and greater level of agreement between the novice raters and the expert rater. Future studies on the development of training methodologies for novice raters may be helpful in the implementation of the FIDO tool.

Additionally, the presence of the expert may have introduced bias by increasing or decreasing the confidence of the novice raters. Novice raters were found to agree more amongst themselves than with an expert when evaluating video recordings of odor-detection dogs training (Munch et al 2019). Evaluation of video recordings can help to control for bias related to the presence of an expert. However, the FIDO tool is intended to be utilized on site. Future studies should also focus on the role of experts/trainers in novice use of the FIDO tool.

With respect to the presence of a caretaker during the approach test, there was no statistically significant difference in the proportion of responses to approach classified as Green, whether a caretaker was present or not. Companion dogs may view their primary caretakers as stress buffers when threatened (Gacsi et al. 2013), however, this may not be the case for shelter dogs. First, dogs housed in animal shelters tend not to display the same degree of social referencing with their caretakers as companion dogs do with their owners (Duranton et al. 2017) as dogs may not have had the time to develop a strong bond with the caretaker, thus they would not serve as a stress buffer (Pullen et al. 2012). Second, the high proportion of Green responses to approach indicate that the tested dogs were not stressed by the presence of the unfamiliar person, hence not needing the social buffer effect of a familiar person.

The approach tests were conducted with the dogs in their home kennels. There are benefits for evaluating a dog within its home environment. For example, this allows for assessment of the environment as well as the dog, which is part of the full FIDO tool (Bauer et al 2017). However, the familiar environment could have played a role in the high level of Green responses to approach. Future studies with a focus on affiliative behaviors and adoptability in the shelter environment may consider evaluating the approach test in the context of a novel or less-familiar environment such as an outdoor enclosure or a meet-and-greet room. In a study evaluating the validity of the FIDO tool in commercial breeding facilities, dogs were evaluated in their home pens both with and without the ability to move



Pet Behaviour Science
2026, Vol. 20, 18 - 36
doi:10.21071/pbs.vi20.18281



UNIVERSIDAD D CORDOBA

Amy E. Bauer¹

Monica Colon²

Mary Jordan³

Traci Shreyer⁴

Candace Croney^{5,*}

outside and away from the evaluator. There was significant agreement in dogs displaying Red and Green responses to approach regardless of the ability to move outside. Dogs displaying Yellow responses to approach when they were allowed the freedom to move outside showed moderate agreement with their response to approach when this ability was restricted (Mugenda et al 2025).

Alternatively, the approach test could be performed first with the dog in an unfamiliar environment and then after a waiting period with the dog in its familiar kennel. Mugenda et al also conducted an approach test in a novel environment (an outdoor arena). In the arena test, dogs that demonstrated Red responses during the approach tests in their kennels were more likely to avoid the observer in the arena than dogs that demonstrated behaviors classed as Green during the approach tests (Mugenda et al 2025).

Future studies should also include larger sample sizes and collect breed, behavioral, and medical information for each dog as it is available. As the prevalence index increases, the sample size needed to detect agreement independent of chance increases. (Sim and Wright 2005, Hallgren 2012). Having each rater pair evaluate a greater number of dogs could help to resolve this problem. Additionally, a larger sample size would allow for incorporation of breed/breed type and behavioral or medical history as variables in the statistical analyses. This study has established that a high prevalence of Green responses can be expected from dogs housed in limited-intake shelters. Awareness of this trend can help in planning future study designs to include sample sizes that are able to control for the effect of high prevalence on the kappa statistic.

The current study adds to the body of literature exploring different validation criteria for the FIDO tool, in both the form presented here and further extended adaptations of it (Barnard et al. 2021). When using this tool to assess dogs from commercial breeding facilities, in addition to confirming an overall good inter-rater reliability across studies, test-retest reliability and construct validity have also been demonstrated (Bauer et al. 2017; Mugenda et al. 2019, Barnard et al. 2021). Finally, as described above, dogs scoring Red in their home pens using the FIDO tool, were also found to be more fearful of strangers in open field tests, suggesting a consistency in dogs' fear-related behavioral responses across different contexts, which further supports the robustness of this test (Mugenda 2019, Mugenda et al 2025). Indeed, it would be worthwhile to further explore and establish other criteria, such as predictive or convergent validity by comparing the FIDO behavioral scorings against long-term indicators of overall welfare. Validating the instantaneous fear responses to stranger approach with



Pet Behaviour Science
2026, Vol. 20, 18 - 36
doi:10.21071/pbs.vi20.18281



UNIVERSIDAD DE CORDOBA

Amy E. Bauer¹

Monica Colon²

Mary Jordan³

Traci Shreyer⁴

Candace Croney^{5,*}

longer-term welfare measures would permit generalization of the findings of a single use of the FIDO tool to an overall description of a dog's welfare, which is currently not possible (Mugenda 2019). However, as welfare is a dynamic state (Broom 1991), use of the FIDO tool may help to identify dogs in need of intervention to help them adapt to the shelter environment. Taken together, the results from this and previous studies suggest that the FIDO tool can be used by individuals without much expertise in canine behavior and attests to the practicality of the tool via a reliable, five-second approach test (Mugenda et al. 2019). Further, these results also suggest that FIDO scoring can help to identify fearful dogs in need of greater socialization towards unfamiliar people.

It should be noted that the FIDO tool is not intended to replace more detailed or long-term assessments of dog welfare, or to be predictive of behavior in a home post-adoption. It is worth noting that findings from this group reported an association between social fear recorded at the kennel of origin, and levels of social fear reported by new owners after these dogs were rehomed (Barnard et al. 2023). Nevertheless, the FIDO tool was designed to be one available tool in the animal welfare assessment toolbox and is limited to dogs' responses to approach, as measured at that time. However, it has the potential to be of benefit to the caretakers of dogs housed in a shelter environment, both in identifying dogs in need of interventions for successful rehoming and evaluating the success of those interventions.

CONCLUSIONS

The present study aimed to test the reliability of the behavioral component of the FIDO tool in the context of animal shelters, compare the level of agreement between novices and an expert in canine behavior and welfare in characterizing a dog's response to approach, and evaluate the impact of caretaker presence in these facilities. This response-to-approach test appears to provide a reliable means of assessment in this dog population, although training of raters is important. With appropriate training, novices can reach a substantial level of agreement with an expert rater. Additional studies are needed to confirm that the FIDO tool can also provide a valid reflection of long-term welfare. Nevertheless, this assessment strategy appears to be helpful for identifying dogs that would benefit from social enrichment and other interventions to improve their welfare in the shelter environment and increase their potential for successful, long-term adoption.



Pet Behaviour Science
2026, Vol. 20, 18 - 36
doi:10.21071/pbs.vi20.18281



UNIVERSIDAD DE CORDOBA

Amy E. Bauer¹

Monica Colon²

Mary Jordan³

Traci Shreyer⁴

Candace Croney^{5,*}

ACKNOWLEDGEMENTS

The authors would like to acknowledge the invaluable contributions of Lynda Mugenda, Andy Pietraniec, and Mikayla Small for assistance with data collection and their expertise in rating the dogs. Sarah T. Boysen, Judi Stella, and Shanis Barnard provided feedback on the manuscript.

Financial support for the research was provided by the World Pet Association and the Pet Food Institute

REFERENCES

- American Society for the Prevention of Cruelty to Animals (ASPCA). 2021. Shelter intake and surrender pet statistics. Accessed March 6, 2024. <https://www.aspc.org/helping-people-pets/shelter-intake-and-surrender/pet-statistics>
- Arhant, C. and Troxler, J. 2014. Approach behaviour of shelter dogs and its relationships with the attitudes of shelter staff to dogs. *Applied Animal Behaviour Science* 160: 116-126. doi:10.1016/j.applanim.2014.08.013
- Barnard, S., Flint, H., Shreyer, T. and Croney, C. 2021. Evaluation of an easy-to-use protocol for assessing behaviors of dogs retiring from commercial breeding kennels. *PloS One* 16(8). e02. 83. doi:10.1371/journal.pone.0255883
- Barnard, S., Flint, H., Diana, A., Shreyer, T., Arrazola, A., Shepell, J. and Croney, C. 2023. Management and behavioral factors associated with rehoming outcomes of dogs retired from commercial breeding kennels. *PloS One* 18(3). e0282459. doi:10.1371/journal.pone.0282459
- Barnard, S., Marshall-Pescini, S., Passalacqua, C., Beghelli, V., Capra, A., Normando, S., Pelosi, A. and Valsecchi, P. 2016. Does subjective rating reflect behavioural coding? Personality in 2-month-old dog puppies: An open-field test and adjective-based questionnaire. *PLOS One* 11: e0149831. doi:10.1371/journal.pone.0149831
- Barnard, S., Pedernera, C., Candeloro, L., Ferri, N., Velarde, A. and Dalla Villa, P. 2015. Development of a new welfare assessment protocol for practical application in long-term dog shelters. *Veterinary Record* 178: 18. doi:10.1136/vr.103336
- Bauer, A.E., Jordan, M., Colon, M., Shreyer, T. and Croney, C.C. 2017. Evaluating FIDO: Developing and pilot testing the Field Instantaneous Dog Observation tool. *Pet Behaviour Science* 4: 1-15. doi:10.21071/pbs.v0i4.5766
- Beerda, B., Schilder, M.B.H., van Hooff, J.A.R.A.M., De Vries, H.W. and Mol, J.A. 1999. Chronic stress in dogs subjected to social and spatial restriction. I. Behavioral responses. *Physiology & Behavior* 66: 233-242. doi:10.1016/S0031-9384(98)00289-3
- Belshaw, Z., Asher, L., Harvey, N.D. and Dean, R.S. 2015. Quality of life assessment in domestic dogs: An evidence-based rapid review. *The Veterinary Journal* 206: 203-212. doi:10.1016/j.tvjl.2015.07.016
- Broom, D.M. 1991. Animal welfare: Concepts and measurement. *Journal of Animal Science* 69: 4167-4175. doi:10.2527/1991.69104167x
- Burn, C.C., Weir, A.A.S. 2011. Using prevalence indices to aid interpretation and comparison of agreement ratings between two or more observers. *The Veterinary Journal* 188: 166-170. doi:10.1016/j.tvjl.2010.04.021



Pet Behaviour Science
2026, Vol. 20, 18 - 36
doi:10.21071/pbs.vi20.18281



UNIVERSIDAD DE CORDOBA

Amy E. Bauer¹

Monica Colon²

Mary Jordan³

Traci Shreyer⁴

Candace Croney^{5,*}

- Coppola, C., Enns, R.M. and Grandin, T. 2010. Noise in the animal shelter environment: Building design and the effects of daily noise exposure. *Journal of Applied Animal Welfare Science* 9: 1-7. doi:10.1207/s15327604jaws0901_1
- Dalla Costa, E., Dai, F., Murray, L.A.M., Guazzetti, S., Canali, E. and Minero, M. 2015. A study on validity and reliability of on-farm tests to measure human-animal relationship in horses and donkeys. *Applied Animal Behaviour Science* 163: 110-121. doi:10.1016/j.applanim.2014.12.007
- Diesel, G., Brodbelt, D. and Pfeiffer, D.U. 2008. Reliability of assessment of dogs' behavioural responses by staff working at a welfare charity in the UK. *Applied Animal Behavior Science* 115: 171-181. doi:10.1016/j.applanim.2008.05.005
- Duranton, C., Bedossa, T. and Gaunet, F. 2017. Do shelter dogs engage in social referencing with their caregiver in an approach paradigm? An exploratory study. *Applied Animal Behaviour Science* 189: 57-65. doi:10.1016/j.applanim.2017.01.009
- Fratkin, J.L., Sinn, D.L., Thomas, S., Hilliard, S., Olson, Z. and Gosling, S.D. 2015. Do you see what I see? Can non-experts with minimal training reproduce expert ratings in behavioral assessments of working dogs? *Behavioural Processes* 110: 105-116. doi: 10.1016/j.beproc.2014.09.028
- Gacsi, M., Maros, K., Sernkvist, S., Farago, T. and Miklosi, A. 2013. Human analogue safe haven effect of the owner: Behavioural and heart rate response to stressful social stimuli in dogs. *PloS One* 8: e58475. doi:10.1371/journal.pone.0058475
- Hallgren, K.A. 2012. Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology* 8: 23-34. doi: 10.20982/tqmp.08.1.p023
- Kiddie, J.L. and Collins, L.M. 2014. Development and validation of a quality of life assessment tool for use in kennelled dogs (*Canis familiaris*). *Applied Animal Behavior Science* 158: 57-68. doi:10.1016/j.applanim.2014.05.008
- Kis, A., Klausz, B., Persa, E., Miklosi, A. and Gacsi, M. 2014. Timing and presence of an attachment person affect sensitivity of aggression tests in shelter dogs. *Veterinary Record* 174: 196. doi:10.1136/vr.101955
- Landis, J.R. and Koch, G.G. 1977. The measure of observer agreement for categorical data. *Biometrics* 33: 159 - 174. doi:10.2307/2529310
- Lindsay, S.R. 2000. *Handbook of Applied Dog Behavior and Training*. In Volume 1, *Adaptation and Learning*. Ames, Iowa: Iowa State Press.
- Lindsay, S.R.. 2001. *Fears and phobias*. (eds). *Handbook of Applied Dog Behavior and Training*. In Volume 2, *Etiology and assessment of behavior problems*. Ames, Iowa: Iowa Press.
- McGuire, B., and Song, A. 2023. Influence of sex of stranger on responses of shelter dogs during canine behavioral evaluations. *Animals* 13(15): 2461. doi:10.3390/ani13152461
- Meagher, R.K. 2009. Observer ratings: Validity and value as a tool for animal welfare research. *Applied Animal Behavior Science* 119: 1-14. doi:10.1016/j.applanim.2009.02.026
- Mugenda, L., Shreyer, T. and Croney, C.C. 2019. Refining canine welfare assessment in kennels: evaluating the reliability of Field Instantaneous Dog Observation (FIDO) scoring. *Applied Animal Behaviour Science* 221: 104874. doi:10.1016/j.applanim.2019.104874



Pet Behaviour Science
2026, Vol. 20, 18 - 36
doi:10.21071/pbs.vi20.18281



UNIVERSIDAD DE CORDOBA

Amy E. Bauer¹

Monica Colon²

Mary Jordan³

Traci Shreyer⁴

Candace Croney^{5,*}

- Mugenda, L., Shreyer, T. and Croney, C.C. 2025. Advancing canine welfare assessment in kennels: evaluating the validity of Field Instantaneous Dog Observation (FIDO) scoring in commercial breeding kennels. *Applied Animal Behaviour Science* 292: 106770. doi:10.1016/j.applanim.2025.106770
- Mugenda, L.W. 2019. Refining and validating an on-site canine welfare assessment tool developed for use in commercial breeding kennels. PhD thesis, Purdue University Graduate School, IN, USA.
- Munch, K.L., Wapstra, E., Thomas, S. Fisher, M., and Sinn, D.L. 2019. What are we measuring? Novices agree amongst themselves (but not always with experts) in their assessment of dog behaviour. *Ethology* 125:203-211. doi:10.1111/eth.12846
- Muri, K., Stubbsjoen, S.M., and Valle, P.S. 2013. Development and testing of an on-farm welfare assessment protocol for dairy goats. *Animal Welfare* 22: 385-400. doi: 10.7120/09627286.22.3.385
- Part, C.E., Kiddie, J.L., Hayes, A., Mills, D.S., Neville, R.F., Morton, D.B., and Collins, L.M. 2014. Physiological, physical, and behavioural changes in dogs (*Canis familiaris*) when kennelled: Testing the validity of stress parameters. *Physiology & Behavior* 133: 260-271. doi:10.1016/j.physbeh.2014.05.018
- Petelle, M. B. and Blumstein, D.T. 2014. A critical evaluation of subjective ratings: Unacquainted observers can reliably assess certain personality traits. *Current Zoology* 60: 162-169. doi:10.1093/czoolo/60.2.162
- Protopopova, A. 2016. Effects of sheltering on physiology, immune function, behavior, and the welfare of dogs. *Physiology & Behavior* 159: 95-103. doi:10.1016/j.physbeh.2016.03.020
- Pullen, A. J., Merrill, R.J.N. and Bradshaw, J.W.S. 2012. The effect of familiarity on behaviour of kennel housed dogs during interactions with humans. *Applied Animal Behaviour Science* 137: 66-73. doi:10.1016/j.applanim.2011.12.009
- Scarlett, J. 2013. Population statistics. In *Shelter Medicine for Veterinarians and Staff*. (eds.) L. Miller and S. Zawistowski. John Wiley & Sons, Inc, Ames, Iowa, USA
- Serpell, J. and Jagoe, J.A. 1995. Early experience and the development of behaviour. In *The Domestic Dog: Its Evolution, Behaviour and Interactions with People* 79-102 (ed.) Serpell, J. Cambridge University Press, Cambridge, UK.
- Sim, J. and Wright, C.C. 2005. The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy* 85:257-268. doi: 10.1093/ptj/85.3.257
- Simmons, K.E. and Hoffman, C.L. 2016. Dogs on the move: Factors impacting animal shelter and rescue organizations' decisions to accept dogs from distant locations. *Animals* 6: 11. doi:10.3390/ani6020011
- Sinn, D.L., Gosling, S.D. and Hilliard, S. 2010. Personality and performance in military working dogs: Reliability and predictive validity of behavioral tests. *Applied Animal Behavior Science* 127: 51-65. doi:10.1016/j.applanim.2010.08.007
- Tuber, D.S., Miller, D.D., Caris, K.A., Halter R., Linden, F. and Hennessy, M.B. 1999. Dogs in animal shelters: Problems, suggestions, and needed expertise. *Psychological Science* 10: 379-386. doi:10.1111/1467-9280.00173
- Weiss, E., Miller, K., Mohan-Gibbons, H. and Vela, C. 2012. Why did you choose this pet? Adopters and pet selection preferences in five animal shelters in the United States. *Animals* 2(2): 144-159 doi:10.3390/ani2020144