Journal of Veterinary Behavior 10 (2015) 94–102

Contents lists available at ScienceDirect

Journal of Veterinary Behavior

journal homepage: www.journalvetbehavior.com

Research

A test for the evaluation of emotional reactivity in Labrador retrievers used for explosives detection

Barbara L. Sherman [a,*], Margaret E. Gruen [a], Beth C. Case [a], Melanie L. Foster [b], Richard E. Fish [a], Lucia Lazarowski [b], Venita DePuy [c], David C. Dorman [b]

[a] Department of Clinical Sciences, College of Veterinary Medicine, North Carolina State University, Raleigh, North Carolina
[b] Department of Molecular and Biomedical Sciences, College of Veterinary Medicine, North Carolina State University, Raleigh, North Carolina
[c] Bowden Analytics, 104 Strawthorne Court, Apex, North Carolina

ARTICLE INFO

Article history:
Received 24 May 2014
Received in revised form
6 November 2014
Accepted 30 December 2014
Available online 8 January 2015

Keywords:
emotional reactivity test
anxiety scores
selection criteria
behavioral screening
explosives detection dog

ABSTRACT

The United States Marine Corps (USMC) uses Labrador retrievers as improvised explosive device detection dogs (IDDs). Of critical importance is the selection of dogs that are emotionally suited for this highly specialized application. The goal of our study was to develop an emotional reactivity test (ERT) as a screening tool for the selection of IDDs. The ERT included a series of subtasks that expose each dog sequentially to visual, auditory, and experiential stimuli with an associated grading scale used by trained observers to rate individual dog responses. In this study, 16 Labrador retrievers that met initial selection criteria as candidate IDDs were assessed using the ERT, measurement of plasma and salivary cortisol concentrations (pre- and post-ERT), and an independent open-field test of anxiety in response to sound stimuli. Based on the sum of its responses, each dog was assigned an aggregate ERT score. Aggregate ERT scores from independent trained observers were highly concordant [Shrout-Fleiss's intraclass correlation (2,1) = 0.96] suggesting excellent inter-rater reliability. The aggregate ERT scores were also negatively correlated with the dogs' scores on the open-field anxiety test (Spearman rank correlation, n = 16, r = −0.57, $P = 0.0214$). In addition, there were significant increases in salivary (Wilcoxon signed rank, n = 16, S = 38.5, $P = 0.0458$) and plasma (Wilcoxon signed rank, n = 16, S = 68, $P < 0.0001$) cortisol levels after the ERT, compared with baseline, suggesting that exposure to the ERT test elements produced a physiological stress response. We conclude that the ERT is a useful pre-training screening test that can be used to identify dogs with a low threshold of emotional reactivity for rejection, and dogs with a high threshold of emotional reactivity for entry into the IDD training program.

Introduction

Dogs used for the detection of explosives require a high standard of performance, as their success or failure may have profound repercussions. One specialized group is the improvised explosive device detection dog (IDD) used by the United States Marine Corps (USMC). Candidate IDDs are selected from adult Labrador retrievers bred for field trial and hunt competition. The goal of the IDD program is to produce a dog that can work with a USMC handler to conduct directed off-leash searches for improvised explosive devices (Crawford, 2012). To meet this goal, candidate IDDs are sent to a military contractor facility where they undergo a rigorous, approximately 70-day training program that includes physical conditioning, scent training, and behavior modification protocols. Dogs that successfully complete this program are paired with a USMC handler. Each dog/handler team undergoes an intensive 5-week training program at the training facility, followed by further off-leash explosives detection training in desert terrain at a military facility in the United States. Then, each fully trained IDD/handler pair is deployed overseas for active combat duty.

Selecting suitable dogs for the program is critically important. Unsuitable behavioral traits may negatively influence training and contribute to impaired dog/handler team performance. For example, the emotions of fear and anxiety may affect the ability

* Address for reprint requests and correspondence: Barbara L. Sherman, PhD, DVM, DACVB, DACAW, Department of Clinical Sciences, College of Veterinary Medicine, North Carolina State University, 1060 William Moore Drive, Raleigh, NC 27607, USA. Tel.: +1-011-919-513-6141; Fax: +1-001-919-513-6464.
E-mail address: barbara_sherman@ncsu.edu (B.L. Sherman).

http://dx.doi.org/10.1016/j.jveb.2014.12.007
1558-7878/

of individual dogs to learn new tasks and to perform at optimal levels (Haverbeke et al., 2010). Fear responses may alter environmental perception by focusing attention on fear-inducing stimuli, and weakening attention to other salient environmental features, thereby impairing a dogs' ability to work (Blackwell et al., 2013). Anxiety, the anticipation of fear-inducing events, may lead to increased vigilance and avoidance (Araujo et al., 2013). There is compelling evidence that behavioral traits, more so than sensory or physiological capacity, may influence performance in dogs (Beerda et al., 1998, 1999). Selecting dogs that are resistant to debilitating emotional responses, such as fear and anxiety, is an important prerequisite for training dogs for explosives detection. Ideal dogs for training would exhibit modest and transient "emotional reactivity" that would not impair their ability to function under combat or other work situations.

A number of studies support the finding that "personality," "temperament," or "performance" tests may be used to predict behavior (Diederich and Giffroy, 2006; Jones and Gosling, 2005; Swartberg, 2005; Wilsson and Sundgren, 1997). Provocative tests in dogs have been used to elicit behavioral and physiological fear responses (Hydbring-Sandberg et al., 2004). More specific screening tests developed for German shepherd and Malinois breeds reliably predict the outcome for military working dog training test performance at Lackland Air Force Base in the United States (Sinn et al., 2010).

An open-field arena test is another model for measuring fear responses and anxiety states in animals (Prut and Belzung, 2003). In the open field, the animal may be subjected to provocative stimuli, such as relevant sounds, and its behavioral and physiological responses quantified. In dogs, the open-field model has been used to identify Beagles that suffer from thunderstorm anxiety and attenuate this state using pharmacologic and other means (Araujo et al., 2013; Landsberg et al., 2013). The aim of the present study was to validate an emotional reactivity test (ERT) as a screening (selection) tool for candidate IDDs using physiologic measures and an open-field model of anxiety We also evaluated the ERT with respect to inter-rater reliability and established convergent validity of this test to select dogs robust to "stress" effects.

Materials and methods

Dog procurement

A military working dog contractor (K2 Solutions, Inc. [hereafter K2], Southern Pines, NC) procured the dogs for training in the USMC IDD program.

To be considered for procurement, dogs had to be less than 24 months of age at the time of procurement, have started field trial training, and be in apparent good health. Dogs were not considered for procurement if they exhibited human-directed or dog-dog aggression, marked avoidance of the procurement officer, or pronounced submission to their handler. Procured dogs were collected throughout the Unites States and then transported by truck to south-central North Carolina. After a14-day quarantine at an off-site commercial boarding kennel, the dogs were transported to the K2 training facility. A K2 veterinarian performed a comprehensive physical evaluation, which included a retinal examination and evaluation of digital radiographs of the pelvis, lumbar-sacral spine, and elbows. Screening laboratory tests included comprehensive blood chemistry; complete blood count; urine and fecal analysis; tests to assess thyroid, heart worm, *Borrelia burgdorferi*, Ehrlichia, and Anaplasmosis exposure; and genetic testing for exercise-induced collapse. A subcutaneous microchip was placed for individual identification. Procured dogs that passed the veterinary assessment were held at the K2 facility where they were housed in individual kennels and received regular group exercise. The mean time in residence at K2 was 335 days (range: 225-411). Dogs were transported from K2 approximately 130 km to *North Carolina State University* (NCSU) on November 29, 2011. Additional details regarding housing conditions at K2 have been previously described (Lazarowski et al., 2014).

Experimental subjects and animal welfare oversight

The experimental subjects used for this study were drawn from the stock of candidate IDDs that were procured for the IDD program. They were 16 Labrador retriever dogs between 2 and 4 years of age; there were 8 intact males, 5 intact females, and 3 spayed females (Table 1). The coat color of 10 dogs was black and the coat

Table 1
Demographic information and result of ERT, physiologic, and open-field test assessment of Labrador retrievers (n = 16)

Dog Name	Sex[a]	Coat color[b]	Age (days)[c]	ERT score (max: 85)	Salivary cortisol (μg/dL)			Plasma cortisol (μg/dL)			Open-field anxiety Score
					Pre- ERT	Post-ERT	Δ salivary cortisol (post-pre)	Pre-ERT	Post-ERT	Δ plasma cortisol (post-pre)	
Ace	M	B	768	75	0.262	0.186	−0.079	1.42	1.63	0.21	0.0
Annie	F	Y	821	51	0.208	0.256	0.048	1.05	2.21	1.16	2.5
Baxter	M	B	1186	65	0.133	0.228	0.095	1.97	4.55	2.58	2.0
Bullet	M	Y	1347	75	0.119	0.115	−0.004	0.99[d]	2.50	1.51	1.75
Dakota	F	B	863	72	0.169	0.195	0.026	1.27	2.22	0.95	0.0
Honey	F	Y	829	48	0.089	0.166	0.007	2.22	4.15	1.93	2.75
Hunter	M	B	659	71	0.156	0.373	0.217	0.99[d]	1.46	0.47	1.0
Jimmy	FS	B	1025	73	0.110	0.245	0.135	1.68	3.70	2.02	1.5
Macks	M	B	872	63	0.065	0.192	0.127	0.99[d]	1.15	0.16	0.75
Mercy	FS	B	1012	74	0.413	0.103	−0.310	0.99[d]	1.83	0.84	−0.25
Piper	F	Y	741	36	0.049	0.240	0.191	1.05	5.41	4.36	1.0
Reno	M	Y	920	71	0.206	0.158	−0.048	0.99[d]	4.02	3.03	0.75
Rip	M	B	774	68	0.216	0.270	0.054	1.40	2.19	0.79	1.75
Ruby	FS	Y	805	74	0.116	0.139	0.023	1.39	1.83	0.44	0.75
Valentine	F	B	752	58	0.130	0.222	0.092	1.31	5.77	4.46	1.75
Wizard	M	B	821	71	0.090	0.156	0.066	1.28	2.37	1.09	0.5

ERT, emotional reactivity test.
[a] Sex: F = Female, M = Male, FS = Female, spayed; all males were intact.
[b] Coat color: B = black; Y = yellow.
[c] At time the ERT was administered.
[d] A value of 0.99 μg/dL was assigned when the sample concentration was below the limit of detection (<1.00 μg/dL).

color of 6 dogs was yellow. The 16 dogs were housed in an environmentally controlled indoor canine facility at the NCSU College of Veterinary Medicine (NCSU-CVM) Laboratory Animal Resources Unit. Details concerning housing conditions and husbandry at the NCSU canine facility have been previously described (Lazarowski et al., 2014). Unless otherwise noted, the tests and evaluations were performed at NCSU-CVM by NCSU veterinarians and staff. At the conclusion of research studies at NCSU-CVM, all dogs were returned to K2 and subsequently adopted to private homes as pets.

All experimental protocols were reviewed and approved by the NCSU Institutional Animal Care and Use Committee and the US Army Medical Research and Material Command Animal Care and Use Review Office.

ERT protocol

The ERT included a series of subtasks involving visual, sound, and experiential stimuli (Table 2). In total, there were 17 subtasks in the ERT. Twelve subtasks had been used historically by trainers at K2 to assess dogs after procurement since the inception of the IDD program. One subtask, down stairs, was added for convenience, to follow the "up stairs" subtask. Two subtasks were created by dividing 2 previous complex subtasks relating to stranger approach into 2 parts so as to capture specific dog responses. Two subtasks were added based on results from other studies evaluating dog reactivity (Diederich and Giffroy, 2006). For example, umbrella opening (King et al., 2003; Murphy, 1998), and response to and approach toward a novel electronic toy car (Goddard and Beilharz, 1985; Haverbeke et al., 2008) were presented as potentially fear-inducing stimuli. Scores for each ERT subtask could range from 1 to 5, with the low score indicating a poor response and a high score indicating a favorable response. The maximum aggregate score was 85. The ERT test area consisted of an outdoor area surrounded by a chain link fence (subtasks 1-3 and 14-16) and an indoor arena with cement floor (subtasks 4-13 and 17). Owing to firearms restriction at NCSU-CVM, a playback recording of gunfire was used with the dogs in the indoor arena rather than actual gunfire produced in an outside area. Development of the ERT procedures and descriptors (Table 2) relied on the results obtained from a pilot study that used 4 privately owned dogs. This pilot study also provided training for key participants conducting the ERT so that the test protocol would be standardized and safety measures reviewed. Any dog displaying active aggression (growling, lunging, and snapping) in response to stranger approaches was to be immediately removed from the situation and a score of 0 assigned to the subtask.

After 91 days at NCSU-CVM, on February 28, 2012, from 13:00 to 17:00 hours, all dogs were individually tested on the 17 subtasks that comprise the ERT. The subtasks were conducted as described in Table 2 except gunfire exposure (subtask 17). Owing to firearms restriction at NCSU-CVM, we used a playback recording of gunfire while the dogs were in the indoor arena rather than actual gunfire produced in an outside area. The order of subtask presentation was the same for all dogs. All dogs proceeded through the test without interruption. To avoid distracting intersex odor cues, males were tested first, followed by spayed females, then intact females. Within each sex group, testing order was randomized. Each test was recorded on video for future review and scoring verification.

The same experienced handler (MEG), unfamiliar to the dogs, guided and restrained each dog using a leash and plain nylon collar. In general, the handler directed the dog to each task, but remained neutral with regard to interactions with the dog. While following the protocol, dogs that were reluctant to move forward in response to a specific task were given approximately 5 seconds to examine the stimulus, then prompted by the handler speaking to the dog in a nonemotional manner. Neither punishment nor force was applied to the dogs. In cases in which a dog did not respond to prompts, and either emphatically avoided the stimulus or attempted to escape, the handler proceeded to the next stimulus. Although there were several instances of growling at the approach of the unusual stranger, there were no instances of active aggression, which required removal of any dog from a subtask. Additional trained persons participated in specific roles, such as the examiner, the unfamiliar stranger, and the videographer. An experienced observer (BCC) scored the dogs' reactions in real time during the trials using standardized scoring sheets (Table 2). All scores were subsequently verified by video analysis.

Inter-rater reliability of the ERT

Standardization of scoring was evaluated in terms of inter-rater reliability. Three experienced, independent observers (BLS, MEG, and BCC) viewed the 16 video recordings of the ERT, played back on a 17″ computer monitor. The observers scored each dog's response to each ERT subtask using standardized and clearly defined behavioral descriptions associated with a 5-point ordinal scale (Table 2). Each score category was mutually exclusive, so that each observer assigned only 1 score per subtask for each dog. An individual (VD), masked to the identity of the evaluators, tabulated the scores.

Test-retest consistency of the ERT

To evaluate the test-retest consistency attribute of the ERT, aggregate scores for 11 subtasks from tests performed on 12 of the 16 dogs at the K2 training facility shortly after arrival were compared with aggregate scores for the same 11 subtasks of the ERT performed on the same dogs at NCSU, 10-12 months later. The remaining 4 of the 16 dogs had been tested at the K2 training facility, but their scores were not available.

Convergent validity of the ERT

To establish convergent validity of the 16 dogs' scores on the ERT, 2 independent measures of emotional "stress" responses were obtained: (1) measurement of salivary and plasma cortisol concentrations (pre- and post-ERT) and (2) responses on an open-field anxiety test.

Salivary and plasma cortisol sampling

In preparation for salivary and plasma cortisol sampling, all 16 dogs were acclimated to the necessary restraint by trained technicians familiar to the dogs. To facilitate the salivary cortisol sampling protocol, dogs were individually trained to sit on cue for a food reward with minimal restraint. Then, once or twice per day for 5 days, dogs were conditioned to allow one end of a cotton rope to be placed in the lateral commissure of their mouth while the experimenter held a small piece of odoriferous food treat (Pupperoni; DelMonte Foods, San Francisco, CA, USA) in a closed hand in front of the dog. The dog was encouraged to sniff the treat to stimulate salivation. After collecting an adequate sample volume (>0.3 mL), and within 2-3 minutes of the start of sampling, the dog was given the treat. This method facilitated rapid collection of an adequate volume of saliva (Bennett and Hayssen, 2010).

Approximately 1 week before the ERT, baseline saliva and baseline blood plasma samples were collected from each dog using standard sample collection and handling protocols (Haverbeke et al., 2008, Hennessy et al., 1997). Sample collection was

Table 2
The emotional reactivity test (ERT), with 17 subtasks, descriptions, and subtask scores

Dog Name ________________________ Dog ID# _____________ Date ______________

1. Stranger examination 1

The handler and dog approach a stranger. If the dog does not approach after the handler is standing next to the stranger, the stranger shall crouch down and speak to the dog. Stranger performs a cursory examination on the dog (run hands along dog's body). The dog's reaction to the examination and willingness to approach the stranger are scored.

1. Dog actively attempts to escape or growls and threatens.
2. Dog withdraws or shrinks away from person, nervousness. Note if dog is flank shy.
3. Dog accepts examination, indifferent to stranger.
4. Dog accepts examination, attentive to stranger.
5. Dog accepts examination, actively seeks play with stranger, excited.

2. Stairs—Up

The dog is walked up a flight of open riser stairs. The dog's willingness to walk up the stairs is scored.

1. Dog refuses to ascend stairs, cannot be motivated to proceed up stairs.
2. Dog requires active and continuous motivation to proceed up stairs.
3. Dog, with initial handler encouragement, moves up stairs, tentative (low posture).
4. Dog hesitates before or on stepping on stair, and then moves easily.
5. Dog moves onto and up stairs without hesitation.

3. Stairs—Down

The dog is walked down a flight of open riser stairs. The dog's willingness to walk down the stairs is scored.

1. Dog refuses to descend stairs, cannot be motivated to proceed down stairs.
2. Dog requires active and continuous motivation to proceed down stairs.
3. Dog, with initial handler encouragement, moves down stairs, tentative (low posture)
4. Dog hesitates before or on stepping on stair, and then moves easily.
5. Dog moves onto and down stairs without hesitation.

4. Visual startle—Bag drop—Reaction

A black plastic bag (76.2 × 83.8 cm) containing crumpled paper and weights totaling 1.4 kg (3 lb) is hung from the ceiling via a pulley 2 m above the floor. The handler walks the dog toward the bag. The bag is dropped to the floor approximately 4 m in front of the dog as it is walking. The dog's reaction is scored.

1. Dog is fearful, bolts to end of leash facing away or turns away from bag.
2. Dog is startled, steps backward (>2 steps), pronounced movement/retreat, remains facing bag.
3. Dog stops and crouches, may step back (1-2 steps), remains facing bag.
4. Dog stops briefly, or flinches, transient reaction, recovers quickly.
5. Dog shows no fear reaction to being startled.

5. Visual startle—Bag drop—Approach

The dog's willingness to approach and investigate the black plastic bag is scored.

1. Dog refuses to approach object despite handler motivation.
2. Dog requires handler motivation to approach object.
3. Dog approaches hesitantly (start/stop avoidance), angles toward object on curving path.
4. Dog approaches cautiously (may have low posture) but directly.
5. Dog approaches immediately without hesitation.

6. Acoustic startle—Grate drop front—Reaction

Handler walks the dog toward the grates. A metal grate is dropped (by pulling a control rope) in front of the dog within approximately 3 m as it is walking. The dog's reaction is scored.

1. Dog is fearful, bolts to end of leash facing away or turns away from sound.
2. Dog is startled, steps backward (>2 steps), pronounced movement/retreat, remains facing sound.
3. Dog stops and crouches, may step back (1-2 steps), remains facing sound.
4. Dog stops briefly or flinches, transient reaction, recovers quickly.
5. Dog shows no fear reaction to being startled.

7. Acoustic startle—Grate drop—Approach

The dog's willingness to approach and investigate the dropped grate is scored.

1. Dog refuses to approach grate despite handler motivation.
2. Dog requires handler motivation to approach grate.
3. Dog approaches hesitantly (start/stop avoidance), angles toward grate on curving path.
4. Dog approaches grate cautiously (may have low posture) but directly.
5. Dog approaches grate immediately without hesitation.

8. Acoustic startle—Grate drop behind—Reaction

As the dog is walked away from the first grate (Subtask 6 and 7), a second metal grate is dropped behind the dog and the dog's reaction is scored.

1. Dog is fearful, bolts to end of leash.
2. Dog is startled, pronounced movement.
3. Dog stops and crouches.
4. Dog stops briefly or flinches, transient reaction, recovers quickly.
5. Dog shows no fear reaction to being startled.

9. Remote control vehicle—Reaction

A small remote control vehicle (approximately 38 cm in length × 20 cm in width × 20 cm in height).is driven out from behind a barrier toward the dog, and is moved back and forth 2 times. The dog's reaction is scored.

1. Dog retreats behind handler immediately and remains.
2. Dog retreats behind handler after initial approach.
3. Dog retreats (steps back) either immediately or after initial approach, but not further than handler's side, remains facing car.
4. Dog steps back (mild retreat) then approaches or shows intermittent displacement behaviors (lip licking, yawn, sniff, looks away).
5. Dog shows no fear reaction to car movement, no withdrawal.

10. Remote control vehicle—Approach

The dog's willingness to approach and investigate the vehicle is scored.

1. Dog refuses to approach object despite handler motivation.
2. Dog requires handler motivation to approach object.
3. Dog approaches hesitantly (start/stop avoidance), angles toward object on curving path.
4. Dog approaches cautiously (may have low posture) but directly.
5. Dog approaches immediately without hesitation.

(*continued on next page*)

Table 2 (*continued*)

11. Unusual stranger—Fear
The dog is positioned next to the handler. Handler and dog face the unusual stranger (person wearing a burqa). The unusual stranger begins walking toward the dog, pausing every 3 m then stops when within 9 m of the dog. The dog's reaction is scored.
1. Dog escapes behind handler before stranger is half of the distance.
2. Dog escapes behind handler after stranger is half of the distance.
3. Dog retreats (steps back), but never further back than handler's side.
4. Dog shows intermittent displacement behaviors (lip licking, yawn, sniff, and looks away).
5. Dog shows little or no reaction to stranger, no withdrawal.

12. Unusual stranger—Aggression
Any aggression by the dog is scored as the stranger approaches (Subtask 11) and the dog is walked toward the stranger (Subtask 13). If the dog shows any aggression toward the unusual stranger, the dog's forward motion is stopped and the dog is moved further away from the unusual stranger.
1. Dog lunges at stranger before stranger is half the distance to dog.
2.Dog shows repeated aggression with lunges after stranger is half the distance.
3. Dog shows repeated aggressive threats (growling or curling lip), but no lunge.
4. Dog shows mild, intermittent aggression (piloerection, aggressive bark).
5. Dog shows no aggression.

13. Unusual stranger—Approach
The dog's willingness to approach and greet the stranger is scored. If the dog refuses to approach, the unusual stranger shall crouch down and speak to dog.
1. Dog does not approach stranger.
2. Dog approaches after stranger crouches down and speaks to dog.
3. Dog approaches when handler is next to stranger.
4. Dog approaches as soon as handler moves toward stranger.
5. Dog approaches stranger immediately and independently.

14. Stranger examination 2—Approach
Handler and dog approach a stranger (the same stranger as subtask 1). If the dog does not approach after the handler is standing next to the stranger, the stranger shall crouch down and speak to the dog. The dog's willingness to approach the stranger is scored.
1. Dog does not approach stranger.
2. Dog approaches after stranger crouches down and speaks to dog.
3. Dog approaches when handler is next to stranger.
4. Dog approaches as soon as handler moves toward stranger.
5. Dog approaches stranger immediately and independently.

15. Stranger examination 2—Repeat examination
The stranger repeats a cursory examination on the dog (run hands along dog's body). The dog's reaction to the examination is scored.
1. Dog actively attempts to escape or growls and threatens.
2. Dog withdraws or shrinks away from person, nervousness. Note if dog is flank shy.
3. Dog accepts examination, indifferent to stranger.
4. Dog accepts examination, attentive to stranger.
5. Dog accepts examination, actively seeks play with stranger, excited.

16. Umbrella startle
The handler walks the dog toward person with an automatic umbrella orienting the dog to the person. The person holds closed umbrella upright, then opens the umbrella when dog is within 3 ft. The dog's reaction to the umbrella opening is scored.
1. Dog is fearful, bolts to end of leash facing away or turns away from umbrella.
2. Dog is startled, steps backward (>2 steps) or, pronounced movement/retreat, remains facing umbrella, then may slink away.
3. Dog stops and crouches, may step back (1-2 steps), remains facing umbrella.
4. Dog stops briefly or flinches in a transient reaction, recovers quickly.
5. Dog shows no fear reaction to being startled.

17. Gunfire[a]
The handler walks the dog toward a "shooter," a seated person with gun. The shooter fires the first blank when the dog is 100 ft. away. The handler continues waking dog toward shooter. The shooter fires the second blank when the dog is within 75 ft. of the shooter. The handler and dog continue walking and the shooter fires the third blank when the dog is within 50 ft. of the shooter. The dog's reaction is scored.
1. Dog is fearful, bolts to end of leash, marked escape attempts.
2. Dog fearful, some effort to escape/retreat, avoidance.
3. Dog is mildly fearful, may show displacement behaviors (lip licking, yawn, sniff, looks away), crouching, no recovery.
4. Dog stops briefly, orients toward sound, transient reaction, recovers quickly.
5. Dog shows no fear reaction to being startled, may lift head, activity remains normal.

[a] The actual test used in the present report was adapted as follows owing to firearms restrictions at North Carolina State University. The handler walks the dog toward a seated person. A gunfire sound recording of a blank fired 100 ft. away is played to the dog. After a pause, a gunfire sound recording of a blank fired 75 ft. away is played to the dog. After a final pause, the sound of a blank fired 50 ft. away is played to the dog. The dog's reaction is scored.

performed between 13:00 and 16:00 hours to minimize the impact of circadian influences on cortisol levels (Dreschel and Granger, 2009). During baseline sample collection, dogs were handled by familiar technicians who followed the protocol established during the conditioning period. There were no stimulating events or visitors to the kennel during the collection period and at least 2 hours before. Saliva was collected with a 15-cm piece of test-specific cotton rope (Salimetrics, State College, PA) at least 2 hours after a meal. The wet end of the rope was placed in a centrifuge tube and kept on ice until the sample could be extracted. Within 4 hours of sample collection, samples were centrifuged at 4°C for 20 minutes at 1300 $\times g$ to isolate the saliva. Saliva (0.1–1.8 mL) was transferred to a microfuge tube, stored at −20°C, and analyzed (in duplicate) using a Salimetrics high-sensitivity salivary cortisol enzyme immunoassay kit. This competitive immunoassay has a limit of detection of 0.03 μg/dL.

Immediately after the saliva was collected, a blood sample (4-6 mL) was collected from each dog's cephalic vein, using a butterfly catheter and Vacutainer tube with ethylenediaminetetraacetic acid (BD, Franklin Lakes, NJ). The blood sample was centrifuged at 4°C for 15 minutes at 1300 $\times g$ to separate the plasma. The plasma was removed by pipette and placed in microfuge tubes for storage at −20°C until analysis. Plasma cortisol was measured using an Immulite 1000 Cortisol kit (Siemens Healthcare Diagnostics, Tarrytown, NY), a solid-phase, competitive chemiluminescent enzyme immunoassay with a limit of detection of 1 μg/dL. For data analysis purposes, a value of 0.99 μg/dL was assigned when the sample concentration was below the limit of detection.

Samples were collected from each dog using the same protocol as described 10-15 minutes after completion of the ERT. Salivary samples were collected first, followed by plasma samples. The time from start of saliva collection to completion of blood sample collection was variable, ranging from approximately 5-10 minutes. The salivary and plasma cortisol results from baseline and post-ERT samples were statistically compared.

Open-field anxiety scores

To further validate the ERT, an independent open-field anxiety score was generated for each dog (Araujo et al, 2013; Landsberg et al., 2013). The interior open-field arena had a floor size of approximately 3 m × 3 m. Two video cameras (ICD-49 B/W camera, Ikegami Tsushinki, Co., Ltd., Tokyo, Japan) recorded each dog's movement, one placed overhead to detect motor activity and one placed laterally to detect postural changes. The horizontal camera was fitted with an infrared filter and illuminator (IR-ROOM Ultra-Covert 940 nm, Nightvisionexperts.com, Buffalo, NY). All tests were conducted from between 13:00 to 16:00 hours. The arena floor was sanitized (Virkon-S; Dupont, Fayetteville, NC) and allowed to air dry between subjects. To acclimate the dogs to the arena, on Day 1, 2 weeks after completion of the ERT test battery, the dogs were individually placed in the open-field arena for 9 minutes. On Days 2 and 4, each dog was again individually placed in the arena. The first 3 minutes of this day served as a control period, with no sound stimuli exposure. This was followed by 3 minutes of exposure to thunderstorm sound recording (source: CanCog Technologies, Inc., Toronto, CA) played back at mean sound level of 88.8 dB (decibel) through overhead speakers (Day 2) or 3 minutes of exposure to gunfire sound recording (source: K2 Solutions, Inc., Southern Pines, NC) played back at a mean sound level of 95.2 dB (Day 4), and followed by 3 minutes of recovery (no sound), after which the dogs were removed from the arena. On Days 3 and 5, the dogs were again exposed to the arena for 9 minutes without sound exposure.

To assess the dogs' stress responses, the videos were randomized and identified by code. A trained technician (BCC), masked to treatment condition (day and sound condition), applied a dedicated behavioral analysis program (Ethovision; Noldus Technologies, The Netherlands), which tracked the ambulatory movement of the dog in the open field. In addition, from the video recordings, the technician quantified active and inactive anxiety signs for each dog. Active anxiety-associated behaviors included startling; bolting; vigilance; scanning; and active responses, such as pacing, aimless activity, stereotypic circling, retreat/escape attempts, digging, and climbing. Inactive anxiety-associated behaviors included decreased activity, such as freezing; positioning in corners, against the wall, or at door; lowered body postures, such as crouching, tail tucking, and ears back; and autonomic/conflict behaviors, such as panting, shaking, salivating, yawning, lip licking, or elimination. Then, using a predetermined 6-point scoring rubric (Table 3) based on the frequency and duration of active and inactive anxiety-associated behavioral postures and motor activity (Landsberg et al., 2013), an anxiety score was determined for each dog during each of four 3-minutes test conditions: pre-thunderstorm (Day 2), thunderstorm (Day 2), pre-gunfire (Day 4), and gunfire (Day 4). The difference in scores between the treatment period and its pretreatment period for each day was calculated, generating a change in anxiety for thunderstorm sounds and a change in anxiety for gunfire sounds. The open-field anxiety score was the mean of these 2 values (Table 1).

Table 3
The scoring rubric for generation of anxiety scores

Anxiety score	Expression of anxiety behaviors
1	None; No anxiety behaviors
2	Occasional and mild
3	Some of the time and mild/occasional and moderate
4	Most of the time and mild/some of the time and moderate/ occasional and severe
5	Some of the time and severe/most of the time and moderate
6	Most of the time and severe

Statistical analysis

All ERT data were visually inspected before analysis. Because the group size for spayed versus intact female dogs was small, all female dogs were analyzed collectively, irrespective of their reproductive status. When a factor was identified as not statistically significant, the data were pooled appropriately. The Shapiro-Wilk test for normality and visual inspections of histograms were used to determine if the scores were normally distributed. Because subtasks were scored on a 5-point ordinal scale and results of subtasks and aggregate scores were not normally distributed, nonparametric tests (such as Spearman rank correlation and Wilcoxon signed rank test) were used unless otherwise noted. To calculate inter-rater reliability of the modified ERT, we calculated an intraclass correlation (ICC) using Shrout and Fleiss ICC, Model 2, Form 1 (Gourraud et al., 2012; Shrout and Fleiss, 1979). Changes from baseline to post-ERT in cortisol scores were analyzed using the Wilcoxon signed rank test. Relationships between results from different assessments were evaluated using Spearman rank correlations. Changes in mean anxiety scores were compared with control periods using Wilcoxon signed rank tests.

We used SAS v9.2 (SAS, Cary, NC) for statistical analysis. The results were considered statistically significant if $P \leq 0.05$. All results are shown for all dogs ($n = 16$) unless otherwise specified; mean values are provided ± standard deviation.

Results

Emotional reactivity test

All dogs completed all subtasks without interruption. The mean ERT score was 65.3 ± 11.46; range was 36-75 (Table 1). There was not a significant difference in ERT scores by sex (Wilcoxon rank sum, $W = 76$, $P = 0.4412$) or coat color (Wilcoxon rank sum, $n_1 = 10$ black, $n_2 = 6$ yellow, $W = 44$, $P = 0.4897$).

ERT: Inter-rater reliability

The ERT total scores of the individual observers were highly concordant [Shrout-Fleiss's ICC (2,1) = 0.97], indicating a high inter-rater reliability of the ERT scores. The individual ERT subtask scores of the individual observers were also highly concordant (13 of 17 subtasks [Shrout-Fleiss's ICC (2,1) > 0.75]). Three subtask scores were moderately concordant ([Shrout-Fleiss's ICC (2,1) = 0.65–0.75; Table 4), and 1 subtask score was poorly concordant ([Shrout-Fleiss's ICC (2,1) = 0.44, walk down stairs]).

Consistency of the subset ERT scores: Test-retest consistency

Test-retest consistency was calculated for 12 dogs for whom scores on an earlier 11-subtask ERT test conducted 10-12 months previously at K2 were available. For each dog, we compared subset scores available in both the K2-ERT and the NCSU-ERT. Neither the subset K2-ERT scores (Shapiro-Wilk, $n = 12$, $W = 0.82$, $P = 0.0155$) or subset NCSU-ERT scores (Shapiro-Wilk, $n = 12$, $W = 0.86$, $P = 0.0487$) were normally distributed. The variance of the subset

Table 4
Subtasks found in the ERT and evaluation of inter-rater reliability

Subtask	ERT subtasks including order of completion	ERT ICC inter-rater reliability by subtask[a]
1	Stranger examination 1	0.96
2	Stairs—Walk up	0.90
3	Stairs—Walk down	0.44
4	Visual startle (bag drop) reaction	0.98
5	Visual startle (bag drop) approach	0.77
6	Acoustic startle grate drop front reaction	0.87
7	Acoustic startle grate drop front approach	0.94
8	Acoustic startle grate drop behind	0.89
9	Remote control vehicle reaction	0.79
10	Remote control vehicle approach	0.91
11	Unusual stranger reaction	0.88
12	Unusual stranger aggression	0.74
13	Unusual stranger approach	0.98
14	Stranger examination 2—Approach	0.65
15	Stranger Examination 2—Repeat examination	0.71
16	Umbrella startle	0.97
17	Gunfire	0.88
	Total = 17	0.97[b]
	Maximum possible number of points = 85	

ERT, emotional reactivity test; ICC, intraclass correlation.
[a] Shrout-Fleiss's ICC (2,1).
[b] Based on ERT total scores.

K2-ERT was not significantly different from the variance of the subset NCSU-ERT score (variances 139.17 and 178.81, respectively; Folded F test, n = 12, $F = 1.28$, $P = 0.6849$). For each of the 12 dogs, the subset K2-ERT and subset NCSU-ERT aggregate scores were correlated (Spearman rank, n = 12, r = 0.73, $P = 0.0073$). There was strong test-retest reliability between the 2 subset ERT aggregate scores (Cronbach's alpha = 0.89).

Convergent validity of the ERT: Cortisol levels

Baseline and post-ERT salivary and plasma cortisol samples were collected from each dog (Table 1). The change in salivary and plasma cortisol was defined as the difference between the post-ERT and baseline levels. Five dogs had baseline plasma cortisol concentrations below the assay limit of detection (<1.00 μg/dL). For these dogs, a conservative value of 0.99 μg/dL was assigned. After the ERT, there was a statistically significant increase in salivary cortisol levels (Wilcoxon signed rank, S = 38.5, $P = 0.0458$) and plasma cortisol levels (Wilcoxon signed rank, S = 68, $P < 0.0001$), compared with baseline.

For all dogs, there was not a significant correlation between changes in salivary and changes in plasma values (Spearman rank, r = 0.24, $P = 0.3804$). However, if the 5 dogs arbitrarily assigned the same pre-ERT plasma cortisol level were omitted from the analysis because the precise measures of their baseline plasma cortisol concentrations are unknown, then there was a strong correlation between plasma and salivary cortisol levels (Spearman rank, n = 11, r = 0.89, $P = 0.0002$). Although absolute post-ERT plasma cortisol values were not significantly correlated with ERT scores (Spearman rank, r = −0.44, $P = 0.0877$), when the 5 dogs with baseline plasma cortisol values below limit of detection were excluded, there was a significant inverse correlation of change in plasma cortisol values with ERT score (Spearman rank, r = −0.68, $P = 0.0208$), as shown in Figure 1.

As shown in Figure 2, there was a significant inverse correlation of change in salivary cortisol values with ERT score (Spearman rank, r = −0.58, $P = 0.0180$). Absolute post-ERT salivary cortisol values were not significantly correlated with ERT scores (Spearman rank, r = −0.49, $P = 0.0546$).

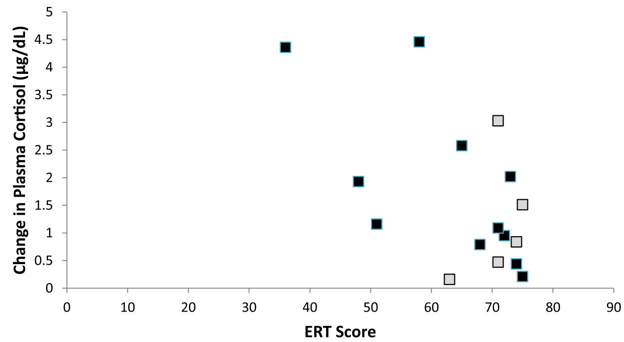

Figure 1. Relationship between ERT Scores and change in plasma cortisol (n = 16). Grey squares indicate changes from baseline values calculated using a nominal baseline value of 0.99 μg/dL, which was assigned when baseline values were below the limit of detection (<1 μg/dL). ERT, emotional reactivity test.

There were no differences by sex in baseline salivary (Wilcoxon rank sum, W = 73, $P = 0.6433$) or plasma (Wilcoxon rank sum, W = 61, $P = 0.4986$) cortisol levels and no differences by sex in changes in salivary (Wilcoxon rank sum, W = 67, $P = 0.9588$) or changes in plasma (Wilcoxon rank sum, W = 57, $P = 0.2875$) cortisol levels.

Convergent validity of the ERT: Anxiety scores in response to sound stimuli

In response to the sound stimuli in the open field, most dogs showed an increase in mean anxiety scores compared with control periods as reflected in an active anxiety score difference (Table 1, Wilcoxon signed rank, S = 51.5, $P = 0.0002$). As shown in Figure 3, the magnitude of the anxiety response was significantly negatively correlated with the dogs' ERT score (Spearman rank, r = −0.57, $P = 0.0214$).

Discussion

The ERT, described herein, was developed in consultation with experienced military working dog trainers and evaluators and United States Department of Defense personnel to establish the needs of the IDD program and verify how the ERT would be executed and used. Based on these discussions and a review of the IDD program's use of a similar 12-subtask instrument, we developed the present ERT, comprising 17 provocative subtasks with associated score descriptions, and evaluated it in terms of inter-rated reliability, test-retest consistency, and convergent validity. If

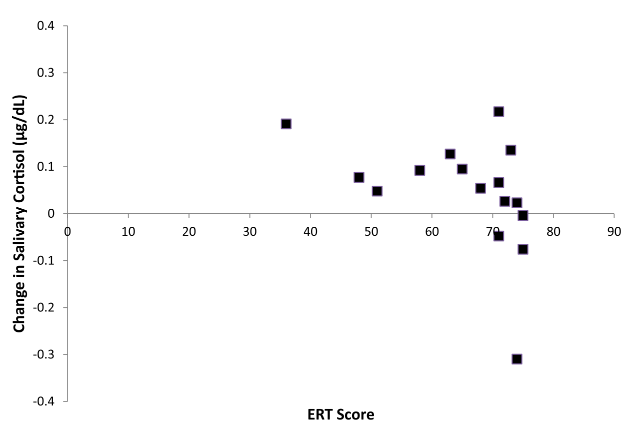

Figure 2. Relationship between ERT scores and change in salivary cortisol (n = 16). ERT, emotional reactivity test.

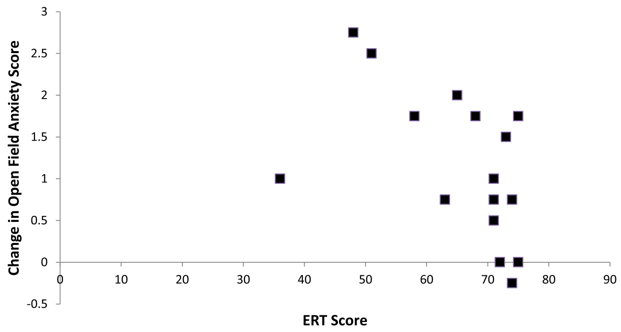

Figure 3. Relationship between ERT scores and change in open-field anxiety scores (n = 16). ERT, emotional reactivity test.

the ERT reflects the behavioral profile of individual dogs, and generalizes to other situations, it could serve as one predictor of IDD performance in the uncertain environments of combat situations.

Inter-rater reliability for the total test and for most of the subtasks was high, indicating the ability of trained observers to score the dogs' performance reliably. Ratings of individual tests showed a high degree of concordance (ICC > 0.70) except for walk down stairs task. These findings indicate that, using the ERT scoring system, the same dog could be evaluated similarly by different trained individuals. Not all published reports that assess dog behavior assess inter-rater reliability (Sinn et al., 2010). However, inter-rater reliability is a critical assurance that the scoring system is well defined and can be replicated.

The ERT scores of the dogs ranged from 36 to 75, of a total possible score of 85, indicating behavioral variation among the dogs tested (Bollen and Horowitz, 2008; van der Borg et al., 2010). A high score represented resilience to the potentially fear-inducing challenges of the test and a low score represented anxiety and fear responses, such as startle, postural changes, and avoidance. Emotional resilience, rapid recovery from fear-inducing stimuli, is a critical trait in dogs performing in combat situations. The correlation of the ERT test-retest results over 10-12 months of elapsed time suggests that the test results reflect behavioral consistency of individual dogs across time (Fratkin et al., 2013).

When a measured trait correlates with a test outcome, the resulting correlation validates the test (Jones and Gosling, 2005). The ERT was validated with physiological and behavioral measures that demonstrated a conceptualization of the construct of emotional resilience. As a physiological measure, we quantified the dogs' plasma and saliva cortisol concentration. Other studies have shown plasma cortisol levels to be useful measures of the physiologic response of working dogs to environmental challenges (Haverbeke et al., 2008), fear provocation (Hydbring-Sandberg et al., 2004), stress (Beerda et al., 1999), and enrichment (Lefebvre et al., 2009). Similarly, elevated salivary cortisol has been shown to be associated with fearfulness in temperament tests in retrievers (Batt et al., 2009) and in response to different types of fearful stimuli (Beerda et al., 1998). We collected both salivary and plasma cortisol samples as baseline, 1 week before, and immediately after the ERT to determine if the test produced stimulation of the hypothalamic-pituitary axis, indicative of emotional challenges (Haverbeke et al., 2008). The data revealed statistically significant elevations in salivary cortisol concentrations after the ERT test compared with baseline values. Elevations in plasma cortisol concentrations after the ERT were significant if the 5 dogs for whom plasma cortisol levels were below the level of test detection were omitted from the analysis. The change in salivary and plasma cortisol was significantly inversely correlated to individual dogs' ERT scores, suggesting that dogs with emotional resilience showed a smaller increase in cortisol as a result of the ERT test battery. Dogs that showed the greatest "emotional reactivity" (i.e., lowest scores on the ERT) also showed the greatest increase in plasma cortisol and salivary cortisol after the ERT compared with baseline values. In addition, these findings suggest that the ERT results reflect a "stress" response and not simply an "arousal" response.

The open-field anxiety test further validates the ERT results. The open-field paradigm has been used in many species to evaluate behavioral responses (Prut and Belzung, 2003). It has been used in dog to identify individuals fearful of sounds and to evaluate the efficacy of anxiolytic drugs and other agents on fear responses (Araujo et al., 2013). This is its first application to military working dog evaluation (Gruen et al., 2015, unpublished data). During the open-field test, dogs were individually subjected to relevant high-amplitude sound stimuli (thunderstorm and gunfire sounds). The dogs' anxiety response, expressed with active or inactive signs of anxiety was calculated (Landsberg et al., 2013; Wilsson and Sinn, 2012) and compared with their behavior during control periods. We found that ERT scores were significantly inversely correlated with the open-field anxiety scores. Dogs with relatively low ERT scores demonstrated larger increases in anxiety scores during open-field sound tests compared with quiet control periods, providing independent convergent validity of the ERT.

There are several important limitations to our study that need to be considered. First, it remains to be determined whether the aggregate ERT predicts a dogs' completion of training, and its response in a working (combat) environment. In the case of an IDD, the most meaningful predictive validation of the ERT for canine emotional reactivity would be obtained by evaluating the dog's effectiveness at detecting improvised explosive devices and their behavioral resilience in combat and other harsh conditions. The USMC has adopted the ERT described in this study and has used this test to assess the most recently trained dogs used for explosive detection in Afghanistan. Ongoing efforts will follow these dogs from procurement to deployment and eventual return. Handler observations may be used to determine their "emotional resilience" over time. However, field recording of dog performance and/or under combat conditions may be incomplete. Second, we were limited in our ability to evaluate test-retest consistency on the ERT. We were able to evaluate individual dog's performance on 11 subtasks that were included in an earlier test used by K2. We found a strong correlation between our test and the K2-ERT test over all subset scores. Repeating the ERT after training and post-deployment would verify the strong test-retest correlation we observed for the ERT. The ERT results for individual dogs might prove useful in identifying combat-related "stress" effects.

Conclusions

Dogs used in combat to detect improvised explosive devices require emotional resilience to fearful stimuli. The purpose of this study was to evaluate the usefulness of an ERT, a series of sequential behavioral challenges used as a selection tool for Labrador retrievers considered for IDD training. We obtained excellent interobserver reliability such that trained individuals independently scored the tests similarly. We also established the convergent validity of the ERT based on physiological and behavioral responses. Although larger sample sizes and in-field performance data are needed for additional verification, these findings suggest the usefulness of the ERT as a behavioral screening test for candidate IDDs.

Acknowledgments

This work was funded by a contract with K2 Solutions, Inc. from the United States Office of Naval Research, which has approved this report for publication. Lisa Albuquerque and Michael Hoglund provided logistical support. The authors thank the North Carolina State University Laboratory Animal Resources staff for animal care and K2 Solutions personnel for facilitating this study. Dr. Margaret E. Gruen receives support from the NIH Ruth L. Kirschstein National Research Service Award T32OD011130.

Conflict of interest

The authors report that they have no conflict of interest. All authors report that they have each made substantial contributions to this article. The idea for the article was conceived by Barbara Sherman. The experiments were designed by Barbara L. Sherman, Margaret E. Gruen, and David C. Dorman. The experiments were performed by Beth C. Case, Melanie L. Foster, Richard E. Fish, and Lucia Lazarowski. The data were analyzed by Venita DePuy. The paper was written by Barbara L. Sherman and Margaret E. Gruen with input from David C. Dorman and Richard E. Fish.

References

Araujo, J.A., de Rivera, C., Landsberg, G., Milgram, N.W., 2013. Development and validation of a novel laboratory model of sound-induced fear and anxiety in Beagle dogs. J. Vet. Behav.: Clin. Appl. Res. 8, 204–212.

Batt, L.S., Batt, M.S., Baguley, J.A., McGreevy, P.D., 2009. The relationships between motor lateralization, salivary cortisol concentrations and behavior in dogs. J. Vet. Behav.: Clin. Appl. Res. 4, 216–222.

Beerda, B., Schilder, M.B.H., van Hooff, J.A., 1998. Behavioral, saliva cortisol and heart rate responses to different types of stimuli in dogs. Appl. Anim. Behav. Sci. 58, 365–381.

Beerda, B., Schilder, M.B.H., Bernadina, W., Van Hooff, J.A., De Vries, H.W., Moi, J.A., 1999. Chronic stress in dogs subjected to social and spatial restriction. II. Hormonal and immunological responses. Physiol. Behav. 66, 243–254.

Bennett, A., Hayssen, V., 2010. Measuring cortisol in hair and saliva from dogs: coat color and pigment differences. Domest. Anim. Endocrinol. 39, 171–180.

Blackwell, E.J., Bradshaw, J.W.S., Casey, R.A., 2013. Fear responses to noises in domestic dogs: prevalence, risk factors and co-occurrence with other fear related behavior. Appl. Anim. Behav. Sci. 145, 15–25.

Bollen, K.S., Horowitz, J., 2008. Behavioral evaluation and demographic information in the assessment of aggressiveness in shelter dogs. Appl. Anim. Behav. Sci. 112, 120–135.

Crawford, K.H., 2012. Marine's Best Friend Shows Explosive-Detecting Capabilities. Available at. http://www.onr.navy.mil/Media-Center/Press-Releases/2012/IED-Explosive-Device-Detect-Dog.aspx. Accessed May 4, 2014.

Diederich, C., Giffroy, J.-M., 2006. Behavioural testing in dogs: a review of methodology in search for standardisation. Appl. Anim. Behav. Sci. 97, 51–72.

Dreschel, N.A., Granger, D.A., 2009. Methods of collection of salivary cortisol measurement in dogs. Horm. Behav. 55, 163–168.

Fratkin, J.L., Sinn, D.L., Patall, E.A., Gosling, S.D., 2013. Personality consistency in dogs: a meta-analysis. PLoS One 8, 1–19.

Goddard, M.E., Beilharz, R.G., 1985. A multivariate analysis of the genetics of fearfulness in potential guide dogs. Behav. Genet. 15, 69–89.

Gourraud, P.-A., Le Gall, C., Puzenat, E., Aubin, F., Ortonne, J.-P., Paul, C.F., 2012. Why statistics matter: limited inter-rater agreement prevents using the psoriasis area and severity index as a unique determinant of therapeutic decision in Psoriasis. J. Invest. Dermatol. 132, 2171–2175.

Haverbeke, A., Diederich, C., Depiereux, E., Giffroy, J.M., 2008. Cortisol and behavioral responses of working dogs to environmental challenges. Physiol. Behav. 93, 59–67.

Haverbeke, A., Rzepa, C., Depiereux, E., Deroo, J., Giffroy, J.-M., Diederich, C., 2010. Assessing efficiency of a human familiarization and training programme on fearfulness and aggressiveness of military dogs. Appl. Anim. Behav. Sci. 123, 143–149.

Hennessy, M.B., Davis, H.N., Williams, M.T., Mellott, C., Douglas, C.W., 1997. Plasma cortisol levels of dogs at a county animal shelter. Physiol. Behav. 62, 485–490.

Hydbring-Sandberg, E., von Walter, L.W., Höglund, K., Svartberg, K., Swenson, L., Forkman, B., 2004. Physiological reactions for fear provocation in dogs. J. Endocrinol. 180, 439–448.

Jones, A.C., Gosling, S.D., 2005. Temperament and personality in dogs (*Canis familiaris*): a review and evaluation of past research. Appl. Anim. Behav. Sci. 95, 1–53.

King, T., Hemsworth, P.H., Coleman, G.J., 2003. Fear of novel and startling stimuli in domestic dogs. Appl. Anim. Behav. Sci. 82, 45–64.

Landsberg. G., Beck, A., Araujo, J.A., Milgram, N.W., 2013. Effect of dog appeasing pheromone collar on a model of sound-induced fear and anxiety. Proceedings of the 9th International Veterinary Behaviour Meeting, Lisbon, PT, Vol. 9, pp. 45–46.

Lazarowski, L., Foster, M.L., Gruen, M.E., Sherman, B.L., Case, B.C., Fish, R.E., et al., 2014. Acquisition of a visual discrimination and reversal learning task by Labrador retrievers. Anim. Cogn. 17, 787–792.

Lefebvre, D., Giffroy, J.M., Diederich, C., 2009. Cortisol and behavioral responses to enrichment in military working dogs. J. Ethol. 27, 255–265.

Murphy, J., 1998. Describing categories of temperament in potential guide dogs for the blind. Appl. Anim. Behav. Sci. 58, 163–178.

Prut, L., Belzung, C., 2003. The open field as a paradigm to measure the effects of drugs on anxiety-like behaviors: a review. Eur. J. Pharmacol. 463, 3–33.

Shrout, P.E., Fleiss, J.L., 1979. Intraclass correlations: uses in assessing rater reliability. Psychol. Bull. 86, 430–438.

Sinn, D.L., Gosling, S.D., Hilliard, S., 2010. Personality and performance in military working dogs: reliability and predictive validity of behavioral tests. Appl. Anim. Behav. Sci. 127, 51–65.

Svartberg, K., 2005. A comparison of behaviour in test and in everyday life: evidence of three consistent boldness-related personality traits in dogs. Appl. Anim. Behav. Sci. 91, 103–128.

van der Borg, J.A.M., Beerda, B., Ooms, M., de Souza, A.S., van Hagen, M., Kemp, B., 2010. Evaluation of behaviour testing for human directed aggression in dogs. Appl. Anim. Behav. Sci. 128, 78–90.

Wilsson, E., Sinn, D.L., 2012. Are there differences between behavioral measurement methods? A comparison of the predictive validity of two ratings methods in a working dog program. Appl. Anim. Behav. Sci. 141, 158–172.

Wilsson, E., Sundgren, P.-E., 1997. The use of a behavior test for the selection of dogs for service and breeding, I: method of testing and evaluating test results in the adult dog, demands on different kinds of service dogs, sex, and breed differences. Appl. Anim. Behav. Sci. 53, 279–295.