JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, January 2015, Vol. 60, No. S1
doi: 10.1111/1556-4029.12626
Available online at: onlinelibrary.wiley.com

TECHNICAL NOTE

GENERAL

Dorothee Rocznik,[1,2] *B.Sc.; David L. Sinn,*[1,3] *Ph.D.; Scott Thomas,*[4] *and Samuel D. Gosling,*[1] *Ph.D.*

Criterion Analysis and Content Validity for Standardized Behavioral Tests in a Detector-Dog Breeding Program*

ABSTRACT: Many working-dog programs assess behavior during a dog's first year of life with the aim of predicting success in the field. However, decisions about which tests to administer are frequently made on the basis of tradition or intuition. This study reports results from a survey given to U.S.A.'s Transportation Security Administration (TSA) detection-dog handlers (N = 34). We categorized and summarized handlers' responses regarding traits they felt were important for work. We used this criterion analysis to examine the content validity of the TSA's puppy tests. Results indicate that 13 of 15 traits that are currently being measured are relevant. However, several traits not currently measured were identified as being highly important, notably "play" and off-duty "calmness." These results provide support that the TSA tests are measuring traits relevant to operational search team performance but also highlight other traits that may be profitable to assess in this and other detection-dog programs.

KEYWORDS: forensic science, detection dogs, animal personality, measurement science, content validity, criterion analysis, questionnaire methods

Working dogs provide a vital role in a wide array of domains that are important to human society [1]. One particularly consequential domain is substance detection, where dogs play a key role in detecting dangerous explosive materials and other illicit substances [2]. As a result of their valuable role, the use of detector dogs is widespread across the globe. As the use of working dogs has increased, programs tasked with breeding, procuring, and training detection dogs have become increasingly interested in identifying factors early in life that can predict subsequent adult success in the field.

Often, a key determinant of adult working success is a dog's "personality" or "temperament" [3]. Accordingly, many working-dog programs administer standardized tests to puppies and young dogs that are designed to assess the behavioral traits thought to play a role in adult field performance [4–9].

[1]Department of Psychology, The University of Texas at Austin, 108 E. Dean Keeton Stop A8000, Austin, TX, 78712-1043.
[2]Department of Psychology, Westfälische Wilhelms-Universität Münster, Fliednerstraße 21, Münster, 48149, Germany.
[3]School of Biological Sciences, University of Tasmania, Private Bag 5, Hobart, TAS 7001, Australia.
[4]Transportation Security Administration Canine Breeding and Development Center, Lackland Air Force Base, San Antonio, TX.

*Funded, in part, by the U.S. Department of Homeland Security, Science & Technology Directorate research contract HSHQDC-10-C-00085: "Improving the Effectiveness of Detector-Dog Selection and Training through Measurement of Behavior and Temperament."

[Correction was added on 27 November 2014 after publication 20 November 2014; a correction was made on Table 3.]

Received 13 June 2013; and in revised form 4 Dec. 2013; accepted 22 Jan. 2014.

Ideally, the selection of standardized behavioral tests would be based on a criterion analysis [10], consisting of a careful description and quantification of the tasks the dogs are called upon to perform in the field and the behavioral traits that help or hinder successful completion of those tasks. Often, however, communication between the field handlers and the organizations responsible for breeding, selecting, and training the dogs is inefficient or lacks clarity. As a result, the selection of traits assessed by breeding and development programs is often determined by precedent set by other working programs and/or the experience and intuition of the individuals charged with developing the breeding and training programs. Typically, once a program is established, traditions are established and there is little opportunity for revising the testing procedures based on feedback from the handlers in the field. As a result, very little is known regarding the extent to which the traits assessed in standardized tests in dog breeding and training programs actually overlap with the traits thought by the handlers to be important during operational performance.

Thus, the goal of the present studyis twofold. First, we aim to identify the traits thought by handlers to be important to their dogs' success in the field; this task can be conceived as a form of criterion analysis. Second, we aim to determine the extent to which those handler-identified traits overlap with the traits assessed in the breeding program that is providing the handlers with dogs; this task can be conceived as determining the content validity of the battery of standardized tests. We examine these questions in the context of dogs bred and deployed as detection dogs by the United States' Transportation and Security Administration (TSA).

This study has the potential to be of strong applied importance in three respects. First, the systematic formal identification of

traits thought to be useful in detection dogs can serve as a useful guide to detector-dog programs around the globe. Second, assessing the content validity of the existing battery of the standardized tests used by the TSA breeding program can identify sources of neglected domains that could potentially be important with regard to establishing the predictive validity of standardized tests; these findings can directly identify new candidate traits for inclusion in subsequent test batteries. Third, this analysis can serve as a model for other working dog programs interested in undertaking criterion analyses and in evaluating the content validity of their assessments.

Methods

Background on TSA-Canine Breeding and Development Center's (TSA-CBDC) Standardized Behavioral Tests

In 2002, staff from the United States' TSA-CBDC began conducting standardized behavioral test assays for its purpose-bred dogs, using a test instrument originally devised by the Australian Customs and Border Patrol detector-dog breeding and development program [11, 12]. Two standardized tests were designed to broadly measure an individual dog's responses in novel environments and during search and retrieve exercises; these domains were chosen because they were thought to be the key criteria for successful detection work in operational conditions. The two tests are given once at 4 different times across the first year of life (at *c*. 3, 6, 9, and 12 months of life), a period thought to be characterized by relatively rapid neural, cognitive, and behavioral development [13–15]. In both test domains, behavioral ratings on five-point Likert scales are made based on observed dog behaviors (Table 1); 10 ratings are given in environment tests and eight ratings are given during search and retrieve tests.

The traits that are measured during environment tests are based on a dog's overall confidence, concentration, responsiveness, initiative, excitability, hearing sensitivity, and body sensitivity during the entire assessment. Dogs are also introduced to a kong during environment tests and are scored according to their observed chase retrieve, independent possession, and physical possession behaviors. The behavior traits measured in the search and retrieve test assess the dog's chasing and retrieving desire, the physical, mental, and independent possession of a scented towel, the dog's ability to find the scented towel in tall grass and underneath empty flowerpot containers, and the dog's overall activity level. The terminology for trait ratings used in the TSA-CBDC standardized tests closely matches that given by the Scientific Working Group on Dog and Orthogonal detector Guidelines (SWGDOG; http://swgdog.fiu.edu/).

Generation of Questionnaire Survey

A questionnaire survey was compiled in 2004 by the TSA-CBDC Program Manager (Scott Thomas [ST], Table 2). Questions were compiled with the aim of determining whether TSA-CBDC standardized behavioral tests that were already in use were relevant for detector dogs working in the field (i.e., questions were designed to assess the content validity of already existing standardized behavioral tests). Questions were also compiled with the goal of identifying behavioral criteria that handlers working in the field found important (i.e., questions were also designed in terms of a criterion analysis).

The questionnaire survey given to TSA handlers consisted of two parts. The first part consisted of eight items focused on the handler background (i.e., not concerning the dogs *per se*). Respondents were asked about the number of years experience they had handling working dogs, detection dogs, explosive-detection dogs, and with handling their present dog. Respondents were also asked about the breed of their present dog and the breeds of detection dogs they had handled previously.

The second part of the survey dealt with *a priori* behavioral categories that the program manager of the TSA-CBDC (ST) intuitively felt to be important; it consisted of nine categories. The

TABLE 1—*Operational definitions of behavioral traits measured during standardized TSA-CBDC environmental sureness (ES) and search and retrieve (SR) tests.*

Behavior rating	Definition	Test domain
Confidence	An environmentally conditioned acceptance of safety; in other words, a measure of lack of fear (high fear = low confidence). Typical signs of fear include extended attention, freezing, and avoidance.	ES
Concentration	The dog's focus on during searches; a measure of the lack of distraction toward objects not related to searches. Typical distraction objects are sticks and leaves on the ground, people, self, and the leash.	ES
Responsiveness	The dog's ability to react to corrections or encouragement (verbal and physical praise) from the handler.	ES
Initiative	A dog's willingness to walk at the end of the leash and investigate the environment on his/her own without being asked by the handler.	ES
Excitability	The dog's enthusiasm during the walk. High excitability is typically characterized by a high degree of activity and movements not related to searching, while low excitability is typically characterized by a lackadaisical approach during testing.	ES
Hearing sensitivity	The dog's reactivity to noise stimulus during environmental testing.	ES
Body sensitivity	The dog's physical reactivity to touch, praise, or correction on a flat collar.	ES
Chase/retrieve	The speed and desire at which the dog runs for a thrown toy.	ES & SR
Independent possession	The dog's willingness to continue to interact and possess the toy independently of the handler.	ES & SR
Physical possession	The dog's desire to play tug-of-war with the handler for the toy, including the force and determination to maintain its grip on the toy.	ES & SR
Mental possession	The dog's ability to focus on the towel, even after it is hidden. Signs of lower mental possession include the dog focusing on the handler more than the towel.	SR
Hidden grass	The dog's enthusiasm and use of olfaction (as opposed to vision) to find a towel hidden from the dog's direct line of sight.	SR
Hunt 1	The dog's concentration, willingness, and ability to move purposefully down a line of upside down flower pots, one of which contains a hidden scented towel. Higher scores are also given to dogs that show a behavioral change when at the scent cone and then self-reward with the found scented towel.	SR
Hunt 2	Identical scores are given during a second hunt search at the flowerpots, when the tester stands at a different location previous to the first hunt search (i.e., at the level of the dog/handler, as opposed to at the end of the flower pot line).	SR
Activity	The dog's ability to use his/her energy effectively.	SR

TABLE 2—*The questionnaire survey given to professional detection dog handlers working for the U.S.A.'s Transportation and Security Administration.*

K9 Handler Experience			
1.	Experience handling working dogs	Years	
2.	Experience handling detection dogs	Years	
3.	Experience handling explosives-detection dogs	Years	
4.	Experience with present dog	Years	
5.	Breed of present dog		
6.	Have you handled previous explosives-detection dogs	YES	NO
7.	If so, please list breeds-		
8.	Please list any additional experience you have in dog training (dog sport, hunting dogs, obedience trails, etc.)		

Observed Behavioral Traits

Please list in order of priority from what you consider to be the most important in each category. Attempt to use one word or short phrase descriptors in describing the observed behavioral traits that make your dog a quality detector. Remember there are no right answers and that we are looking for traits observed and described by the handler.

Environmental Stability—Please describe what makes your canine partner capable of interacting in the distractions of a working environment.
1.
2.
3.
4.
5.

Play Behavior—Please describe your canine partner′s behavior when playing outside the work environment.
1.
2.
3.
4.
5.

Interaction with Handler—Please describe what makes your canine partner capable in interacting with his handler.
1.
2.
3.
4.
5.

Retrieve/Reward Behavior—Please describe how your canine partner behaves when interacting with his reward.
1.
2.
3.
4.
5.

Search Behavior—Please describe your canine partner′s behavior during a search.
1.
2.
3.
4.
5.

Off-Duty Behavior—Please describe the traits of your canine partner′s off-duty behavior.
1.
2.
3.
4.
5.

Physical Abilities—Please describe the physical traits that make your canine partner a quality detector dog.
1.
2.
3.
4.
5.

Aberrant Behavior—Please list unique (unlearned, unrewarded) behavior exhibited by your canine partner.
1.
2.
3.
4.
5.

Please list additional traits you think should be considered

first eight categories asked the participant to list in a free-response format, in order of priority within that category, behavioral traits he/she considered important for a working detector dog. The *a priori* categories given in the survey were "Environmental Stability," "Play Behavior," "Interaction with Handler," "Retrieve/Reward Behavior," "Search Behavior," "Off-duty Behavior," "Physical Abilities," and "Aberrant Behavior." The survey contained a short explanation of each category (Table 2). Participants were able to answer freely, but were asked to rank their responses from 1 to 5 from the most to the least important. The category "Environmental Stability" was chosen because it was one of the domains that was tested during TSA-CBDC standardized tests. The categories

"Retrieve/Reward Behavior" and "Search Behavior" were chosen because they were thought to correspond with the search and retrieve standardized testing domain. The category "Interaction with Handler" was chosen because it was thought to be a component of both the environment and the search and retrieve tests. The four other categories ("Play Behavior," "Off-duty Behavior," "Physical Abilities," and "Aberrant Behavior") were chosen in an attempt to identify seemingly unrelated behavior that also might be important for predicting a dog's working ability. Finally, a ninth category used a free-text format to obtain a list of any additional traits the handler considered to be important in working explosive-detector dogs.

Survey participants did not have any specific time restrictions. The only instruction was to keep the answers short and to list any behavioral traits that were deemed important within each category for working performance. Before filling out the survey, ST presented an overview of the TSA-CBDC program to the participants. During this presentation, participants were also told about the motivations behind the development of the survey as described above.

Human Participants

Thirty-four dog handlers filled out the survey at a TSA conference on September 2, 2004. All participants had been trained by the TSA to handle working dogs. Moreover, at that time participants were all actively serving as local law enforcement officers in partnership with the TSA. Participants' experience with handling working dogs was 8.8 years on average (SD = 7.1), with a minimum of 2 years experience and a maximum of 26 years experience. Participants' mean experience with handling detector dogs was 7.5 years (SD = 6.2), with a minimum of 2 years and a maximum of 25 years. Participants had on average 3.4 years (SD = 2.0) experience with their current dog, with a minimum of 6 months and a maximum of 10 years.

Thirteen participants (38.2%) were currently handling a German shepherd dog, ten (29.4%) were currently handling a Labrador, nine (26.5%) a Belgian Malinois, one (2.9%) a Chesapeake Bay retriever, and one (2.9%) a Belgian Tervuren.

Eighteen (52.9%) participants had handled explosive-detector dogs previous to their current dog. Most of these 18 people (40.7%) had handled a Belgian Malinois previously. Seven participants (25.9%) had handled a German shepherd dog, six (22.2%) a Labrador, two (7.4%) a German shorthair pointer and one (3.7%) a golden retriever. Six handlers had handled more than one of these breeds of explosive-detector dogs previously. Sixteen handlers (47.1%) had not handled explosive-detector dogs previously. All participants were currently handling an explosives detection dog.

Eleven participants (32.4%) listed additional experience in dog training outside of the TSA workplace. Six (17.6%) had additional experience in obedience training, five (14.7%) in hunting-dog training, three (8.8%) in narcotic-detection dog training, three (8.8%) in patrol-dog training, and two (5.9%) in pet training. One (2.9%) had additional experience in therapy-dog training, one (2.9%) was a certified trainer, and one (2.9%) listed "other" training that did not fit into any of these categories.

Criterion Analysis

The criterion analysis consisted of three phases. First, to understand the relative importance of dog behavioral traits within each of the eight *a priori* categories in the survey, a judge (DR) aggregated each of the handler responses within each category into nominated traits[1]. For example, the judge examined the handlers′ responses listed in the category Environmental Stability and searched for similar answers; similar answers were considered to be the same behavioral trait. The judge then examined the handlers′ answers in the category Play Behavior and searched for similar answers within this category; similar responses were grouped again into the same nominated traits. Handler responses within each survey category were grouped into narrower nominated traits in this manner by the judge for all eight categories (Table 3). In some instances, the same nominated trait emerged across *a priori* survey categories. For example, the nominated trait "playfulness" was recorded under both the Interactions with Handler and Off-duty Behavior categories because handlers listed the answer "playful" or similar answers under each category. In the history of personality research, there have been many debates on what counts as a trait [16–19]. For the sake of efficiency of communication, we use the term "trait" here to refer to aggregate handlers' nominations but acknowledge that the behaviors referred to in some of these nominations may not meet some classical definitions of "trait".

Inspection of the responses revealed that participants sometimes listed responses under the Environmental Stability category that actually fell under a different survey category, suggesting that participants did not realize there would subsequently be a more appropriate category under which to report the trait (i.e., Environmental Stability was the first category listed on the survey). In such cases, the trait nominations were assigned to the conceptually appropriate category. There were 12 responses that were reassigned from the category Environmental Stability to the categories Interaction with Handler (11 responses) and Physical Abilities (one response).

Some responses were unique (e.g., "hard feet pads" listed by a handler under the category Physical Abilities). To account for these unique singleton responses, we created an "other" nominated behavioral trait grouping within each survey category. One hundred and three responses (16.9% of the raw data) ranging from 6 to 34 answers per category were categorized as "other" across the eight *a priori* categories. Handler responses classified as "other" were all unique, so we did not consider this nominated trait category further.

Only nine participants made use of the ninth survey category, in which handlers were asked to list any additional traits in a free-response format, not previously listed, that he/she considered to be important in working detector dogs. We interpreted the lack of responses in the ninth category as indicating that the previous eight survey categories had successfully captured the vast majority of traits that handlers considered to be important to detector-dog success. As a result of the low numbers of nominations in the ninth category and the lack of common ground across responses, these data were excluded from the analysis.

In a second phase, we took steps to ensure the assignments of handler responses to nominated behavioral traits were generalizable and not idiosyncratic to the judge who made the initial assignments. Specifically, a second judge (with 5 years experience of working with dogs) independently sorted the handlers' responses into the nominated behavioral traits identified by the first judge. The first judge first gave a written explanation of each nominated behavioral trait within each survey category to the second judge and had a discussion to clarify nominated trait definitions. Next, the second judge was given "dummy data" which resembled handler responses and was asked to sort the data into the nominated behavioral traits. After this training, the

TABLE 3—*Survey categories, nominated behavioral traits, and types of responses given by Transportation Security Administration detection dog handlers that defined nominated behavioral traits.*

Category	Nominated Behavioral Traits	Types of Responses by Handlers
Environmental Stability	General drive/alertness	Willingness, vigor, and/or enthusiasm to engage in work.
	Concentration	A lack of distraction during searching. Typical signs of distraction were inanimate objects (sticks, leaves, etc) and/or other people.
	Confidence	The ability to know when to act on its abilities; anticipation that the behavior could be accomplished safely.
	Hunt drive	General drive to chase objects or other animals.
	Responsiveness	Reactions to corrections or praise. Signs of responsiveness could include those to correction (flat collar) or to encouragement (verbal and physical praise).
	Acclimatization	Ability to settle into new environments; adjustment to a new circumstance.
Play behavior	Sniffing	Use of olfaction while playing.
	Roughness	Intensity of physical contact while playing.
	Barking	Vocalization while playing.
	Endurance	Ability to continue activity through time during play.
	Activity/excitability	Intensity of energy display during play.
	Initiative	Willingness to search and initiate play independently.
	Sociability	Desire to be in close physical proximity with humans and other animals.
	Playfulness	Degree of display of voluntary, intrinsically motivated activities normally associated with pleasure and enjoyment.
Interaction with handler	Focus on reward	Intensity of concentration on reward objects.
	Human focus	Intensity of concentration on gaining attention or positive reinforcement from the handler.
	Independence	Ability to work without the handler′s attention.
	Concentration	A lack of distraction during searching. Typical signs of distraction were inantimate objects (sticks, leaves, etc) and/or other people.
	Human-perceived relationship	Handler perceptions of the relationship with the dog. Typical perceived characteristics attributed to the relationship included love, rapport, trust, respect, loyalty, and bond.
	Responsiveness	Reactions to corrections or praise. Signs of responsiveness could include those to correction (flat collar) or to encouragement (verbal and physical praise).
	Playfulness	Degree of display of voluntary, intrinsically motivated activities normally associated with pleasure and enjoyment.
Retrieve and reward behavior	Endurance	Ability to continue activity through time during retrieval.
	Independent possession	Desire to interact and possess a reward independently of the handler (e.g., carrying, shaking, chewing, throwing to self).
	Physical possession	Desire to hold, chew, and maintain physical possession of a reward during tug-of-war with the handler.
	General possession	Desire for general possession of a reward object. This category included handler answers that referred to possession, but do not specify further (i.e., independent, physical, or mental).
	General drive	Willingness, vigor, or enthusiasm to engage in work.
	Mental possession	Focus on a reward that was previously in sight, but that has been hidden.
	Playfulness	Degree of display of voluntary, intrinsically motivated activities normally associated with pleasure and enjoyment.
	Hunt drive	General drive to chase objects or other animals.
	Chase retrieve	Desire to pursue and pick up a thrown toy.
	Activity/excitability	Intensity of energy display during retrieval of a thrown object.
Search behavior	Concentration	A lack of distraction from searching. Typical signs of distraction were objects on the ground (sticks, leaves, etc) and other people.
	Confidence	The ability to know when to act on its abilities; anticipation that the behavior can be accomplished safely.
	Search drive	General drive to search for a hidden object.
	Activity/excitability	Intensity of energy display during searching for a hidden object.
	Speed	Length of time it takes to find a hidden object.
	Independence	Ability to work and find objects without the handler′s attention.
	Sniffing	Use of olfaction while searching.
	Endurance	Ability to continue activity through time during searching.
Off-duty behavior	Calmness	Ability to sleep, lay in a kennel, and be quiet during off-duty hours.
	Independence	Lack of attachment to the handler outside of work hours.
	Sociability	Desire to be in close physical proximity with humans and other animals.
	Playfulness	Degree of display of voluntary, intrinsically motivated activities normally associated with pleasure and enjoyment.
	Protectiveness	Desire to defend kennel and human family.
Physical abilities	Activity/excitability	Intensity of energy dog exhibits.
	Athleticism	Agility while negotiating obstacles.
	Endurance	Ability to continue activity through time.
	Speed	Quickness to run and climb objects.
	Sniffing	Use of olfaction.
	General health	Susceptibility to disease and physical symptoms of sickness.
	Weight	Body mass and build (e.g., thin, thick, heavy, light).
	Size	Includes whole body size and the size of particular parts of the body, for example, legs.
	Strength	Muscular quality and ability to exert force.
Aberrant behavior	Carrying	Tendency to grab and carry items and small objects for extended lengths of time.
	Eating	Quantity of food eaten and attention to foraging behavior.
	Independent play	Quality of play when left alone.
	Destructiveness	Desire to destroy material things.
	Protectiveness	Desire to defend kennel and human family.
	Barking	Tendency to produce vocalizations.

second judge was given the real data and the second judge independently classified the actual responses into the nominated traits within each category. We used Cohen's Kappa to estimate the inter-rater reliability across the two judges' categorizations [20]. Cohen's Kappa is composed of the sum of the relative frequencies of the corresponding assignments between judges, adjusted by the frequency of corresponding assignments that would result from an assignment by chance. Thus, the result is the amount of corresponding assignments that are above chance [21]. The values of Cohen's kappa can be between −1 and +1. Values above 0.75 are considered to reflect excellent agreement, values of 0.74–0.60 indicate good agreement, values of 0.59–0.40 indicate fair agreement, and values lower than 0.40 indicate poor agreement [22].

The third and final phase of the criterion analysis was to estimate the handler-ranked importance of each behavioral trait. Handlers were asked to rank their responses within each *a priori* category according to their importance (one being the most important, five being the least important), so we were able to compute a mean importance score for each nominated trait within each category. Multiplying the mean score by the frequency of occurrence across different handlers allowed us to compute a weighted importance score. To generate a score in which higher numbers reflected more importance, the ranks were reverse-coded such that a "5" was assigned to the highest rank and a "1" was assigned to the lowest. These three phases allowed us to assess what handlers' thought was important and to formally identify the criterion that the TSA-CBDC's standardized tests are trying to predict to.

Content Validity

To evaluate content validity, we examined the degree of overlap between the nominated behavioral traits identified in the criterion analysis and the behavioral traits assessed during search and retrieve and environmental sureness standardized TSA-CBDC tests. We assessed content validity both qualitatively (e.g., Did the traits measured by the TSA-CBDC tests emerge as traits in the handler survey?) and quantitatively (e.g., Using weighted "importance" scores from the survey are traits currently measured by TSA-CBDC tests thought to be important to detection dog handlers?).

Results

Criterion Analysis

The mean number of nominated traits generated from handler responses per survey category was 8.36 (SD = 6.82). Forty-seven unique nominated behavioral traits were used to classify handler responses from the survey. Six nominated behavioral traits were generated within the "Environmental Stability" category, eight within the "Play Behavior" category, seven within the "Interaction with Handler" category, ten within the "Retrieve/Reward Behavior" category, eight within the "Search Behavior" category, five within the "Off-duty Behavior" category, nine within the "Physical Abilities" category, and six traits within the "Aberrant Behavior" survey category. The nominated behavioral traits are given in Table 4 in order of their weighted importance scores, with the most important listed first.

The estimates of inter-rater reliability indicated good to excellent agreement among the judges regarding how to categorize the handlers' responses into nominated behavioral traits. The Kappa reliability estimates for assigning responses to nominated traits within each of the eight survey categories ranged from 0.63 to 0.80 (mean Kappa = 0.71), reinforcing the idea that the nominated behavioral traits were useful for categorizing the free-text responses by handlers.

Weighted importance scores for nominated traits ranged from 5.00 to 145.08 with an overall grand mean of 32.87 (SD = 28.94). Seven nominated traits in five survey categories had a weighted importance score that was more than one standard deviation away from the overall grand mean weighted importance score. These included concentration (Environmental Stability), playfulness and activity/excitability (Play Behavior), human-perceived relationship and human focus (Interactions with Handler), search drive (Search behavior), and calmness (Off-Duty Behavior).

Other notable nominated traits had high average importance scores (within 10 units of the 1SD cut-off used above). These included general drive/alertness (Environmental Stability), sociability (Play Behavior), physical possession (Retrieve and Reward Behavior), concentration and independence (Search Behavior), and athleticism (Physical Abilities).

The remaining 46 nominated traits all had weighted importance scores >10 units away from the 1SD cut-off, with a mean of 20.22 and a SD of 9.77. The survey category Aberrant Behavior was the only category to not contribute any nominated traits that had high importance scores based on the 1 SD criterion.

Twelve nominated behavioral traits were used to classify responses in more than one survey category, five of these were used in more than two survey categories. Playfulness emerged in four categories (Play Behavior, Interaction with Handler, Retrieve and Reward Behavior, and Off-duty Behavior), endurance and activity/excitability in four (Play Behavior, Retrieve and Reward Behavior, Search Behavior, and Physical Abilities), sniffing appeared in three categories (Play Behavior, Search Behavior, and Physical Abilities), and concentration also appeared in three categories (Environmental Stability, Interaction with Handler, and Search Behavior).

Content Validity

Most of the traits evaluated in the TSA-CBDC standardized tests also emerged as nominated traits from the survey data. Specifically, the TSA-CBDC environment test traits "confidence," "concentration," and "responsiveness" emerged in the survey in the category Environmental Stability. "Concentration" and "responsiveness" also emerged from the survey category Interaction with Handler, and "confidence" and "concentration" emerged from the survey category Search Behavior. The TSA-CBDC environment test trait "initiative" emerged from the survey category Play Behavior, and an analogue to initiative ("independence") emerged in the categories Interaction with Handler, Search Behavior, and Off-duty Behavior. The TSA-CBDC environment test trait "excitability" emerged in the survey categories Play Behavior, Retrieve and Reward Behavior, Search Behavior, and Physical Abilities. The mean importance score given to traits currently measured in TSA-CBDC standardized environment tests was 33.20, SD = 24.44. Only two traits that are measured in the TSA-CBDC environment standardized test were not identified in the survey data, namely hearing and body sensitivity.

The TSA-CBDC search and retrieve test traits "chase retrieve," "mental possession," "physical possession," and "independent possession" were replicated from handlers′

TABLE 4—*Frequency of handler responses, mean priority rank, and weighted importance scores from the questionnaire survey given to detection dog handlers.*

Survey Category	Nominated Behavioral Trait	Frequency of Responses Classified Under Nomination	Mean Priority Rank Given by Handlers	SD of Priority Rank	Weighted Importance of Nominated Behavioral Trait
Environmental stability	Concentration	23	3.83	1.14	88.09
	General drive/alertness	15	3.87	1.06	58.05
	Confidence	6	4.50	0.84	27.00
	Acclimation	6	4.50	0.84	27.00
	Hunt drive	6	3.83	1.17	22.98
	Responsiveness	3	3.67	1.53	11.01
	Other	6			
Play behavior	Playfulness	27	3.78	1.12	102.06
	Activity/excitability	17	4.29	0.98	72.93
	Sociability	13	3.92	0.86	50.96
	Roughness	4	4.75	0.50	19.00
	Endurance	4	3.75	1.50	15.00
	Barking	3	2.67	1.53	8.01
	Sniffing	2	4.00	0.00	8.00
	Initiative	2	2.50	0.71	5.00
	Other	12			
Interaction with handler	Human-perceived relationship	27	4.41	0.97	119.07
	Human focus	23	3.96	1.19	91.08
	Concentration	10	3.50	1.18	35.00
	Responsiveness	8	3.75	1.28	30.00
	Playfulness	5	4.00	1.00	20.00
	Independence	5	3.80	1.30	19.00
	Focus on reward	3	3.67	0.58	11.01
	Other	8			
Retrieve and reward behavior	Physical possession	14	4.07	0.73	56.98
	Playfulness	12	3.83	0.94	45.96
	Activity/excitability	11	3.82	1.40	42.02
	General drive	9	4.44	1.33	39.96
	General possession	7	3.71	1.38	25.97
	Hunt drive	5	4.40	0.89	22.00
	Chase retrieve	6	3.50	1.23	21.00
	Independent possession	3	3.33	1.15	9.99
	Mental possession	2	4.50	0.71	9.00
	Endurance	3	2.67	1.53	8.01
	Other	13			
Search behavior	Search drive	17	4.18	0.81	71.06
	Concentration	14	4.07	0.99	56.98
	Independence	14	3.57	1.45	49.98
	Speed	6	4.50	0.84	27.00
	Endurance	6	3.50	1.64	21.00
	Sniffing	6	3.17	1.17	19.02
	Activity/excitability	5	3.40	1.82	17.00
	Confidence	3	4.00	1.00	12.00
	Other	14			
Off-duty behavior	Calmness	31	4.68	0.54	145.08
	Playfulness	7	4.00	0.82	28.00
	Sociability	8	3.25	1.28	26.00
	Independence	3	5.00	0.00	15.00
	Protectiveness	4	3.25	0.96	13.00
	Other	8			
Physical abilities	Athleticism	11	4.27	1.01	46.97
	Size	9	4.00	1.00	36.00
	Sniffing	9	3.22	1.48	28.98
	Weight	7	3.86	1.07	27.02
	General health	6	3.83	0.98	22.98
	Endurance	6	3.67	0.82	22.02
	Strength	5	4.20	1.09	21.00
	Activity/excitability	4	4.25	1.50	17.00
	Speed	3	3.00	1.73	9.00
	Other	34			
Aberrant behavior	Independent play	6	4.00	0.89	24.00
	Carrying	4	4.50	0.58	18.00
	Destructiveness	3	5.00	0.00	15.00
	Barking	3	4.00	1.00	12.00
	Eating	3	3.33	2.08	9.99
	Protectiveness	2	3.50	2.12	7.00
	Other	10			

responses in the survey category Retrieve and Reward Behavior, while an analogue to "mental possession," "focus on reward" was replicated in Interaction with Handler. The trait "activity" was replicated in Play Behavior, Retrieve and Reward Behavior, Search Behavior, and Physical Abilities. Analogues to "hidden grass," "hunt one," and "hunt two" TSA-CBDC standardized measurements (i.e., "search drive" and "hunt drive") emerged in the survey categories Environmental Stability, Retrieve and Reward Behavior, and Search Behavior. The mean importance score given to traits currently measured in TSA-CBDC standardized search and retrieve tests was 31.08, SD = 23.59. All traits that are measured in the TSA-CBDC search and retrieve test were identified in the survey data.

Discussion

Many working dog programs procure, breed, and train dogs according to behavioral criteria that are thought to be important to operational work (e.g., [7, 8]). However, in many cases, operational characteristics required of successful working dogs are not completely known to breeders and other humans charged with the care of juvenile dogs. Here, we performed a criterion analysis to determine what behavioral traits operational detection-dog handlers thought were important to search team performance in operational conditions. Based on handler frequency of response and ranking of importance, the top seven nominated traits were playfulness, human-perceived relationship as deemed by the handler, human focus, calmness, concentration, activity/excitability, and search drive. Other notable traits also deemed important by real-life handlers included sociability, physical possession, independence/initiative, athleticism, and to a lesser extent, endurance and use of olfaction ("sniffing").

Most of the nominated traits identified here have previously been identified by working dog programs as being important to search team performance (e.g., concentration, calmness, search drive, activity/excitability, endurance, athleticism, sociability, physical possession, and use of olfaction, see: [23]). However, in our questionnaire, handlers identified characteristics related to play and calmness during off-duty periods as being two of the most important behavioral traits associated with the working capabilities of their dogs. In general, reasons why animals play are not entirely understood, but available evidence suggests that play may be an important influence on species-typical neurologic growth and development in the mammalian cerebellum during ontogeny [24, 25]. In other words, having the opportunity to play appears to be an important component of neurobiological organization in the mammalian brain. Related to this point, calmness, or the ability to manage energy effectively during off-duty hours (i.e., the ability to relax), is another behavioral trait of working dogs that has to date received scant attention in the scientific literature. Our results suggest that, in addition to traits normally thought to be important for the successful performance of detection dog work, off-duty calmness and a general willingness and ability to play may be just as important for working dog companions. To our knowledge, no detection dog working organizations currently incorporate explicit measures of play or off-duty calmness into their behavioral assessment criteria, although we acknowledge that some measurements involving possession of reward objects (a typical measure used in many working dog programs) may incorporate aspects of a dog's desire to play. Future work assessing how to reliably measure "play," "calmness," and aspects of these traits that have predictive validity with regard to working outcomes is needed. The idea here is that both working behaviors and off-duty behaviors, such as play and calmness, may covary with one another and with subsequent search team performance (i.e., the idea of personalities and behavioral carryovers, see: [26]).

Overall, our content validity results suggest that the TSA-CBDC standardized tests appear to be measuring behaviors thought important by operational TSA working dog handlers. Thirteen of 15 behavioral traits currently being used during TSA-CBDC standardized tests were deemed important by operational handlers. While overall this is a promising result, it is also worth noting that the TSA-CBDC standardized tests do not measure aspects of the human–dog relationship identified here (e.g., human focus and perceived human relationship) and elsewhere as a critical component of successful detection dog work (e.g., [27, 28]). Understanding handler effects on detection dog efficacy and choice remains one of the outstanding issue in studies on working dogs [29], and we expect human influences on dog behavior to be ubiquitous in working dog programs. From a practical standpoint, our results suggest that dropping measurements of hearing and body sensitivity, while adding measurements of play, off-duty calmness, and interactions with humans would refine the ability of the TSA-standardized tests given earlier in juvenile dog life to predict to later adult working dog behavior.

These results, which represent some of the first systematic criterion analysis of working dog handlers (see also [23]), are potentially informative. Nonetheless, they are subject to a number of limitations and should be considered exploratory and treated with caution. First, the sample size of handlers was modest ($N = 34$). Second, the analyses were potentially limited by the framework of our questionnaire; that is, instead of permitting an entirely pure free-response format, the survey contained *a priori* survey categories in which handlers inserted their answers. As a result, the characteristics generated by handlers may have been more limited than if they had been generated using a less structured survey format. Future research could build on the present research by making several key improvements to the design. In addition to assessing how the format of different questionnaires may influence handler responses, future surveys should also record additional demographic information about dogs (e.g., age and sex: see [5]). Such information could be particularly informative because dog sex can influence the development of sex-specific cognitive processes that may be important in detection work, such as detecting size constancy violations of common objects [30]. Of course, as with any scientific endeavor, replication of our results is needed, and further detailed questionnaire surveys combined with criterion analyses and content validity in this and other working dog programs would be welcome as well.

In conclusion, our analyses provide preliminary evidence that the TSA-CBDC standardized tests given to dogs during their first year of life demonstrate good content validity with regard to handler perceptions of dog traits that are important to operational work. Nevertheless, we did identify a number of traits that were considered important by handlers that are not included in the current TSA-CBDC test protocol. These traits were as follows: off-duty calmness, human-perceived relationship, playfulness, human focus, and to a lesser extent sociability and athleticism. Useful efficiencies could be made to this and other working dog development programs by incorporating assessments and opportunities for social play, off-duty relaxation, and human impacts on behavior during working dogs' first year of life.

Acknowledgments

Jamie Fratkin provided support by providing independent judgments regarding classifications of handler responses into trait categories. Two anonymous reviewers and the editor provided helpful comments on an earlier version of this manuscript.

References

1. Helton WS. Canine ergonomics: the science of working dogs. London, U.K.: CRC Press, 2009.
2. Furton KG, Myers LJ. The scientific foundation and efficacy of the use of canines as chemical detectors for explosives. Talanta 2001;54:487–500.
3. Graham LT, Gosling SD. Temperament and personality in working dogs. In: Helton WS, editor. Canine ergonomics: the science of working dogs. London, U.K.: CRC Press, 2009;63–81.
4. Wilsson E, Sundgren P-E. Behaviour test for eight-week old puppies – heritabilities of tested behaviour traits and its correspondence to later behaviour. Appl Anim Behav Sci 1998;58:151–62.
5. Wilsson E, Sundgren P-E. The use of a behaviour 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 1997;53:279–95.
6. Maejima M, Inoue-Murayama M, Tonosaki K, Matsuura N, Kato S, Saito Y, et al. Traits and genotypes may predict the successful training of drug detection dogs. Appl Anim Behav Sci 2007;107:287–98.
7. Wilsson E, Sinn DL. 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 2012;141:158–72.
8. Sinn DL, Gosling SD, Hilliard S. Personality and performance in military working dogs: reliability and predictive validity of behavioral tests. Appl Anim Behav Sci 2010;127:51–65.
9. Duffy DL, Serpell JA. Predictive validity of a method for evaluating temperament in young guide and service dogs. Appl Anim Behav Sci 2012;138:99–109.
10. Wiggins JS. Personality and prediction: principles of personality assessment. Reading, MA: Addison-Wesley Pub. Co., 1973.
11. Champness KA. Development of a breeding program for drug detector dogs [dissertation]. Melbourne, Vic., Australia: The University of Melbourne, 1996.
12. Champness KA. Revised breeding and rearing strategy for FAA detector dogs. San Antonio, TX: Federal Aviation Administration, 2000.
13. Fox MW. Neurobehavioral development and the genotype-environment interaction. Q Rev Biol 1970;45:131–47.
14. Scott JP, Fuller JL. Genetics and the social behavior of the dog. Chicago, IL: University of Chicago Press, 1965.
15. Scott JP, Bielfelt SW. Analysis of the puppy testing program. In: Pfaffenberger CJ, Scott JP, Fuller JL, Ginsburg BE, Bielfelt SW, Scott SF, editors. Guide dogs for the blind: their selection, development, and training. New York, NY: Elsevier, 1976;39–75.
16. Allport GW. The person in psychology: selected essays by Gordon W. Allport. Boston, MA: Beacon Press, 1968.
17. McAdams DP. What do we know when we know a person? J Pers 1995;63:365–96.
18. Pervin LA. A critical analysis of current trait theory. Psych Inquiry 1994;5:103–13.
19. Wiggins JS. In defense of traits. In: Hogan R, Johnson J, Briggs JR, editors. Handbook of personality psychology. New York, NY: Academic Press, 1997;97–115.
20. Kaufman AB, Rosenthal R. Can you believe my eyes? The importance of interobserver reliability statistics in observations of animal behaviour. Anim Behav 2009;78:1487–91.
21. Burghardt GM, Bartmess-LeVasseur JN, Browning SA, Morrison KE, Stec CL, Zachau CE, et al. Perspectives – minimizing observer bias in behavioral studies: a review and recommendations. Ethology 2012;118:511–7.
22. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess 1994;4:284–90.
23. Rooney NJ, Bradshaw JWS, Almey H. Attributes of specialist search dogs – a questionnaire survey of UK dog handlers and trainers. J Forensic Sci 2004;49:300–06.
24. Spinka M, Newberry RC, Bekoff M. Mammalian play: training for the unexpected. Q Rev Biol 2001;76:141–68.
25. Siviy SM, Panksepp J. In search of the neurobiological substrates for social playfulness in mammalian brains. Neurosci Biobehav Rev 2011;35:1821–30.
26. Bensky MK, Gosling SD, Sinn DL. The world from a dog's point of view: a review and synthesis of dog cognition research. In: Brockmann HJ, Roper TJ, Naguib M, Mitani JC, Simmons LW, Barrett L, editors. Adv Study Behav. London, U.K.: Elsevier, 2013;209–406.
27. Lit L, Schweitzer J, Oberbauer A. Handler beliefs affect scent detection dog outcomes. Anim Cogn 2011;14:387–94.
28. Gazit I, Goldblatt A, Terkel J. The role of context specificity in learning: the effects of training context on explosives detection in dogs. Anim Cogn 2005;8:143–50.
29. Rooney NJ, Cowan S. Training methods and owner-dog interactions: links with dog behaviour and learning ability. Appl Anim Behav Sci 2011;132:169–77.
30. Müller CA, Mayer C, Dorrenberg S, Huber L, Range F. Female but not male dogs respond to a size constancy violation. Biol Lett 2011;7:689–91.

Additional information and reprint requests:
David L. Sinn, Ph.D.
Department of Psychology
The University of Texas at Austin
108 E. Dean Keeton Stop A8000
Austin, TX 78712-1043
E-mail: david.sinn@utexas.edu