Applied Animal Behaviour Science 138 (2012) 99–109

Contents lists available at SciVerse ScienceDirect

Applied Animal Behaviour Science

journal homepage: www.elsevier.com/locate/applanim

Predictive validity of a method for evaluating temperament in young guide and service dogs

Deborah L. Duffy, James A. Serpell*

Department of Clinical Studies, School of Veterinary Medicine, University of Pennsylvania, 3900 Delancey Street, Philadelphia, PA 19104-6010, USA

ARTICLE INFO

Article history:
Accepted 12 February 2012
Available online 13 March 2012

Keywords:
Behavior
Dogs
Canine
C-BARQ
Temperament

ABSTRACT

As part of a collaborative project involving five guide and service dog organizations in the USA (Canine Companions for Independence, Guide Dogs for the Blind, Guiding Eyes for the Blind, Leader Dogs for the Blind and The Seeing Eye), volunteer puppy raisers provided information about the behavior of the guide and service dogs in their care via a questionnaire (the Canine Behavioral Assessment and Research Questionnaire, or C-BARQ©; www.cbarq.org). The surveys were completed online when the puppies were 6 months old and again at 12 months of age. Dogs were tracked through training and those that successfully completed training and were matched with a blind/disabled handler or were selected as breeders were classified as "successful" while dogs rejected from the program due to behavioral issues were classified as "released" (dogs rejected for medical reasons were excluded from analysis). A total of 11,997 C-BARQ evaluations for 7696 dogs were analyzed. Generalized linear modeling for each of the five schools revealed that dogs that successfully completed training scored more favorably on 27 out of 36 C-BARQ traits at both 6 and 12 months of age compared to those that were released from the programs. The most predictive trait at both age levels was 'pulls excessively hard on leash,' for which each unit increase in score was associated with a 1.4 increase in the odds of being released from the program. The ability of the C-BARQ to discriminate between dogs that were later successful or released differed across organizations ($P = 0.001$ and $P < 0.0001$ for 6- and 12-month surveys, respectively), most likely due to differences in the procedures used when making decisions about whether or not to release dogs. These findings provide convincing evidence that the C-BARQ is able to discriminate between dogs that are behaviorally suited for guide or service work and those that are not and may provide trainers with useful information about potential training or breeding candidates as early as 6 months of age.

1. Introduction

Behavioral issues are the primary reasons for the rejection of dogs from training among many guide and service dog organizations (Goddard and Beilharz, 1982; Serpell and Hsu, 2001). Based on information provided by the guide/service dog organizations that participated in the present study, training failure rates of 50–70% are typical, and behavioral reasons are cited as the primary reason for release in 63–87% of these cases. Assessment techniques that provide early detection of behavioral problems, and that identify dogs that are likely to be unsuitable for the work, are therefore being urgently sought after by working dog groups.

While many guide/service dog organizations have developed or adapted various forms of behavioral assessment methods for their own purposes, until now there has been no single standardized procedure for which both reliability and validity has been established (Jones and Gosling, 2005). Methods of behavioral assessment

* Corresponding author. Tel.: +1 215 898 1004; fax: +1 215 746 2090.
E-mail address: serpell@vet.upenn.edu (J.A. Serpell).

0168-1591/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.applanim.2012.02.011

commonly used at many organizations range from test batteries (Goddard and Beilharz, 1986; Ruefenacht et al., 2002; Tomkins et al., 2011; Wilsson and Sundgren, 1997) to observations by experienced trainers under naturalistic or semi-naturalistic conditions (Goddard and Beilharz, 1984; Maejima et al., 2007; Murphy, 1998). One drawback of behavioral tests is that generalizations are made about how the dog would respond in other environments based on what are essentially samples of very limited duration compared to the dog's entire behavioral repertoire; therefore, thorough validation of such tests is paramount but often not performed (Taylor and Mills, 2006). Evaluations based upon observations by experienced trainers likewise need to be thoroughly validated, particularly with respect to inter-observer reliability (agreement between trainers conducting independent assessments). Relying on the observations of a select few individuals necessitates that those individuals have extensive experience and training regarding the selection of dogs. Such personnel can be difficult to replace when needed and the evaluation of large numbers of dogs can become very taxing on their time.

An alternative approach to behavioral testing and assessment by trainers, is to rely on the experiences of people, such as puppy-raisers, who have lived with the dogs for extended periods of time. Surveys can ask such individuals to consider the dogs' behaviors in a variety of naturally occurring contexts. One such survey instrument is the Canine Behavioral Assessment and Research Questionnaire, or C-BARQ© (http://www.cbarq.org) (Hsu and Serpell, 2003). The C-BARQ is currently used by numerous working dog organizations worldwide but its ability to discriminate effectively between successful and unsuccessful dogs remains unproven.

Recently, Batt et al. (2009) used a portion of the C-BARQ in a pilot study exploring the potential of puppy raisers to predict the future training success of guide dog puppies. Using data obtained from 110 puppy raisers, stepwise logistic regression analysis indicated that the puppy raisers' own estimates of their dogs' likelihood of success in training and the presence of other dogs in the home, were the best predictors of success in training while none of the behavioral questions based on the C-BARQ met the criteria for inclusion in their final model (Batt et al., 2009). Unfortunately, a variety of methodological flaws raise doubts about the validity of these preliminary findings. In particular, less than a quarter of the questions from the C-BARQ were included in the study and, of these, six were modified from their original format (e.g., rewording of questions and changing the scales of possible responses). Additional questions were also added, addressing issues of dog distraction and socialization. Furthermore, puppy raisers' opinions about their dogs' probability of success in the program were unlikely to be independent of their responses to a relatively objective behavioral questionnaire, such as the C-BARQ. And, because a stepwise logistic regression was employed, once responses to the question regarding the puppy raisers' opinions were entered into the model, the related C-BARQ questions would become redundant and, thus, would be unlikely to meet the criteria for inclusion. While it is encouraging that puppy raisers' opinions about their dogs' potential for success in training appear to have some predictive value, and no doubt would be of great interest to guide and service dog organizations, more detailed and objective information about the dogs' behavioral phenotypes are needed for research in this area to move forward. For example, genotypic studies are best served by well-defined phenotypes and there is considerable interest in the genetic basis of behavior and working performance in dogs (Freimer and Sabatti, 2003; Spady and Ostrander, 2008).

The purpose of the present paper is to determine the degree to which behavioral assessments based on puppy-raiser responses to the C-BARQ (in its complete and currently validated form) can discriminate between dogs that later go on to successfully complete training in several different guide/service dog programs and those that are released for behavioral reasons.

2. Methods

2.1. Participants

This study was part of an on-going collaboration between the University of Pennsylvania School of Veterinary Medicine (Philadelphia, PA) and five guide and service dog schools in the USA: Canine Companions for Independence (headquartered in Santa Rosa, CA), Guide Dogs for the Blind (San Rafael, CA), Guiding Eyes for the Blind (Yorktown Heights, NY), Leader Dogs for the Blind (Rochester, MI) and The Seeing Eye (Morristown, NJ).

All dogs were provided by each organization via their own independent breeding programs. According to each organization's protocol, puppies were placed in the homes of volunteer puppy raisers at approximately 8 weeks of age (range 7–12 weeks). Puppy raisers were asked by their respective guide/service dog schools to complete an online questionnaire (C-BARQ) when the puppies were 6 months old and again at 12 months of age. The dogs were turned into each organization's training facility at approximately 15 months of age (range 12–18 months), at which time they were housed in kennels and entered formal training as guide/service dogs. All dogs were sexually intact at the time when the surveys were completed by puppy raisers. They were spayed/neutered after they returned to the organizations for training (unless selected for breeding).

Dogs were tracked through training and those that successfully completed training and were matched with a blind or disabled handler (usually at approximately 2 years of age) were classified as "successful" (see Table 1). Dogs selected as breeders were also included in the successful category because behavioral traits are strongly considered when selecting breeding stock. Dogs rejected from the program due to behavioral issues that were incompatible with the work were classified as "released" (including dogs that were released during either the puppy raising period or formal training). Dogs rejected for medical conditions were excluded from analysis. During the data collection period, decisions by the different organizations to release or successfully graduate dogs were "blinded" in the sense that they were made without reference to the dogs' C-BARQ assessments. The breeds included primarily Labrador and

Table 1
Number of dogs from each organization for which C-BARQ data and training outcomes were available.

	Successful		Released		
	Field Service	Breeding	Medical[a]	Behavioral	Total
Organization 1	1161	164	425	980	2730
Organization 2	903	145	415	824	2287
Organization 3	527	94	343	684	1648
Organization 4	449	76	127	869	1521
Organization 5	486	28	180	270	964

[a] Dogs released for medical reasons were omitted from analyses.

Golden retrievers and crosses between the two along with German Shepherds (Table 2).

2.2. C-BARQ

The C-BARQ was developed by Hsu and Serpell (2003) and has been shown to meet acceptable standards of reliability and validity (Duffy and Serpell, 2008; Hsu and Serpell, 2003). The internal reliability of all of the subscales meets or exceeds the generally accepted threshold (Cronbach's alpha $\geq$0.70) (Duffy and Serpell, 2008; Nunnally, 1978). The test–retest reliability of the subscales, as measured by comparing guide dog puppy raisers' evaluations of their dogs at 6 and 12 months of age, ranged from $r = 0.25$ to 0.56, with an average correlation coefficient of $r = 0.47$ (Duffy and Serpell, 2008). The inter-rater reliability was assessed using a population of pet owners ($N = 75$ pairs). The average percentage agreement (quadratic weighted formula) for the items composing each subscale ranged from 82% to 97% (Duffy and Serpell, 2008). The methods used to validate the main C-BARQ subscales has been described in detail elsewhere (Duffy and Serpell, 2008; Hsu and Serpell, 2003).

The questionnaire consists of 100 items that ask respondents to use a series of 5-point ordinal rating scales (from 0 to 4) to indicate their dogs' typical responses to a variety of everyday situations during the recent past (see Appendix). The scales rate either the severity (aggression, fear and excitability subscales with 0 indicating no sign of the behavior and 4 indicating a severe form of the behavior) or frequency (all remaining subscales and miscellaneous items with 0 indicating "never" and 4 indicating "always") of the behaviors. Participants were instructed to answer all questions. However, if they were unable to answer a question because they had never observed the dog in the specified situation they had the option to select "not observed/not applicable" and the item was treated as a missing value during statistical analysis.

Using factor analysis, 78 of the original items were condensed into 14 behavioral subscales (Table 3) that have been found to be remarkably consistent irrespective of breed, sex or geographic location (Duffy and Serpell, 2008; Hsu and Serpell, 2003; Hsu and Sun, 2010; Nagasawa et al., 2011; van den Berg et al., 2006, 2010). Twenty-two miscellaneous items were also included as stand-alone behavioral measures. High scores are less favorable for all items and subscales with the exception of 'trainability', for which high scores are more desirable. For the purposes of analysis, subscale scores are calculated as the average of the scores for the questionnaire items pertaining to that subscale. For all calculations of averages, cases that had missing values for more than 20% of the relevant items were excluded from the calculation and a missing value was recorded.

2.3. Statistical analysis

Data were analyzed using SPSS 17.0 for Windows (SPSS, Inc.). Separate analyses were used for 6-month and 12-month C-BARQ evaluations (see Table 4). The 14 subscale scores and 22 miscellaneous items were standardized by converting them into z-scores prior to analysis to account for the fact that the subscales are composites of multiple questions. Data were analyzed using Logistic Generalized Linear Models with a logit link function. In order to ascertain how each of the 36 C-BARQ scores could predict training outcome as a stand alone item, separate GLM analyses were performed for each score with the z-score included as a predictor variable along with breed and sex of the dog, each nested within 'organization,' while 'training outcome' (released vs. successful) served as the dependent variable. Therefore, 72 separate nested GLM analyses were performed (36 for the 6-month scores and 36 for the 12-month scores).

In order to offset the increased risk of Type I errors associated with multiple tests and to determine whether effects with P values less than 0.05 were false discoveries,

Table 2
Frequencies of different breeds from each organization that were included in analyses (with percentage of males in parentheses).

Breed	Organization					
	1	2	3	4	5	Total
Labrador retriever	683 (53)	1632 (50)	1169 (53)	295 (44)	557 (52)	4336 (51)
Golden retriever	448 (52)	50 (46)	62 (55)	67 (45)	170 (49)	797 (51)
German shepherd	694 (49)	50 (36)	52 (52)	0	45 (58)	841 (49)
Lab x golden cross	454 (45)	130 (48)	19 (63)	1032 (51)	0	1635 (49)
Other	26 (58)	10 (40)	3 (0)	0	11 (64)	50 (52)
Total	2305 (50)	1872 (49)	1305 (53)	1394 (49)	783 (51)	7659 (50)

Table 3
Factor and item structure of the C-BARQ.

1 (Subscale): Trainability (frequency scale)
- Dog returns immediately when called while off leash
- Dog obeys a "sit" command immediately
- Dog obeys a "stay" command immediately
- Dog seems to attend to or listen closely to everything the owner says or does
- Dog is slow to respond to correction or punishment
- Dog is slow to learn new tricks or tasks
- Dog is easily distracted by interesting sights, sounds, or smells
- Dog will fetch or attempt to fetch sticks, balls, and other objects

2 (Subscale): Stranger-directed aggression (severity scale)
- Dog acts aggressively
  - When approached directly by an unfamiliar adult while being walked or exercised on a leash.
  - When approached directly by an unfamiliar child while being walked or exercised on a leash
  - Toward unfamiliar persons approaching the dog while it is in the owner's car
  - When an unfamiliar person approaches the owner or a member of the owner's family at home
  - When an unfamiliar person approaches the owner or a member of the owner's family away from home
  - When mailmen or other delivery workers approach the home
  - When strangers walk past the home while the dog is in the yard
  - When an unfamiliar person tries to touch or pet the dog
  - When joggers, cyclists, roller skaters, or skateboarders pass the home while the dog is in the yard
  - Toward unfamiliar persons visiting the home

3 (Subscale): Owner-directed aggression (severity scale)
- Dog acts aggressively
  - When verbally corrected or punished by a member of the household
  - When toys, bones, or other objects are taken away by a member of the household
  - When bathed or groomed by a member of the household
  - When approached directly by a member of the household while it is eating
  - When food is taken away by a member of the household
  - When stared at directly by a member of the household
  - When stepped over by a member of the household
  - When a member of the household retrieves food or objects stolen by the dog

4 (Subscale): Dog rivalry (severity scale)
- Dog acts aggressively
  - Towards another (familiar) dog in your household.
  - When approached at a favorite resting/sleeping place by another household dog
  - When approached while eating by another household dog
  - When approached while playing with/chewing a favorite toy, bone, object by another household dog

5 (Subscale): Stranger-directed fear (severity scale)
- Dog acts anxious or fearful
  - When approached directly by an unfamiliar adult while away from the home
  - When approached directly by an unfamiliar child while away from the home
  - When unfamiliar persons visit the home
  - When an unfamiliar person tries to touch or pet the dog

6 (Subscale): Nonsocial fear (severity scale)
- Dog acts anxious or fearful
  - In response to sudden or loud noises
  - In heavy traffic
  - In response to strange or unfamiliar objects on or near the sidewalk
  - During thunderstorms, firework displays, or similar
  - When first exposed to unfamiliar situations
  - In response to wind or wind-blown objects

7 (Subscale): Dog-directed aggression (severity scale)
- Dog acts aggressively
  - When approached directly by an unfamiliar male dog while being walked or exercised on a leash
  - When approached directly by an unfamiliar female dog while being walked or exercised on a leash
  - Toward unfamiliar dogs visiting the home
  - When barked, growled or lunged at by an unfamiliar dog

8 (Subscale): Dog-directed fear (severity scale)
- Dog acts anxious or fearful
  - When approached directly by an unfamiliar dog of the same or larger size
  - When approached directly by an unfamiliar dog of a smaller size
  - When unfamiliar dogs visit the home.
  - When barked, growled or lunged at by an unfamiliar dog

9 (Subscale): Touch sensitivity (severity scale)
- Dog acts anxious or fearful
  - When examined or treated by a veterinarian
  - When having its nails clipped by a household member
  - When groomed or bathed by a household member
  - When having feet toweled by a household member

Table 3 (*Continued*)

10 (Subscale): Separation-related behavior (frequency scale)
Shaking, shivering or trembling when left or about to be left on its own
Excessive salivation when left or about to be left on its own
Restlessness/agitation/pacing when left or about to be left on its own
Whining when left or about to be left on its own
Barking when left or about to be left on its own
Howling when left or about to be left on its own
Chewing or scratching at doors, floor, windows, and curtains when left or about to be left on its own.
Loss of appetite when left or about to be left on its own
11 (Subscale): Excitability (severity scale)
When a member of the household returns home after a brief absence
When playing with a member of the household
When the doorbell rings
Just before being taken for a walk
Just before being taken on a car trip
When visitors arrive at its home
12 (Subscale): Attachment/attention-seeking (frequency scale)
Dog displays a strong attachment for a particular member of the household
Dog tends to follow a member of household from room to room about the house
Dog tends to sit close to or in contact with a member of the household when that individual is sitting down
Dog tends to nudge, nuzzle, or paw a member of the household for attention when that individual is sitting down
Dog becomes agitated when a member of the household shows affection for another person
Dog becomes agitated when a member of the household shows affection for another dog or animal
13 (Subscale): Chasing (frequency scale)
Dog acts aggressively toward cats, squirrels, and other animals entering its yard
Dog chases or would chase cats given the opportunity
Dog chases or would chase birds given the opportunity
Dog chases or would chase squirrels, rabbits and other small animals given the opportunity
14 (Subscale): Energy level (frequency scale)
Dog is playful, puppyish, and boisterous
Dog is active, energetic, and always on the go
Miscellaneous (frequency scales)
15 Escapes or would escape home or yard given a chance
16 Rolls in animal droppings or other 'smelly' substances
17 Eats own or other animals' droppings or feces
18 Chews inappropriate objects
19 Mounts objects, furniture, or people
20 Begs persistently for food when people are eating
21 Steals food
22 Nervous or frightened on stairs
23 Pulls excessively hard when on the leash
24 Urinates against objects/furnishings in your home
25 Urinates when approaches, petted, handled or picked up
26 Urinates when left alone at night, or during the daytime
27 Defecates when left alone at night, or during the daytime
28 Hyperactive, restless, has trouble settling down
29 Stares intently at nothing visible
30 Snaps at (invisible) flies
31 Chases own tail/hind end
32 Chases/follows shadows, light spots, etc.
33 Barks persistently when alarmed or excited
34 Licks him/herself excessively
35 Licks people or objects excessively
36 Displays other bizarre, stranger or repetitive behavior(s)

Table 4
Number of surveys included in analyses for each school, separated by age at evaluation and training outcome (not all dogs had surveys completed at both time points).

Organization	6-month scores		12-month scores	
	Successful	Released	Successful	Released
1	849	622	1302	909
2	788	684	912	596
3	506	577	474	529
4	314	481	471	794
5	385	211	398	195

we estimated the *Q* value for each test using the positive False Discovery Rate (pFDR) method (Benjamini and Hochberg, 1995; Storey, 2002). This approach estimates the proportion of false positives (e.g., Type I errors) amongst the tests that indicate significant results (i.e., number false positives/number significant tests). The *Q* values are calculated based on the *P* values generated by the multiple tests performed and estimate the proportion of false positives incurred when a given test is called significant. Thus, *Q* values are the pFDR analog to the *P* value (Storey, 2002). The *Q* values were estimated using freely available Q-VALUE software (http://genomics.princeton.edu/storeylab/qvalue/). *Q* values were calculated based on the following four groups of tests: tests of the main effects of C-BARQ *z*-scores at each age of evaluation (36 *P* values each for 6 and 12 months), effects of C-BARQ *z*-scores nested within organization at each age of evaluation for which a significant main effect was found (135 *P* values each for 6 and 12 months). We chose a conservative *Q* value cutoff of 0.01; therefore, amongst the tests for which we determined a significant effect was present, fewer than 1% of them are estimated to be false positives.

Finally, to determine whether the C-BARQ as a whole could predict training outcomes, we performed hierarchical logistic regression analyses that fitted all standardized C-BARQ subscales/miscellaneous items in a single model. C-BARQ subscales/miscellaneous items that had greater than 10% missing values were excluded in order to maximize the sample size. Breed and sex were included as potential confounders in the first block and all eligible C-BARQ *z*-scores were entered as the second block. The full model (containing all C-BARQ *z*-scores, breed, and sex) was compared to the empty model (containing the intercept only) to determine the degree to which training outcomes could be predicted. The full model was also compared to the reduced model (containing only breed and sex) to determine whether the addition of C-BARQ scores significantly improved the model compared to a model containing only breed and sex. Separate logistic regression models were created for each organization and age at evaluation, thus 10 separate regression models were fitted. The fit of the models were assessed using the Hosmer–Lemeshow goodness-of-fit test (Hosmer and Lemeshow, 2000) and by examining the receiver operating characteristic curve (ROC; i.e., a plot of the true positive rate (sensitivity) versus the false positive rate (1 – specificity)). More precisely, the area under the ROC (AUROC) was calculated as an estimate of each model's ability to accurately classify a dog as "released" or "successful," wherein an AUROC of 1.0 would indicate that the model predicts training outcome perfectly and an AUROC of 0.5 is equal to a random guess (Bewick et al., 2004).

3. Results

3.1. Missing values

For all questions, a response of 'not observed/not applicable' (indicating that the puppy raiser had not had an opportunity to observe the dog in the context described) was coded as a missing value and all other responses were regarded as valid answers. For the 6- and 12-month surveys, 90.2% and 96.0% of respondents provided valid answers to 90 or more questions, respectively. 'Dog rivalry,' 'escaping,' and 'chasing' topped the list of the items/subscales that had the highest percentage of missing values (23% each for 'dog rivalry' and 'escapes' and 20% for 'chasing' amongst 6-month surveys, and 19%, 14% and 12% for the 12-month surveys, respectively). Other items/subscales that had missing values of 10% or greater included 'shadow/light spot chasing' (12%), 'rolling in strong smelling odors, etc.' (11%) and 'dog-directed fear' (10%), all from the 6-month survey (the remaining items/subscales from the 12-month survey had fewer than 10% missing values).

3.2. Comparison of successful and released dogs

Significant differences were found between successful and released dogs for 27 of the 36 C-BARQ items at each age of evaluation (see Tables 5 and 6), with successful dogs scoring more favorably than released dogs in every case. With the exception of six items ('emotional urination,' 'urination when left alone,' 'compulsive staring,' 'snapping at (invisible) flies,' 'allogrooming' and 'dog-directed fear'), the same C-BARQ items were significant at both 6- and 12-month time points. For six C-BARQ items, there were no significant differences between successful and released dogs at either age ('rolling in strong smelling odors, etc.,' 'coprophagia,' 'urine marking in the home,' 'defecation when left alone,' 'shadow/light chasing' and 'self-grooming').

There were significant differences across organizations in the overall number of C-BARQ items that were predictive of training outcomes ($\chi^2 = 17.9$, df = 4, $P = 0.001$ and $\chi^2 = 47.6$, df = 4, $P < 0.0001$ for 6 and 12 month surveys, respectively).

Logistic regression models indicated that the C-BARQ was able to discriminate between successful and released dogs at both 6 and 12 months of age (Table 7). The models performed better than chance (e.g., null hypothesis: area = 0.50) at discriminating training outcomes as indicated by the area under the ROC curves. However, models containing breed and sex were not significantly improved by the addition of C-BARQ scores for Organization 3. For the other four organizations, the addition of the C-BARQ scores significantly improved models containing only breed and sex.

4. Discussion

The present study collected behavioral data on potential guide dog puppies from five different USA guide/service dog organizations using a standardized and validated behavioral survey (C-BARQ©) that was completed by puppy raisers at two time points 6 months apart. The dogs were then tracked through training and the C-BARQ scores of dogs that subsequently were successful or released from these programs were compared. C-BARQ scores of successful dogs were significantly more favorable than those of released dogs at both 6 and 12 months of age.

Table 5
Results of 36 individual logistic GLMs based on 6-month survey data. For each GLM, a single C-BARQ item/subscale, sex and breed were nested within organization and served as the independent variables while training outcome (successful or released) was the dependent variable.

Organization											
C-BARQ item (standard deviation)	Main effect	1		2		3		4		5	
	P	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Trainability (0.411)	**<0.0001**[e,f]	**0.84**[c]	**(0.76, 0.93)**	**0.76**[d]	**(0.68, 0.85)**	0.89	(0.79, 1.01)	**0.73**[d]	**(0.63, 0.85)**	**0.72**[c]	**(0.60, 0.87)**
Stranger-directed aggression (0.195)	**<0.0001**[e]	**1.11**[b]	**(1.04, 1.19)**	**1.22**[a]	**(1.02, 1.50)**	0.99	(0.86, 1.13)	1.28	(0.97, 1.79)	**1.36**[b]	**(1.13, 1.65)**
Owner-directed aggression (0.175)	**<0.0001**[e,f]	**1.18**[d]	**(1.09, 1.27)**	**1.71**[d]	**(1.42, 2.10)**	1.12	(0.97, 1.31)	1.26	(0.93, 1.77)	**1.66**[d]	**(1.32, 2.20)**
Dog rivalry (0.307)	**<0.0001**[e,f]	1.04	(0.96, 1.14)	**1.25**[a]	**(1.05, 1.52)**	1.00	(0.87, 1.15)	**1.46**[b]	**(1.12, 1.98)**	**1.53**[c]	**(1.19, 1.97)**
Stranger-directed fear (0.211)	**0.048**[e,f]	1.08	(0.98, 1.19)	1.15	(0.99, 1.36)	1.04	(0.92, 1.18)	1.11	(0.97, 1.30)	1.12	(0.98, 1.29)
Nonsocial fear (0.453)	**0.004**[e,f]	**1.13**[a]	**(1.02, 1.26)**	**1.15**[a]	**(1.03, 1.29)**	1.06	(0.94, 1.20)	1.14	(0.97, 1.33)	1.17	(0.96, 1.42)
Dog-directed aggression (0.386)	**0.002**[e,f]	**1.11**[a]	**(1.02, 1.21)**	**1.22**[a]	**(1.04, 1.45)**	1.08	(0.98, 1.21)	1.15	(0.90, 1.50)	1.21	(0.97, 1.50)
Dog-directed fear (0.463)	0.136[e,f]										
Touch sensitivity (0.473)	**0.001**[e,f]	**1.14**[b]	**(1.04, 1.26)**	**1.17**[a]	**(1.01, 1.36)**	1.03	(0.92, 1.14)	**1.23**[a]	**(1.05, 1.46)**	1.15	(0.95, 1.40)
Separation-related problems (0.346)	**<0.0001**[e,f]	**1.22**[d]	**(1.11, 1.34)**	**1.18**[b]	**(1.05, 1.32)**	**1.20**[b]	**(1.06, 1.38)**	**1.28**[b]	**(1.07, 1.55)**	1.09	(0.89, 1.32)
Excitability (0.704)	**<0.0001**[e,f]	**1.13**[a]	**(1.02, 1.26)**	1.10	(0.98, 1.24)	1.11	(0.97, 1.26)	**1.33**[b]	**(1.12, 1.59)**	**1.32**[b]	**(1.10, 1.58)**
Attachment/attention-seeking (0.665)	**0.002**[e]	**1.23**[d]	**(1.11, 1.37)**	1.01	(0.90, 1.13)	1.11	(0.98, 1.24)	1.06	(0.91, 1.23)	1.04	(0.88, 1.23)
Chasing (0.850)	**<0.0001**[e,f]	**1.19**[b]	**(1.06, 1.33)**	**1.27**[c]	**(1.11, 1.45)**	1.03	(0.91, 1.17)	**1.20**[a]	**(1.01, 1.43)**	1.07	(0.88, 1.29)
Energy (0.900)	**<0.0001**[e,f]	**1.22**[c]	**(1.08, 1.36)**	**1.15**[a]	**(1.02, 1.29)**	**1.16**[a]	**(1.03, 1.31)**	**1.34**[c]	**(1.14, 1.57)**	**1.25**[a]	**(1.04, 1.51)**
Escapes (0.941)	**0.003**[e,f]	0.90	(0.89, 1.10)	**1.23**[b]	**(1.08, 1.40)**	**1.17**[a]	**(1.01, 1.36)**	1.12	(0.94, 1.33)	1.19	(0.96, 1.46)
Rolls in feces (0.479)	0.117[e,f]										
Coprophagia (1.001)	0.874[e,f]										
Chews (1.007)	**0.0002**[e,f]	1.09	(0.98, 1.22)	**1.14**[a]	**(1.02, 1.27)**	**1.20**[b]	**(1.06, 1.36)**	1.16	(1.00, 1.34)	1.19	(0.99, 1.42)
Mounts (0.796)	**0.0003**[e,f]	**1.15**[b]	**(1.05, 1.27)**	**1.17**[b]	**(1.04, 1.31)**	1.07	(0.96, 1.21)	**1.27**[a]	**(1.04, 1.57)**	1.12	(0.93, 1.36)
Begs (0.751)	**0.015**[e,f]	1.07	(0.97, 1.17)	1.14	(0.99, 1.31)	0.89	(0.79, 1.01)	**1.23**[a]	**(1.03, 1.48)**	1.08	(0.90, 1.28)
Steals food (0.820)	**<0.0001**[e,f]	**1.22**[d]	**(1.11, 1.33)**	**1.23**[b]	**(1.08, 1.41)**	1.05	(0.92, 1.19)	**1.49**[c]	**(1.21, 1.87)**	1.14	(0.96, 1.35)
Nervous on stairs (0.714)	**0.001**[e,f]	**1.17**[b]	**(1.05, 1.29)**	1.07	(0.96, 1.20)	1.01	(0.90, 1.14)	**1.26**[b]	**(1.08, 1.50)**	1.21	(0.99, 1.48)
Pulls on leash (0.992)	**<0.0001**[e,f]	**1.21**[c]	**(1.09, 1.34)**	**1.30**[d]	**(1.16, 1.47)**	**1.24**[c]	**(1.09, 1.41)**	**1.58**[d]	**(1.32, 1.91)**	**1.63**[d]	**(1.35, 1.98)**
Urine marking (0.206)	0.38[e,f]										
Emotional urination (0.433)	0.184[e,f]										
Urination when left alone (0.431)	**0.043**[e,f]	0.93	(0.85, 1.01)	1.11	(0.99, 1.25)	0.93	(0.80, 1.08)	1.17	(0.98, 1.42)	1.09	(0.93, 1.28)
Defecation when left alone (0.353)	0.11[e,f]										
Hyperactive (0.800)	**<0.0001**[e,f]	**1.18**[c]	**(1.07, 1.30)**	**1.25**[d]	**(1.12, 1.40)**	1.07	(0.96, 1.21)	**1.26**[b]	**(1.06, 1.52)**	**1.29**[b]	**(1.09, 1.53)**
Compulsive staring (0.534)	0.429[e,f]										
Snaps at flies (0.496)	**0.01**[e,f]	1.03	(0.95, 1.13)	1.04	(0.92, 1.18)	**1.24**[c]	**(1.09, 1.43)**	1.10	(0.91, 1.35)	1.15	(0.94, 1.41)
Tail chasing (0.942)	**0.006**[e,f]	**1.20**[c]	**(1.08, 1.32)**	1.07	(0.96, 1.19)	1.03	(0.91, 1.16)	1.10	(0.95, 1.28)	1.08	(0.92, 1.28)
Shadow chasing (0.707)	0.124[e,f]										
Barks persistently (0.723)	**<0.0001**[e,f]	**1.27**[d]	**(1.16, 1.40)**	**1.29**[d]	**(1.13, 1.47)**	1.04	(0.93, 1.17)	**1.29**[a]	**(1.06, 1.59)**	**1.32**[b]	**(1.08, 1.60)**
Self grooming (0.636)	0.122[e,f]										
Allo grooming (0.935)	**0.044**[e,f]	1.03	(0.93, 1.14)	1.04	(0.94, 1.16)	1.11	(0.98, 1.25)	**1.22**[b]	**(1.05, 1.41)**	1.08	(0.91, 1.28)
Other stereotyped behavior (0.594)	**0.026**[e,f]	**1.15**[b]	**(1.04, 1.27)**	1.05	(0.92, 1.19)	0.96	(0.83, 1.12)	1.17	(1.00, 1.40)	0.95	(0.79, 1.13)

Items in boldface are statistically significant based on a False Discovery Rate (*Q* value) of 0.01.

[a] $P < 0.05$.

[b] $P < 0.01$.

[c] $P < 0.001$.

[d] $P < 0.0001$.

[e] Breed was a significant main effect on training outcomes at $P < 0.05$ and $Q < 0.01$.

[f] Sex was a significant main effect on training outcomes at $P < 0.05$ and $Q < 0.01$.

Table 6
Results of 36 individual logistic GLMs based on 12-month survey data. For each GLM, a single C-BARQ item/subscale, sex and breed were nested within organization and served as the independent variables while training outcome (successful or released) was the dependent variable.

Organization											
C-BARQ item (standard deviation)	Main effect	1		2		3		4		5	
	P	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Trainability (0.416)	**<0.00001**[e]	**0.79**[d]	**(0.73, 0.86)**	**0.79**[d]	**(0.71, 0.89)**	0.88	(0.77, 1.00)	**0.74**[d]	**(0.65, 0.84)**	**0.72**[c]	**(0.60, 0.87)**
Stranger-directed aggression (0.227)	**<0.00001**[e]	**1.19**[d]	**(1.11, 1.27)**	**1.37**[b]	**(1.13, 1.68)**	1.10	(0.94, 1.29)	**1.40**[c]	**(1.17, 1.72)**	**1.29**[c]	**(1.11, 1.52)**
Owner-directed aggression (0.152)	**<0.00001**[e]	**1.22**[d]	**(1.14, 1.31)**	**1.45**[c]	**(1.18, 1.83)**	1.13	(0.97, 1.33)	**1.42**[a]	**(1.11, 1.93)**	1.20	(1.01, 1.48)
Dog rivalry (0.308)	**0.0002**[e]	**1.12**[b]	**(1.04, 1.20)**	**1.37**[b]	**(1.11, 1.70)**	1.04	(0.88, 1.23)	**1.31**[a]	**(1.06, 1.66)**	1.01	(0.82, 1.23)
Stranger-directed fear (0.246)	**<0.00001**[e]	**1.23**[d]	**(1.13, 1.36)**	**1.29**[c]	**(1.12, 1.50)**	1.02	(0.89, 1.17)	**1.34**[c]	**(1.16, 1.59)**	1.15	(1.00, 1.34)
Nonsocial fear (0.426)	**<0.00001**[e]	**1.13**[b]	**(1.04, 1.22)**	**1.33**[d]	**(1.18, 1.49)**	1.12	(0.98, 1.27)	**1.28**[c]	**(1.12, 1.47)**	**1.37**[b]	**(1.13, 1.68)**
Dog-directed aggression (0.455)	**<0.00001**[e]	**1.14**[c]	**(1.06, 1.22)**	**1.42**[d]	**(1.20, 1.68)**	1.06	(0.93, 1.21)	**1.44**[b]	**(1.12, 1.90)**	**1.29**[b]	**(1.08, 1.56)**
Dog-directed fear (0.456)	**<0.00001**[e]	**1.16**[c]	**(1.07, 1.25)**	**1.21**[c]	**(1.08, 1.36)**	1.01	(0.88, 1.16)	1.12	(0.99, 1.28)	**1.22**[a]	**(1.04, 1.45)**
Touch sensitivity (0.510)	**<0.00001**[e]	**1.15**[c]	**(1.07, 1.24)**	**1.27**[c]	**(1.10, 1.47)**	1.06	(0.95, 1.19)	**1.24**[b]	**(1.08, 1.44)**	1.15	(0.95, 1.38)
Separation-related problems (0.347)	**<0.00001**[e]	**1.21**[d]	**(1.12, 1.31)**	1.13[a]	(1.00, 1.27)	1.06	(0.92, 1.22)	**1.39**[d]	**(1.19, 1.64)**	1.08	(0.90, 1.30)
Excitability (0.756)	**<0.00001**[e]	**1.18**[c]	**(1.08, 1.29)**	**1.16**[a]	**(1.03, 1.31)**	1.13	(0.99, 1.29)	**1.42**[d]	**(1.23, 1.64)**	1.16	(0.97, 1.39)
Attachment/attention-seeking (0.694)	**0.0003**[e]	**1.21**[d]	**(1.10, 1.32)**	1.04	(0.93, 1.17)	0.99	(0.88, 1.12)	1.14a	(1.00, 1.30)	1.10	(0.93, 1.31)
Chasing (0.865)	**0.0002**[e]	**1.13**[b]	**(1.03, 1.23)**	**1.20**[b]	**(1.06, 1.37)**	1.12	(0.98, 1.26)	1.15[a]	(1.01, 1.32)	1.15	(0.96, 1.39)
Energy (0.954)	**<0.00001**[e]	**1.18**[c]	**(1.07, 1.29)**	**1.21**[c]	**(1.08, 1.36)**	**1.30**[c]	**(1.13, 1.49)**	**1.38**[d]	**(1.21, 1.57)**	**1.37**[b]	**(1.13, 1.66)**
Escapes (0.987)	**<0.00001**[e]	**1.10**[a]	**(1.01, 1.19)**	**1.16**[a]	**(1.02, 1.33)**	1.17[a]	(1.00, 1.37)	**1.21**[b]	**(1.07, 1.39)**	1.21	(0.99, 1.48)
Rolls in feces (0.519)	0.343[e]										
Coprophagia (1.041)	0.462[e]										
Chews (1.008)	**0.001**[e]	**1.12**[b]	**(1.03, 1.22)**	1.00	(0.90, 1.12)	1.07	(0.93, 1.22)	**1.24**[c]	**(1.10, 1.41)**	1.02	(0.85, 1.23)
Mounts (0.741)	**0.00002**[e]	**1.11**[b]	**(1.03, 1.20)**	**1.23**[b]	**(1.08, 1.40)**	1.12	(1.00, 1.27)	1.17	(1.00, 1.39)	**1.25**[a]	**(1.02, 1.52)**
Begs (0.762)	**0.038**[e]	**1.13**[c]	**(1.05, 1.22)**	1.04	(0.90, 1.21)	0.96	(0.84, 1.09)	0.96	(0.84, 1.10)	1.05	(0.87, 1.27)
Steals food (0.821)	**<0.00001**[e]	**1.19**[d]	**(1.11, 1.29)**	**1.18**[a]	**(1.03, 1.35)**	1.11	(0.97, 1.28)	**1.28**[b]	**(1.10, 1.50)**	**1.33**[b]	**(1.11, 1.59)**
Nervous on stairs (0.614)	**0.00002**[e]	**1.11**[b]	**(1.03, 1.21)**	1.08	(0.94, 1.24)	1.12[a]	(1.00, 1.26)	**1.29**[d]	**(1.14, 1.47)**	1.12	(0.90, 1.38)
Pulls on leash (0.997)	**<0.00001**[e]	**1.23**[d]	**(1.13, 1.34)**	**1.19**[b]	**(1.05, 1.34)**	**1.40**[d]	**(1.21, 1.61)**	**1.55**[d]	**(1.32, 1.83)**	**1.50**[d]	**(1.24, 1.81)**
Urine marking (0.232)	0.492[e]										
Emotional urination (0.309)	**0.001**[e]	**1.09**[b]	**(1.03, 1.15)**	1.15	(0.88, 1.53)	1.17	(0.96, 1.50)	1.08	(0.91, 1.31)	**1.28**[b]	**(1.10, 1.52)**
Urination when left alone (0.264)	0.884[e]										
Defecation when left alone (0.267)	0.263[e]										
Hyperactive (0.843)	**<0.00001**[e]	**1.21**[d]	**(1.12, 1.31)**	**1.20**[c]	**(1.08, 1.35)**	**1.17**[a]	**(1.02, 1.33)**	**1.36**[d]	**(1.17, 1.60)**	**1.24**[a]	**(1.05, 1.47)**
Compulsive staring (0.549)	**0.009**[e]	1.08[a]	(1.00, 1.17)	**1.14**[a]	**(1.02, 1.28)**	1.08	(0.95, 1.23)	1.16[a]	(1.01, 1.35)	1.05	(0.85, 1.27)
Snaps at flies (0.505)	0.117[e]										
Tail chasing (0.906)	**<0.00001**[e]	**1.22**[d]	**(1.12, 1.33)**	1.01	(0.91, 1.13)	1.03	(0.91, 1.16)	**1.20**[b]	**(1.05, 1.36)**	1.15	(0.97, 1.37)
Shadow chasing (0.663)	0.394[e]										
Barks persistently (0.759)	**<0.00001**[e]	**1.27**[d]	**(1.18, 1.38)**	**1.36**[d]	**(1.20, 1.55)**	1.11	(0.97, 1.28)	**1.47**[d]	**(1.27, 1.72)**	**1.26**[a]	**(1.04, 1.52)**
Self grooming (0.690)	0.021[e]	1.11[a]	(1.02, 1.20)	1.09	(0.97, 1.22)	1.11	(0.98, 1.26)	1.09	(0.95, 1.26)	1.10	(0.91, 1.32)
Allo grooming (0.896)	0.208[e]										
Other stereotyped behavior (0.650)	**0.00009**[e]	**1.14**[b]	**(1.05, 1.23)**	**1.16**[b]	**(1.05, 1.29)**	1.16	(0.99, 1.38)	1.16[a]	(1.01, 1.35)	1.03	(0.86, 1.22)

Items in boldface are statistically significant based on a False Discovery Rate (*Q* value) of 0.01.

[a] $P < 0.05$.

[b] $P < 0.01$.

[c] $P < 0.001$.

[d] $P < 0.0001$.

[e] Breed was a significant main effect on training outcomes at $P < 0.05$ and $Q < 0.01$.

Table 7
Results of the 10 separate logistic regression models, with the effects of adding the C-BARQ *z*-scores to a model containing breed and sex (full vs reduced models) and the effects of the full model (containing C-BARQ *z*-scores, breed and sex) contrasted with an intercept-only model.

Organization	*N*	Full vs reduced			Full vs intercept-only			Fit of full model	
		Chi-square	df	*P* value	Chi-square	df	*P* value	Hosmer–Lemeshow (*P* value)	AUROC[a]
6 months evaluation									
Org 1[b]	783	62.392	31	0.0007	111.342	35	<0.00001	0.575	0.72
Org 2	907	53.96	31	0.006	58.044	35	0.008	0.493	0.64
Org 3[b]	777	39.702	31	0.136	58.002	35	0.009	0.177	0.64
Org 4[b,c]	540	47.147	31	0.032	61.972	34	0.002	0.290	0.70
Org 5	503	49.578	31	0.018	55.31	34	0.012	0.288	0.69
12 months evaluation									
Org 1[b]	1445	85.936	33	<0.00001	147.671	37	<0.00001	0.307	0.69
Org 2	1048	81.608	33	<0.00001	85.879	37	<0.00001	0.914	0.66
Org 3[c]	758	34.97	33	0.375	57.199	37	0.018	0.255	0.64
Org 4	910	98.309	33	<0.00001	104.727	36	<0.00001	0.412	0.70
Org 5	527	63.274	33	0.001	65.304	36	0.002	0.300	0.71

[a] AUROC = Area Under the Receiver Operating Characteristic curve; all statistically greater than the null hypothesis of AUROC = 0.50 at the 95% confidence level.
[b] Breed was a significant predictor of training outcome at the 95% confidence level.
[c] Sex was a significant predictor of training outcome at the 95% confidence level.

Twenty-seven C-BARQ items/subscales were able to discriminate between successful and released dogs at the 95% confidence level, and 21 of those were consistent between the 6- and 12-month evaluations.

High scores for energy level, hyperactivity, and pulling on the leash were significantly related to reduced training success across all five organizations. For the 6-month evaluation, owner-directed aggression (i.e., aggression directed toward members of the household) was one of the most sensitive indicators of the likelihood of training failure. For example, for every 0.175 increase in owner-directed aggression, the odds of being released increased by 1.7 for both organizations 2 and 5. This means that for these two organizations, a dog that scored 0.5 for owner-directed aggression at the age of 6 months had approximately 4:1 odds against successfully completing training compared to a dog that scored a zero.

Owner-directed aggression was predictive at the 12-month evaluation but less so compared to the 6-month survey. For every 0.152 increase in owner-directed aggression score, the odds of being released from training increased by a factor ranging from 1.22 to 1.45 for organizations 1, 2 and 4. In other words, a dog scoring a 0.5 for owner-directed aggression at 12 months of age had approximately two to three times the odds against successfully completing training compared to a dog with a score of zero. Pulling on the leash had the highest odds ratios, on average, for the 12-month survey (ranging from 1.19 to 1.55). A dog that scored the maximum (4) for pulling on the leash had two- to six-fold higher odds against successfully completing training compared to a dog that scored a zero.

The ability of the C-BARQ scores to discriminate between successful and released dogs varied across organizations. This may be due, in part, to differences in sample sizes of the different organizations and, by extension, variation in statistical power. However, sample sizes alone cannot account for the paucity of statistically significant relationships for Organization 3. While there are some differences across organizations in regards to the proportion of different breeds used, breed was controlled for in the analysis and is not likely an explanation for the observed differences across organizations in the number of significant effects. There are some substantial procedural differences between Organization 3 and the remaining four schools in terms of how and when decisions are made to release dogs from the program. For example, Organization 3 is the only organization in our study that relies substantially on puppy testing (when puppies are 7–9 weeks of age) with a release rate of approximately 20%. If it is assumed that the puppy test results are valid predictors of future training success, this organization would be identifying the dogs with the behavioral phenotypes that are least suitable for successful training and removing them from their program prior to placement with puppy raisers and C-BARQ evaluations. This would, in theory, remove one tail of the distribution of behavioral phenotypes in this population and make it more difficult for an instrument such as the C-BARQ to discriminate from amongst the remaining dogs. However, the predictive validity of early puppy testing is not well-established (Beaudet et al., 1994; Goddard and Beilharz, 1986; Svobodová et al., 2008; Wilsson and Sundgren, 1998), therefore, we can only speculate regarding how this practice affects the distribution of behavioral phenotypes. Further research is needed to determine the validity of puppy testing in working dog programs.

Organization 3 also relies much more heavily than the other four organizations on behavioral testing when dogs leave their puppy raisers' homes and arrive at the training centers, releasing approximately 16% of dogs based on their test performance. In our study, dogs released based upon these "in-for-training" (IFT) tests were included in analyses as part of the 'released' group. If the IFT test is valid, we would expect there to be agreement between C-BARQ scores and IFT test scores; thus, improving the ability of the C-BARQ to predict the odds of being rejected from the program. However, if the test does not accurately reflect the dogs' behavioral phenotypes, and dogs are released based upon factors idiosyncratic to the testing paradigm, it would interfere with the ability of the C-BARQ to identify

dogs that are at higher risk of being released from the program. Therefore, validation of IFT testing is necessary to determine its influence on the ability of other behavioral measures to predict success in training.

Our findings indicate that the C-BARQ is sufficiently sensitive to discriminate between dogs that are well suited for guide and assistance work and those that are not. However, while performing significantly better than chance, the areas under the ROC curves (0.64–0.72) suggest that the accuracy with which the C-BARQ as a whole can correctly identify a successful versus an unsuccessful dog may still be somewhat low to be used as the sole criterion for releasing a dog.

Accurately predicting training outcomes based upon behavioral phenotypes is difficult for multiple reasons. For one, some phenotypic measures, such as the C-BARQ, are completed many months before the fate of a dog is finally determined. Also, dogs that are released from training are not really a cohesive phenotypic group that all share some intrinsic attribute. Rather, the released group is a heterogeneous collection of dogs rejected for a variety of unrelated behavioral reasons. For example, dogs that were released for being too easily distracted would not necessarily be expected to score more poorly for aggression than those that successfully completed training. Unfortunately, most organizations do not record detailed behavioral information about the reasons for a dog's release, and this currently prevents an assessment of the C-BARQ's ability to predict specific behavioral issues later during training. In addition, judgements about releasing dogs tend to be based on their performance in training relative to one another such that the rate at which dogs are matched to blind or disabled handlers is rather fixed. In other words, the decision to reject or train a given dog depends upon the quality of the other dogs in training at the time. Training staff may invest a great deal of resources in a dog that is deemed "marginal" if the overall quality of dogs entering training at the same time is similarly marginal. Conversely, this same dog may be released early on if the other dogs currently available show considerably more promise. This process ensures that training programs produce the best dogs possible to meet current demand. However, it also means that what constitutes a "released" dog is flexible and therefore more difficult to predict with accuracy. Our current research includes a standardized format 'Behavioral Checklist' that is completed by trainers/instructors during the training process. This may allow more specific comparisons in the future between behaviors observed during the puppy raising period and those that arise during training.

In addition to being a potentially valuable aid to phenotyping and selecting working dogs, routine C-BARQ evaluations of young dogs may have other important uses. As a standardized behavioral measurement tool with fixed parameters, the C-BARQ represents a potentially valuable long-term yardstick for monitoring real behavioral changes in working dog populations over time in response to genetic selection or developmental interventions. Furthermore, by providing snapshots of behavior at specific points in a dog's first year of life, C-BARQ assessments can also serve to alert working dog organizations to the onset of behavior problems such as aggression or shyness that may be amenable to appropriate remedial interventions.

5. Conclusions

Valid and effective procedures for predicting working ability in guide and service dogs are currently needed. Standardized measures of well-defined behavioral phenotypes that can be used to identify genetic markers for either preferred or undesirable behavioral traits would also be of great value to working dog organizations. The present study indicates that the C-BARQ not only serves as a validated standardized measure of behavioral phenotypes in guide and service dogs, it also provides predictive value with respect to the probability of success in training.

Acknowledgments

The authors thank their collaborators at the participating guide and service dog organizations, in particular Dr. Eldin Leighton, Dr. Dolores Holle, and Peggy Gibbon (The Seeing Eye); Jane Russenberger and Barbara Havlena (Guiding Eyes for the Blind), Samantha Ziegenmeyer (Leader Dogs for the Blind), Michele Pouliot, Sarah Netoff and Marina Hall Phillips (Guide Dogs for the Blind); Paul Mundell and Kerinne Levy (Canine Companions for Independence). The authors also gratefully acknowledge the Arthur L. & Elaine V. Johnson Foundation for financial support.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.applanim.2012.02.011.

References

Batt, L.S., Batt, M.S., Baguley, J.A., McGreevy, P.D., 2009. The value of puppy raisers' assessments of potential guide dogs' behavioral tendencies and ability to graduate. Anthrozoos 22, 71–76.
Beaudet, R., Chalifoux, A., Dallaire, A., 1994. Predictive value of activity level and behavioral evaluation on future dominance in puppies. Appl. Anim. Behav. Sci. 40, 273–284.
Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate – a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B: Methodol. 57, 289–300.
Bewick, V., Cheek, L., Ball, J., 2004. Statistics review 13: receiver operating characteristic curves. Crit. Care 8, 508–512.
Duffy, D.L., Serpell, J.A., 2008. Behavioral assessment of guide and service dogs. J. Vet. Behav. Clin. Appl. Res. 3, 186–188.
Freimer, N., Sabatti, C., 2003. The human phenome project. Nat. Genet. 34, 15–21.
Goddard, M., Beilharz, R., 1986. Early prediction of adult behavior in potential guide dogs. Appl. Anim. Behav. Sci. 15, 247–260.
Goddard, M.E., Beilharz, R.G., 1984. The relationship of fearfulness to, and the effects of, sex, age and experience on exploration and activity in dogs. Appl. Anim. Behav. Sci. 12, 267–278.
Goddard, M.E., Beilharz, R.G., 1982. Genetic and environmental-factors affecting the suitability of dogs as guide dogs for the blind. Theor. Appl. Genet. 62, 97–102.
Hosmer, D.W., Lemeshow, S., 2000. Applied Logistic Regression. John Wiley & Sons.
Hsu, Y., Serpell, J.A., 2003. Development and validation of a questionnaire for measuring behavior and temperament traits in pet dogs. J. Am. Vet. Med. Assoc. 223, 1293–1300.
Hsu, Y., Sun, L., 2010. Factors associated with aggressive responses in pet dogs. Appl. Anim. Behav. Sci. 123, 108–123.

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.

Maejima, M., Inoue-Murayama, M., Tonosaki, K., Matsuura, N., Kato, S., Saito, Y., Weiss, A., Murayama, Y., Ito, S., 2007. Traits and genotypes may predict the successful training of drug detection dogs. Appl. Anim. Behav. Sci. 107, 287–298.

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

Nagasawa, M., Tsujimura, A., Tateishi, K., Mogi, K., Ohta, M., Serpell, J.A., Kikusui, T., 2011. Assessment of the factorial structures of the C-BARQ in Japan. J. Vet. Med. Sci. 73, 870–875.

Nunnally, J.C., 1978. Psychometric Theory. McGraw-Hill, New York.

Ruefenacht, S., Gebhardt-Henrich, S., Miyake, T., Gaillard, C., 2002. A behavior test on German shepherd dogs: heritability of seven different traits. Appl. Anim. Behav. Sci. 79, 113–132.

Serpell, J.A., Hsu, Y.Y., 2001. Development and validation of a novel method for evaluating behavior and temperament in guide dogs. Appl. Anim. Behav. Sci. 72, 347–364.

Spady, T.C., Ostrander, E.A., 2008. Canine behavioral genetics: pointing out the phenotypes and herding up the genes. Am. J. Hum. Genet. 82, 10–18.

Storey, J.D., 2002. A direct approach to false discovery rates. J. Roy. Stat. Soc., Series B 64, 479–498.

Svobodová, I., Vápeník, P., Pinc, L., Bartoš, L., 2008. Testing German shepherd puppies to assess their chances of certification. Appl. Anim. Behav. Sci. 113, 139–149.

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

Tomkins, L.M., Thomson, P.C., McGreevy, P.D., 2011. Behavioral and physiological predictors of guide dog success. J. Vet. Behav. Clin. Appl. Res. 6, 178–187.

van den Berg, L., Schilder, M.B.H., de Vries, H., Leegwater, P.A.J., van Oost, B.A., 2006. Phenotyping of aggressive behavior in golden retriever dogs with a questionnaire. Behav. Genet. 36, 882–902.

van den Berg, S.M., Heuven, H.C.M., van den Berg, L., Duffy, D.L., Serpell, J.A., 2010. Evaluation of the C-BARQ as a measure of stranger-directed aggression in three common dog breeds. Appl. Anim. Behav. Sci. 124, 136–141.

Wilsson, E., Sundgren, P.E., 1997. The use of a behavior test for the selection of dogs for service and breeding. 1. 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.

Wilsson, E., Sundgren, P.E., 1998. Behavior test for eight-week old puppies – heritabilities of tested behavior traits and its correspondence to later behavior. Appl. Anim. Behav. Sci. 58, 151–162.