



ELSEVIER

Applied Animal Behaviour Science 72 (2001) 347–364

APPLIED ANIMAL
BEHAVIOUR
SCIENCE

www.elsevier.com/locate/applanim

Development and validation of a novel method for evaluating behavior and temperament in guide dogs

James. A. Serpell*, Yuying Hsu

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

Accepted 2 December 2000

Abstract

Most guide and service dog organizations would benefit from the development of accurate methods for the early evaluation of canine temperament traits. This paper describes the development and validation of a novel questionnaire method for assessing behavior and temperament in 1-year-old guide dogs. Volunteer puppy-raisers scored a total of 1097 prospective guide dogs on a series of 40 semantic differential-type, behavioral rating scales. Principle components factor analysis of these scores extracted eight stable and interpretable common factors: stranger-directed fear/aggression, non-social fear, energy level, owner-directed aggression, chasing, trainability, attachment, and dog-directed fear/aggression. Three of these eight factors exhibited moderate internal consistency (Cronbach's $\alpha \geq 0.72$), while the reliabilities of the remaining factors were relatively low (Cronbach's $\alpha = 0.53$ – 0.61). The eight factors were then validated against the guide dog school's own criteria for rejecting dogs for behavioral reasons. The results of this analysis confirmed the construct validity of the puppy raisers' questionnaire assessments of their dogs, and suggested that such methods can provide a useful and accurate means of predicting the suitability of dogs for guiding work. Various modifications to the original questionnaire are proposed in order to enhance its overall reliability. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Behavioral assessment; Temperament; Guide dog

1. Introduction

Behavior or temperament problems represent one of the most commonly reported reasons for disqualifying dogs from serving as guides for blind or partially-sighted persons (Goddard and Beilharz, 1982, 1983). Rearing and training prospective guide

* Corresponding author. Tel.: +1-215-898-1004; fax: +1-215-573-6050.

E-mail address: serpell@vet.upenn.edu (J.A. Serpell).

dogs is costly in terms of time, effort, and resources, and it is generally more difficult to find alternative homes for adult dogs that fail to qualify in training than is the case for younger animals. It is, therefore, desirable to develop methods for determining the temperamental and behavioral suitability of these dogs for guiding work at the earliest possible age.

Previous efforts to assess working dog temperament and behavior have tended to focus on the predictive value of puppy temperament or 'aptitude' tests, usually conducted when pups are between 6 and 12 weeks of age (e.g. Scott and Bielfelt, 1976; Goddard and Beilharz, 1984, 1986; Wilsson and Sundgren, 1998; Slabbert and Odendaal, 1999). Testing typically involves exposing each pup to a limited range of controlled test situations that are considered useful for evaluating the behavioral traits of interest (e.g. aggression, fear, confidence, trainability, and so on), and assigning scores to its responses. The advantages of testing are that all pups are exposed to identical test situations at the same early age, and are scored by the same trained observer(s). The principal disadvantages are that the tests must be performed when the animals are still developmentally immature, and the results can strictly only be said to reflect a puppy's performance at the time of testing. In addition, and in spite of how frequently puppy tests have been used as a method of selecting working dogs, a surprising dearth of evidence exists that early testing provides a reliable means of forecasting adult performance traits (e.g. Scott and Bielfelt, 1976; Goddard and Beilharz, 1986; Wilsson and Sundgren, 1998).

The emphasis on early puppy testing as a method for selecting potential guide dogs arises from the fact that most organizations foster their pups out to the homes of volunteer puppy-raisers (or puppy-walkers) at about 7 weeks of age. Since these households are often widely scattered geographically, pups are generally considered difficult or impossible to evaluate behaviorally until they return to their training schools when they may be anything from 12 to 24-month-old. In addition, some schools favor early neutering, and are, therefore, interested in identifying potential breeding animals as early as possible.

This paper describes the development of an alternative method for evaluating the behavior of young, prospective guide dogs using puppy-raisers as primary sources of behavioral information. Despite its obvious potential, the acquired knowledge and experience of puppy-raisers concerning the behavior of their charges has been largely unexploited as a means of predicting the suitability of dogs for guiding work, primarily because the accuracy and reliability of such information has not been demonstrated. Using a simple 40-item behavioral questionnaire, the present study investigated the extent to which subjective behavioral ratings assigned by puppy-raisers can be used to provide accurate and reliable evaluations of the suitability of dogs for guiding work. Comparable questionnaire survey methods that use dog owners as primary sources of information regarding their pets' behavior and temperament have been used in a number of recent studies (e.g. Serpell, 1983, 1996; Serpell and Jagoe, 1995; Jagoe and Serpell, 1996; Podberscek and Serpell, 1996, 1997a,b; Goodloe and Borchelt, 1998; Podberscek et al., 1999). However, the psychometric properties of these owner-derived assessments have rarely been established, and their validity (i.e. their ability to measure what they purport to measure) is unclear. Recently, Hewson et al. (1988) attempted to validate owner-derived ratings of canine compulsive disorder by first asking dog owners to record the frequency,

duration, and context of their dogs' compulsive behaviors while the latter were undergoing treatment for the disorder. They found significant associations between these owner reports and subsequent scores on the rating scales administered at the end of each trial. However, as the authors note, these two sets of data were not independent (since both were derived from the same owners), and the fact that owners were required to closely observe the behavior of their dogs may have artificially increased the validity of the outcome measures. In addition, it is unclear whether the evaluations obtained from these owners accurately measured the dogs' behavior.

The present study attempts to overcome some of these methodological problems by comparing owner- or keeper-derived questionnaire evaluations with independent assessments made by guide dog trainers and handlers. We believe that this study represents the first time such methods have been appropriately validated.

2. Materials and methods

2.1. Questionnaire development

The questionnaire approach to the collection of behavioral data on dogs is based on the assumption that the owner (or primary keeper) of a dog usually knows more about its typical behavior than anyone else. The methodological difficulty lies in extracting this knowledge from the human informant in a form that is accurate, quantitative, and reliable. The method developed in this study employed a format similar to those developed in temperament studies of children in which parents are asked to rate their children's recent responses to a variety of specific environmental events and stimuli (e.g. Hegvik et al., 1982). The canine equivalent employed here uses a series of five-point, semantic differential-type rating scales (Osgood et al., 1957) that allow puppy-raisers to provide quantitative assessments of their dogs' typical responses to a range of common situations and events. For example, the questionnaire item for 'response to traffic' is presented in the following format:

Seems unperturbed
by heavy traffic 1.....2.....3.....4.....5 Is obviously alarmed or
frightened by heavy traffic

Puppy-raisers are asked to rate dogs by circling or underlining a number on the scale that best describes their dog's typical behavior in this context. The choice of situations and events to include in the initial questionnaire was determined from interviews with guide dog instructors and their supervisors, and from approximately 18 h of field observations of guide dogs in training. They also reflected the most common behavior problems observed in disqualified guide dogs. An initial prototype questionnaire was critically reviewed by a panel of guide dog trainers and puppy-placement supervisors, whose comments and suggestions were then incorporated into the final 40-item instrument employed in the study. The instrument comprises 12 items on aggression, eight items on fear and anxiety, eight items on training and obedience, four items on attachment and separation distress, and eight 'general behavior' items (these items are listed in Tables 3 and 4).

2.2. Subjects

All subjects for the study were recruited via The Seeing Eye Inc. (TSE), a large, national guide dog organization based in New Jersey, USA. Volunteer puppy-raisers were routinely requested to evaluate their pups at 6 and 12 months of age using the standard 40-item questionnaire. The 12 month's questionnaire data on a total of 1067 prospective guide dogs were collected in this way, and were used for the present analysis. Results from the 6 month's questionnaires are currently being analyzed and will be the subject of a separate report. The dogs evaluated comprised four different breed categories: German Shepherd Dog ($N = 293$), Labrador Retriever ($N = 369$), Golden Retriever ($N = 264$), and Labrador \times Golden Retriever crosses ($N = 140$). All but 10 of the dogs were sexually intact at the time of questionnaire evaluation.

2.3. Statistical analyses

A separate and independent measure of guide dog performance was needed in order to estimate the construct validity of the 12 month's questionnaire evaluations. For this purpose, we employed TSE's own pre-existing system for classifying dogs that are rejected which was developed independently many years before the initiation of the present study. TSE determines whether or not a dog will become a successful guide through a process of continuous assessment involving the input of veterinarians, training instructors, and their supervisors. This process occurs when the dogs return to the training school for final instruction when they are between 14 and 24 months of age. A dog may be rejected at any point during the 4 months training program, depending on the severity of its problem(s). Every dog that is rejected is assigned a single, primary "reason for rejection" by TSE. A list of the most common behavioral reasons for rejection, together with their formal definitions, is provided in Table 1 (only reasons for rejection that were assigned to more than 10 individual dogs were used for validation purposes).

To simplify the task of comparing TSE's "reasons for rejection" with puppy-raisers' evaluations, the 40-item questionnaire scores were subjected to principal-components factor analysis, using Kaiser's eigenvalue rule and varimax rotation, in order to identify a set of stable behavioral sub-scales that measured different aspects of canine temperament (Devellis, 1991). The returned surveys were first divided randomly into three groups, two of which were used to identify common factors. The third group was then used to test the stability of the factors identified from the first two groups. To avoid the possible biasing effects of common parentage, individuals bred from the same dams and/or sires were randomly distributed among the three groups. In addition, the grouping method did not result in any significant bias in male:female ratio ($\chi^2_2 = 0.86$, $P = 0.65$, $N = 1064$) or breed composition ($\chi^2_6 = 1.24$, $P = 0.97$, $N = 1065$) among the three groups.

Cronbach's α (Devellis, 1991) was used to assess the internal consistency (reliability) of the extracted factors, and Mann-Whitney U -tests (Siegel and Castellan, 1988) were used to determine whether dogs that were rejected for particular behavioral reasons showed correspondingly high or low scores on appropriate factors derived from the questionnaire data.

Table 1
Primary behavioral reasons assigned to the dogs rejected from guide dog training^a

Fail reason	N (%)	Definition
Suspicious of people	58 (18.0)	Fear of strange people; the behavior is characterized by raised hackles, flight posture, flight, growling, and perhaps fear biting if cornered
Lack of confidence	54 (16.7)	Is characterized by general nervousness when asked to perform stressful tasks; this can be exhibited by tail tuck, trembling, salivating, flight toward home, and over-solicitous behavior toward the handler
Distraction	44 (13.6)	Unable (or unwilling) to focus on the task of guiding; distracters could be flying leaves, other animals, children, etc
Aggressive/suspicious toward dogs	40 (12.4)	Barks/growls at dogs; fear may or may not be the cause of the behavior
Self-right	29 (9.0)	A dominant dog that will not yield to discipline or difficult training tasks
Fear of thunder, traffic, noise	20 (6.2)	Shows fearful behavior (flees, trembles, panics) toward thunder, traffic, or unknown noises
Person aggression/protective	19 (5.9)	Is over-devoted to its owner and defends her/him against strangers; the aggressiveness can sometimes stem from suspiciousness of strangers
Excitable	15 (4.6)	High energy, has a very difficult time settling down and behaving quietly for extended periods of time
Fear of stairs	10 (3.1)	Is afraid to approach stairs or, once on the staircase, rushes to get off them
Emotional urinating	10 (3.1)	Urinate when approached
Other behavior	24 (7.4)	All fail reasons with $N < 10$
Total	323	

^a The definitions, and the frequency of dogs rejected for each reason are also provided (% = $N/323$).

3. Results

3.1. Factor analyses

Questionnaire response rates were high (>96%) for 38 out of the 40 questionnaire items. The two exceptions were “aggression toward other dogs in the household” (74.6%) and “tendency to chase cats” (85.9%), reflecting the absence of either other dogs in the household, or opportunities to chase cats. Although we aimed to include all behavior scores for data analyses, the score for “aggression toward other dogs in household” was excluded because it would have severely reduced the sample size.

The results of the factor analyses suggested that 25 out of the remaining 39 behavioral scores formed eight stable and interpretable common factors for the individuals in groups 1 and 2. To confirm this structure, factor analyses were performed on the subset of these 25 behavioral scores of the individuals in groups 1 and 2. These analyses produced eight factors accounting for 66.3 and 63.9% of the common variance for groups 1 and 2, respectively. The eight factors were designated as follows: “stranger-directed fear/aggression” — five items; “non-social fear” — four items; “energy level” — two items; “owner-directed aggression” — three items; “chasing” — three items; “trainability” — three items; “attachment” — three items; and “dog-directed fear/aggression” — two items (see Table 2). As shown in Table 2, behavioral scores loaded moderately to highly (≥ 0.47) on their designated factors, with little or no cross-loading on other factors.

3.1.1. Stability of the factor structure

We used the same 25 behavior scores of the individuals in group 3 to test the replicability of the factor structure revealed from the behavior scores of individuals in groups 1 and 2. Eight factors that accounted for 63.9% of the common variance were extracted (Table 2). The factor structure was nearly identical to those obtained from the behavior scores of the individuals in groups 1 and 2. All behavior scores had moderate to high loadings (≥ 0.41) on their designated factors with a few moderate cross-loadings on other factors: “growls/snaps when food, bone, or toys taken away” also received a loading of -0.44 on the “attachment” factor, and “easily distracted by interesting sights, sounds, smells, etc.” and “becomes wildly excited when owner comes home” also received loadings of 0.56 and 0.43 on the “chasing” factor, respectively.

Factor analysis of the same 25 behavioral scores of the entire population extracted the same eight factors that accounted for 63.3% of the common variance. All behavior scores had moderate to high loadings (≥ 0.54) on their designated factors with little or no cross-loadings on other factors (Table 2).

The stability of the factor structure over sex and breed was also tested by performing factor analyses on male–female and breed sub-groups. Three breeds (Labrador Retriever, German Shepherd Dog, and Golden Retriever) with sample sizes greater than 200 were included in the breed sub-group analysis. The same eight factors, accounting for 63.9–65.4% of the common variance, were extracted from the male/female sub-groups and the three breed sub-groups (Table 3). The eigenvalues for most of these factors were greater than one (Kaiser’s eigenvalue rule) with a few exceptions: the eigenvalue for the last factor for the Golden Retriever breed sub-group was 0.99 and the eigenvalues for the last two

Table 2
Moderate to high factor loadings for the behavioral items comprising the eight identified factors for the three random sub-groups and the entire population

Behavior item	1 (273, 66.3) ^a	2 (278, 63.9)	3 (267, 63.9)	Pooled (818, 63.3)
Factor 1 — stranger-directed fear/aggression				
Fearful when approached by unfamiliar children	−0.84	0.84	0.77	−0.83
Barks/growls when approached by unfamiliar children	−0.81	0.83	0.75	−0.80
Fearful when approached by unfamiliar adults	−0.76	0.82	0.79	−0.78
Barks/growls when approached by unfamiliar adult	−0.78	0.71	0.83	−0.77
Barks/growls at unfamiliar visitors	−0.69	0.66	0.78	−0.71
Factor 2 — non-social fear				
Obviously alarmed or disturbed by sudden or loud noises	−0.81	−0.79	−0.80	0.81
Obviously alarmed or frightened by heavy traffic	−0.78	−0.74	−0.74	0.76
Is ‘spooked’ by odd or unexpected things or objects	−0.75	−0.78	−0.76	0.76
Anxious or fearful in unfamiliar situations	−0.67	−0.65	−0.66	0.66
Factor 3 — energy level				
Playful; puppyish	0.85	0.87	0.86	0.86
Active, energetic	0.86	0.83	0.83	0.84
Factor 4 — owner-directed aggression				
Objects (growls, etc.) to being manhandled/groomed	0.77	−0.47	−0.78	0.70
Objects (growls, etc.) to being corrected/punished	0.79	−0.67	−0.62	0.68
Growls/snaps when food, bones, or toys taken away	0.72	−0.77	−0.42	0.75
Factor 5 — chasing				
Chase birds or squirrels (would like to)	0.79	−0.76	0.76	−0.81
Chase cats (would like to)	0.52	−0.63	0.54	−0.63
Chase cars, cyclists, or joggers (would like to)	0.55	−0.63	0.72	−0.66
Factor 6 — trainability				
Learns new things quickly	−0.81	−0.75	0.76	−0.78
Shows initiative at solving problems on her/his own	−0.72	−0.71	0.76	−0.75
Attentive; not easily distracted	−0.49	−0.62	0.52	−0.54

Table 2 (Continued)

Behavior item	1 (273, 66.3) ^a	2 (278, 63.9)	3 (267, 63.9)	Pooled (818, 63.3)
Factor 7 — attachment				
Very strongly attached to one particular person in household	0.62	0.65	−0.75	0.74
Extremely demanding of attention	0.65	0.75	−0.59	0.66
Wildly excited when owner comes home	0.74	0.62	−0.41	0.62
Factor 8 — dog-directed fear/aggression				
Fearful when approached by unfamiliar dogs	−0.72	−0.69	0.63	−0.72
Barks/growls when approached by unfamiliar dogs	−0.78	−0.80	0.73	−0.83

^a Values in parentheses represent *N*, variance accounted for (in %).

Table 3

Moderate to high factor loadings for the behavioral items comprising the eight identified factors for male and female sub-groups, and the three breed sub-groups

Behavior item	Sex ^a		Breed		
	M (401, 63.9)	F (416, 64.1)	LAB (294, 65.4)	GSD (218, 65.4)	GR (205, 64.7)
Factor 1 — stranger-directed fear/aggression					
Fearful when approached by unfamiliar children	0.85	−0.77	0.75	0.83	−0.82
Barks/growls when approached by unfamiliar children	0.77	−0.75	0.58	0.80	−0.86
Fearful when approached by unfamiliar adults	0.82	−0.80	0.85	0.85	−0.69
Barks/growls when approached by unfamiliar adult	0.77	−0.80	0.69	0.78	−0.78
Barks/growls at unfamiliar visitors	0.66	−0.76	0.75	0.73	−0.65
Factor 2 — non-social fear					
Obviously alarmed or disturbed by sudden or loud noises	−0.82	0.79	−0.77	−0.78	0.84
Obviously alarmed or frightened by heavy traffic	−0.74	0.79	−0.74	−0.76	0.76
Is ‘spooked’ by odd or unexpected things or objects	−0.74	0.76	−0.79	−0.68	0.77
Anxious or fearful in unfamiliar situations	−0.61	0.70	−0.63	−0.56	0.72
Factor 3 — energy level					
Playful; puppyish	0.87	−0.84	0.88	−0.82	0.84
Active, energetic	0.84	−0.84	0.81	−0.87	0.85
Factor 4 — owner-directed aggression					
Objects (growls) to being manhandled/groomed	0.73	−0.59	0.70	0.56	0.68
Objects (growls) to being corrected/punished	0.66	−0.64	0.68	0.70	0.66
Growls/snaps when food, bones, or toys taken away	0.76	−0.78	0.72	0.81	0.69
Factor 5 — chasing					
Chase birds or squirrels (would like to)	0.84	−0.70	0.76	0.79	−0.74
Chase cats (would like to)	0.64	−0.54	0.68	0.61	−0.52
Chase cars, cyclists, or joggers (would like to)	0.69	−0.60	0.67	0.66	−0.68
Factor 6 — trainability					
Learns new things quickly	0.83	−0.73	0.81	0.82	−0.72
Shows initiative at solving problems on her/his own	0.66	−0.79	0.78	0.50	−0.74
Attentive; not easily distracted	0.56	−0.57	0.52	0.57	−0.63

Table 3 (Continued)

Behavior item	Sex ^a		Breed		
	M (401, 63.9)	F (416, 64.1)	LAB (294, 65.4)	GSD (218, 65.4)	GR (205, 64.7)
Factor 7 — attachment					
Very strongly attached to one particular person in household	0.68	0.73	−0.80	−0.75	−0.35
Extremely demanding of attention	0.70	0.65	−0.65	−0.61	−0.84
Wildly excited when owner comes home	0.65	0.53	−0.58	−0.55	−0.62
Factor 8 — dog-directed fear/aggression					
Fearful when approached by unfamiliar dogs	−0.70	0.68	0.59	0.59	−0.65
Barks/growls when approached by unfamiliar dogs	−0.85	0.80	0.71	0.80	−0.84

^a Values in parentheses represent *N*, variance accounted for (in %).

factors of the female sub-group were 0.97 and 0.94, respectively. The factor structures of these sub-groups were nearly identical to each other and to those obtained from groups 1, 2 and 3. All behavior scores had moderate to high loadings on their designated factors. Only two behavior scores (“objects to be manhandled or groomed” and “easily distracted by interesting sights, sounds, smells, etc.”) in the German Shepherd Dog sub-group also had moderate cross-loadings (−0.45 and −0.44, respectively) on the factors “trainability” and “chasing”, respectively (Table 3). These results indicate that the factor structure is stable between the male and female sub-groups, and among the three primary breeds of dogs trained by TSE.

3.1.2. Excluded behaviors

A high proportion of the training/obedience related behavior scores (four out of eight) were excluded by factor analysis (Table 4). All of these behavior scores, except “pulls excessively hard on leash”, showed moderate but inconsistent factor loadings on both the “owner-directed aggression” factor (obedient: 0.25–0.57; sensitive to correction: 0.00–0.56; willingness: 0.19–0.56; response to praise: 0.03–0.40) and the “trainability” factor (obedient: 0.25–0.52; sensitive to correction: 0.02–0.52; willingness: 0.14–0.70; response to praise: 0.08–0.62). Indeed, they tended to oscillate between these two factors from sub-group to sub-group (individually but not as a group). The score for “pulls excessively hard on leash” did not seem to belong to any of the eight factors and had only low to moderate loadings on the eight factors in all sub-groups.

Table 4
Behavior items not included in the eight identified factors

Behavior item
Aggression related
1. Barks/growls at approaching strangers when inside the car
2. Barks loudly and persistently when doorbell rings or mail is delivered
3. Becomes agitated or aggressive if owner gives attention to other person or animal
4. Boisterous or rough when playing
Fear related
6. Is anxious or nervous about going up or down stairs
Training/obedience related
7. Obedient
8. Sensitive; easily upset by correction or punishment
9. Eager to please; willing
10. Pulls weakly or inconsistently on leash
11. Loves to be praised
Attachment related
12. Hates being left alone; noisy, destructive or makes a mess
Other behavior
13. Always picky about food
14. Becomes wildly excited on car journeys
15. Excitable

Four out of the 11 aggression-related behavior scores could not be grouped with any of the 8 factors. The score for “boisterous or rough when playing” loaded moderately on both the “Owner-directed aggression” factor (0.01–0.49) and the “trainability” factor (0.02–0.47) but its relationship with both factors was unstable. Similarly, although the item “becomes agitated or aggressive if owner gives attention to other person or animal” loaded moderately on both the “owner-directed aggression” (0.16–0.51) and the “attachment” factors (0.26–0.74), these relationships were unstable. The relation between the items “barks/growls at approaching strangers when inside the car” and “barks loudly and persistently when doorbell rings or mail is delivered” is not clear. These two behavior scores stood out as one factor in some sub-groups. However, association between these two behavior scores was not stable and could not be detected in other sub-groups. These two behavior scores had only low to moderate loadings on the eight factors in all sub-groups.

Only one out of the eight fear-related behavior scores was not grouped into any of the eight factors: “is anxious or nervous about going up or down stairs”. The score for this behavior stood out as a separate factor (9th factor) in all three random sub-groups if included in the factor analysis (with a loading of 0.78, 0.93, and 0.85 in the sub-groups after varimax rotation). However, we decided to extract 8 instead of 9 factors because the eigenvalues for the 9th factor were lower than one in all three sub-groups (0.87, 0.92, and 0.96). In addition, the score for stair anxiety was the only score that loaded highly on this 9th factor, which defeats the purpose (condensing information) of factor analysis. It is worth noting, however, that all non-social types of fears, apart from stair anxiety, loaded highly on the “non-social fear” factor. This suggests that fear of staircases may develop independently from the fear of other non-social stimuli such as noise, traffic, unexpected objects, and unfamiliar situations.

One out of the four attachment-related behavior scores was not grouped into any factor: “hates being left alone. Noisy, destructive, or makes a mess.” The score for this behavior had low loadings on the identified factors.

The score for “always picky about food” had low loadings on the eight identified factors, as did the score for “becomes wildly excited on car journeys”. The score for “excitable” had consistent moderate loadings on the “energy level” factor (0.56, 0.53, and 0.40 for the three random sub-groups). However, it was excluded because it would reduce the internal consistency (Cronbach’s α) of the factor.

3.2. Internal consistency (reliability) of the sub-scales

The internal consistency of each sub-scale was estimated using Cronbach’s α (Devellis, 1991). Despite the small number of items in all sub-scales, three out of the eight sub-scales had moderate α -values (stranger-directed fear/aggression: 0.84; non-social fear: 0.77; energy level: 0.72). However, the alphas of the other five sub-scales appeared to be relatively low (owner-directed aggression: 0.57; chasing: 0.61; trainability: 0.60; attachment: 0.53; dog-directed fear/aggression: 0.58), suggesting that more behavioral items measuring these constructs should be added to the sub-scales to increase their reliabilities. Although fearful and aggressive responses toward unfamiliar people were highly correlated with each other, and came out as one sub-scale in the factor analysis (“stranger-directed fear/aggression”), some of TSE’s “reasons for rejection” consist of fearful

responses only, or aggressive responses only (see Table 1). It, therefore, seemed appropriate to separate fear and aggression components and examine their validity independently. Scores for “fearful when approached by unfamiliar children” and “fearful when approached by unfamiliar adults” were consequently combined as a “stranger-directed fear” sub-scale, and scores for “barks, growls when approached by unfamiliar children”, “barks, growls when approached by unfamiliar adults” and “barks/growls at unfamiliar visitors” were combined as a “stranger-directed aggression” sub-scale. Even though subdivided from one factor, these two sub-scales appeared to have reasonable internal consistency as measured by Cronbach’s α (stranger fear: $\alpha = 0.77$, stranger aggression: $\alpha = 0.76$).

3.3. Construct validity of the sub-scales

The Seeing Eye’s “reasons for rejection” (Table 1) were used to examine the convergent validity (evidence of similarity between measures of theoretically related constructs) and discriminant validity (absence of correlation between measures of unrelated constructs; Devellis, 1991) of the eight sub-scales. One hundred and seventy out of the 963 dogs with known training outcomes were rejected for health-related reasons and were excluded from subsequent analyses, and 323 of the remaining 793 dogs failed because of behavioral problems and were assigned a reason for rejection. If the behavior scores derived from puppy-raisers’ questionnaire evaluations are valid measures of the future performance of dogs (convergent validity), the following correspondences would be expected between particular reasons for rejection by TSE and dogs’ behavioral sub-scale scores on the questionnaire:

- dogs rejected for “suspicious of people” would be expected to score higher on both “stranger fear” and “stranger aggression”;
- dogs rejected for “lack of confidence” would be expected to score higher on “stranger fear” and “non-social fear”;
- dogs rejected for “distraction” would be expected to score lower on “trainability”, and higher on “energy level” and possibly “chasing”;
- dogs rejected for “aggressive/suspicious toward dogs” would be expected to score higher on “dog fear/aggression”;
- dogs rejected for “self right” would be expected to score higher on “owner-directed aggression” and lower on “trainability”;
- dogs rejected for “fear of thunder/traffic/noise” would be expected to score higher for “non-social fear”;
- dogs rejected for “aggressive to strangers/protective” would be expected to score higher on “stranger fear”, “stranger aggression” and “attachment”;
- and dogs rejected for “excitable” would be expected to score higher for “energy level” as well as “attachment” (since this factor contains the item “wildly excited when owner comes home”).

In addition, there should be little or no correspondence between behavior sub-scales and unrelated reasons for rejection (discriminant validity).

Observed relationships between the TSE’s reasons for rejection and the puppy-raiser’s behavior scores are summarized in Table 5. The expected relationship between rejection

Table 5

Results (*P*-values) of Mann-Whitney *U*-statistics testing the relationship between the primary reasons assigned to the dogs that were rejected and their behavioral sub-scale (factor) scores^a

	Primary reason for rejection:							
	Suspicious of people	Lacking confidence	Distraction	Aggressive/suspicious toward dogs	Self-right	Fear of thunder/traffic/noise	Aggressive to strangers/protective	Excitable
Behavioral sub-scale								
Stranger fear	<u><0.001</u> ^b	<u>0.035</u> ^b	0.455	0.386	0.858	0.554	<u><0.001</u> ^b	0.707
Stranger aggression	<u><0.001</u> ^b	0.278	0.771	0.803	0.959	0.783	<u><0.001</u> ^b	0.028 ^b
Non-social fear	0.286	<u>0.009</u> ^b	0.909	0.133	0.990	<u>0.012</u> ^b	0.185	0.980
Energy level	0.816	0.376	<u>0.062</u> ^b	0.314	0.878	0.424	0.274	<u>0.057</u> ^b
Owner aggression	0.699	0.429	0.509	0.457	<u>0.006</u> ^b	0.874	0.151	0.705
Chasing	0.871	0.692	<u>0.859</u>	0.514	0.815	0.710	0.100	0.605
Trainability	0.672	0.985	<u>0.254</u>	0.024 ^b	<u>0.019</u> ^c	0.539	0.744	0.675
Attachment	0.003 ^b	0.442	0.752	0.185	0.619	0.868	<u>0.021</u> ^b	<u>0.009</u> ^b
Dog fear/aggression	0.399	0.186	0.355	<u>0.011</u> ^b	0.168	0.921	0.203	0.578

^a Predicted associations between behavioral sub-scales and reasons for rejection are underlined.

^b Indicates that dogs rejected for a specific reason scored higher.

^c Indicates that dogs rejected for a specific reason scored lower.

for “suspicious of people” and both the “stranger fear” and “stranger aggression” sub-scales was strongly confirmed by the analysis (Mann-Whitney *U*-tests). The additional association with the “attachment” sub-scale was unexpected, and would suggest some overlap or confusion between TSE’s two rejection categories: “suspicious of people” and “aggressive to strangers/protective” (see below).

Dogs rejected for “lack of confidence” scored higher on the “stranger fear” and “non-social fear” sub-scales as expected. Dogs rejected for “distraction” scored higher for the “energy level” factor as anticipated, but the lack of association with the “chasing” and “trainability” factors was unexpected and hard to explain. Further analyses revealed that dogs rejected for “distraction” scored significantly lower for the “attentive, not easily distracted” item of the “trainability” factor (Mann-Whitney *U*-statistic, $P = 0.026$), but not for the “learns new things quickly” and “problem solving” items ($P = 0.598$ and 0.609 , respectively). These results may indicate that although less attentive dogs may show moderately reduced learning and problem solving abilities, dogs that were rejected for “distraction” were not necessarily slow at learning tasks or solving problems. It is also possible that easily distracted dogs which also display resistance to learning tasks and solving problems were assigned a different primary reason for rejection (i.e. self right) by TSE.

As expected, dogs that were rejected for “aggressive/suspicious toward dogs” scored higher for the “dog-directed fear/aggression” sub-scale. In addition, however, they also scored higher on the “trainability” sub-scale. One possible explanation for this unexpected association might be that dogs that are aggressive or fearful toward other dogs are also more socially orientated toward humans, in which case we would predict a positive correlation between the “dog-directed fear/aggression” sub-scale and the “trainability” sub-scale. This explanation can be rejected, however, since the actual correlation between these two sub-scales is significantly negative (product-moment correlation coefficient $r = -0.176$, $N = 1038$, $z = -5.72$, $P < 0.001$). In other words, although dogs rejected by TSE for “aggressive/suspicious toward dogs” scored higher for the “trainability” sub-scale, dogs that were evaluated by puppy-raisers as being aggressive or fearful toward other dogs tended to score lower for the “trainability” sub-scale. Dogs rejected by TSE for “aggressive/suspicious toward dogs” scored significantly higher for only one of the three behavioral items comprising the “trainability” sub-scale: “attentive, not easily distracted” (Mann-Whitney *U*-test, $P < 0.001$). Again, however, the general relationship between the “dog-directed fear/aggression sub-scale” and the item score for “attentive, not easily distracted” was negative (product-moment correlation coefficient $r = -0.094$, $N = 1051$, $z = -3.06$, $P = 0.002$). Thus, it remains unclear to us why dogs rejected for “aggression/suspicious toward other dogs” were more attentive than other dogs.

Dogs rejected for “self-right” scored, as expected, higher for “owner-directed aggression” and lower for “trainability”. The expected relationship between rejection for “fearful of thunder/traffic/noise” and high scores for the “non-social fear” factor was also confirmed.

The definition given for “aggressive to strangers/protective” is similar to the one given for “suspicious of people” (Table 1), except that the former is supposedly motivated by a desire to ‘protect’ the owner. A dog rejected for its tendency to be “aggressive to strangers/protective” would, therefore, be expected to score higher for the factors “stranger-directed

fear”, “stranger-directed aggression”, and “attachment”. These three relationships were strongly confirmed.

The relationships between rejection for “excitable” and scores for “energy level” and “attachment” were confirmed, although the stronger observed association with the “stranger aggression” sub-scale was unexpected. Since excitability is also a juvenile trait, it is possible that these additional associations reflect behaviorally immature animals that tend to bark excitedly at strangers, as well as being more dependent on, and attached to, their puppy-raisers.

4. Discussion

The results of the analysis confirm the overall construct validity of a questionnaire method for evaluating the behavior of this population of 1-year-old, prospective guide dogs. Fourteen out of the 16 expected associations between rejection reasons and behavior factors were statistically significant or marginally significant in the predicted direction (convergent validity), in spite of the fact that the source of validation (reasons for rejection) occurred from 2 to 12 months after questionnaire administration. The only expected associations that were not significant were between rejection for “distraction” and the sub-scales “trainability” and “chasing”. Conversely, only 3 out of 56 predicted non-associations were statistically significant (discriminant validity). Unexpectedly, the dogs that failed for being “suspicious of people” scored significantly higher for the “attachment” sub-scale; dogs that failed for being “aggressive/suspicious toward dogs” scored significantly higher for “trainability” sub-scale; and dogs rejected for “excitable” scored higher for the “stranger aggression” sub-scale. Some of these significant effects may be false results arising from the 5% type I error allowed. Future study will be necessary to confirm these relationships. The main conclusion to be drawn from these findings is that the puppy raisers’ behavioral assessments of their dogs provided useful predictive information regarding the suitability of these dogs for guiding work. To the best of the authors’ knowledge, this is the first time that such a behavioral survey method has been properly validated.

Since the questionnaire used in this study was developed to evaluate a particular population of working dogs at a particular age, its general validity as a tool for assessing the behavior of dogs derived from other populations, age-groups, and contexts is unknown. The findings nevertheless illustrate the potentially broad applicability of similar research methods for measuring the behavior of dogs — especially working dogs — in situations where puppy-raisers (or their equivalents) are used, and where behavioral observation and/or testing is difficult, or when the behaviors of interest are too rare or context-specific to be readily detected by other more conventional means. Questionnaire methodologies of the kind described here have the added advantage of being relatively simple and inexpensive to conduct, and the information they supply could be used to substantially reduce the overhead costs of organizations that breed and train dogs for working roles. In addition, they may help the guide and service dog industries to move toward a more standardized classification system for canine temperament and behavior (Knol et al., 1988).

With respect to the structure of the questionnaire itself, the findings suggest that additional questions should be added to the five behavioral sub-scales that have low

reliabilities. However, despite the low reliabilities, all sub-scales appear to have good validities. The questionnaire only weakly predicted the tendency of dogs to be rejected for “distraction”. Because a relatively high proportion of dogs (13.6%) were rejected for “distraction”, a new set of questions should be added to the questionnaire to better measure this tendency. Dogs rejected for “suspicious of people” scored significantly higher for the “attachment” sub-scale unexpectedly. The definition for “suspicious of people” is somewhat similar to the definition of “aggressive to strangers/protective” (Table 1), and the individuals rejected for these two reasons scored similarly on the behavioral sub-scales. It is possible that these two reasons are used interchangeably by TSE when assigning primary reasons for rejection. It is also possible that dogs who are simultaneously fearful and aggressive toward strangers tend to appear more attached to their owners.

Although only a handful of dogs were rejected for “fear of stairs”, these dogs scored significantly higher for the “stair anxiety” item in the puppy-raiser survey than the rest of the healthy dogs (Mann-Whitney *U*-statistics, $P < 0.001$). Because this item could not be grouped with other fearful responses toward environmental factors but is still useful in predicting the outcome of guide dog training, more items measuring the same behavior should be added to the survey in future studies in the hope of extracting a stair anxiety factor with adequate reliability characteristics.

Although fearful and aggressive responses toward unfamiliar people were highly correlated and emerged as one factor in the factor analysis, they were associated with different rejection reasons. For example, prospective guide dogs rejected for “lack of confidence” scored significantly higher on the “stranger fear” sub-scale than the rest of the population but not on the “stranger aggression” sub-scale. Conversely, dogs rejected for “excitable” scored significantly higher on the “stranger aggression” sub-scale than the rest of the population but not on the “stranger fear” sub-scale. These findings suggest that although aggression is a common response to fear-evoking stimuli in dogs, both traits or tendencies may also appear in the absence of the other, and should, therefore, be treated and measured as two separate aspects of temperament. This suggestion is also consistent with the results of Goodloe and Borchelt (1998) who, using a wider range of questions measuring either aggressive or fearful responses toward strangers, also extracted two separate factors “aggression to strangers” and “fear or avoidance of strangers”.

The results of the present study also indicate that the fearful and aggressive responses of dogs are relatively object-specific. Fearful responses toward unfamiliar people, dogs, environmental factors (thunder, traffic, noise, strange objects, etc.), and stairs appear to develop independently, as do aggressive responses toward owners, unfamiliar people, and other dogs. These findings are also generally in agreement with the conclusions of Goodloe and Borchelt (1998), although fearful responses toward dogs and environmental factors were not evaluated in their study.

Acknowledgements

This research was supported and assisted by The Seeing Eye, Inc. The authors thank especially Ken Rosenthal, Doug Roberts, Eldin Leighton, Dolores Holle and Christine

Dickson for their help with various aspects of the study, as well as acknowledging the contributions of the many puppy-raisers who provided the data.

References

- Devellis, R.F., 1991. Scale Development. SAGE publications, Newbury Park, CA, 121 pp.
- 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.
- Goddard, M.E., Beilharz, R.G., 1983. Genetics of traits which determine the suitability of dogs as guide dogs for the blind. *Appl. Anim. Ethol.* 9, 299–315.
- Goddard, M.E., Beilharz, R.G., 1984. A factor analysis of fearfulness in potential guide dogs. *Appl. Anim. Behav. Sci.* 12, 253–265.
- Goddard, M.E., Beilharz, R.G., 1986. Early prediction of adult behaviour in potential guide dogs. *Appl. Anim. Behav. Sci.* 15, 247–260.
- Goodloe, L.P., Borchelt, P.L., 1998. Companion dog temperament traits. *JAAWS* 1, 303–338.
- Hegvik, R.L., McDevitt, S.C., Carey, W.B., 1982. The middle childhood temperament questionnaire. *Dev. Behav. Pediat.* 3, 197–200.
- Hewson, C.J., Luescher, U.A., Ball, R.O., 1988. Measuring change in the behavioural severity of canine compulsive disorder: the construct validity of categories of change derived from two rating scales. *Appl. Anim. Behav. Sci.* 60, 55–68.
- Jagoe, J.A., Serpell, J.A., 1996. Owner characteristics and interactions and the prevalence of canine behaviour problems. *Appl. Anim. Behav. Sci.* 47, 31–42.
- Knol, B.W., Roozendaal, C., Bogaard, L.V.D., Bouw, J., 1988. The suitability of dogs as guide dogs for the blind: criteria and testing procedures. *Vet. Q.* 10, 198–204.
- Osgood, C.E., Suci, G.J., Tannenbaum, P.H., 1957. *The Measurement of Meaning*. University of Illinois Press, Urbana.
- Podberseck, A.L., Serpell, J.A., 1996. The English Cocker Spaniel: preliminary findings on aggressive behaviour. *Appl. Anim. Behav. Sci.* 47, 75–89.
- Podberseck, A.L., Serpell, J.A., 1997a. Environmental influences on the expression of aggressive behaviour in English cocker spaniels. *Appl. Anim. Behav. Sci.* 52, 215–227.
- Podberseck, A.L., Serpell, J.A., 1997b. Aggressive behaviour in English cocker spaniels and the personality of their owners. *Vet. Rec.* 141, 73–76.
- Podberseck, A.L., Hsu, Y., Serpell, J.A., 1999. Evaluation of clomipramine as an adjunct to behavioural therapy in the treatment of separation problems in dogs. *Vet. Rec.* 145, 365–369.
- Scott, J.P., Bielfelt, S.W., 1976. Analysis of the puppy testing program. In: Pfaffenberger, C.J., Scott, J.P., Fuller, J.L., Ginsburg, B.E., Bielfelt, S.W. (Eds.), *Guide Dogs for the Blind: Their Selection, Development and Training*. Elsevier, Amsterdam, pp. 39–75.
- Siegel, S., Castellan, N.J., 1988. *Non-parametric Statistics for the Behavioral Sciences*. 2nd Edition. McGraw-Hill, London, 399 pp.
- Serpell, J.A., 1983. The personality of the dog and its influence on the pet-owner bond. In: Katcher, A.H., Beck, A.M. (Eds.), *New Perspectives on Our Lives with Companion Animals*. University Pennsylvania Press, Philadelphia, pp. 57–63.
- Serpell, J.A., 1996. Evidence for an association between pet behaviour and owner attachment levels. *Appl. Anim. Behav. Sci.* 47, 49–60.
- Serpell, J.A., Jagoe, J.A., 1995. Early experience and the development of behaviour. In: Serpell, J.A. (Ed.), *The Domestic Dog*. Cambridge University Press, pp. 80–102.
- Slabbert, J.M., Odendaal, J.S.J., 1999. Early prediction of adult police dog efficiency — a longitudinal study. *Appl. Anim. Behav. Sci.* 64, 269–288.
- Wilsson, E., Sundgren, P.E., 1998. Behaviour test for 8-week-old puppies-heritabilities of tested behaviour traits and its correspondence to later behaviour. *Appl. Anim. Behav. Sci.* 58, 151–162.