Journal of Veterinary Behavior (2008) 3, 143-151

ELSEVIER

Journal of Veterinary Behavior
Clinical Applications and Research

RESEARCH

Factors associated with success in guide dog training

Lara S. Batt, Marjolyn S. Batt, John A. Baguley, Paul D. McGreevy

Faculty of Veterinary Science, University of Sydney, NSW, Australia.

KEYWORDS:
guide dog;
dog;
temperament;
motor lateralization;
cortisol

Abstract Tests of motor laterality and behavioral reactivity, as well as salivary cortisol concentrations, were examined in this pilot study to identify dogs best suited to guide dog work. Over a 14-month period, lateralization tests were conducted and cortisol concentrations were determined on 3 separate occasions, and temperament testing was performed on 2. Potential guide dogs (N = 43) involved in this study were 5 golden retrievers (4 males, 1 female) and thirty-eight Labrador retrievers (8 black males, fifteen yellow males, 5 black females, and ten yellow females). Results from these tests were then compared with the ultimate success of the dogs in the Guide Dogs NSW/ACT training program. This comparison produced evidence that motor lateralization (particularly the rate at which both paws were used during the Kong Test and the lateralization index during the Tape Test), reactions to an unfamiliar dog, the latency for dogs to drop and rest during an uninterrupted period, and the dog's color and breed were predictive of ultimate success. This study also identified 14 months of age as a more accurate time to assess dogs for these traits than either 6 months of age or at the age at which they completed their training (ranging from 14 to 20 months of age).

Introduction

Over the past 30 years, a variety of studies have been conducted to investigate various aspects of guide dog selection and training. These studies have largely focused on the selection and training of dogs (Peel, 1975; Weiss and Greenberg, 1997; Wilsson and Sundgren, 1997; Coppinger et al., 1998; Rooney et al., 2004), genetics and breeding programs (Goddard and Beilharz, 1974; Goddard and Beilharz, 1982; Goddard and Beilharz, 1983; Goddard and Beilharz, 1984b), the raising of pups (Koda, 2001a; Koda, 2001b; Serpell and Hsu, 2001; Kikkawa et al., 2005), and predicting and assessing dogs for their suitability for guide dog programs (Goddard and Beilharz, 1984a; Goddard and Beilharz, 1986; Knol et al., 1988; Murphy, 1995; Vincent and Leahy, 1997; Murphy, 1998; Kikkawa, et al., 2005). Although these studies have produced interesting findings, they have failed to assess the various factors simultaneously. Therefore, a study that investigated the effect of several factors within one trial may further add to the industry's understanding of the factors that influence guide dog success. Some of the areas of greatest interest include the relationships between guide dog success and results from temperament and lateralization tests as well as cortisol profiles.

Temperament tests have been used by many studies to determine if dogs' behavioral traits make them suitable for guiding (Goddard and Beilharz, 1984a; Goddard and Beilharz, 1984b; Goddard and Beilharz, 1986; Knol, et al., 1988; Murphy, 1995), re-homing from shelters (van der Borg et al., 1991; Ledger et al., 1995; Ledger and Stephen, 2004; De Palma et al., 2005) or police (or other service dog) work (Weiss and Greenberg, 1997; Wilsson and Sundgren, 1997; Slabbert and Odendaal, 1999). Goddard and Beilharz

Address for reprint requests and correspondence: Lara S. Batt, Faculty of Veterinary Science (B19), University of Sydney, NSW 2006, Australia; Phone: +612403988554
E-mail: larab@vetsci.usyd.edu.au

1558-7878
doi:10.1016/j.jveb.2008.04.003

(1984a; 1986) conducted temperament tests on 102 dogs of different breeds and found that these tests were predictive of fearfulness (shown best by avoidance responses) and that the predictive ability of the tests was greater when applied to older dogs. Knol et al. (1988) examined the use of temperament tests in a number of guide dog organizations and reported that, although temperament tests were commonly used, they had not been subjected to rigorous testing or peer-reviewed scrutiny. In the study by Serpell and Hsu (2001), temperament was assessed by administering questionnaires to 1097 puppy raisers to determine how the pup behaved in a variety of contexts. They reported that this method of assessment was highly predictive of success and failure.

Motor lateralization tests in dogs have shown that ambilateral dogs (ie, those without a significant left or right paw preference) or those with weak lateral strength (those with a low absolute value of the lateralization index |LI|) are more likely to show signs of distress in response to thunder, fireworks (Branson and Rogers, 2006), and loud metallic noises (unpublished data). Lateralization in other species also has been associated with differences in social status, with right-handed macaques being less prone to attacks by conspecifics and spending more time in proximity with conspecifics than their left-handed counterparts (Westergaard et al., 2003). Right-handed macaques were also found to be less reactive during social and spatial restriction (Westergaard et al., 2001), and right-handed marmosets were also found to be more bold, entering unfamiliar rooms quicker and touching more objects than their left-handed counterparts (Cameron and Rogers, 1999). Given that high reactivity and distress responses are undesirable traits in guide dogs, lateralization tests have potential predictive value for guide dog success or failure.

Cortisol concentrations have been studied extensively in dogs (Beerda et al., 1996; Hennessy et al., 1997; Beerda et al., 1998; Hennessy et al., 2001; Bergeron et al., 2002; Kobelt et al., 2003; Hiby et al., 2006). However, although IgA (Kikkawa, et al., 2005) and heart rate (Vincent and Leahy, 1997; McGreevy et al., 2005) have been studied in guide dogs, the authors could find only 1 paper that examined cortisol in potential guide dogs (unpublished data). Elevated concentrations are often associated with distress responses (Beerda et al., 1999; King et al., 2003; Hydbring-Sandberg et al., 2004; Jones and Josephs, 2006), and given that Goddard and Beilharz (1984a) found that fear was the primary reason for potential guide dogs failing, the authors hypothesized that elevated cortisol concentrations will be associated with fear and therefore failure in potential guide dogs. Haubenhofer et al. (2005) found that dogs that underwent intensive training to become therapy dogs did not show elevated (salivary) cortisol concentrations. However, Haubenhofer and Kirchengast (2006) found that the therapy work itself did result in an elevation of cortisol concentrations.

Within the guide dog industry, the influence of coat color on success is often discussed in anecdotal terms. However, the authors could find no published empirical evidence to suggest that this factor contributed to a dog's ability to be used as a guide. Because of the lack of empirical evidence, this study incorporated the factor of coat color into the analysis to help determine whether it influenced guide dog success.

This is a pilot study designed to identify whether measures of temperament, lateralization, cortisol activity, or color can be used to determine the probability of guide dog success.

Materials and methods

Subjects

Potential guide dogs (N = 43) were randomly selected at 6 months of age from a pool of 105 trainee dogs. Of these 43 dogs, 5 were golden retrievers (GR) (4 males, 1 female) and 38 were Labrador retrievers (LR) (8 black males, 15 yellow males, 5 black females and 10 yellow females). All dogs were raised in family homes as part of a puppy-walking program run by Guide Dogs NSW/ACT (2-4 Thomas Street, Chatswood, Australia) until approximately 14 months of age. At 6 months of age, dogs were brought into the association's training center for neutering, and it was at this stage that the first round of testing took place. Dogs were sampled for saliva, and the temperament and lateralization tests were conducted. These tests were repeated again at approximately 14 months of age, when they returned to the center to commence their training. A third series of lateralization tests were conducted and saliva samples were taken between 14 and 20 months of age after the trainers had determined whether or not the dogs were suitable for guiding.

Temperament tests

The temperament test took approximately 25 minutes to complete and consisted of 6 components: Social Contact Test; Passive Test; Chase Test; Noise Test; Dog Distraction Test; and the Sudden Appearance Test. Each of these tests was conducted in an objective manner and is described in detail below.

Social contact test

This test was based on the social contact component outlined in the temperament test by Svartberg (2002). This study differed from that previously published, as it measured the time for the dogs to approach rather than describing the dog's behavior. In the current study, an assistant brought the dog from the kennels to the testing room (approximately 5 m by 6 m). To reduce any threat that the dog may have perceived, the tester was waiting in the room, turned to the side, and reading a book to avoid eye contact. The assistant entered the room, closed the door, and released the dog. The latency for the dog to approach the tester after the leash was removed was recorded. If the dog did not approach immediately, the tester continued reading (without speaking to or making eye contact with the dog).

Passive test

This test was based on the passive component of the temperament test outlined by Svartberg (2002). This study differed from that previously published, as it measured the latency of the dog to sit, lie down, and rest (place its head on the ground) rather than describing the dog's level of activity during that period. In keeping with Svartberg (2002), dogs in the current study were leashed for the duration of this test. However, the time allowed for this test in the current study was extended from 3 minutes to 6 minutes. The tester sat and read (from a text on her lap, to avoid eye contact) for the duration of the test while the assistant observed from an adjoining room and recorded the aforementioned latencies of interest.

Chase test

This test was based on the Chase Test component outlined by Svartberg (2002) and the Novel Object Test conducted by King et al. (2003). Similar to the test conducted by King et al. (2003), the current study measured the latency for the dog to catch the object and the duration for which it was held once caught (rather than describing the dog's reactions to it). During this test, the leash was removed from the dog and the tester placed a "rolling ferret" (a battery-operated ball that moves in random patterns with a feather boa tail attached (Dah Yang Toy Industrial Co., Ltd., China) on the floor. The assistant remained in the adjoining room and recorded the latency for the dog to catch the ferret and the duration for which the dog held the ferret, once caught. A time limit of 5 minutes was placed on the test.

Noise test

This test was based on a combination of the play and metallic noise tests developed by Svartberg (2002); the Bowl test described in Goddard and Beilharz (1984a), which involved dropping an aluminum food bowl behind the dog; and the Loud Noise Test from Wilsson and Sundgren (1997). In the current study, the tester engaged the dog in play with a squeaky ball. Once the dog was engaged in the game (ie, fetching or playing chase with the ball, wagging, or play bowing) the assistant (still in the adjoining room) dropped a metal plate from a height of 1 meter onto a concrete floor. The latency for the dog to resume the game with the tester was recorded. This was repeated on 5 occasions.

Dog distraction test

This test was broadly based on that described by van der Borg et al. (1991), except that it did not involve any confrontation between the dogs and used a consistent unfamiliar dog throughout. This test occurred outside the testing area used for the other challenges. The dog to be tested was attached to a vertical post by a 2.5-meter lead. The assistant walked with an unfamiliar dog (the same dog was used consistently throughout all the tests) past the dog being tested at a distance of 5 meters from the post. If the dog being tested attempted to approach, it could get within 2.5 meters of the unfamiliar dog, but it could get as far as 7.5 meters away from the dog if it traveled in the opposite direction. The unfamiliar dog was a spayed (desexed) black female Labrador-cross that was trained to walk along a designated path with the handler without making eye contact with those dogs being tested. Records from this test involved simply ticking boxes to indicate if the test dog wagged its tail, pulled on the leash toward the unfamiliar dog, moved away from the unfamiliar dog (avoid), whined, barked, growled, raised hackles, adopted a lowered body posture, or bared teeth.

Sudden appearance test

This test was based on the Sudden Appearance Test developed by Svartberg (2002), the Startling Test by King et al. (2003), subtest 29 from Netto and Planta (1997) and Startling Test 1 from Wilsson and Sundgren (1997). Like the test conducted by King et al. (2003), the latencies to approach (rather than descriptions of the dog's reactions) were recorded. The test was conducted in a 2 meter x 5 meter fenced run adjacent to the testing area. The dog entered the run with the tester, who then shut the gate. Lying on the ground was a mask hinged to a meter-length of timber (1000 mm x 300 mm x 10 mm) with a broad base (500 mm x 500 mm) that was held flat with 2 weights, each measuring 2.5 kg. Once in the enclosed area, the tester held the dog's leash until it looked toward the device lying on the ground. At this time, the tester erected the device by pulling on the string (attached to the end of the timber furthest from the hinge) and simultaneously dropped the dog's leash. The latency for the dog to approach the mask (2.5 m away) was recorded by the assistant, who remained outside the enclosed area filming. If the dog had not sniffed/touched the mask from within a distance of 100 mm within 1 minute, the tester approached the mask and knelt beside it. If, after another minute, the dog still had not approached, the tester would touch the mask and encourage the dog to approach.

Lateralization tests

Two paw-preference tests were administered to all dogs during the 3 testing periods.

Tape test

This test was based on that described by Quaranta et al. (2004). The methods involved in the collection of these data are detailed in Batt et al. (2007). In this test, PVC duct tape (manufactured by Cling Adhesive products, Moorebank, NSW, Australia) measuring 15 mm x 50 mm was applied to the midline dorsal surface of the dog's nasal planum. The paw used during each attempt to remove the tape was recorded directly. In the study by Quaranta et al. (2004), testing was conducted over a 2-minute period. The current study deviated from method by collecting a total of 26 observations rather than setting a limit of 2 minutes for the task. If the dog removed the tape prior to the recording of 26 attempts, a new piece of tape was applied and the process continued.

Kong test

This test was based on that described by Branson et al. (2003) and Branson and Rogers (2006). The current study modified the content of the Kong to accommodate the appetite of the LRs and GRs. The mixture used to fill the Kongs and the methods involved in data collection are detailed in Batt et al. (2007) and involved filling the Kong with a mixture of highly palatable foods prior to freezing it. Unimanual paw use to "hold" the Kong steady was recorded directly (ie, during testing) until a total of 100 data points were obtained. Data on the use of both paws (B%) has been reported by Branson and Rogers (2006), and the current study was modified after initial data collection to permit comparisons with this work. The authors retrospectively recorded the number of times the dogs were observed using both paws simultaneously and divided this number by the amount of time for which valid footage ran. Video footage was counted as valid when both paws could be seen or when the Kong was completely visible. This method produced the rate of usage of both paws. Unfortunately, the data available on the videotapes were limited as the dog's body obscured the camera's view or the dog was not 'in shot.' This occurred for an average of 72.10% ± 1.81 of the time that it took for the dog to complete the Kong Test.

Salivary cortisol

During Tests 1 and 2, saliva samples for measurement of pretest cortisol concentration were collected after the dogs had completed the Social Contact Test component of the temperament testing (ie, after the dog entered the testing facility and approached the tester), and the posttest sample was collected between the Tape and Kong Tests. In contrast, during Test 3, there was a slight change in the procedures in that the pretest saliva sample was collected prior to the commencement of testing. Nevertheless, saliva samples were still collected after the dog had entered the facility and approached the tester.

Samples were collected from each dog by swabbing the mouth with a cotton pad (Swispers make-up pads, Accantia Health and Beauty Pty, Ltd, Auckland, New Zealand). The pad was then placed in a 10-mL collection tube with a small piece of drinking straw (approximately 10 mm in length) in the bottom to facilitate the separation of saliva from the swab. Samples were stored for up to 8 hours on ice. If the dogs produced a seemingly insufficient amount of saliva for analysis, they were stimulated to salivate by being offered the opportunity to sniff at a bag containing dog food. To the authors' knowledge, this has not been done previously. However, given that the dogs did not obtain the food prior to the collection of saliva and that the flow rate of saliva does not affect cortisol concentrations (Kirschbaum and Hellhammer, 1994; Aardal and Holm, 1995), it was not anticipated that this would influence the results. The samples were transported to the laboratory, where the saliva and swabs were separated by centrifuging at 3000 rpm for 15 minutes. The swab and straw were removed and discarded, and the saliva was transferred into a 1.5 mL Eppendorf tube and stored at –80°C until further analysis was conducted. Cortisol was assayed, using the salivary cortisol DSL-10-67100 enzyme immunoassay (Diagnostic Systems Laboratory Inc; Gladesville, Australia) which has been previously validated for use on dog samples by Coppola et al. (2006) and Jones and Josephs (2006).

Determining success and failure

Guide dog trainers began assessments of dogs when they returned to the Glossodia training center (NSW, Australia) at approximately 14 months of age. Dogs could be failed at any stage of the 20-week training program for a number of reasons. Reasons for failure included health, anxiety, body sensitivity, poor concentration, dog distraction, dominance, excitability, pulling, scent distraction, submissiveness, suspicion, temperament, toileting or "other." With the exception of health, these reasons were subjective and the trainers applied them at their own discretion. Although attempts at definitions of these reasons are available (Harrison, 2006, pp 347-374), their application to individual dogs is likely to differ among trainers. In the sample of dogs used for this study, 19 passed and 25 failed. Of those that failed, 4 failed for dog distraction, 6 failed for anxiety, 4 failed for excitability, 4 failed for health, 4 failed for suspicion, and 3 failed for body sensitivity. Given that there are such low numbers for each of these reasons for failure, there were insufficient data to examine individually the reasons for failure; all reasons for failure were pooled, with the exception of dogs that failed for health. Those failing for health concerns were removed from the study, because the current measures were not designed to identify these problems.

Statistical analysis

Results from both the temperament and lateralization tests were compared with the dogs' ultimate success in the program, a binary outcome (the manager's decision that a dog had successfully completed training or not). Results from Tests 1, 2, and 3 were separated and analyzed independently to determine the predictors of success and failure at each stage of testing. A logistic regression, starting with a full model and using backward elimination, was performed. Given that this is a pilot study, significance was set at $P =$ .10 to provide an indication of which factors were worthy of further investigation. A pseudo R^2 was calculated by dividing the mean deviance by the total deviance to estimate the accuracy with which the predictions could be made. All variables were assessed for normality using an inspection of the residuals, and Genstat 9th edition (VSN International Ltd, Hemel Hempstead, UK) was used for all analyses. The

Table 1 Steps of the logistic regression during the first testing phase (approximately 6 months of age) to determine significant contributors to guide dog success (n=43)

Stage	Variables included in model	P value	Variables removed from the model	Variable remaining in the model
Stage 1	Jumping up (Dog Distraction Test)	**0.005**		✓
	Wag (Dog Distraction Test)	0.540	✓	
	Bark (Dog Distraction Test)	**0.072**		✓
	Strength of lateralization in Kong Test (|LI|)	0.696	✓	
	Strength of lateralization in Tape Test (|LI|)	0.549	✓	
	Latency to Drop (Passive Test)	**0.002**		✓
	Latency to Rest (Passive Test)	<**0.001**		✓
	Latency to Catch (Chase Test)	0.910	✓	
Stage 2	Latency to Rest (Passive Test)	**0.005**		✓
	Latency to Drop (Passive Test)	**0.002**		✓
	Jumping up (Dog Distraction Test)	**0.027**		✓
	Bark (Dog Distraction Test)	0.242	✓	
	Duration of hold (Chase Test)	0.933	✓	
	Occurrence of whining (Dog Distraction Test)	0.883	✓	
	Latency to sit (Passive Test)	0.902	✓	
	Latency to approach tester (logged) (Social Contact Test)	0.287	✓	
Stage 3	Latency to Rest (Passive Test)	**0.062**		✓
	Latency to Drop (Passive Test)	**0.056**		✓
	Jumping up (Dog Distraction Test)	**0.003**		✓
	Degree of lateralization in Tape Test (LI)	0.158	✓	
	Sex of dogs	0.112	✓	
	Degree of lateralization in Kong Test (LI)	0.702	✓	
	Post cortisol concentrations (logged)	0.153	✓	
	Pre cortisol concentrations (logged)	0.802	✓	
Stage 4	Latency to Rest (Passive Test)	0.005	✓	
	Latency to Drop (Passive Test)	0.024	✓	
	Jumping up (Dog Distraction Test)	0.002	✓	
	Latency to recover from Noise 1	**0.340**		✓
	Latency to recover from Noise 2	0.041	✓	
	Latency to recover from Noise 3	**0.254**		✓
	Latency to recover from Noise 4	**0.111**		✓
	Latency to recover from Noise 5	**0.487**		✓
	Coat color	**0.404**		✓
Stage 5	Latency to Rest (Passive Test)	0.008	✓	
	Latency to Drop (Passive Test)	0.012	✓	
	Jumping up (Dog Distraction Test)	<0.001	✓	
	Latency to recover from Noise 2	**0.414**		✓
	Occurrence of Pulling (Dog Distraction Test)	**0.125**		✓
	Rate of both paw usage (Kong Test)	0.075	✓	
	Latency to recover from Noise Tests (average)	**0.839**		✓
Stage 5	Latency to Rest (Passive Test)	0.006	✓	
	Latency to Drop (Passive Test)	0.005	✓	
	Jumping up (Dog Distraction Test)	0.005	✓	
	Rate of both paw usage (Kong Test)	**0.189**		✓

regression analysis was conducted in stages, as there were a large number of observations that needed to be included.

Lateralization index (LI) (often referred to as the degree of lateralization) was obtained using the equation [(R-L)/(R+L)] x 100 (where L represents the number of times the left paw was used and R represents the number of times the right paw was used). The strength of the bias was expressed as laterality index regardless of its direction (ie, without a sign, |LI|). In addition, a binomial *z* test was used to categorize animals as Left (L), Right (R) or Ambilateral (A) as used by Branson and Rogers (2006). Animals with a *z* score < -1.96 were classified as left pawed, and those with a *z* score > 1.96 were classified as right pawed. Animals with a *z* score between –1.96 and +1.96 were classified as ambilateral.

Table 2 Steps of the logistic regression during the second testing phase (approximately 14 months of age) to determine significant contributors to guide dog success (n=43)

Stage	Variables included in model	P value	Variables removed from the model	Variable remaining in the model
Stage 1	Jumping up (Dog Distraction Test)	NA – no dogs jumped	✓	
	Wag (Dog Distraction Test)	0.497	✓	
	Bark (Dog Distraction Test)	0.986	✓	
	Strength of lateralization in Kong Test (|LI|)	0.620	✓	
	Strength of lateralization in Tape Test (|LI|)	0.679	✓	
	Latency to Drop (Passive Test)	0.396	✓	
	Latency to Rest (Passive Test)	0.377	✓	
	Latency to Catch (Chase Test)	0.616	✓	
Stage 2	Duration of hold (Chase Test)	0.606	✓	
	Occurrence of whining (Dog Distraction Test)	0.828	✓	
	Latency to sit (Passive Test)	0.656	✓	
	Latency to approach tester (logged) (Social Contact Test)	0.269	✓	
Stage 3	Latency to recover from Noise 1	0.111	✓	
	Latency to recover from Noise 2	**0.012**		✓
	Latency to recover from Noise 3	**0.087**		✓
	Latency to recover from Noise 4	0.140	✓	
	Latency to recover from Noise 5	0.163	✓	
Stage 4	Latency to recover from Noise 2	**0.035**		✓
	Latency to recover from Noise 3	0.965	✓	
	Sex of dogs	**0.019**		✓
	Degree of lateralization in Kong Test (LI)	0.607	✓	
	Post cortisol concentrations (logged)	0.421	✓	
	Pre cortisol concentrations (logged)	0.309	✓	
Stage 5	Latency to recover from Noise 2	0.996	✓	
	Sex of dogs	0.997	✓	
	Occurrence of Pulling (Dog Distraction Test)	**0.004**		✓
	Rate of both paw usage (Kong Test)	<**0.001**		✓
	Latency to recover from Noise Tests (average)	0.990	✓	
	Degree of lateralization in Tape Test (LI)	**0.008**		✓
	Coat colour	**1.000**		✓
Stage 6	Degree of lateralization in Tape Test (LI	**0.001**		✓
	Rate of both paw usage (Kong Test)	<**0.001**		✓
	Occurrence of Pulling (Dog Distraction Test)	<**0.001**		✓

The "rate of both" used for the Kong Test was determined by dividing the number of times the dogs were observed using both paws simultaneously by the period for which the dogs were visible. Video footage was counted as valid when both paws could be seen or when the Kong was completely visible (i.e. the dog was not obscuring the Kong).

Results

During Test 1, 3 factors were found to significantly influence the probability of success. These were the latency to drop (Passive Test) (estimate = -0.014, s.e. = 0.006, P = .005), latency to rest (Passive Test) (estimate = 0.014, s.e. = 0.006, P = .006), and the occurrence of jumping (Dog Distraction Test) (estimate = -10.9, s.e. = 20.7, P = .005). A shorter latency to drop, a greater latency to rest, and the absence of jumping maximized the probability of passing. This method predicted the outcome with an accuracy of 27.13%. The equation for predicting probability of success (as determined from the estimate values) was: $\frac{1}{1+1/(0.459\times 0.986^{Drop}\times 1.014^{Rest}\times 0.00002^{Jump})}\times 100$. The steps taken during the regression are outlined in Table 1.

During Test 2, 3 factors were found to significantly influence the probability of success. These were the LI from the Tape Test (estimate = 2.28, s.e. = 9.5, P = .001), the rate at which both paws were used during the Kong Test (estimate = -410, s.e. = 1677, $P < .001$), and the occurrence of pulling on the lead during the Dog Distraction Test (estimate = -165, s.e. = 679, $P < .001$). A higher LI during the Tape test, a lower rate of both paw usage during the Kong Test, and a lack of pulling during the Dog Distraction Test maximized the probability of passing. This method predicted the outcome with an accuracy of 100%

Table 3 Steps of the logistic regression during the third testing phase (approximately 14-20 months of age) to determine significant contributors to guide dog success (n=43)

Stage	Variables included in model	P value	Variables removed from the model	Variable remaining in the model
Stage 1	Strength of lateralization in Kong Test (|LI|)	0.984	✓	
	Strength of lateralization in Tape Test (|LI|)	0.501	✓	
	Degree of lateralization in Tape Test (LI)	0.502	✓	
	Sex of dogs	0.189	✓	
	Degree of lateralization in Kong Test (LI)	0.984	✓	
	Coat color	**0.032**		✓
	Rate of both paw usage (Kong Test)	**0.055**		✓
Stage 2	Coat color	**0.02**		✓
	Rate of both paw usage (Kong Test)	**0.023**		✓

from the pseudo R^2. The equation for predicting probability of success (as determined from the estimate values) was: $\frac{1}{1+1/(3.598\times10^{85}\times2.195\times10^{-72^{Pull}}\times9.777^{TapeLI}\times8.695\times10^{-179^{RateofBoth}})}\times100$. The steps taken during the regression are outlined in Table 2.

During Test 3, 2 factors were found to significantly influence the probability of success. These were the rate at which both paws were used during the Kong Test (estimate = -1.416, s.e. = 0.756, P = .023), and the dog's color (for black dogs, estimate = -1.97, s.e. = 1.13; for yellow dogs, estimate = -0.34, s.e. = 0.744; and for golden retrievers, estimate = -10.27, s.e. = 23.9, P = .02). A shorter latency to drop, a greater latency to rest and the absence of jumping maximized the probability of dogs successfully completing their training to become a guide dog. This method predicted the outcome of results with an accuracy of 32.44% from the pseudo R^2. The equation for predicting probability of success (as determined from the estimate values) was: $\frac{1}{1+1/(0.1395\times4.1206^{RateofBoth})}\times100$ for black LRs, $\frac{1}{1+1/(0.7118\times4.1206^{RateofBoth})}\times100$ for yellow LRs, and $\frac{1}{1+1/(0.00003\times4.1206^{RateofBoth})}\times100$ for GRs. The steps taken during the regression are outlined in Table 3.

Discussion

Previous studies have examined the relationship between lateralization, temperament, and cortisol tests as well as the reliability of the observations and the stability of traits over time and across sex, color, and breed (unpublished data). Comparing our results from these tests with the training outcomes reported by Guide Dogs NSW/ACT has enabled us to estimate the tests' predictive merit. Jones and Gosling (2005) suggest that, in general, creators of canine temperament tests have tended to overlook the need to demonstrate that their tests truly reflect behavioral tendencies. Taylor and Mills (2006) added that such validation was best achieved by comparing test results with a definitive measure or outcome. However, they acknowledged that reliable, definitive training or working outcomes are rare. In the current study, the training outcome used to validate our test results was the dogs' performance at the culmination of the guide dog training program. Unfortunately, this measure cannot be considered definitive, as it includes subjective assessment by different trainers (Murphy, 1998) and undoubtedly reflects factors such as the supply and demand for guide dogs (Murphy, 1995). That said, it was the most relevant outcome available for validation. Greater dog numbers would have enabled the reasons for failure to be assessed in more detail. Nevertheless, the current pilot study has identified a suite of tests that are predictive of overall failure.

Although many failures may be caused by fearfulness, a study that involved more dogs and a more objective system of describing reasons for failure would clarify the predictions that can be drawn from the proposed suite of tests. Although the current study has identified both lateralization tests (Kong and Tape Tests) as being worth pursuing, it should be remembered that some inconsistencies in the use of the Tape Test can exist (unpublished data). These inconsistencies should prompt caution with the use of the Tape Test and suggest that the Kong Test is preferred and, furthermore, has been used in more canine ethology laboratories than the Tape Test. Other predictive measures of success include the latency of the dogs to drop and rest during the Passive Test, the occurrence of pulling and jumping during the Dog Distraction Test, as well as the dogs' physical characteristics. Using only those tests with the most predictive merit and focusing on dogs of 1 breed and color makes a follow-up study more feasible (Taylor and Mills, 2006). This method would then enable more dogs to be tested in a given period.

The current study has also identified 14 months of age (when the dogs return for their assessment and training) as a more suitable time for testing than 6 months. This may be because the dogs are closer to social maturity; social maturity is achieved at 12-36 months of age (Overall and Dunham, 2002), and behavioral traits are likely to be more stable than at 6 months (just prior to having reached sexual maturity at 6-9 months of age (Case, 1999). This finding aligns with those of Goddard and Beilharz (1984a), who found that the accuracy of predictions increased with the dog's age, as one would expect.

The final group of tests was conducted after the dogs had passed or failed the program. However, success and failure were not identified at a single point in the training process. Dogs could be failed at any stage during the 20-week training program. Consequently, our assessments of dogs after official determination of their success or failure was influenced by the period they had spent in the program (and so the amount of training they had received). It may be for this reason that tests conducted at the completion of assessment and training (ranging from 14 to 20 months of age) were less indicative of success than at 14 months of age.

In addition, it would be impractical to use the results of tests conducted after 14 months of age, as doing so would convey no advantage to the guide dog organization (since their training and assessments would have already commenced). The strategic timing of tests may differ in other guide dog organizations, for example, if the age at which dogs entered the program or the period for which they were trained differed considerably. Results of this study (and for this guide dog school) indicate that 14 months of age was the most appropriate time for conducting temperament and lateralization tests as predictors of success. This study has assessed the predictive ability of a range of tests to detect success in guide dogs. Specifically, the Passive Test, the Kong and Tape Tests, and the Dog Distraction Test show particular merit as being predictive of guide dog success. The authors recommend conducting further studies into these behavioral domains using a larger cohort of dogs.

Acknowledgments

The authors would like to thank Dr John Black and the team at Guide Dogs NSW/ACT Glossodia training center, ans Lisa Tomkins and Sylvia Johnson for their help with and support of this study, which was funded by Guide Dogs NSW/ACT and the Australian Research Council.

References

Aardal, E., Holm, A.C., 1995. Cortisol in saliva—reference ranges and relation to cortisol in serum. Eur. J. Clin. Chem. Clin. Biochem. 33, 927-932.

Batt, L., Batt, M., McGreevy, P., 2007. Two tests for motor laterality in dogs. J. Vet. Behav.–Clin. Appl. Res. 2, 47-51.

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

Beerda, B., Schilder, M.B.H., Janssen, N.S.C.R.M., Mol, J.A., 1996. The use of saliva cortisol, urinary cortisol, and catecholamine measurements for a noninvasive assessment of stress responses in dogs. Horm. Behav. 30, 272-279.

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

Bergeron, R., Scott, S.L., Emond, J.P., Mercier, F., Cook, N.J., Schaefer, A.L., 2002. Physiology and behavior of dogs during air transport. Can. J. Vet. Res.-Rev. Can. Rech. Vet. 66, 211-216.

Branson, N.J., Rogers, L.J., 2006. Relationship between paw preference strength and noise phobia in *Canis familiaris*. J. Comp. Psychol. 120, 176-183.

Branson, N.J., Rogers, L.J., Kaplan, G.L., Johnson, D.L., 2003. The impact of brain lateralization on canine behavioural disorders: preliminary results. In: Seksel, K., Perry, G., Mills, D., Frank, D., Lindell, E., McGreevy, P.D. (Eds.), Proceedings of the 4th International Veterinary Behaviour Medicine, pp. 221-224.

Cameron, R., Rogers, L.J., 1999. Hand preference of the common marmoset (Callithrix jacchus): Problem solving and responses in a novel setting. J. Comp. Psychol. 113, 149-157.

Case, L.P., 1999. The Dog: Its Behaviour, Nutrition and Health. Iowa State University Press, Ames, Iowa.

Coppinger, R., Coppinger, L., Skillings, E., 1998. Observations on assistance dog training and use. J. Appl. Anim. Welf. Sci. 1, 133-144.

Coppola, C.L., Grandin, T., Enns, R.M., 2006. Human interaction and cortisol: Can human contact reduce stress for shelter dogs? Physiol. Behav. 87, 537-541.

De Palma, C., Viggiano, E., Barillari, W., Palme, R., Dufour, A.B., Fantini, C., Natoli, E., 2005. Evaluating the temperament in shelter dogs. Behaviour. 142, 1307-1328.

Goddard, M.E., Beilharz, R.G., 1974. A breeding programme for guide dogs. In: 1st World Congress on Genetics applied to Livestock Production; 371-376.

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., 1984a. A factor-analysis of fearfulness in potential guide dogs. Appl. Anim. Behav. Sci. 12, 253-278.

Goddard, M.E., Beilharz, R.G., 1984b. 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., 1986. Early prediction of adult behavior in potential guide dogs. Appl. Anim. Behav. Sci. 15, 247-260.

Harrison, T.R., 2006. Guide Dogs Current Practice Volume 1. Guide Dogs NSW/ACT, Sydney.

Haubenhofer, D., Mostl, E., Kirchengast, S., 2005. Cortisol concentrations in saliva of humans and their dogs during intensive training courses in animal-assisted therapy. Wien. Tierarz. Monats. 92

Haubenhofer, D.K., Kirchengast, S., 2006. Physiological arousal for companion dogs working with their owners in animal-assisted activities and animal-assisted therapy. J. Appl. Anim. Welf. Sci. 9

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

Hennessy, M.B., Voith, V.L., Mazzei, S.J., Buttram, J., Miller, D.D., Linden, F., 2001. Behavior and cortisol levels of dogs in a public animal shelter, and an exploration of the ability of these measures to predict problem behavior after adoption. Appl. Anim. Behav. Sci. 73, 217-233.

Hiby, E.F., Rooney, N.J., Bradshaw, J.W.S., 2006. Behavioural and physiological responses of dogs entering re-homing kennels. Physiol. Behav. 89, 385-391.

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

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

Jones, A.C., Josephs, R.A., 2006. Interspecies hormonal interactions between man and the domestic dog (*Canis familiaris*). Horm. Behav. 50, 393-400.

Kikkawa, A., Uchida, Y., Suwa, Y., Taguchi, K., 2005. A novel method for estimating the adaptive of guide dogs using salivary sIgA. J. Vet. Med. Sci. 67, 707-712.

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

Kirschbaum, C., Hellhammer, D.H., 1994. Salivary cortisol in psychoneuroendocrine research: recent developments and applications. Psychoneuroendocrinology. 19, 313-333.

Knol, B.W., Roozendaal, C., Vandenbogaard, L., Bouw, J., 1988. The suitability of dogs as guide dogs for the blind: Criteria and testing procedures. Vet. Q. 10, 198-204.

Kobelt, A.J., Hemsworth, P.H., Barnett, J.L., Butler, K.L., 2003. Sources of sampling variation in saliva cortisol in dogs. Res. Vet. Sci. 75, 157-161.

Koda, N., 2001a. Development of play behavior between potential guide dogs for the blind and human raisers. Behav. Processes. 53, 41-46.

Koda, N., 2001b. Inappropriate behavior of potential guide dogs for the blind and coping behavior of human raisers. Appl. Anim. Behav. Sci. 72, 79-87.

Ledger, R.A., Baxter, M., McNicholas, J., 1995. Temperament testing dogs in a rescue shelter: Improving owner-dog compatibility. In: Rutter, S.M., Rushen, J., Randle, H.D., Eddison, J.C. (Eds), 29th International Congress of the ISAE; 101-102.

Ledger, R.A., Stephen, J.M., 2004. Reducing dog return rates at rescue shelters: applying science for animal welfare. Anim. Welf. 13, S247.

McGreevy, P.D., Righetti, J., Thomson, P.C., 2005. The reinforcing value of physical contact and the effect on canine heart rate of grooming in different anatomical areas. Anthrozoos. 18, 236-244.

Murphy, J.A., 1995. Assessment of the temperament of potential guide dogs. Anthrozoos. 8, 224-228.

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

Netto, W.J., Planta, D.J.U., 1997. Behavioural testing for aggression in the domestic dog. Appl. Anim. Behav. Sci. 52, 243-263.

Overall, K.L., Dunham, A.E., 2002. Clinical features and outcome in dogs and cats with obsessive-compulsive disorder: 126 cases (1989-2000). J. Am. Vet. Med. Assoc. 221, 1445-1452.

Peel, B.W., 1975. The training of guide dogs for the blind. NZ Vet. J. 23, 269-272.

Quaranta, A., Siniscalchi, M., Frate, A., Vallortigara, G., 2004. Paw preference in dogs: relations between lateralised behaviour and immunity. Behav. Brain Res. 153, 521-525.

Rooney, N.J., Bradshaw, J.W.S., Almey, H., 2004. Attributes of specialist search dogs: A questionnaire survey of UK dog handlers and trainers. J. Forensic Sci. 49, 300-306.

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.

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.

Svartberg, K., 2002. Shyness-boldness predicts performance in working dogs. Appl. Anim. Behav. Sci. 79, 157-174.

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

van der Borg, J.A.M., Netto, W.J., Planta, D.J.U., 1991. Behavioural testing of dogs in animal shelters to predict problem behaviour. Appl. Anim. Behav. Sci. 32, 237-251.

Vincent, I.C., Leahy, R.A., 1997. Real-time non-invasive measurement of heart rate in working dog: A technique with potential applications in the objective assessment of welfare problems. Vet. J. 153, 179-183.

Weiss, E., Greenberg, G., 1997. Service dog selection tests: Effectiveness for dogs from animal shelters. Appl. Anim. Behav. Sci. 53, 297-308.

Westergaard, G.C., Chavanne, T.J., Lussier, I.D., Houser, L., Cleveland, A., Suomi, S.J., Higley, J.D., 2003. Left-handedness is correlated with CSF monoamine metabolite and plasma cortisol concentrations, and with impaired sociality, in free-ranging adult male rhesus macaques (Macaca mulatta). Laterality. 8, 169-187.

Westergaard, G.C., Lussier, I.D., Suomi, S.J., Higley, J.D., 2001. Stress correlates of hand preference in rhesus macaques. Dev. Psychobiol. 38, 110-115.

Wilsson, E., Sundgren, P.E., 1997. The use of a behaviour 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.