



Assessment of Canine Temperament: Predictive or Prescriptive?

Mary R. Burch¹

¹ *American Kennel Club*

Historically, canine temperament testing has been linked to the predictability of future behavior. A *predictive* model of canine temperament testing assumes that a dog's behavior in one situation will likely be similar to its behavior in a variety of other situations. An alternative model is proposed for a canine temperament test that could identify areas in which a dog might fail to perform certain test items, but, by using modern behavior analysis techniques, behaviors could be modified through a *prescriptive* approach. This article describes the AKC Temperament Test (ATT), which is the first prescriptive canine temperament test. The ATT is designed to provide pet dog owners with information about potential problem areas that can be modified through training.

Keywords: temperament, temperament tests, canine temperament tests, predictive validity, predictability, prescriptive

The assessment of canine temperament is a growing area of interest to researchers, breeders, dog trainers, animal behaviorists, and dog owners. Although temperament is sometimes described with single words such as make-up, personality, disposition, character, and constitution, there are also more expansive definitions of temperament. As early as 1941, C. S. Hall used a somewhat nonscholarly approach and described temperament as, "the raw stuff of individuality." Mason (1984) defined temperament as the individual's basic stance towards environmental change and challenge. Other researchers also recognized the role of the environment when considering temperament. Ruefenacht et al. (2002) defined temperament as "physical flexibility and intensity of reaction to different environmental stimuli."

Acknowledging that temperament is inherited, Buss (1995) and Goldsmith et al. (1987) described temperament as inherited early tendencies that continue throughout the life of the individual and lay the foundation for personality. For the purpose of conducting a review and evaluation of past research, Jones and Gosling (2005) used the terms "personality" and "temperament" interchangeably, however, they acknowledged the distinctions made by others regarding personality and temperament. For example, research on animals and human infants has referred to "temperament" while research on human children and adults has often used the term "personality" (Jones & Gosling, 2005). John and Gosling (2000) described personality as "personality traits" which are "consistent patterns in the way individuals behave, feel, and think" (John & Gosling, 2000, p. 140).

For the AKC Temperament Test (ATT) program that provides prescriptive materials for remediation if a dog fails a temperament test item, the American Kennel Club defined temperament as "an individual's natural predisposition to react in a certain way to a stimulus. Behaviors related to temperament may be modified over time with exposure and learning" (American Kennel Club, 2019). Diederich and Giffroy (2006) suggested the word "temperament" should only be used when characterizing the dog's behavior as a whole and that traits related to temperament should be described objectively and with respect to science.

Canine Temperament Tests

The importance of assessing canine temperament was described as early as 1910 when Colonel Konrad

Most, a German dog trainer, wrote about the reactions to stimuli and “character” of police dogs, military working dogs, and service dogs for people who were blind (Most, 1954). During World War I (1914 to 1918), as a result of soldiers returning from the war with blindness, the first guide dog school was established in Germany. Soon thereafter, guide dog programs, including the Seeing Eye in 1929 and Guide Dogs for the Blind (1934), began in the United States (International Guide Dog Federation, 2019).

In 1924, a program designed to breed dogs with a desired temperament was implemented at the Fortunate Fields Project in Switzerland. German Shepherd Dogs were bred to become police, herding, and guide dogs with the intent of maintaining working potential along with the temperament traits of courage, adaptability, and judgment (Lindsay, 2001). Similar attention was given to temperament in the 1930s when W. M. Dawson of the U.S. Department of Agriculture advocated breeding dogs not just for their appearance but for temperament traits including intelligence and cooperation with trainers (Dawson, 1937).

The knowledge pertaining to the assessment of canine temperament was significantly advanced in the 1940s to 1960s when John Paul Scott and John L. Fuller conducted what would become the first comprehensive study of canine genetics. In a laboratory setting at the Jackson Labs in Bar Harbor, Maine, Scott and Fuller studied 300 dogs that included both purebred and mixed breed dogs. To study the genetics and social behaviors of the dogs, a test was developed that included items such as social attraction/handling, dominance, retrieving, following, climbing, negotiating barriers, stair-climbing, mazes, motor skills (such as climbing a ladder), obedience tasks such as staying on a platform and walking on a leash, manipulating objects, trailing, and spatial orientation. Temperament traits that were evaluated included confidence and fear (Scott & Fuller, 1965). Lindsay (2001) credited the basic research by Scott and Fuller (1965) as the foundation on which several later temperament tests were developed. Some of these included the testing procedures used by the BioSensor Research Team of the U.S. Army Super Dog Program (Lindsay, 2001) and temperament test items suggested by Michael Fox (1972).

Following the 1965 publication of *Genetics and Social Behavior of the Dog* (Scott & Fuller, 1965), practitioners who included dog trainers, behaviorists, veterinarians, kennel club trainers, and breeders began to develop temperament tests. Initially, many of these tests were thought to be predictive. In a predictive test model, if a dog does poorly on certain test items as a puppy, the dog is determined to be not suitable for certain roles as an adult dog (e.g., stable family pet, police dog, service dog). In temperament tests that were thought to be predictive, puppy characteristics such as being outgoing or timid were thought to remain consistent throughout the dog’s lifetime (Fox, 1972).

The temperament test that brought Scott and Fuller’s work out of the laboratory was developed by Clarence Pfaffenberger for Guide Dogs for the Blind. Pfaffenberger collaborated with John Paul Scott and John Fuller at their Behavior Research Laboratory in Bar Harbor to develop a test that predicted which puppies would make better Guide Dogs than others. The Guide Dogs for the Blind test included practical items such as sit, come, fetch, and heel. Tasks on the test related to guide dog work were staying close to the handler, street crossing, and checking for traffic. Dogs were also scored on body and ear sensitivity and having an intelligent response and willing temperament (Pfaffenberger, 1963).

In 1946, at Guide Dogs for the Blind, trainers placed 109 German shepherds with their guide dog users and found that with these dogs, there was “little more than 9% success” (Pfaffenberger, 1963, p. 84). Although the total number of dogs was not reported, in a subsequent evaluation of nine litters, through selective breeding combined with testing, nearly all (90-95%) of the puppies succeeded in their work as Guide Dogs (Pfaffenberger, 1963). This high rate of success validated the importance of temperament testing for potential guide dogs.

Following the work of Pfaffenberger, *Understanding Your Dog* (Fox, 1972) was one of the first texts to introduce puppy temperament testing to pet dog owners. In *Understanding Your Dog*, Michael Fox wrote,

The basic temperamental characteristics of young pups are the ones that remain with them throughout life. Those pups rated as outgoing, aggressive, passive or timid and dependent at four or six weeks of age tend to conform to these early ratings when fully mature. (Fox, 1972, p. 55).

Fox's test included test items such as handling puppies 1 to 3 weeks of age and assessing the puppy's reaction to mild pain, sensory, and motor tests including isolation and novel stimuli (visual and auditory), sociability, fear of strangers, problem solving, and leash tests. While Fox (1972) advocated the use of temperament tests conducted when puppies were 6 to 8 weeks old for the purpose of matching puppies and new homes, he stated that, "although general temperament can be clearly assessed at six weeks, it is not fully mature in the dog until one or one and a half years" (Fox, 1972, p. 56).

William E. Campbell's *Puppy Behavior Test* (1975) included test items similar to those of Scott and Fuller (1965). The *Puppy Behavior Test* evaluated social attraction, following, restraint dominance, social dominance, and the puppy's reaction to being elevated (Campbell, 1975). While Campbell (1975) differentiated between the ability of a temperament test to predict specific adult canine behavior versus the ability of temperament tests to predict behavioral tendencies (Campbell, 1975), research on the predictive ability of tests similar to Campbell's test did not come until later with studies such as Asher et al. (2013) and Robinson et al. (2016).

Asher et al. (2013) used a temperament test similar to the tests developed by Campbell (1975) and Volhard (1981). In the Asher et al. (2013) study, test items that included retrieving, stroking/stimulus, stroking by tester, response to toy squirrel, and negotiating a ramp were used to evaluate the potential for future success in guide dog training. Responses of the dogs were scored on a 7-point scale for stimuli that included retrieve, gentle restraint, noise, stroking, moving toy squirrel, tunnel, and a ramp. Five of the stimuli (retrieve, stroking by a stimulus, stroking by the assessor, squirrel toy, and ramp) were associated with success in guide dog training. Robinson et al. (2016) adapted Campbell's (1975) and Lindsay's (2001) tests to create their own test and found that while puppy temperament tests were reliable in predicting test scores of two of the study's eight dogs, overall the tests were not reliable for predicting the adult temperament of the dogs.

Also developed in the 1970s, the Volhard Puppy Aptitude Test (PAT), includes 10 test categories such as social attraction, social dominance, touch, and sound and sight sensitivity. The PAT was developed by adding additional test items to existing temperament tests (Volhard, 2007). The test uses a scoring system from 1-6 and is administered to puppies for the purpose of selecting the right dog for the right home (Volhard, 1981). Volhard (1981) stated that it was possible to predict the future behavioral traits of adult dogs by evaluating puppies at 49 days of age and that testing before or after that age could affect the accuracy of the results. As with the Campbell (1975) test, controlled research was not conducted on a test similar to Volhard's puppy test until years later.

While the majority of early temperament tests focused on puppies, the American Temperament Test Society (ATTS) test evaluates dogs who are at least 18 months old. According to Tom Szebenyi, ATTS Chief Tester (personal communication, May 4, 2019), in 1977, using the Federacion Canofila Mexicana temperament test as a model, Alfons Ertelt developed the American Temperament Test Society (ATTS) temperament test. Although the ATTS was originally designed to screen and predict the suitability of adult dogs for the working dog sport of Schutzhund, the ATTS is now open to all breeds. Test items include reacting to a neutral stranger, reacting to a friendly stranger, hidden clattering (i.e., chain in a metal pail), reaction to gunshot, opening an umbrella, walking on plastic footing and a grate, reacting to a potential threat, recognizing a threat, and reacting to a threatening stranger who approaches aggressively. The scoring system is a rating scale from 0 to 10 including an additional category for No Response for each item. The ATTS describes the test as a "means of evaluating temperament and giving owners insight into their dog's behavior" (American Temperament Test Society, 2019). The two test items including reacting to gunshot and reacting to an aggressive stranger relate to the original purpose of the test and are particularly relevant for screening working police dogs for protective

behaviors and the ability to handle a threat.

Canine Temperament Test Research

Following the research of Scott and Fuller (1965), early temperament tests were developed primarily by practitioners who were not researchers. As a result, there are limited controlled studies in the scientific literature pertaining to these early practitioner-developed tests. In the last few decades, researchers have conducted numerous studies in the area of canine temperament testing. These studies can be broadly categorized into general research, which includes literature reviews, and research that focuses on the nature of the tests, including how dogs responded to test items and whether or not there was predictive validity. Predictive validity (i.e., the likelihood of test scores to predict future performance) was a frequently addressed topic within the studies. Test items across the various temperament tests included many items that were the same from test to test (e.g., walk on an unusual surface such as a wire grid). Some temperament tests included items that were unique to a particular test, such as walking on a pegboard placed over an air mattress in the AKC Temperament Test (ATT). Temperament tests generally screened for desirable traits such as cooperation (i.e., does not refuse to do the task) and the ability to recover when startled, as well as undesirable traits such as fear, shyness, or aggression.

Jones and Gosling (2005) conducted an extensive review and organized the literature into methods of assessment, breeds examined, the purpose of the research, dog ages at testing, breeds, rearing environment, and whether or not dogs were spayed or neutered. Of the 51 studies reviewed, 33% tested the reactions of dogs to stimuli, 18% used owner ratings, 18% used breed expert ratings, 16% used observational tests (i.e., a natural setting), and 18% of the studies reviewed used a combination of evaluation methods such as questionnaires and direct observations (Jones & Gosling, 2005).

Other literature reviews have been conducted in the context of addressing a specific question. McGarrity et al. (2015) reviewed 47 studies to determine the behavioral dimensions that were measured in puppy temperament tests. Of 264 subtests that were identified in the 47 studies, the findings showed that the dimensions most frequently measured were sociability (26%), fearfulness/nervousness (21%), and activity (14%). Other subtests included boldness/self-assuredness (11%), exploration (11%), reactivity (9%), aggressiveness (8%), submissiveness (4%), and trainability/responsiveness (3%).

Although not specifically a literature review, Robinson et al.'s (2016) study on the predictability of puppy temperament assessments included an in-depth review of the temperament literature. One conclusion was that more research is needed on interrater reliability related to dog caregivers in canine temperament and behavior studies.

Some reviews focus on a specific segment of the literature. Brady et al. (2018) reviewed 16 papers. Behavior tests were grouped based on their relationship with either positive core affect or negative core affect. Positive core affect tests included a dog's willingness to work, human-directed social behavior, and object-directed play activities. In contrast, negative core affect tests addressed human-directed aggression, approach withdrawal tendencies, and sensitivity to aversive stimuli. This review concluded that there continues to be a need for standardization in canine behavior tests and when addressing temperament, those tests should focus on traits related to emotionality rather than cognitive processes such as sociality. Similar to the Robinson et al. (2016) conclusion related to the need for increased interrater reliability in studies, Brady et al. (2018) reported on the lack of information pertaining to reliability and validity of behavioral assessment measures.

Predictive Tests and Predictive Validity

To date, temperament tests have been predictive and research has utilized direct observation, questionnaires, or a combination of both to assess canine temperament. Predictive validity has been addressed in most studies, showing the widespread interest in the ability of canine temperament tests to predict the future

behavior of the dogs in the studies.

Slabbert and Odendall (1999) addressed predictive validity in a study with 167 puppies who were bred at the South African Police Service Dog Breeding Centre. These puppies were designated to be police or military working dogs. Dogs were observed on eight tests that included negotiating obstacles, retrieving (at 8 and 12 weeks), responding to startle stimuli (at 12 and 16 weeks), gunshot, and provoked aggression (at 6 and 9 months). In this longitudinal study, dogs were observed from 8 weeks to 2 years. In addition to direct observations on the temperament test items, dogs were also observed as they went on walks (e.g., on polished floors, through a noisy workshop, etc.), took rides in a vehicle, went swimming, and met novel adults and children. Score sheets were kept by the researchers, and the passing score for dogs to be accepted as police dogs was 80%. The most significant tests with regard to predicting adult police dog success were retrieving at 8 weeks ($p < 0.001$) and aggression (responding to an assailant) at 9 months ($p < 0.001$). At 9 months of age, of the 96 dogs that got low scores on aggression, 82.2% did not become police dogs. Of the dogs that got high scores on aggression, 54.9% (71 dogs) became police dogs. There was one unexpected finding in Slabbert and Odendaal's (1999) study, which was that a puppy's reaction to gunshot was not a predictor of police dog success. One possible reason for this finding is that with experience and repeated exposure, puppies could become desensitized over time and no longer startle at the sound of a gun.

To evaluate the reactivity of 32 German Shepherds for the purpose of predicting adult police dog performance, Sforzini et al. (2009) used seven test items that included a tunnel test, staring into the dog's eyes, noise, retrieving a ball, problem solving, bowl removal, and approaching the food bowl. Tests were video-recorded with a digital recorder. For each test item, behaviors were recorded such as ear and tail position, attentiveness, fear posture, restlessness, stillness, barking, tunnel, latency, execution time, and a maximum time. Dogs were tested at 5, 7, 9, and 24 months of age. There was variability in test results related to the age of the dog. The dogs that were tested at 24 months achieved the highest scores on a majority of the test items. Dogs became more confident as they aged, suggesting that behaviors related to temperament may change over time, and this could make predictions difficult.

Svobodova et al. (2008) also studied temperament testing with potential police dogs. In a study with 206 puppies, factors associated with passing the certification for police dog work were the puppy was willing to chase, catch, fetch a ball, and follow a rag while having a limited response to distracting noises and show low activity while negotiating objects and moving with the handler. Puppies were tested at 7 weeks of age. The reactions of puppies were scored from 0 to 5 points. Of the 206 puppies, 148 (72%) passed and 58 (28%) failed the test. Svobodova et al. (2008) concluded that puppy testing can predict the puppy's ability to do police work as an adult dog. The probability that puppies would pass the police dog test was tested by a logistic regression.

Two different rating systems (behavioral and subjective) were used in a study involving 496 German Shepherds (aged 15 to 18 months) that were potential police dogs (Wilsson & Sinn, 2012). After the behavioral assessment was completed, training leaders used subjective ratings with scores from 1 to 5. This method yielded high levels of predictive validity in the areas of confidence ($z = 2.51, p < 0.05$) and engagement ($z = 3.35, p < 0.001$). These areas were found to be the strongest predictors of dogs completing the Swedish Armed Forces training. In this study, both the behavioral tests and subjective ratings correctly identified dogs that did or did not complete training (Wilsson & Sinn, 2012).

Robinson et al. (2016) used the Campbell (1975) and Lindsay (2001) puppy tests to evaluate whether puppy temperament tests could predict the temperament of adult dogs. These puppy tests were found to be reliable for identifying puppy breeds and AKC groups. However, the puppy tests did not predict the temperament of adult dogs. The total adult temperament test scores ranged from 18 to 32 ($SE = 0.58$). None of the original puppy scores on the temperament test correlated with corresponding adult scores. Results of this study showed that temperament testing is more likely to be reliable when tests are done for working dogs (e.g., police and service dogs) that are usually tested at an older age (Robinson et al., 2016).

A large scale study conducted at the Swedish Dog Training Centre (SDTC) with 1,310 German Shepherds and 797 Labrador Retrievers showed that complex behaviors in dogs can be evaluated by experienced testers for the purpose of identifying suitable working dogs that could be used as military, police, or service dogs (Wilsson & Sundgren, 1997b). In a previous study also at the SDTC, Wilsson and Sundgren (1997a) conducted temperament tests including handling, approachability, startling, loud nose, threat, attack on handler, and reaction to gunfire. Rating scales were used to measure courage, sharpness, defense drive, prey drive, nerve stability, reaction to gunfire, energy level, hardness, ability to cooperate, and affability.

Researchers have also conducted studies related to evaluating the temperament of service, assistance, and guide dogs. A number of these studies have addressed the topic of predictive validity. In 1984, Goddard and Beilharz used a factor analysis to assess fearfulness in potential guide dogs. Results showed that the ability to predict fearfulness increased with age. In this study, a moderate correlation between scores at 6 and 12 months for fearfulness during walks was 0.36 (Goddard & Beilharz, 1984).

A subsequent study (Goddard & Beilharz, 1986) tested potential guide dogs between 4 weeks and 6 months of age on test items that were designed to observe the dog's response to a human handler, sounds and unusual objects, simple training, and while being walked on a leash. Dogs came from 32 litters that were raised in kennels until they were 12 weeks old. As an example of the observation method and scoring in this study, in the handling test, dogs were scored on a rating scale from 1 to 9. *Approach* and *avoid* scores, a score for tail position (TPH), and an overall rating that was categorized as *activity* were calculated. This study evaluated the ability of tests to predict fearfulness, activity, and learning ability when puppies became adults. The findings showed that consistent individual differences in fearfulness could be seen at 8 weeks of age, and the ability to predict fearfulness increased with age (Goddard & Beilharz, 1986). Test items were similar to those used by Scott and Fuller (1965).

Guide Dogs for the Blind developed and has employed the use of the Puppy Profiling Assessment (PPA) to assess the potential of guide dog puppies before placing them in homes with puppy walkers (Evans et al., 2015). In the Evans et al. (2015) study, test items on the PPA were scored by an assessor on a rating scale that ranged from 1 (least confident) to 7 (most confident). One finding of this study was that the behavioral test results could be used for determining breeding stock.

Asher et al. (2013) also used the PPA to study potential guide dog puppies. Similar to Volhard (1981) and Campbell's (1975) temperament tests, the PPA has a more detailed scoring procedure. In this study, a group of 587 potential guide dogs were tested at the ages of 6-8 weeks on test items that included retrieving, following, tolerating gentle restraint, reacting appropriately to noises, stroking, a toy squirrel, and encouragement to go over a ramp. Dogs were scored by three assessors who conducted inter- and intraobserver reliability observations using video recordings of dogs being tested. A 7-point scale was used to determine response to a stimulus or recovery. Five of the test items, including retrieving, stroking/stimulus, stroking by tester, response to simulated squirrel, and a ramp, were shown to relate to success in guide dog training. Two additional factors were also related to success as a guide dog. These included breed and having been bred by Guide Dogs for the Blind. Asher et al. (2013) concluded that, with adjustments, the test used in this study had the potential to predict suitability for guide dog work.

Another study that examined the predictive validity of tests for guide dogs evaluated 93 puppies at the ages of 5 and 8 months (Harvey et al., 2016). Puppies were tested on 11 stimuli including meeting a stranger, obedience tests such as sit, down, and wait, body sensitivity, and animal and human distractions. Five measures at 8 months were associated with predicting the probability of a dog qualifying for or being withdrawn from guide dog training. The measures included not displaying a low posture, having no problems with distractions (human and animal), no fear/anxiety, and low reactivity. A composite regression model classified whether a dog would qualify for or be withdrawn from the program for 79.7% of dogs at 5 months and 87.3% for the dogs at 8 months.

In a study of heritabilities of tested behavior traits, Wilsson and Sundgren (1998) found that puppy test results did not predict adult dog suitability for service dog work. When 630 puppies that were 8 weeks old were tested on nine categories and retested as adult dogs, only three significant regressions were found at the $p < 0.05$ level. These were for affability, the ability to cooperate, and prey drive (fetch). The only two significant regressions at the $p < 0.01$ level were for defense drive and prey drive (retrieve).

Valsecchi et al. (2009, 2011) conducted two similar studies with shelter dogs. Both studies used direct observation to test shelter dogs that were adopted. Dogs were tested both in the shelter and in the home setting with owners after they were adopted. These were the first studies to provide this level of follow-up and to demonstrate that temperament traits documented in the shelter setting correlate with behaviors observed when dogs are tested with owners in their new homes. A positive correlation was found (between test items related to handling the dog) administered 20 days after the dog was admitted to the shelter and a test given post-adoption ($r_s = 0.44$, $N = 34$, $p < 0.007$) (Valsecchi et al., 2011).

Weiss and Greenberg (1997) also studied shelter dogs. Nine shelter dogs were provided with training on a retrieval task. Results of this study showed that there was no correlation between the dog's performance on the shelter dog selection test and the ability to later perform the retrieval task (Weiss & Greenberg, 1997). In this study, the shelter dog selection test was not a good predictor of behavior. However, in the Valsecchi et al. (2009, 2011) studies of shelter dogs, behaviors seen in dogs at the shelter correlated with behaviors seen at a later time in the home setting. The inconsistency in findings in the Weiss and Greenberg (1997) study and the Valsecchi et al. (2009, 2011) studies could be attributed to the dissimilarity of tests used to evaluate the dogs. Weiss and Greenberg (1997) used an 11-item test that included sideways approach, initial contact, touch, approach to touch, stare, quick approach, cage exit, on-leash behavior, a room test, an umbrella test, and a pinch test. Valsecchi et al. (2009, 2011) used a temperament test with 22 subtests in 10 areas that included behavior in kennel, human sociability, docility to leash, human sociability (handling), cognitive skills, playfulness, reactivity (food possession), intraspecific sociability, reactivity (to sounds), and docility to leash (returns to kennel).

Stellato et al. (2017) used direct observation with 31 pet dogs to determine which behaviors were associated with fear in response to social stimuli (a mildly threatening stranger) and nonsocial stimuli (a garbage bag filled with crumpled newspaper that is dropped). Dogs were scored as fearful if they exhibited behaviors associated with fear, such as avoidance and reducing body posture. There were increases in these fear-related behaviors for a significant number of dogs for both the stranger appearance (74% of the dogs, $p = 0.0001$) and the garbage bag appearance (42%, $p = 0.01$). The test with the threatening stranger resulted in more fear-related responses than the nonsocial bag (Stellato et al., 2017).

Another study related to predictive validity and canine temperament testing for pet dogs was conducted by Riemer et al. (2014). A neonate test was administered to 99 Border Collie puppies when the puppies were 2-10 days old. During this neonatal phase, puppies were assessed on activity levels, vocalizations when left alone, and sucking force. Puppies were retested at 40-50 days in breeder homes on more difficult tasks such as approaching, greeting, and ball play. Follow-up testing was completed with the owners when the dogs were 1.5 to 2 years old. Results showed that the only behavior that was correlated between puppy and adult tests was exploratory activity ($p = 0.008$), indicating that the predictive validity in puppy tests for predicting traits in adult dogs is limited (Riemer et al., 2014).

One of the most large-scale studies involving direct observation to assess canine temperament was conducted by Svartberg and Forkman (2002) who used a factor analysis to identify broad personality traits in pet dogs. A total of 15,329 dogs of 164 breeds were tested by trained observers on the Dog Mentality Assessment (DMA). Factor analyses (both primary and secondary order) were conducted with 25 randomly selected dogs of 47 different breeds (1,175 dogs). Four of the five primary factors (playfulness,

curiosity/fearfulness, chase-proneness, and sociability) had loadings of > 0.40 (0.54 - 0.74). There were low loadings for aggressiveness on the secondary factor (0.03 - 0.20). The personality traits that were identified included playfulness, curiosity/fearlessness, chase proneness, sociability, and aggressiveness. While Svartberg and Forkman (2002) were able to study more than 15,000 dogs by using direct observations, conducting a study of this magnitude is not typically possible, and researchers need another method for gathering large amounts of data. One solution to this problem is using a questionnaire for data collection.

While many canine temperament studies have relied upon the use of direct observation techniques, questionnaires have been used to assess canine temperament and determine predictive validity. A benefit of questionnaires is that they make it possible to conduct research with very large sample sizes in an efficient manner (Wiener & Haskell, 2016). Questionnaires have been used to evaluate the predictive validity of guide dog and service dog temperament tests. As an example, in a study with 537 Belgian assistance dogs, questionnaires were effectively used for screening orthopedic and behavioral problems. Results showed a 92% success rate when screening addressed both orthopedic and behavioral characteristics (Bogaerts et al., 2019).

The Canine Behavioral Assessment and Research Questionnaire (C-BARQ) was used to evaluate the temperament of 7,696 young guide and service dogs. Volunteer puppy raisers were surveyed when puppies were 6 months and 12 months old (Duffy & Serpell, 2012). Dogs were tracked throughout training and were categorized as successful or released. Pulling excessively hard on the leash was the most predictive trait ($p = 0.0001$) at both age levels for being released from the program. Results showed that dogs that successfully completed guide dog training scored better on 27 of 36 C-BARQ traits (Duffy & Serpell, 2012). The C-BARQ discriminated between those dogs that were suited for guide and service work and those dogs that were not good candidates (Duffy & Serpell, 2012).

Foyer et al. (2014) studied 71 prospective military working dogs in their first year of life. An amended C-BARQ and a temperament test administered when dogs were 17 months old evaluated behaviors such as trainability and fear. Dogs who scored higher on the C-BARQ on trainability also scored significantly higher on the temperament test ($p < 0.001$). Dogs that scored high on stranger-directed fear, nonsocial fear, and dog-directed fear showed a significantly lower success rate on the temperament test ($p < 0.05$). Conclusions were that experiences during the first year of life can determine the future behavior and temperament of military working dogs (Foyer et al., 2014).

The C-BARQ was also used in a study with pet dogs who were being evaluated for aggression. Direct observation was used to evaluate the reactions of 34 pet dogs to model (toy) dogs and childlike dolls (Barnard et al., 2012), and, in addition, a C-BARQ questionnaire was used to obtain information from owners about the dog's aggression history. The direct observation testing took 20 min per dog, and a video analysis was completed by the experimenter. Behaviors were analyzed by duration of occurrence, and event recording software was used. The reactions of the dogs were compared to C-BARQ scores, and correlations included $r = 0.48$, $p = 0.004$ for dog-directed aggression/fear and $r = 0.58$, $p = 0.001$ for stranger-directed aggression.

Sheppard and Mills (2002) also used a questionnaire for the purpose of assessing the emotional predisposition of pet dogs. The questionnaire included 45 items, and owner reports were used to measure individual differences in positive or negative activation. A total of 358 pet owners returned the first questionnaire that was sent to dog owners. The negative activation scale included items relating to fearful and relaxed states, responses to changing environments, unfamiliar environments, habituation, and startle response. The negative activation scale showed correlations that ranged from $r = 0.08$ to $r = 0.49$. The positive activation scale (which was related to energy and interest, persistence, and excitement) showed correlations ranging from $r = 0.05$ to $r = 0.54$ (Sheppard & Mills, 2002). Thirteen of forty-five correlations for this scale fell outside of the desired range of 0.15 to 0.50.

The Socially Acceptable Behaviour Test (SAB) was developed in the Netherlands to evaluate canine temperament. Planta and De Meester (2007) used the SAB test to evaluate the aggressive biting behavior of

330 dogs. Aggressive biting was defined as bites, snaps, and lunging (but failed) to bite a person. The correspondence between the history of biting and aggressive biting behavior during the test was 82%. A questionnaire was administered at least one year after the test to determine the predictability of aggressive biting behavior towards people. Predictability was calculated by comparing/combining the results of the SAB with the occurrence of aggressive biting behavior at least one year after the test. The agreement between the occurrence of biting behavior in and after the test was 81.8% (kappa coefficient of 0.420, $p < 0.0001$). Findings indicated that the SAB is a valid tool for testing the aggressive behavior tendencies toward unfamiliar humans.

Another study using the SAB to evaluate aggression and fearful behavior evaluated owner perceptions of their dogs' aggressive behavior with the C-BARQ (Dalla Villa et al., 2017). Dogs that demonstrated aggression on the SAB obtained significantly higher scores on the C-BARQ subscales of stranger directed aggression ($p < 0.001$), owner directed aggression ($p = 0.03$), and familiar dog aggression ($p = 0.006$) than dogs who did not react aggressively. The findings showed that the SAB can reliably assess aggression toward unfamiliar people.

The SAB and C-BARQ were also used in a study related to a national breeding policy. From 2001 to 2009, a breeding policy excluded Dutch Rottweilers from obtaining a pedigree certificate if the dogs had been identified on the SAB test as fearful or aggressive. The C-BARQ was used by van der Borg et al. (2017) to determine the efficacy of implementing a breed-related policy on a large scale.

When designing the data collection methods for a study, some researchers have combined direct observation with questionnaires. Kobayashi et al. (2013) used a combination of direct observations by trainers and questionnaires to evaluate 158 Labrador Retrievers that were candidates for a guide dog program. Puppy raisers completed a questionnaire when puppies were 5 months old. Trainers who directly observed the dogs completed assessments when the dogs were 15 months old and had completed 3 months of training. Distraction was used as a behavioral index for the early prediction of guide dog qualification. In this study, distraction points were lower in the 28 dogs that were successful than the 82 failed dogs ($U = 508.5$, $p < 0.0001$) (Kobayashi et al., 2013).

The Mira Foundation is an organization that provides service dogs to children with autism and individuals who have motor or visual disabilities. Using a combination of direct observations completed by trainers and questionnaires delivered to foster families, 37 years of behavioral data was analyzed. In a sample size of 5,340 dogs, 3,210 dogs (60%) qualified as service dogs, and 2,020 (38%) dogs did not qualify. Of the dogs that did not qualify, 1,261 were disqualified due to behavior or temperament issues. There were 759 dogs that were disqualified due to health. The study showed that fear and reactivity are indicators that dogs will be disqualified from the program (Dollion et al., 2019).

Personality consistency is an important factor when selecting both working dogs and companion dogs for families (Fratkin et al., 2013). With the premise that personality consistency implies predictability of behavior, Fratkin et al. (2013) studied personality consistency in dogs. Following a review of the literature, 31 studies were selected for a meta-analysis. In this study, personality consistency was higher in older dogs showing that age can affect the predictive validity, and consistency between tests administered to puppies as compared to adult dogs is higher when dogs are tested as adults (Fratkin et al., 2013). The average weighted adult personality consistency estimate ($r = 0.51$) was higher than the puppy personality consistency estimate ($r = 0.30$). Personality consistency was also higher when assessment intervals were shorter and when the same tool was used from one test to the next. The meta-analysis also showed that there was no difference in personality consistency in dogs tested first as puppies and later as adults when compared to dogs that were tested as puppies (< 12 months old) and retested while still puppies (Fratkin, et al., 2013). In this study, for puppies, aggression and submissiveness were the most consistent dimensions. Behaviors that did not show a correlation between puppy and adult tests included sociability, fearfulness, and responsiveness to training (Fratkin et al., 2013).

While some studies show the predictive value of temperament testing for particular tasks or dimensions, there are also studies that clearly failed to find predictive validity. Beaudet et al. (1994) used Campbell's (1975) puppy test to evaluate 39 puppies at 7 weeks of age and again at 16 weeks of age. Only 5 of the 39 puppies (13%) increased the value for social tendencies from 7 to 16 weeks. The results showed that the test had no predictive value for social tendencies and did not predict the future behavior of puppies (Beaudet et al., 1994).

To summarize, six studies reviewed in this article concluded that certain aspects of temperament tests did not demonstrate predictive validity (Beaudet et al., 1994; Brady et al., 2018; Riemer et al., 2014; Robinson et al., 2016; Weiss & Greenberg, 1997; Wilsson & Sundgren, 1998). In these studies, age was a critical factor, and results indicated that puppy tests did not predict adult suitability for service dog (Wilsson & Sundgren, 1998) or police work (Brady et al., 2018). It was also found that there was no predictive value when pet dogs were first tested as puppies and later as adults (Beaudet et al., 1994; Riemer et al., 2014).

Even though Riemer et al.'s (2014) study showed that early temperament tests have poor predictability with regard to the future behavior of pet dogs, the study also showed there was a correlation between puppy and adult tests for exploratory behavior, demonstrating the importance of carefully reviewing all components of a given study before coming to a generalized conclusion. Robinson et al. (2016) studied puppies provided by dog breeders and found that while puppy temperament assessments could predict breed and AKC group, they could not predict adult temperament. In a study with shelter dogs, Weiss and Greenberg (1997) found that a shelter dog selection test was not a good predictor of behavior, although it was a good tool for predicting fear and submission.

Research related to service and guide dogs has shown that temperament testing can be used to predict future success in training programs (Asher et al., 2013; Goddard & Beilharz, 1984, 1986; Harvey et al., 2016). Further, there are certain behaviors or traits that result in dogs being dismissed from training programs. These include pulling excessively on the leash (Duffy & Serpell, 2012), distraction (Kobayashi et al., 2013), and fear and reactivity (Dollion et al., 2019).

A Prescriptive Test: The AKC Temperament Test (ATT)

The idea of using a temperament test as a prescriptive model is new. In all of the studies above, there was no mention of using temperament tests as a diagnostic tool that could be used to improve a dog's behavior. To date, temperament testing has been predominantly used for predictive purposes.

Developed by the American Kennel Club, the ATT is an all-breed temperament test designed to test the reaction of companion (pet) dogs to stimuli in six categories that include social, auditory, visual, tactile, proprioceptive, and an unexpected stimulus. Within each category, there are 4 possible test items. Dogs are scored on 3 of the 4 test items in each of the six categories (a total of 18 items). The three test items are selected by the evaluator. As an example of test items, the proprioceptive category includes a cavaletti (PVC ladder), intersecting hoops, a low teeter, and a low platform. The ATT is scored from 0 to 5 using a behaviorally anchored rating system (BARS). With a BARS system, along with a numerical rating scale, there is also a written description of performance (Daniels & Bailey, 2014).

The ATT is meant to provide a prescriptive approach to temperament testing. Dog owners arrive at an ATT test and, using a breed temperament guide, fill out a form on which they describe their breed's temperament (based on descriptions provided by each breed's national parent club). The dog owner enters the ring, briefly discusses the dog's breed temperament with the evaluator, and then takes the dog through the temperament test. If dogs do not pass the test, information is provided to dog owners on how to correct the problem. Recommendations for remediation are based on an applied behavior analysis approach using

techniques such as shaping, backward chaining, desensitization, and reinforcement (Burch & Bailey, 1999; Thyer & Curtis, 1983; Wolpe, 1973).

If a dog fails a test item, the dog owner is guided to begin the training process by conducting a reinforcer sampling assessment to identify preferred reinforcers that can be used in training. This basically involves offering the dog a number of food items that can be used as reinforcers (e.g., steak, dried liver, small commercial soft training treats, string cheese, etc.). By observing the dog's reaction to the food, the dog owner identifies the items that the dog prefers. Next, the dog owner is provided with a training protocol. For example, in the visual category, a specific test item is, "The helper circles the dog and handler (at a distance of 3 ft) with a wagon, roller bag, etc." If the dog startles, hides behind the handler, and does not recover, it does not pass the test item. A score of 0 on any test item results in the dog not passing the test. The prescriptive training protocol for this test item is:

1. Start with the handler standing beside the dog. The dog is on leash. For each of the steps below, the handler will give the dog a food reward if the dog's reaction is acceptable.
2. Helper pulls wagon in front of the dog at 20 ft away.
3. Helper pulls wagon in front of the dog at 15 ft away.
4. Helper pulls wagon in front of the dog at 10 ft away.
5. Helper pulls wagon in front of the dog at 5 ft away.
6. Helper pulls wagon in front of the dog at 3 ft away.
7. Repeat steps 2-6 with the helper circling the dog with the wagon.

Throughout the prescriptive training, the idea is to start at a distance the dog can tolerate and gradually move the visual stimulus closer to the dog.

For dogs who do not pass the ATT, dog owners may provide training and retest after the dog can perform the test item. While a dog's basic temperament may not have changed as a result of training, behaviors related to temperament can be modified for pet dogs. The ATT is an educational tool that may be used by dog owners to identify problems discovered as a result of temperament testing. Following testing, dog owners can address any problems (e.g., the dog refuses to walk on unfamiliar surfaces).

Discussion

Two models for the assessment of canine temperament have been presented in this paper. The first model is the traditional *predictive* temperament test that is used to assess canine behavior traits such as being playful, aggressive, fearful, or shy to determine the possible future behavior of the dog. Often, predictive tests are used to evaluate the dog's suitability as a police dog, military dog, or service dog. The second model is a new *prescriptive* temperament test that is designed to assess the dog's reaction to a set of stimuli and provide behavioral protocols for addressing problem areas (e.g., fearful or startles with auditory stimuli, refuses to walk on novel surface).

A key point of this article is extending previously published studies by introducing the concept of a *prescriptive* temperament test for pet dog owners, the ATT, which employs a prescriptive approach to temperament testing. In the ATT, dog owners are encouraged to use the test to identify problems related to their dogs' reactions to specific stimuli and then remediate these problems using specially prepared training materials that are based on learning theory and applied behavior analysis techniques.

It is important to note that when a prescriptive training protocol is successfully completed after the dog has failed an item on the ATT, the dog's temperament has not been permanently changed because it can now perform a test item correctly. Rather, specific behaviors that are related to the temperament test have been modified. While it is unlikely that the fearful or tentative dog will never again show any hesitancy or fear with regard to new tasks, training can result in the dog not exhibiting fearful behaviors in a variety of practical

situations (e.g., walking on unusual surfaces). The prescriptive temperament test model is important for dogs, because it can improve their ability to cope and function in real-world situations. An example related to human behavior would be an extremely shy person acquiring the skills needed to become an excellent public speaker through training and coaching. The person is likely to always be a basically shy person, but the acquisition of adaptive, practical skills can result in overcoming the shyness for the targeted setting (i.e., public speaking). In the laboratory setting, Schneider et al. (1991) studied the early temperament characteristics (e.g., fearfulness) of nursery-reared rhesus monkey infants that were reared either in laboratory cages or enriched environments. During the first month of life and at 8 months of age, tests were administered that included the assessment of temperamental capabilities and responses. Results of this study indicated that any adverse effects may have been “partially attenuated by environmental enrichment” (Schneider et al., 1991, p. 137).

In the early history of temperament testing, temperament tests were viewed as having a predictive nature. Recent research has emphasized the significance of predictive validity to predict the future behavior of the dog. Predictive validity is an important measure and has benefits when evaluating the results of temperament tests. For example, if a dog passed a temperament test for guide dogs when it was a puppy, knowing if the dog was going to be suitable as a guide dog when it was an adult would be advantageous. Similarly, if a young dog who was a potential police dog was afraid of loud noises, when considering a career for the dog, it would be helpful to know if the dog would continue to balk at gunshot throughout its adult life.

There are also some challenges related to conducting research related to predictive temperament tests or predictive validity. For example, many studies take a long time to complete, so administering temperament tests to potential guide dogs or police dogs when they are puppies is only the first step. There is a need for the dogs to be evaluated again as adults. With pet dogs or shelter dogs, it may be difficult or impossible to gain access to the dogs for follow-up.

The question, “Can early temperament tests predict the behavior of adult dogs?” is far more complex than a simple yes-no question. There are conflicting findings, and the results from predictive studies depend on which variables (such as the age of the dog) are being evaluated (Fratkin et al., 2013; Robinson et al., 2016). Temperament may be more likely to be predictive when testing is done with certain categories of dogs, such as police/military dogs and service/guide dogs. Possible reasons for this include that these dogs are often tested at a later age, they have often received training that is administered by a very specific protocol, and they are often raised in a kennel or by trained puppy raisers where socialization and training procedures are consistent day after day. This is very different than what happens with a litter of pet dogs, who, as early as 8 weeks old, may be sent by a breeder to eight different homes where they are raised, trained (or not) under widely disparate circumstances. Service dogs and police/military dogs are unique in that they must be well-screened to optimize chances of success, they require extensive training by skilled trainers, and the process to prepare them for their work is costly. For these dogs, any process, such as a temperament test that is predictive, is one measure that can be combined with others to ensure potential problems are identified before an organization invests considerable time and resources in training a dog. For pet dogs, in some respects, the stakes are not quite as high. Because training is not as costly, the owners of pet dogs do not have to be as efficient with regard to time as specialty trainers (e.g., police and guide dog trainers). As a result, dog owners can take the time to train their pet dogs and provide rehabilitation and training to rescue and shelter dogs.

There are several limitations of this paper. First, the scope of the research articles reviewed included only frequently cited studies from the past 20 years (1999-2019), in addition to three other important articles which included Goddard and Beilharz (1986), Wilsson and Sundgren (1997a), and Slabbert and Odendaal (1999) and several earlier papers that were related to the definition of temperament. While the goal was to review papers from the last two decades to show current trends in temperament research, other studies might have provided additional insight with regard to the predictive validity of temperament testing. Second, some of the studies might have had different findings if recent research techniques were applied. For example, the use of meta-analysis to evaluate studies is particularly beneficial. Third, data were not presented for the ATTS and ATT. The ATT data will be presented in a future study.

Future research in the area of canine temperament testing should address standardizing temperament testing practices for specialty areas, such as police/military and service dog work, and evaluating the effectiveness of easy to administer real-world tests [e.g., Urban Canine Good Citizen (CGC)] that are based on trained skills rather than temperament. Ideally, future studies on temperament testing would include inter-observer reliability that is missing from some existing studies (Taylor & Mills, 2006), sufficient sample sizes, and the writing of results so that the study can be replicated. Further, the results in the area of canine temperament testing are often expressed as being statistically significant. While this standard is certainly of value, in an applied area such as canine temperament, researchers should also consider results that are socially significant (Bailey & Burch, 2017).

With regard to social significance, Bogaerts et al. (2019) and Diederich and Giffroy (2006) recommended that canine temperament tests include practical, realistic behaviors such as testing in the community. Similarly, King et al. (2012) suggested that rather than focus on undesirable traits such as fear and aggression, behavioral assessments should address traits such as being calm, friendly, obedient, and safe around children. Such programs currently exist, but they are not considered temperament tests. For example, the AKC Community Canine test (American Kennel Club, 2019) includes test items for the dog such as walks on a leash through a crowd, does not shy away from a person carrying something such as a backpack, allows a person to approach and pet it, and exits a doorway in a controlled manner. The Urban CGC program (American Kennel Club, 2019) evaluates dogs in a city setting as they respond appropriately to distractions such as sirens, skateboards, and surfaces such as sidewalk grates. In Urban CGC, dogs must also ride an elevator or escalator under control and demonstrate a down-stay in a public space.

There are multiple beneficiaries when temperament testing has a prescriptive function. These include dog owners, trainers, breeders, and the community. Dogs are easier to handle when problematic behaviors have been resolved. The community benefits when dogs do not engage in aggression or react in a fearful manner to stimuli. And certainly, a beneficiary of the prescriptive model of temperament testing is the dog. Dogs who are afraid to greet unfamiliar people, walk on unusual surfaces, and become extremely alarmed at unexpected auditory and visual stimuli to the point that they do not recover during testing are likely to have lower quality lives that are severely impacted by fear. Helping dogs overcome problem behaviors that have been identified on temperament tests can result in both an improvement in their welfare and an enhanced relationship with their owners.

The development of behaviors related to an individual dog's temperament will depend on the environment in which the dog is raised and the socialization and training that is provided. With the exception of dogs that are housed and raised in kennels for specialty training or research purposes, training history, socialization experiences, and methods of raising are likely to vary greatly from one dog to the next. Expanding our knowledge of temperament testing will help us better understand the dogs who serve us as police, military and service dogs, and as our family pets.

References

- American Kennel Club. (August, 2019). AKC family dog program. Retrieved from <https://www.akc.org/sports/akc-family-dog-program/>, September 1, 2019.
- American Temperament Test Society. (August, 2019). *About canine temperament*. Retrieved from <https://atts.org/about-atts/>, September 12, 2019.
- Asher, L., Blythe, S., Roberts, R., Toothill, L., Craigon, P. J., Evans, K. M., Green, M. J., & England, G. C. W. (2013). A standardized behavior test for potential guide dog puppies: Methods and association with subsequent success in guide dog training. *Journal of Veterinary Behavior*, 8, 431-438. <https://doi.org/10.1016/j.jveb.2013.08.004>
- Bailey, J. S. & Burch, M. R. (2017). *Research methods in applied behavior analysis*. (2nd ed.). Routledge.

- Barnard, S., Siracusa, C., Reisner, I., Valsecchi, P., & Serpell, J. (2012). Validity of model devices used to assess canine temperament in behavioral tests. *Applied Animal Behaviour Science*, *138*, 79-87.
- Beaudet, R., Chalifoux, A., & Dallaire, A. (1994). Predictive value of activity level and behavioral evaluation on future dominance in puppies. *Applied Animal Behavior Science*, *40*, 273-284.
- Bogaerts, E., Moons, C. P. H., Van Nieuwerburgh, F., Peelman, L., Saunders, J. H., & Broeckx, B. J. G. (2019). Rejections in an non-purpose bred assistance dog population: Reasons, consequences, and methods for screening. *PLoS One*, *14*(6), e 0118339. <https://doi.org/10.1371/journal.pone.0218339>
- Brady, K., Cracknell, N., Zulch, H., & Mills, D. S. (2018). A systematic review of the reliability and validity of behavioural tests used to assess behavioural characteristics important in working dogs. *Frontiers in Veterinary Science*, *5*, 103. <https://doi.org/10.3389/fvets.2018.00103>
- Burch, M. R. & Bailey, J. S. (1999). *How dogs learn*. Macmillan.
- Buss, A. H. (1995). *Personality: Temperament, social behavior, and the self*. Allyn and Bacon.
- Campbell, W. E. (1975). *Behavior problems in dogs*. American Veterinary Publications, Inc.
- Dalla Villa, P., Barnard, S., Di Nardo, A., Ianetti, L., Vulpiani, M. P., Trentini, R., Serpell, J., & Siracusa, C. (2017). Validation of the socially acceptable behaviour (SAB) test in acentral-Italy pet dog population. *Veterinaria Italiano*, *53*(1), 61-70.
- Daniels, A. C. & Bailey, J. S. (2014). *Performance management: Changing behavior that drives organizational effectiveness*. Performance Management Publications.
- Dawson, W. M. (1937). Heredity in the dog. *U.S. Department of Agriculture Yearbook*, pp. 1351-1349.
- Diederich, C. & Giffroy, J-M. (2006). Behavioural testing in dogs: A review of methodology in search of standardization. *Applied Animal Behaviour Science*, *97*, 51-72. <https://doi.org/10.1016/j.applanim.2005.11.018>
- Dollion, N., Paulus, A., Champagne, N., St-Pierre, N., St-Pierre, E., Trudel, M., & Plusquellec, P. (2019). Fear/reactivity in working dogs: An analysis of 37 years of behavioural data from the Mira Foundation's future service dogs. *Applied Animal Behaviour Science*. <https://doi.org/10.1016/j.applanim.2019.104864>
- Duffy, D. L. & Serpell, J. A. (2012). Predictive validity of a method for evaluating temperament in young guide and service dogs. *Applied Animal Behaviour Science*, *138*, 99-109. <https://doi.org/10.1016/j.applanim.2012.02.011>
- Evans, K. M., Lewis, T. W., Asher, L., Blythe, S., Bottomley, M., Tootill, L., Blott, S. C. (2015). Genetic evaluation of traits in a standardized behavioral test for potential guide dog puppies using crossbreed models. *Journal of Veterinary Behavior*, *10*(6), 459-464.
- Fox, M. W. (1972). *Understanding your dog*. St. Martin's Press.
- Foyer, P., Bjallerhag, N., Wilsson, E., & Jensen, P. (2014). Behaviour and experiences of dogs during the first year of life predict the outcome in a later temperament test. *Applied Animal Behaviour Science*, *155*, 93-100. <https://doi.org/10.1016/j.applanim.2014.03.006>
- Fratkin, J. L., Sinn, D. L., Patall, E. A., & Gosling, S. D. (2013). Personality consistency in dogs: A meta-analysis. *PLoS One*, *8*(1), e 54907. <https://doi.org/10.1371/journal.pone.0054907>.
- Goddard, M. E., & Beilharz, R. G. (1984). A factor analysis of fearfulness in potential guide dogs. *Applied Animal Behaviour Science*, *12*, 253-265.
- Goddard, M. E., & Beilharz, R. G. (1986). Early prediction of adult behavior in potential guide dogs. *Applied Animal Behaviour Science*, *15*, 247-260.
- Goldsmith, H. H., Buss, A. H., Plomin, R., Rothbart, M. K., Thomas, A., Chess, S., Hinde, R. A., McCall, R. B. (1987). Roundtable: What is temperament? Four approaches. *Child Development*, *58*, 505-529.
- Hall, C. S. (1941). Temperament: A survey of animal studies. *Psychological Bulletin*, *38*, 909-943.
- Harvey, N. D., Craigon, P. J., Sommerville, R., McMillan, C., Green, M., England, C.W., & Asher, L. (2016). Test retest reliability and predictive validity of a juvenile guide dog behavior test. *Journal of Veterinary Behavior*, *11*, 65-76. <https://doi.org/doi.org/10.1016/j.jveb.2015.09.005>
- International Guide Dog Federation. (August, 2019). *History of guide dogs*. <https://www.igdf.org.uk/about-us/facts-and-figures/history-of-guide-dogs/>, September 12, 2019.
- John, O. P. & Gosling, S. D. (2000). Personality Traits. In A. R. Kazdin (Ed.), *Encyclopedia of psychology* (Vol. 6, pp. 140-144). American Psychological Association.
- Jones, A. C. & Gosling, S. D. (2005). Temperament and personality in dogs (*Canis familiaris*): A review and evaluation of past research. *Applied Animal Behaviour Science*, *95*, 1-53. <https://doi.org/10.1016/j.applanim.2005.04.008>
- King, T., Marston, L. C., & Bennett, P. C. (2012). Breeding dogs for beauty and behavior: Why scientists need to do more to develop valid and reliable behavior assessments for dogs kept as companions. *Applied Animal Behaviour Science*, *127*, 1-12. <https://doi.org/10.106/j.applanim.2011.11.016>
- Kobayashi, N., Arata, S., Hattori, A., Kohara, Y., Kiyokawa, Y., Takeuchi, Y., & Mori, Y. (2013). Association of puppies' behavioral reaction at five months of age assessed by questionnaire with their later 'distraction' at 15 months of age, an important behavioral trait for guide dog qualification. *Journal of Veterinary Medicine Science*, *75*(1), 63-67.

- Lindsay, S. R. (2001). *Applied dog behavior and training: Etiology and assessment of behavior problems* (Vol. 2). Iowa State University Press.
- Mason, W. A. (1984). Animal learning: Experience, life modes, and cognitive style. *Verh. Dtsch. Zool. Ges.*, 77, 45-56.
- McGarrity, M. E., Sinn, D. L., & Gosling, S. D. (2015). Which personality dimensions do puppy tests measure? A systematic procedure for categorizing behavioral assays. *Behavioural Processes*, 110, 117-124. <https://doi.org/10.1016/j.beproc.2014.09.029>
- Most, K. (1954). *Training dogs: A manual*. Popular Dogs Press.
- Pfaffenberger, C. (1963). *The new knowledge of dog behavior*. Howell Book House.
- Planta, J. U. D. & De Meester, R. H. W. M. (2007). Validity of the Socially Acceptable Behavior (SAB) test as a measure of aggression in dogs towards non-familiar humans. *Vlaams Diergeneeskundig Tijdschrift*, 76, 359-368.
- Riemer, S., Muller, C., Viryani, Z., Huber, L., & Range, F. (2014). The predictive value of early behavioural assessments in pet dogs – A longitudinal study from neonates to adult. *PLoS One*, 9(7): e101237. <https://doi.org/10.1371/journal.pone.0101237>
- Robinson, L. M., Thompson, R. S., & Ha, J. C. (2016). Puppy temperament assessments predict breed and American Kennel Club group but not adult temperament. *Journal of Applied Animal Welfare Science*, 19(2), 101-114. <https://doi.org/10.1080/10888705.2015.1127765>
- Ruefenacht, S., Gebhardt-Henrich, S., Miyake, T., & Gaillard, C. (2002). A behavior test on German shepherd dogs: heritability of seven different traits. *Applied Animal Behaviour Science*, 79, 113-132.
- Schneider, M. L., Moore, C.F., Suomi, S. J., & Champoux, M. (1991). Laboratory assessment of temperament and environmental enrichment in Rhesus Monkey infants (*Macaca mulatta*). *Am. J. Primatol*, 25, 137-155.
- Scott, J. P., & Fuller, J. L. (1965). *Genetics and social behavior of the dog*. University of Chicago Press.
- Sforzini, E., Michelazzi, M., Spada, E., Ricci, C., Carezzi, C., Milani, S., Luzi, F., & Verga, M. (2009). Evaluation of young and adult dogs' reactivity. *Journal of Veterinary Behavior*, 4(1), 3-10. <https://doi.org/10.1016/j.jveb.2008.09.035>
- Sheppard, G., & Mills, D. S. (2002). The development of a psychometric scale for the evaluation of the emotional predispositions of pet dogs. *International Journal of Comparative Psychology*, 15, 201-222.
- Slabbert, J. M., & Odendaal, J. S. J. (1999). Early prediction of adult police dog efficiency—a longitudinal study. *Applied Animal Behaviour Science*, 64, 269-288. [https://doi.org/10.1016/S0168-1591\(99\)00038-6](https://doi.org/10.1016/S0168-1591(99)00038-6)
- Stellato, A. C., Flint, H. E., Widowski, T. M., Serpell, J. A. & Niel, L. (2017). Assessment of fear-related behaviours displayed by companion dogs in response to social and non-social stimuli. *Applied Animal Behaviour Science*, 188, 84-90. doi.org/10.1016/j.applanim.2016.12.007
- Svartberg, K., & Forkman, B. (2002). Personality traits in the domestic dog (*Canis familiaris*). *Applied Animal Behaviour Science*, 79, 133-155.
- Svobodova, I., Vapenik, P., Pinc, L., & Bartos, L. (2008). Testing German shepherd puppies to assess their chances of certification. *Applied Animal Behaviour Science*, 113(1/3), 139-149.
- Taylor, K. D., & Mills, D. S. (2006). The development and assessment of temperament for adult companion dogs. *Journal of Veterinary Behavior*, 1, 94-108. <https://doi.org/10.1016/j.jveb.2006.09.002>
- Thyer, B. A., & Curtis, G. (1983). The repeated pretest-posttest single subject experiment: A new design for empirical clinical practice. *Journal of Behavior Therapy and Experimental Psychiatry*, 14(4), 311-315.
- Valsecchi, P., Barnard, S., Stefanini, C., & Normando, S. (2009). Validation of a new temperament test as a practical tool for adoptions of sheltered dogs. *Journal of Veterinary Behavior*, 4(2), 75-76.
- Valsecchi, P., Barnard, S., Stefanini, C., & Normando, S. (2011). Temperament test for re-homed dogs validated through direct behavioral observation in shelter and home environment. *Journal of Veterinary Behavior*, 6, 161-177. <https://doi.org/10.1016/j.jveb.2011.01.002>
- Van der Borg, J. A. M., Graat, E. A. M., & Beerda, B. (2017). Behavioural testing based breeding policy reduces the prevalence of fear and aggression related behaviour in Rottweilers. *Applied Animal Behaviour Science*, 195, 80-86. <https://doi.org/10.1016/j.applanim.2017.06.004>
- Volhard, W. (1981). *Puppy Aptitude Testing*. Dog Writer's Association of America video.
- Volhard, W. (2007). *Choosing your puppy (PAT)*. Retrieved from <https://www.volhard.com/pages/pat.php>, August 28, 2019.
- Weiss, E., & Greenberg, G. (1997). Service dog selection tests: Effectiveness for dogs from animal shelters. *Applied Animal Behaviour Science*, 53(4), 297-308. [https://doi.org/10.1016/50168-1591\(96\)01178-8](https://doi.org/10.1016/50168-1591(96)01178-8)
- Wiener, P., & Haskell, M. J. (2016). Use of questionnaire-based data to assess dog personality. *Journal of Veterinary Behavior*, 16, 81-85. <https://doi.org/10.1016/j.jveb.2016/10.007>
- Wilsson, E., & Sinn, D. L. (2012). Are there differences between behavioral measurement methods? A comparison of the predictive validity of two ratings methods in a working dog program. *Applied Animal Behavior Science*, 141, 158-172. <https://doi.org/10.1016.j.applanim.2012.08.012>

- Wilsson, E., & Sundgren, P. E. (1997a). The use of a behavior test for selection of dogs for service and breeding, I: Method of testing and evaluating test results in the adult dog, demands on different kinds of service dogs, sex and breed differences. *Applied Animal Behaviour Science*, *53*, 279-295.
- Wilsson, E., & Sundgren, P. E. (1997b). The use of a behaviour test for selection of dogs for service and breeding, II. Heritability for tested parameters and effect of selection based on service dog characteristics. *Applied Animal Behaviour Science*, *54*(2), 235-241. [https://doi.org/10.1016/S0168-1591\(96\)01175-6](https://doi.org/10.1016/S0168-1591(96)01175-6)
- Wilsson, E., & Sundgren, P. E. (1998). Behavior test for eight-week old puppies- heritabilities of tested behaviour traits and its correspondence to later behaviour. *Applied Animal Behavior Science*, *58*, 151-162. [https://doi.org/10.10/50168-1591\(97\)00093-2](https://doi.org/10.10/50168-1591(97)00093-2)
- Wolpe, J. (1973). *The practice of behavior therapy*. Pergamon.

Financial conflict of interest: No stated conflicts.
Conflict of interest: No stated conflicts.

Submitted: October 9th, 2019
Resubmitted: January 14th, 2020
Accepted: May 3rd, 2020