

Temperament Tests: What we do and do not know about them.

Sara L. Bennett, DVM
Maddie's® Animal Behavior Resident
Purdue Animal Behavior Clinic

Temperament tests are a form of behavior assessment tool. There are several different definitions, as can be seen below:

A tool to help determine a dog's overall temperament and its threshold for aggression.¹

Tests that take a measure of a dog's behavior and use it to describe various behavioral traits in an attempt to predict the dog's behavior in another but similar situation.²

Tests that are used as evaluation tools to classify individual animals.³

A series of standardized experimental situations where stimuli serve to elicit behavior that is then statistically compared to other individuals placed in the same situation in order to classify the subject tested.⁴ This, however, is actually a definition of only one type of temperament test, Test Battery, which will be elaborated on below.⁵

Why do we need them?

These have historically been used to select dogs for use as working dogs, service dogs, quality breeding stock, to prevent adoption of aggressive dogs (major public safety concern); to identify at-risk dogs and to help make educated adoption recommendations, or to set up rehabilitation if an option; use as a diagnostic tool to look for indicators of disease (CDS) or poor welfare.^{3,5} The goal of these tests, regardless of their specific indicators, is to try to select the most suitable individual for the circumstances or to match the potential owners to a suitable dog.

There are 4 types of temperament test that have been classified.⁵

Test Battery: expose a dog to a variety of specific stimuli artificially elicited and to record the dog's reactions.

Rating of Individual Dog: Information is gathered about an individual dog's behavior and history from an "informant". This can be the owner, former owner, handler, or caretaker.

Expert Rating of Breed Prototype: "experts" on dogs, such as veterinarians, trainers, and judges, are asked to describe, rank, or rate breeds of dogs as a whole rather than individual dogs.

Observational Test: Dog is exposed to a natural less controlled environment and the behavior exhibited is recorded to describe broad conclusions. This has been used most often when trying to select service dogs.⁵

What is temperament? How is it different from behavior, character, and personality? Do the differences matter?

Many researchers, as the ones previously noted, used these words interchangeably, but are they truly the same thing?

Svartburg used personality and temperament synonymously in 2002 and 2005 while looking at stable traits across dogs, and suggested that there are parallels between human personality traits and other species, specifically dogs.^{6,7}

Ley in 2009 also defined personality as a stable and characteristic behavioral tendency of an adult individual dog while creating a canine personality questionnaire.^{8,9} Her groups were modeled after the big 5 human personality traits, as those of Svartburg.⁶⁻⁹

In another review, Diederich noted that the words temperament, character, and personality were often used interchangeably in the same studies, and interpreted that most implied that the characters identified, however the researchers chose to classify them, were present at an early age, would be relatively stable over time, and expected that the same response would be elicited in a similar set of situations.³ It was also noted that some felt that these traits were made up or influenced by genetics, the prenatal environment, mother-offspring relationship, early experience, and learning capabilities.³ She even noted that one study used the word temperament to name a specific temperament trait.³

When looking to clarify these distinctions, Segurson broke each one down with additional definitions.¹⁰

Character: is considered the product of the animal's environment and experiences. The behaviors seen are the animal's habits of responding to stimuli. These develop secondary to experiences, a *learning history*. It is influenced by changing environment.

Temperament: the pet's natural or innate responses to its environment, its *emotional responses*. It is considered stable over time, and is the foundation of personality. The character can influence it but cannot change it.

Personality: combination of temperament and character, an animal's overall way of responding to the environment. It is based on past experience and genetic tendencies.¹⁰

What are we really testing and what are we testing for?

More recently, most of the research has been done in shelters to predict aggression in potential adoption candidates, and also to screen for specific behavior problems such as separation anxiety and food or resource guarding. Much of the rest has been used to predict a dog's success as a working dog.

In its simplest form, a temperament test is testing a behavioral response to a specific stimulus at that time in that environment by that individual dog. It is a behavior evaluation, a snapshot in time. The attempt is then made to correlate that specific response in that situation as a predictor of how that same animal may act or respond in a similar situation at another time, therefore predicting temperament and/or personality.

Things that we do not know about these tests include:

1. What is the effect of stress of the shelter environment on the displayed behavior? Several researchers have shown that cortisol, a hormone elevated during stress, is elevated at least initially in dogs when entering the shelter environment. Some feel that dogs should be given a period of adjustment prior to testing,¹⁰ while others feel that

testing a dog when stressed will be more likely to identify aggression in those individuals at risk for showing it.¹

2. Are we testing behavior, character, temperament, or personality? When reviewing the development of temperament tests as a whole, Taylor had concerns that a valid assessment of temperament could ever be determined in the overwhelming and unnatural environment of a shelter,² while Jones and Gosling believed after their review of temperament testing literature that there was evidence that some studies showed that constructs (behavior responses) are stable over time and therefore theoretically could measure temperament.⁵

When we focus in on applications in shelter medicine, most often the test battery format is used. And most of the tests are set up within 2 different categories:

Evaluate a response in situations or interactions that are commonly done by the average pet owner or would be found in the average home.

Try to provoke a pet in a highly challenging or stressful situation.¹⁰

But Christensen determined that some types of aggression, specifically intra-species, predatory, and territorial may not be identified through their test battery.¹¹ At this time, we do not know whether these tests are truly predictive. As Segurson pointed out, if there are false positive results, dogs are getting condemned needlessly, if there is a false negative result, dogs are being adopted out that pose a true public safety risk.¹⁰ Also, tests of young puppies and their behavior have not shown to be a good predictor of adult behavior.^{5,12,13}

How do we use the data?

Most experts agree that at this time, the information gained from a temperament test should be used as an ancillary tool that should be added to the information gathered from other sources.

Additional sources of information might include:

Owner relinquishment questionnaires- public or confidential; are they valid? One study showed owners who were told that relinquishment questionnaires would be confidential were more likely to report their dog had owner directed aggression, and fear of strangers but did not change how they reported about other behavior problems. Therefore, sometimes this information may be inaccurate, but should be still considered useful when combined with other information about the pet.¹⁴

Staff experiences with the animal- essentially an informal observational test. One study showed that staff observations were better at predicting the absence of a problem than predicting the presence of a problem.¹⁵

Are these tests standardized between facilities, evaluators? How do we do that?

To be a good test, the subtests should be standardized, with the only variable being the animal tested. And they should be reliable, getting the same results if the test is repeated.

Most of the tests used in shelters, whether using Emily Weiss' SAFER, Sue Sternberg's Assess A Pet, or a permutation of those or others, are aimed primarily at predicting aggression. These tests tend to look at different traits such as sociability, handle-ability, playfulness/ play style/ arousability, food or possessive aggression, reaction to a stranger and/or a child, reaction to another dog and/or another animal species.¹⁰

They should measure the dog's behavioral response using an objective description of the behavior (ideally an ethogram) rather than a subjective interpretation of the dog. A variety of scoring systems have been created to try to correlate with anticipated behavioral responses. These have ranged from pass/ fail, to numerical scoring (1-5), or letter scoring (A-F).¹⁰

Are they valid? Are they truly testing what we think they are testing?

There are varying types of validity with any test. The main question to be asked is: are we really measuring what we think we are measuring?

When broken down, we look at:

External (predictive) validity: Is the test a measure of the predicted behavior?³ Does the factor vary alongside a more established measure that is supposed to measure the same construct (or behavior) and is usually measured at the same time?² For example, do the 2 tests that attempt to show a fear response both really create fear responses?

Internal (content) validity: How well do the different variables of the test represent the desired trait to be evaluated?³ Is there consistency within the components of the test designed to assess the same behavior?² For example, does the scary stimulus you are exposing the dog to create a fear response?

Internal (construct) validity: What is the extent to which a scale measures what it claims to measure?³ What is the extent to which items within a test measure the construct (behavior trait) they were designed to measure?²

Inter-rater reliability: Are different people getting the same result? Is there consistency between observers?² This can be done with different people simultaneously scoring or watching video of the test.³

Intra-rater reliability: does the same person scoring the same test at different times get the same or similar results?³ Is there consistency within the observer?² This can be done by this person watching video back at a later date.

Test- retest reliability: Is the same dog scoring the same at a different time?³

No tests so far have been published in a peer reviewed journal that have passed all of these qualifications.^{3,10}

Examples of tests:

The SAFER test consists of 7 different subtests¹⁶:

1. Look Test- gently hold the dog's head in your hands and look in his (her) eyes using soft eye contact.

2. Sensitivity Test- A firm kneading motion is made over the skin on the dog's side from the shoulder to hip and back three times by the assessor's hand.
3. Tag Test- The assessor tries to engage the dog using an excited voice and play movement, then touches the dog lightly with a finger and tries to initiate play.
4. Squeeze Test- The assessor says "squeeze", then runs a hand down the dog's leg, then gently squeezes between the toes using the pads of her fingers, and repeats.
5. Food Behavior Test- The dog is given a mix of canned and dry food, allowed to start eating, then the assessor asks for the food and places the Assess-A-Hand into the food bowl, pulls it away from the dog, then repeats by pushing the dog's head away from the food bowl gently or stroke the dog's head and neck.
6. Toy Behavior Test- The dog is given a toy, then the assessor asks for the toy and reaches for it with the Assess-A-Hand.
7. Dog to Dog Behavior Test- The subject dog enters the testing room on leash with the assessor where an assistant waits with a stable dog on leash. The dogs do not touch or greet each other. Only the initial approach is evaluated.

Modified Assess A Pet¹ consists of 9 different subtests:

1. Cage presentation- evaluator gives 5 seconds of non-threatening eye contact while standing in front of the kennel or run.
2. Sociability- The dog is taken out of the kennel into the evaluation room on a leash. The evaluator ignores the dog for 30 seconds then speaks to the dog in a friendly voice for 10 seconds, then strokes the dog 3 times along its back.
3. Teeth examination- evaluator attempts to lift the dog's lips five times in succession and hold that position for 5 seconds each time.
4. Handling- evaluator strokes the dog's back then picks up the back leg and touches the foot, touches the tail, handles both ears, wipes the body with a towel, tugs on the collar, applies pressure to the shoulders, and hugs the dog.
5. Arousal- evaluator engages the dog in play to determine play style, level of arousal, and ability to calm down once play stops.
6. Food bowl- evaluator stands near and strokes the dog's back, reaches into the food bowl several times using a fake hand while the dog eats a bowl of food.
7. Possessions- evaluator stands near and strokes the dog's back, reaches toward the valued item (eg. Rawhide or pig ear) that the dog is in possession of.
8. Stranger- an unfamiliar person knocks and enters the room, gives 10 seconds of non-threatening eye contact, steps toward the dog with an outstretched hand, then bends down with a sideways body posture and solicits attention with a calm voice.
9. Dog introduction- test dog is introduced to an unfamiliar dog. Both dogs are on a leash.

Pre-adoption screening and matching programs

ASPCA Meet Your Match- goal to match individual dogs with potential adopter's lifestyles and expectations.¹⁶

Dogs are grouped into a Canine-ality personality profile through behavior assessment and potential owners are grouped based on a self-completed survey. Each individual pet is labeled with a color and corresponding description. The potential owners are also given a color based on their survey answers and are encouraged to choose a pet with a matching color. The same type of program has been started with cats, called the Feline-ality.¹⁰ No data on the success of this program has been published yet.

Acknowledgments:

Andrew Luescher, Dr. med vet, PhD, DACVB ECVBM-CA. Director, Purdue Animal Behavior Clinic. Purdue University Veterinary Teaching Hospital, Purdue University. West Lafayette, IN

Maddie's Shelter Medicine Program is underwritten by a grant from Maddie's Fund, The Pet Rescue Foundation (www.maddiesfund.org), helping to fund the creation of a no-kill nation.

References:

¹Bollen KS, Horowitz J. Behavioral evaluation and demographic information in the assessment of aggressiveness in shelter dogs. *Applied Animal Behaviour Science* 2008; 112: 120-135.

²Taylor D, Mills DS. The development and assessment of temperament tests for adult companion dogs. *Journal of Veterinary Behavior* 2006; 1: 94-108.

³Diederich C, Giffroy JM. Behavioural testing in dogs: A review of methodology in search for standardization. *Applied Animal Behaviour Science* 2006; 97: 51-72.

⁴Serpell JA, Hsu Y. Development and validation of a novel method for evaluating behavior and temperament in guide dogs. *Applied Animal Behaviour Science* 2001; 72, 347-364.

⁵Jones AC, Gosling SD. Temperament and personality in dogs (*Canis familiaris*): A review and evaluation of past research. *Applied Animal Behaviour Science* 2005; 95: 1-53.

⁶Svartberg K, Forkman B. Personality traits in the domestic dog (*Canis familiaris*). *Applied Animal Behaviour Science* 2002; 79: 133-155.

⁷Svartberg K. A comparison of behavior in test and everyday life: evidence of three consistent boldness-related personality traits in dogs. *Applied Animal Behaviour Science* 2005; 91: 103-128.

⁸Ley JM, Bennett PC, Coleman GJ. A refinement and validation of the Monash Canine Personality Questionnaire (MCPQ). *Applied Animal Behaviour Science* 2009; 116: 220-227.

⁹Ley JM, McGreevy P, Bennett PC. Inter-rater and test-retest reliability of the Monash Canine Personality Questionnaire- Revised (MCPQ-R). *Applied Animal Behaviour Science* 2009; 119: 85-90.

¹⁰Segurson SA. Managing and rehoming the rescue dog and cat. Ch. 23 in: BSAVA Manual of Canine and Feline Behavioural Medicine, 2nd ed. British Small Animal Veterinary Association, Quedgeley, Gloucester, Ch 23 pp. 270-280.

¹¹Christensen E, Scarlett J, Campagna M, et al. Aggressive behavior in adopted dogs that passed a temperament test. *Applied Animal Behaviour Science* 2007; 106: 85-95.

¹²Henessey MB, Voith VL, Mazzei SJ, et al. Behavior and cortisol levels of dogs in a public shelter, an exploration of the ability of these measures to predict problem behaviour after adoption. *Applied Animal Behaviour Science* 2001; 73: 217-233.

¹³Goddard ME, Beilharz RG. A factor analysis of fearfulness in potential guide dogs. *Applied Animal Behaviour Science* 1984; 12: 253-265.

¹⁴Segurson SA, Serpell JA, Hart BL. Evaluation of a behavioral assessment questionnaire for use in the characterization of behavioral problems of dogs relinquished to animal shelters. *Journal of the American Veterinary Medical Association* 2005; 227: 1755- 1761.

¹⁵Van der Borg JAM, Netto WJ, Planta DJU. Behavioural testing of dogs in animal shelters. *Applied Animal Behaviour Science* 1991; 32: 237-251.

¹⁶Weiss, E. Meet Your Match SAFER manual and training guide. ASPCA. meetyourmatch@apca.org.