BEHAVIOR ASSESSMENT TOOLS FOR DOGS IN A SHELTER SETTING

Sara L. Bennett, DVM, MS, DACVB
Outline

- **Temperament Tests**
  - Why do we need to assess behavior?
  - What is temperament?
  - What are we testing with behavior assessments?
  - Test classification

- Requirements for useful test
  - Validity (an aside on models)
  - Reliability
  - Standardization
  - Feasibility (an aside on food guarding)

- How do we use the data?
Why Do We Need to Assess Behavior?

- Behavior problems, including aggression, is a common reason for dogs to be relinquished to shelters
  - 3772 owners interviewed
    - 28.8% due to behavior of dogs (non-aggression)
    - 9.8% aggression to people
    - 7.8% aggression to animals
    - 11.9% of dogs had bitten a person

- Shelters need to identify animals at risk for behavior problems if adopted

Why Do We Need to Assess Behavior?

- **Prevent adoption of dangerous dogs**
  - Major public safety concern

- **Identify at-risk dogs**
  - Make educated adoption recommendations
  - Set up rehabilitation if an option
  - Screen for specific behavior problems

- **Diagnostic tool (?)**
  - Look for indicators of disease or poor welfare. ³,⁵

³ Diederich, Giffroy AABS 2006
⁵ Jones, Gosling AABS 2005
What is Temperament?

- Temperament
  - Pet’s natural or innate responses to its environment, *its emotional responses*
  - Considered stable over time
  - The foundation of personality
  - Character can influence it but cannot change it
What is Temperament?

- Assume that characters identified, however classified, are:
  - Present at an early age
  - Relatively stable over time
  - Expect the same response would be elicited in a similar set of situations
What is Really Being Tested?

- Simplest form:
  - Test a behavioral response
  - To a specific stimulus
  - At that time
  - In that environment
  - By that individual dog

- It is a behavioral evaluation, a snapshot in time
  - A “known”
What is Really Being Tested?

- Extrapolation:
  - Attempt 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

- Predictive or external validity
  - Not always a “known”
What is Really Being Tested?

☐ Things we do not know:

☐ What is the effect of stress of the shelter environment on the displayed behavior?
  ■ Cortisol, a hormone elevated during stress, is elevated at least initially in dogs when entering the shelter environment
    ■ Dogs should be given a period of adjustment prior to testing
      ■ How long? What if they continue to deteriorate?
What is Really Being Tested?

- Things we do not know:
  - Are we testing temperament, or something else?
    - Evidence some studies showed constructs (behavior responses) stable over time
      - Theoretically could measure temperament
    - Can valid assessment of temperament ever be determined in overwhelming and unnatural environment of a shelter?
    - Term “temperament test” has fallen out of favor
    - Behavior assessments themselves falling out of favor?
What is Really Being Tested?

- False positive results
  - Dogs are getting condemned needlessly

- False negative results
  - Dogs that pose a true public safety risk are still getting adopted out

10 Segurson BSAVA 2009, oakhilssgoldenslogo.jpg
Test Classification

- 4 Types of Test Classification:
  - Test Battery
    - Expose a dog to a variety of standardized stimuli that artificially replicate real life situations and record the dog’s reactions
  - Rating of Individual Dog
    - Information gathered about an individual dog’s behavior and history from an informant
      - Can be the owner, former owner, handler, or caretaker
Test Classification

- 4 Types of Test Classification:
  - Expert Rating of Breed Prototype
    - Experts on dogs (veterinarians, trainers, judges) are asked to describe, rank, or rate breeds of dogs as a whole rather than individual dogs
  - Observational Test
    - Dog exposed to a natural less controlled environment and behavior exhibited is recorded to describe broad conclusions
      - Used most often to select service dogs

5 Jones, Gosling AABS 2005
Test Classification

- Some types of aggression may not be identified through test battery in shelters\textsuperscript{11}
  - Intra-species, predatory, and territorial
- Tests of young puppy behavior not been shown to be a good predictor of adult behavior\textsuperscript{5,12,13}

5 Jones, Gosling AABS 2005
11 Christensen et al AABS 2007
12 Henessey et al AABS 2001
13 Goddard, Beilharz AABS 1984

www.tumblr.com
Outline

- Temperament Tests
  - Why do we need to assess behavior?
  - What is temperament?
  - What are we testing with behavior assessments?
  - Test classification

- Requirements for useful test
  - Validity (an aside on models)
  - Reliability
  - Standardization
  - Feasibility (an aside on food guarding)

- How do we use the data?
- Performing behavior assessments safely & fairly
Validity

- Are we really measuring what we think we are measuring?

- Validity
  - External (predictive) validity
    - Is the test a measure of the predicted behavior?³
      - Is an animal going to show a fear response in a similar situation in real life?

³ Diederich, Giffroy AABS 2006
freshpicsblogspot.com
Validity

- Internal validity
  - How well do the different variables of the test represent the desired trait to be evaluated and to what extent?
  - What is the extent to which a scale measures what it claims to measure?
  - Does the scary stimulus you are exposing the dog to create a fear response?
  - Do all dogs tested show fear? Or just a subset of them?
A Word on Models

- Repeated exposure to a threatening animal can be stressful and dangerous
  - Lead to learned fear and possibly aggression
  - A welfare issue
- Can we use a model to test for reaction to various stimuli?
  - Stuffed dog
  - Stuffed cat
  - Toddler or infant doll
  - Rubber hand on a stick
Models

  - Statistically significant correlation between behavior history and aggression to a model
    - Child directed aggression and stranger directed aggression with aggression to a doll device and overall avoidance
    - Dog directed aggression/ fear with aggression to a fake dog (at a distance)

- 88% (34) of dogs aggressive to children showed same behavior directed to doll
- 65% (37) dogs historically good with children showed no negative reaction to doll
- Dogs with fear aggression more likely to show fear to doll
- Dogs with fear aggression or owner directed aggression more likely to show aggression to fake hand

Still high level of false positives and false negatives
Models

- Reid and Collins. 2012 AVSAB/ ACVB Veterinary Behavior Symposium
  - ASPCA Anti-Cruelty Behavior Team, Urbana, USA
  - Missouri dog fighting raid
- Lifesize plush dog (Labrador retriever)
- Fake dog a valid stimulus
  - Dogs that reacted with aggression towards real dogs highly likely to display aggression toward the fake dog
  - Testing with model dog, then dog of opposite sex identified ~90% of aggression
Models

- Shabelansky, et al. 2015. Appl Anim Beh Sci
- Agreement between responses for fake dog versus real dog
- Shelter dog population
  - Friendly- good agreement
  - Fearful- moderate agreement
  - Aggressive- little agreement
- Importance of prevalence of problem in sample population
Reliability

- Test-retest reliability
  - Is the same dog scoring the same at a different time?  
  - Do you get the same result if the test is repeated?  
  - Measure the dog’s behavioral response  
  - Objective description of the behavior rather than a subjective interpretation of the dog  
    - Ideally an ethogram
Reliability

- Inter-rater reliability
  - Is there consistency between observers? ²
  - Different people simultaneously scoring or watching video of the test³

- Intra-rater reliability
  - Does the same person scoring the same test at different times get the same or similar results? ³
  - This person watching video back at a later date
Standardization

- The only variable is the animal being tested
  - Everyone is performing the tests the same way
  - Same equipment
  - Same environment

- Scoring systems
  - Correlate with anticipated behavioral responses
    - Pass/ fail
    - Numerical scoring (1-5)
    - Letter scoring (A-F)
Feasibility

- Time to perform
  - Efficient application in busy shelter setting

- Resources needed
  - Inexpensive
  - Easily obtained
  - Safe for dog and handler
Feasibility

- Scoring system
  - Clear and straightforward
  - Objective using ethogram
  - Not so simple lose vital information

- Staff training
  - Easily accessible
    - Training seminars, handbooks, video, webinars
  - Certification
Preliminary Investigation of Food Guarding Behavior in Shelter Dogs in the United States

Heather Mohan-Gibbons ¹, Emily Weiss ²,* and Margaret Slater ³

¹ Shelter Research and Development, Community Outreach, American Society for the Prevention of Cruelty to Animals (ASPCA®), Ojai, CA 93024, USA; E-Mail: heather.mohan-gibbons@aspca.org
² Shelter Research and Development, Community Outreach, American Society for the Prevention of Cruelty to Animals (ASPCA®), 6260 N. Hillside, Wichita, KS 67219, USA
³ Shelter Research and Development, Community Outreach, American Society for the Prevention of Cruelty to Animals (ASPCA®), 50 Stone Ridge Drive, Northampton, MA 01602, USA; E-Mail: margaret.slater@aspca.org

* Author to whom correspondence should be addressed; E-Mail: emily.weiss@aspca.org; Tel.: +1-207-644-8050.
96 dogs identified to guard food in shelter
- Placed in home with behavior modification plan
- 3 weeks
  - Only 6 adopters reported guarding
- 3 months
  - Only one guarding incident (rawhide)

Return rate
- 5% for guarding + dogs
- 9% for guarding – dogs

All adopters non-compliant with at least one part of behavior plan
Food-related aggression in shelter dogs: A comparison of behavior identified by a behavior evaluation in the shelter and owner reports after adoption

After adoption. The detection of FA+ via a behavior evaluation should be interpreted with caution, since a positive finding in the shelter evaluation does not consistently indicate that the behavior will occur in the home nor that a dog is unsuitable for adoption.

**Abstract**

Twenty (20.6%) dogs evaluated were deemed FA+ in the shelter, and slightly more than half (11/20; 55%) of them were later reported by adopters as exhibiting FA+ behavior in the home after adoption, whereas out of the 77 dogs that were deemed to be FA− in the shelter, 17 (22%) were reported to be FA+ by adopters; conversely, the majority (60/77; 78%) of dogs identified as FA− in the shelter were reported by adopters as not having exhibited FA+ behavior in the home ($P = 0.004$). Most adopters, including those whose dogs were reported FA+ in the home, did not consider FA+ behavior to be a challenge to keeping the dog as a pet. In conclusion, in this sample of shelter dogs, the observation of FA+ behavior during a standardized dog behavior evaluation was associated with FA+ behavior in the home following adoption, however, an almost equal number of dogs observed to be FA+ on the behavior evaluation did not show food aggression after adoption. Failure to
Outline

- Temperament Tests
  - Why do we need to assess behavior?
  - What is temperament?
  - What are we testing with behavior assessments?
  - Test classification

- Requirements for useful test
  - Validity (an aside on models)
  - Reliability
  - Standardization
  - Feasibility (an aside on food guarding)

- How do we use the data?
How Do We Use The Data?

- No tests so far have been published in a peer reviewed journal that have passed all of these qualifications\(^3,10\)
  - Validity
    - Predictive Value
    - Internal validity
  - Reliability
    - Test-retest
    - Inter-rater
    - Intra-rater
  - Standardization
  - Feasibility
- Important area for future research
RESEARCH

Temperament test for re-homed dogs validated through direct behavioral observation in shelter and home environment

Paola Valsecchi\textsuperscript{a}, Shanis Barnard\textsuperscript{a}, Cinzia Stefanini\textsuperscript{b}, Simona Normando\textsuperscript{c}

Abstract  This study was the first attempt to carry out a validation of a temperament test (TT) for shelter dogs that addressed the topics of inter- and intra-raters agreements, test-retest reliability, and validity. The TT consisted of 22 subtests. Each dog was approached and handled by an unfamiliar person and made to interact with a same- and an opposite-gender conspecific. Dogs were tested twice in the shelter and once in their new homes 4 months after adoption to evaluate consistency in behavioral assessment. Playfulness, trainability, problem solving abilities, food possessiveness, and reactivity to sudden stimuli were also evaluated. Testers scored dogs’ responses in terms of confidence, fearfulness, and aggressiveness. Results highlighted strengths and limits of this TT that was devised to help shelter staff in matching dogs’ personality and owners’ expectations. Methodological constraints when working with sheltered dogs are unavoidable; however, the test proved to be overall feasible, reliable, and valid although further studies are needed to address the critical issues that emerged.

© 2011 Elsevier Inc. All rights reserved.
How Do We Use The Data?

- Tests used in shelters
  - Untested permutations of other tests
  - Unvalidated

- Tests not being performed in a standardized fashion between staff members or organizations
  - Who, what, when, where, how scored
  - Australian shelter survey
How Do We Use The Data?

- Australian shelter study
  - 75% (26 staff members from 11 shelters) received some type of training
    - Confidence in ability to accurately assess a dog and suitability of test not related to experience/training
  - Improvements
    - Expand test to assess more behaviors
    - More time for assessment
    - 2 people present for assessments
  - 100% wanted scientifically validated and standardized protocol

Mornement et al JAAWS 2010
How Do We Use The Data?

The Use and Perception of Canine Behavioral Assessments In Sheltering Organizations

S. D’Arpino, S. Dowling-Guyer*, A. Shabelansky, A.R. Marder, G.J. Patronek

Center for Shelter Dogs, Animal Rescue League of Boston, MA, USA

* Corresponding author: sdowling@arlboston.org

AVSAB/ACVB Annual Symposium 2012

- 59.2 % reported dog’s behavior could limit admission to their organization
- 60.5% reported shelter resource restrictions (such as space) would limit admission
- 27.4% reported health issues could limit admission
  - Emphasize concern on what to do with dogs with behavior problems
The Use and Perception of Canine Behavioral Assessments In Sheltering Organizations

S. D’Arpino, S. Dowling-Guyer*, A. Shabelansky, A.R. Marder, G.J. Patronek

Only 28.4% of respondents reported their organization uses a formal behavior evaluation

- 59.8% of those using one designed themselves

Respondents want:

- Evaluations that predict dog’s behavior in home
- Identify problematic behaviors
- Evaluations to provide clear guidelines on suitability for adoption
Investigating behavior assessment instruments to predict aggression in dogs

Sara L. Bennett a, Annette Litster a,*, Hsin-Yi Weng b, Sheryl L. Walker a, Andrew U. Luescher c

 a Department of Veterinary Clinical Sciences, College of Veterinary Medicine, Purdue University, Lynn Hall, 625 Harrison Street, West Lafayette, IN 47907, USA
 b Department of Comparative Pathobiology, College of Veterinary Medicine, Purdue University, VPH, 785 Harrison Street, West Lafayette, IN 47907, USA
 c Fondation Barry, 1928 Martigny, Switzerland

ABSTRACT

The study aimed to assess the effectiveness of different behavior assessment instruments in predicting aggressive behavior in dogs. The instruments evaluated included the SAFE and mAAP tests. The results showed that SAFE was more sensitive and specific than mAAP. The odds ratio indicated that SAFE was 4.1 times more likely to identify an aggressive dog compared to mAAP. When the assessments were divided into multiple categories, SAFE showed a statistically significant but weak correlation of 0.34 (P = 0.005) with historical aggression categories. SAFE testing was unable to identify dogs with moderate aggression that could potentially be addressed with behavior modification.
Comparison of SAFER Behavior Assessment Results in Shelter Dogs at Intake and After a 3-Day Acclimation Period

Sara L. Bennett,1 Hsin-Yi Weng,2 Sheryl L. Walker,3 Margaret Placer,3 and Annette Litster3

In this study, it was hypothesized that different results would be obtained by canine behavior assessments performed within 24 hr of shelter intake (Day 0) and after a 3-day acclimation period (Day 3). Safety Assessment for Evaluating Behoming assessments were performed on 33 dogs at 2 municipal shelters. Agreements between Day 0 and Day 3 varied among subtests, and no consistent temporal patterns were observed. Weighted kappa statistics for each subtest ranged from .28 to .78, and percentage discordance was 0% to 18%. In a 2nd analysis, subtests skipped due to serious aggression were replaced with scores corresponding to serious aggression, and missing values for the Food subtest were replaced with scores for no aggression if the dog did not eat. For subtests skipped due to severe aggression, more than 50% of the dogs had scores indicating low aggression on the other assessment. Eight of 16 dogs who did not eat on Day 0 ate on Day 3; 2 showed aggression. Until the ideal time to test can be identified, it should be based on the individual dog’s welfare status, and testing of dogs showing severe stress should be avoided.

Keywords: acclimation, behavior assessment instrument, shelter intake, dogs, temperament test
How Do We Use The Data?

- Used as an ancillary tool added to information from other sources
  - Owner relinquishment questionnaires
    - Public vs confidential
    - Confidential: owner directed aggression, fear of strangers\(^\text{14}\)
  - Staff experiences with the animal
    - Informal observational test
    - Better at predicting absence of a problem\(^\text{15}\)

14 Segurson et al JAVMA 2005
15 VanDerBorg et al AABS 1991
How Do We Use The Data?

Canine Research

No better than flipping a coin: Reconsidering canine behavior evaluations in animal shelters

Gary J. Patronek a,*, Janis Bradley b

a Center for Animals and Public Policy, Cummings School of Veterinary Medicine, Tufts University, North Grafton, Massachusetts
b The National Canine Research Council, Amenia, New York
How Do We Use The Data?

Abstract

Use of behavior evaluations for shelter dogs has progressed despite their lack of scientific validation as reliable diagnostic tools. Yet results of these evaluations are often used to make life-and-death decisions. Despite acknowledging the significant limitations of evaluations, most authors suggest that the solution is to continue to attempt to remedy deficiencies. We take a contrary position and use existing data and principles of diagnostic test evaluation to demonstrate that reliably predicting problematic behaviors in future adoptive homes is vanishingly unlikely, even in theory, much less under the logistical constraints of real-world implementation of these evaluations in shelters. We explain why it would be difficult, if not impossible, to calculate robust values for sensitivity and specificity of a shelter canine behavior evaluation as required for any valid diagnostic test. We further explain the consequences of disregarding the effect of prevalence on the predictive value of a positive test (e.g., eliciting biting or warning behavior from the dog in the behavior evaluation). Finally, we mathematically demonstrate why, for any plausible combination of sensitivity, specificity, and prevalence of biting and warning behaviors, a positive test would at best be not much better than flipping a coin, and often be much worse, because many of the dogs who test positive will be false positives. Shelters already screen from adoption obviously dangerous behavior and may already be selecting against behaviors that are difficult to evaluate.
How Do We Use The Data?

- Sensitivity and Specificity
  - **Sensitivity**: ability to correctly identify those that are diseased
    - Measures level of false negatives
  - **Specificity**: ability to correctly identify those that are healthy
    - Measures level of false positives

<table>
<thead>
<tr>
<th></th>
<th>Disease present</th>
<th>Disease absent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test positive</td>
<td>a (TP)</td>
<td>b (FP)</td>
</tr>
<tr>
<td>Test negative</td>
<td>c (FN)</td>
<td>d (TN)</td>
</tr>
<tr>
<td><strong>Sensitivity</strong>:</td>
<td>a/ (a+c)</td>
<td><strong>Specificity</strong>:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>d/ (b+d)</td>
</tr>
</tbody>
</table>

TP: True positive, FP: False positive, FN: False negative, TN: True negative
How Do We Use The Data?

- **Predictive value** - based on prevalence of problem in the population
  - [https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2636062](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2636062)

- **Positive predictive value**
  - Percentage of patients with positive test who actually have the disease
  - \( \frac{a}{a+b} \) (true positive + false positive)

- **Negative Predictive Value**
  - Percentage of patients with negative test who do not have the disease
  - \( \frac{d}{c+d} \) (false negative + true negative)
For every 100 shelter dogs tested, if ~16% of the population expressed behaviors of concern and test sensitivity=92%, specificity=36%

16 dogs would exhibit problematic behavior in the home if adopted
~15 dogs would test positive (true positive)
~1 dog would test negative (false negative)

84 dogs would NOT exhibit problematic behavior in the home if adopted
~54 dogs would test positive (false positive)
~30 dogs would test negative (true negative)

Result: For every 10 dogs who tested positive and were not adopted, this is what behavior in the home would look like if adopted.

For every 100 shelter dogs tested, if ~16% of the population expressed behaviors of concern and test sensitivity=85%, specificity=85%

16 dogs would exhibit problematic behavior in the home if adopted
~14 dogs would test positive (true positive)
~2 dogs would test negative (false negative)

84 dogs would NOT exhibit problematic behavior in the home if adopted
~13 dogs would test positive (false positive)
~71 dogs would test negative (true negative)

Result: For every 10 dogs who tested positive and were not adopted, this is what behavior in the home would look like if adopted.

---

**Figure 2.** Results of a behavior evaluation using realistic values* (top) and unrealistically optimistic values** (bottom) for key attributes of a diagnostic test. *Values used are those from a meta-analysis of instruments predicting violent offending in people (Fazel et al., 2012). **This combination of values exceeds what is commonly reported for many validated human diagnostic tests, which usually involve trade-offs between sensitivity and specificity. *Numbers listed are approximate due to rounding of fractions. With a positive predictive value = 22%, 78% (n = 54) of the 69 dogs testing positive will be false positives.
Figure 4. Summary of problems with canine behavior evaluations in animal shelters. *See text for details on how value was derived. ‡Assuming a validated test was available for shelter; see Figure 3 for values.
How Do We Use The Data?

- What do you think about their recommendations?
“Shelters already screen from adoption obviously dangerous dogs during the intake process.”

- Do they all do this?
- If so, how do they do this?
- Is it a valid method?
How Do We Use The Data?

- Where do we go from here?
  - More research!

- If suggest eliminating formal behavior assessments
  - Rely on other forms of data collection

- Must ensure these are valid behavior measures
  - Predict presence or absence of behavior in the home
  - Cannot assume
How Do We Use The Data?

- Collect behavior history at relinquishment
  - Owner
  - Good Samaritan
  - Animal Control Officer

- Collect behavior observations at intake exam
  - Medical exam is intensive interaction!

- Collect behavior observations during daily care and enrichment
  - Must log data!
  - Objective behavior descriptions- positive and negative behaviors
How Do We Use The Data?

- Use formal behavior assessment as form of “behavioral physical exam” if resources allow and concerns noted
  - Eg. dog to dog with fake dog if concern for dog aggression
  - Ensure staff are thinking about what they want to know and what they are really measuring
    - Does data collected change what I am going to do with pet?
  - Find abnormality- compare with other data already have, determine how to collect additional information in order to make working diagnosis
Conclusions

- Need to consider what we are measuring and understand assumptions being made
- Need to recognize consequences of false positive and false negative results and importance of interpreting in face of population prevalence
- If decide to no longer perform formal behavior assessments, need to have a way to measure and observe behavior safely
  - Use additional information to help guide decision making process for ultimate plan for shelter dog
  - Prior history, staff observations, observational behavior assessment?
- First priority- Safety of handlers and dog
  - Observe and record behavior, not interpretations
Thank You for Your Time!

- Sara L. Bennett, DVM, MS, DACVB
  - Veterinary Behavior Specialist
    - sbennettdvm@gmail.com
  - Behavior consultation, client or patient inquiries
    - DrB.Behavior@gmail.com
    - P: 812-550-1033
    - F: 888-726-9034
  - www.DrBennettBehavior.com