Journal Pre-proof

Assessment of a behavioral scale for the measurement of fear, anxiety and stress in dogs visiting the veterinary practice

Emma Gatehouse, Annika Bremhorst, Sagi Denenberg, Loni Loftus

PII: S1558-7878(25)00084-X

DOI: https://doi.org/10.1016/j.jveb.2025.08.008

Reference: JVEB1735

To appear in: *Journal of Veterinary Behavior*

Received date: 11 May 2025
Revised date: 27 August 2025
Accepted date: 30 August 2025

Title page

Title: Assessment of a behavioral scale for the measurement of fear, anxiety and stress in dogs visiting the veterinary practice

Author names

Emma Gatehouse[a, b], Annika Bremhorst[c, d], Sagi Denenberg[b, e], Loni Loftus[a, f]

[a] The Royal (Dick) School of Veterinary Studies and The Roslin Institute, Easter Bush Campus, Midlothian, EH25 9RG

[b] Langford Vets, Bristol, England

[c] Dogs and Science, Switzerland

[d] University of Bern, Switzerland

[e] North Toronto Veterinary Behaviour Specialty Clinic, Toronto, Canada

Corresponding author:

- Emma Gatehouse (emma.gatehouse@bristol.ac.uk)
- Langford Vets, Bristol, BS40 5DU

Email addresses of authors:

- Annika Bremhorst annikahubercdl@gmail.com
- Sagi Denenberg sagidvm@gmail.com
- Loni Loftus loni.loftus@ed.ac.uk

Abstract

Validated, reliable instruments for assessing canine stress behaviors during veterinary visits are crucial in mitigating escalating behavioral responses, thereby maximizing patient welfare, facilitating accurate diagnoses and enhancing staff safety. Although various behavioral assessment tools have been developed, few have been evaluated specifically for the veterinary context. This study addresses this deficit by evaluating the Spectrum of Fear, Anxiety and Stress, (FAS, Fear Free® 2022) an eight-item scale, graduating from 0 to 5, designed to assess canine fear, stress and anxiety during veterinary visits. An online survey featuring 14 video recordings of dogs undergoing mock veterinary examination was distributed to 79 participants, including dog owners, veterinarians, behavior experts and trainers. This study assessed inter- and intra-rater reliability, concurrent validity, and ease of use. Scores at either end of the spectrum had the highest percentage of correct responses (FAS 0 Relaxed; 51.90% incorrect, FAS 4 Severe signs; 44.30% incorrect), while mid-range scores were more challenging for participants to correctly identify (FAS 0-1 Perked/Interested/Anxious; 72.78% incorrect). Behavior experts and owners significantly differed in their ability to assess some moderate and severe signs (FAS 3 Moderate signs; $p = 0.0025$, FAS 4a Flight; $p = 0.0355$), suggesting that experience in assessing dog behavior may impact the ability to identify fear, stress and anxiety correctly. Inter-rater reliability was excellent (ICC = 0.99 with 95% confidence interval [0.99–1.00]), intra-rater reliability was very good (ICC = 0.83 with 95% confidence interval [0.80-0.86]), and a strong correlation was found between participant FAS scores and cumulative scores on the Lincoln Canine Anxiety Scale ($\rho = 0.811$, $p =< 0.001$, $n = 79$), suggesting good concurrent validity. Thematic analysis praised the scale's visual aspects, but revealed challenges related to overlapping categories and unfamiliar numbering. The results of this research support further development, including some minor design adjustments and sufficient participant training, of the FAS Spectrum as a valid and reliable behavioral assessment tool for evaluating acute stress in dogs visiting the veterinary practice, in order to provide reliable behavioral assessment to facilitate stress reduction in clinic.

Keywords

Behavior assessment; Veterinary practice; Stress; Anxiety; Fear; Dog

Introduction

Veterinary visits are a highly stressful experience for up to 78.5% of dogs (Döring et al., 2009; Edwards et al., 2019), given the exposure to potentially aversive stimuli such as unfamiliar people and animals, owner separation, aversive noises, novel odors, pain, injury to physical well-being and diagnostic interventions (Horwitz and Mills, 2012; Overall, 2013; Mills et al., 2020). “Stress,” for the purpose of this study, unless otherwise defined, refers to high arousal emotional state with negative valence (Pfaff et al., 2007).

The effect of fear and anxiety, defined as emotional responses to immediate or impending threats (Steimer, 2002), on the development of pre-pathological and pathological states and their subsequent effect on the quality of life, is increasingly recognized within veterinary medicine (Mcauliffe et al., 2022). The examination, diagnosis and treatment of animals exhibiting high levels of stress raises concerns for both human caregivers and the animal's welfare (Campbell, 1975), limiting the ability to adequately treat the individual (Döring et al., 2009). The presence of stress-related behaviors may also hinder the detection of subtle disease signs, especially pain (Lind et al., 2017). Early recognition and intervention to improve the emotional state of patients is vital for successful patient outcomes, staff safety and satisfaction and positive clinic-client relationships (Döring et al., 2009; Csoltova et al., 2017; Lefman and Prittie, 2019; Riemer et al., 2021).

Behavioral assessment of fear and anxiety in dogs has typically been attempted using various methods including the use of observational, subjective categories (Stanford, 1981; Stephen and Ledger, 2005), recording the frequency of behaviors shown, (Döring et al., 2009) ranking systems, cumulative stress scores or ordinal scales (Hauser et al., 2020; Kim et al., 2022; Korpivaara et al., 2022). However, poorly defined behavioral categories and the continuous, progressive nature of these behaviors, which can occur concurrently, often leads to ambiguous, subjective descriptions of behaviors (Hauser et al., 2020). While some behavioral scales have been validated for situational use (Mills et al., 2020), accessible, reliable, and validated tools specifically tailored for use in general veterinary practice are currently lacking (Diederich and Giffroy, 2006). Given that behavioral responses are often stressor and context-specific (Moberg and Mench, 2000), it is beneficial to develop, validate and utilize instruments explicitly for the environment in which they will be applied. While multiple publications have looked to identify the presence of physiological and behavioral manifestations of stress in the veterinary practice, many focus on hospitalized (Hekman et al., 2014; Hauser et al., 2020), painful or post-surgical dogs (Siracusa et al. 2008; Hekman et al., 2012), with non-validated scales, limiting their generalization and reliability. Existing veterinary-specific scales, such as the Clinic Dog Stress Scale (Overall, 2013), offer low (King et al., 2022) to moderate inter-rater reliability (Mercier et al., 2023; Jokela et al., 2023). Tools designed to measure physiological changes associated with stress, such as temperature changes (Travain et al., 2014) may also present problems, such as the effect of utilizing the often-invasive equipment required to monitor changes, whereas behavioral observation can occur with limited equipment in almost any scenario. Individual physiological parameters cannot adequately determine stressful state alone (National Research Council [US], 2008) and may deviate in response to non-stress states (Beerda et al., 2000) including appetitive stimuli (Koolhaas et al., 2011), meaning that assigning valence to deviations in these physiological parameters can be problematic. Multi-modal assessment is the preferred strategy in research but may not be practical when time is short or when a tool is being used by owners. Therefore, assessing the reliability and validity of existing behavioral scales and refining these where necessary may be a pertinent starting point to establish a unified framework for evaluating animal stress and well-being in clinical practice.

The Fear Free® Spectrum of Fear, Anxiety and Stress (Fear Free, 2022) is an eight-point ordinal scale designed, through expert consensus, to assess the behavior of dogs visiting the veterinary practice. The spectrum represents an ethogram of canine behaviors, depicted with illustrations, descriptions and a color-coded traffic light system, observed during an acute stress response, allowing users to assign a FAS score. A low FAS score (0) is represented by green, and a high FAS score (5) is represented by red (see Table 1 for detailed FAS scores and supplementary material S1 for the full FAS spectrum).

The FAS spectrum is widely disseminated by veterinary clinics, shelters and individuals under the Fear Free® Certification scheme. With a growing number of Fear Free® certified professionals within the UK, the FAS spectrum provides a useful tool. However, while reliability has been assessed for use by behavioral specialists (Mercier et al., 2023) suggesting that the spectrum can be used reliably by this subset of people, accuracy, wider reliability (specific to the intended veterinary context) and validity are yet to be assessed. A behavioral assessment tool should be standardized, reliable, and valid (Bateson and Martin, 2021). Several forms of validity including content (whether all aspects of a construct are covered by an assessment tool), construct (how well a tool measures the theoretical construct intended) and criterion (how well the tool measures against an external "gold standard" metric) validity(Bateson and Martin, 2021) . Assessment of criterion validity is challenging within the assessment of behavioral instruments, where standards are continually updated and adapted and where a gold standard does not yet exist (Bellamy, 2015). Taylor and Mills (2006) proposed that concurrent validity, defined as how well one test compares to another, more established measure of the same construct, may be a more suitable assessment of validity of a behavioral tool. Therefore, this study aimed to assess validity through comparison to an existing validated scale. The Lincoln Canine Anxiety Scale (LCAS, Mills et al., 2020), has been validated for assessment of anxiety and fear responses specifically to fireworks. However, Mills et al. (2020) suggest that the scale may be applicable to broader contexts, providing a comparator for assessment of the concurrent validity of a veterinary-specific scale.

This study aims to provide a scientific approach to address the paucity of available instruments by evaluating the FAS spectrum via assessment of: (1) inter-rater reliability; (2) intra-rater reliability; (3) concurrent validity (via evaluation with the LCAS) and (4) ease of use of the FAS spectrum. It was hypothesized that behavior experts and dog trainers would be more likely than veterinarians or dog owners to correctly assign FAS scores and that displayed behaviors indicative of FAS scores at extreme ends of the spectrum would be easier for all participants to identify.

Materials and methods

Study design

This two-phase study utilized a mixed-methods approach (for the full study design, see Figure 2). In Phase One, the researcher (EG) collected and scored videos of dogs during a standardized veterinary examination. Selected videos from this phase were then scored in Phase Two by dog owners, dog trainers, veterinarians and behavior experts using FAS scoring.

Phase One; Video recording

Eighteen dogs (eight male neutered, two male entire, six female neutered, two female entire) of various breeds (five Labrador retrievers, two spaniel crosses, one boxer, one Dalmatian, one Boston terrier, one Chihuahua, one Jack Russell terrier, one pug, one beagle, one husky, one Clumber spaniel, one golden retriever and one large crossbreed) ranging between eight months and twelve and a half years old, participated in Phase One of the study. Recruitment was conducted using convenience sampling of dog owners registered with Langford Vets (LV), Langford, United Kingdom (UK), attending routine veterinary appointments. Inclusion criteria were age (>8 weeks), time with the owner (>1 week), received first puppy vaccinations and were clinically healthy at the time of participation.

All video recordings took place at LV in a dedicated consultation room (4.5 m x 3.7 m; see Supplementary Material S2). Two video cameras (Canon Legria HFR206 and Canon Legria HF200) were used to ensure maximum video data capture. To maintain anonymity, the cameras were positioned to avoid filming the owner's upper body. See Supplementary Materials S2 and S3 for examples of camera angles.

Step 1 - Habituation

Dogs were habituated to the novel consultation room and researcher with a five-minute habituation phase (following Uccheddu et al. (Uccheddu et al., 2022) and Souza et al. (2023). During this time, demographic information, including the breed, age and sex of the participating dogs, was recorded. The owner was instructed to interact normally with their dog, while the researcher avoided interaction.

Step 2 – Standardized vet examination

Following habituation, a standardized mock veterinary examination, following an adapted methodology from previous studies examining related research questions (Godbout et al., 2007; Csoltova et al., 2017) was performed on the floor by the primary researcher (EG, Figure 1). The examination included the following standardized steps: examination of the eyes, ears, teeth and oral mucosa, mandibular lymph nodes, cardiac auscultation, abdominal palpation, and limb palpation. Refusal of examination, determined by struggling for three seconds, backing away or reluctance to approach the researcher, was allowed to utilize recommended low-stress examination techniques (Fear Free, 2023)to prevent eliciting or influencing the dog's FAS score.. In case of refusal, the researcher used the dog's name and gentle verbal encouragement to see if they could motivate the dog to move back towards them to continue the examination. If a dog refused an examination step three times, the researcher moved on to the next examination step (Figure 1). If the dog refused to engage with the researcher at all (e.g., refused the first examination step and then refused to reapproach), the examination was terminated after three refusals for each body part stage.

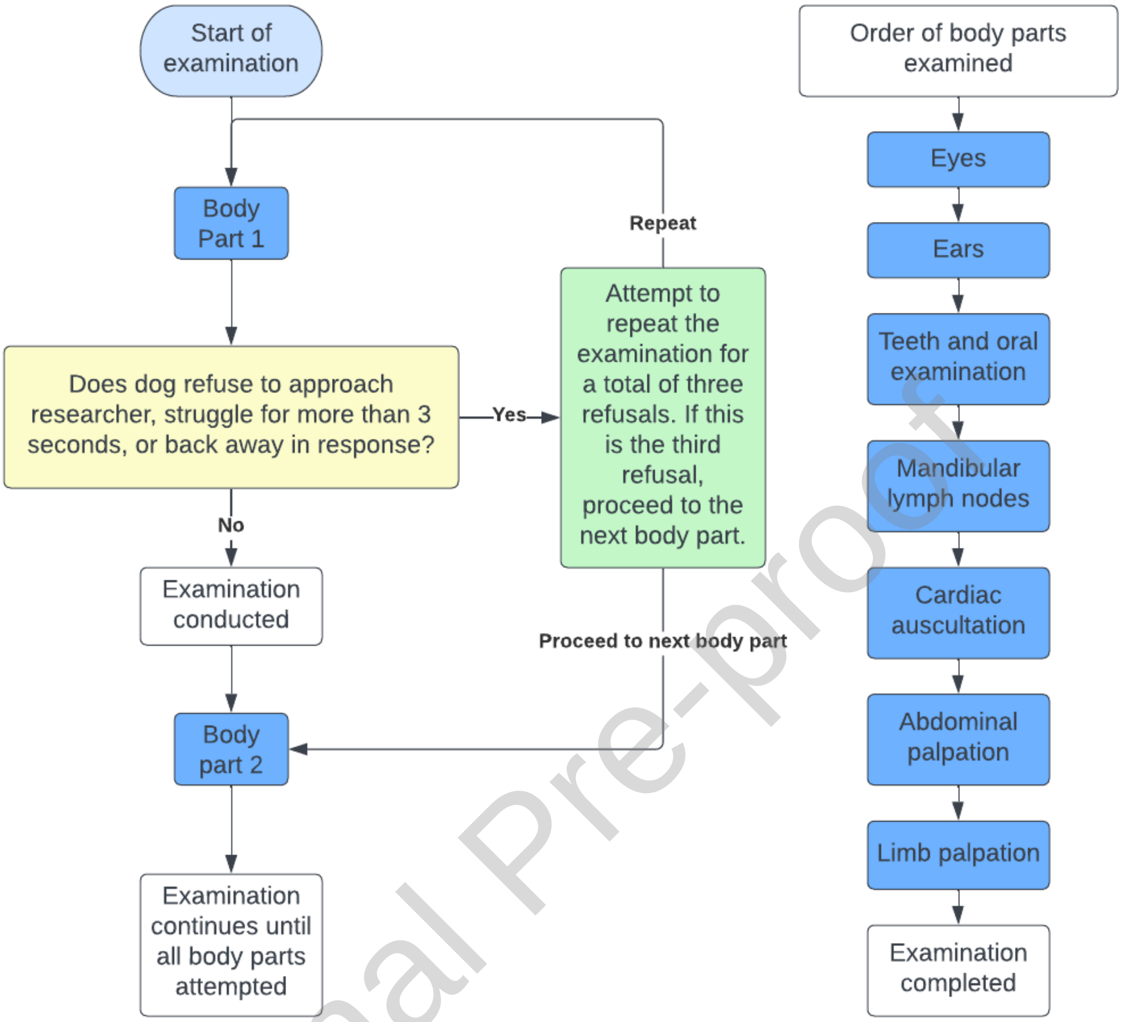

Figure 1: Flow charts detailing the standardized mock veterinary examination process, acknowledgement of refusals and body parts assessed.

Video selection and expert scoring

Thirty-six videos were collected (two for each participating dog) and segmented into 324 (10-30-second) video clips of distinct behaviors. For example, the dog remained in a similar state of emotional arousal, and unsuitable clips were removed. Exclusion criteria for video clips included obscuring the dog, the dog exiting the camera view for more than five seconds, lighting problems, the caregiver's interaction with the dog (for example if they dramatically influenced their behavior), or behaviors not indicated on the FAS spectrum. Example video clips are provided within the supplementary material (S4). Two qualified observers (EG and LL), working in clinical animal behavior, independently assigned an FAS score to each video clip, termed "actual FAS score." Videos with 100% inter-rater agreement (178 clips) in FAS score were selected for possible use within Phase Two of this study; all other video clips were excluded.

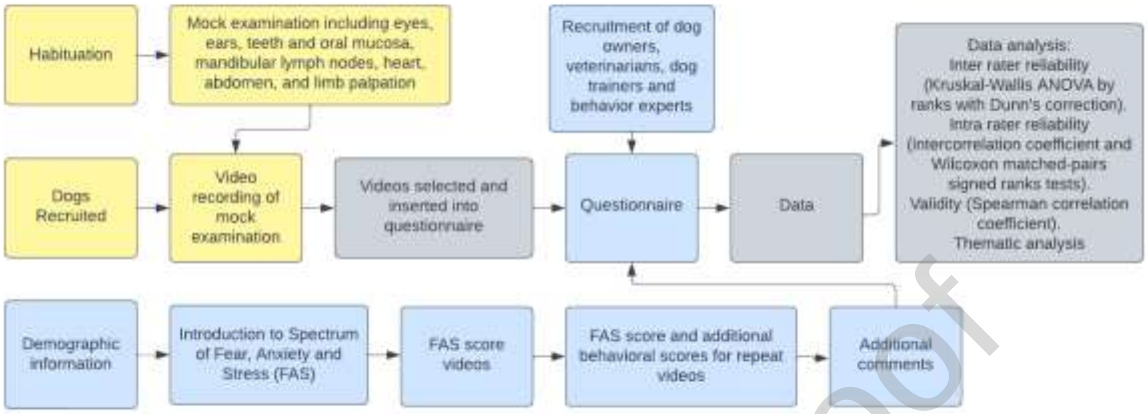

Figure 2: Flowchart depicting the experimental design, including Phase One (yellow): Video sampling of dogs during standardized veterinary examination, and Phase Two (blue): FAS scoring via questionnaire dissemination. Grey boxes represent data and processing steps.

Phase Two – Participant assessment

Video preparation

Two videos from each FAS score 0-4 (Table 1), totaling 14 videos (Figure 3) were randomly selected, by listing the video numbers and using the randbetween function within Microsoft Excel, from those collected in Phase One and inserted into Section 3 of the questionnaire. No videos demonstrated behaviors classified as FAS score 5, likely due to ethical concerns associated with recruiting dogs for the study (see Discussion).

From the 14 (seven pairs) selected videos, one video from each pair was randomly selected (using the randbetween function in Microsoft Excel) and inserted into the repeat video scoring section 4 of the questionnaire (Figure 3).

Table 1. The distribution of selected videos into FAS categories.

FAS Category	Videos
Relaxed (FAS 0)	5 and 12
Perked/Interested/Anxious (FAS 0-1)	9 and 14
Alert/Excited/Anxious (FAS 0-1)	3 and 8
Mild/subtle signs (FAS 1)	2 and 7
Moderate signs (FAS 2)	6 and 11
Moderate signs (FAS 3)	1 and 4
Severe signs - Flight (FAS 4a)	10

Severe signs - Freeze/ Fret (FAS 4b)	13
Severe signs - Fight/Aggression - Offensive (FAS 5a)	None displayed
Severe signs - Fight/Aggression - Defensive (FAS 5b)	None displayed

Survey

A survey (See supplementary material S5), including demographic information questions, an introduction to the Spectrum of Fear, Anxiety, and Stress (FAS), two video scoring sections and a participant feedback section, was developed and distributed using an online survey platform (Jisc Online v2) (Figure 3). Participants were provided with study information and data protection statements and asked to give their consent before completing the survey.

During the first video scoring section (Section 3, Figure 3) participants were presented with 14 video clips selected during Phase One (see Phase One: Video Selection and Expert Scoring and Phase Two; Video preparation). For each video, participants were asked to assign a FAS Score, including a subcategory e.g. FAS 4; 4a Flight or 4b Freeze/Fret, if relevant. When repeat scoring a selection of videos in Section 4, participants were shown seven repeated video clips (one from each FAS score 0 to 4) randomly selected from the two clips of that FAS score used in Section 3 and asked to rescore these videos using the FAS spectrum. These seven videos were further assigned an average cumulative score using eight items from the LCAS (Mills et al., 2020), which crossed over with the Spectrum of FAS (see Supplementary Material S5). These eight items included Hiding (e.g., under furniture or behind owner), Cowering (e.g. tucking tail and flattening ears), Restlessness/pacing, Aggression (e.g., growling, snapping or biting), Freezing to the spot, Panting, Owner-seeking behavior, and Shaking or Trembling. These scores were then used to provide a cumulative LCAS score for each video.

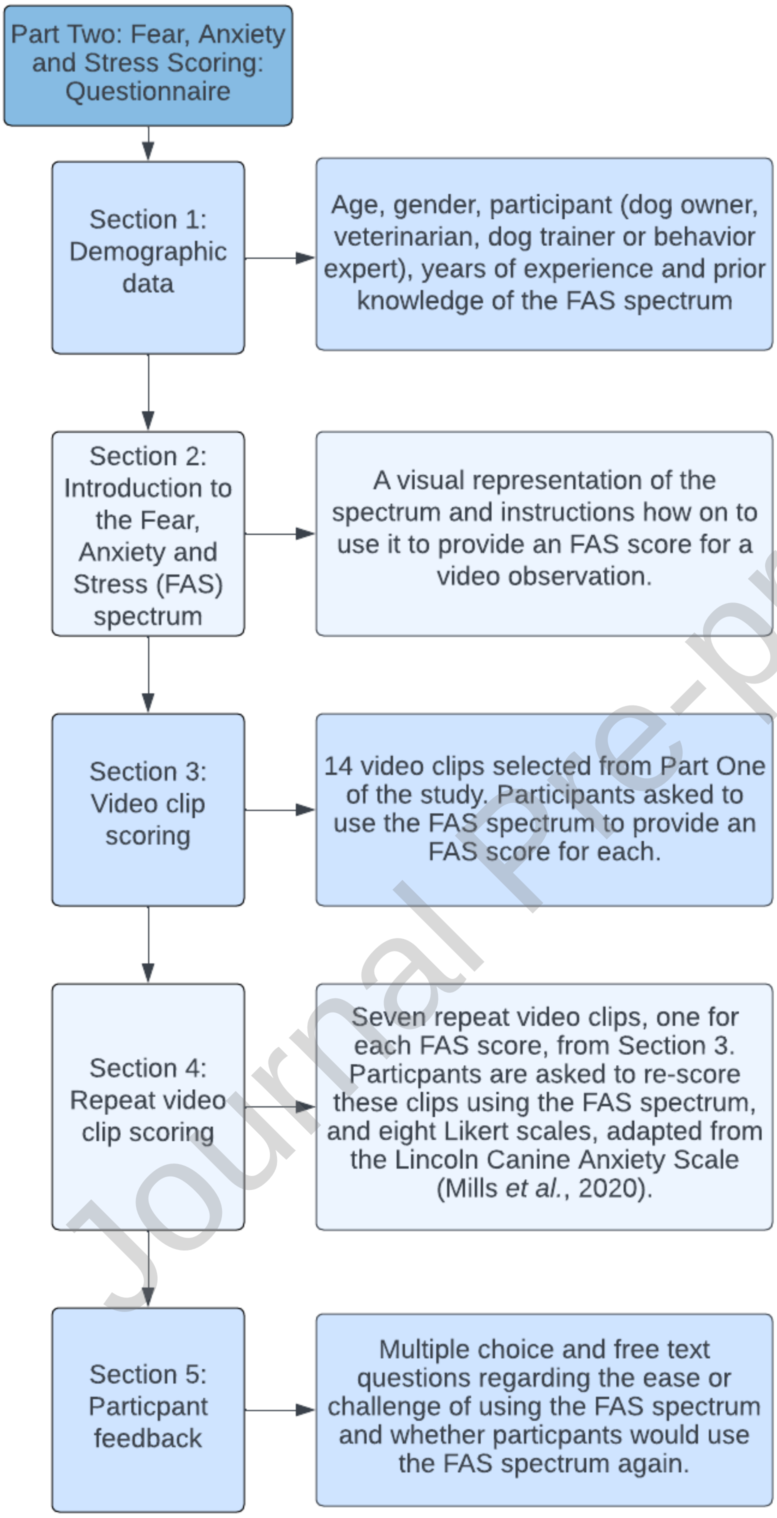

Figure 3: Flowchart depicting the sections of the survey disseminated to participants. Boxes on the left of the flow chart diagram represent the sections of the survey participants worked through, with descriptions of what each section involved on the right-hand side of the diagram. The survey was unidirectional, and participants could not go back to previous pages after they had left them.

Subjects

Participants (subjects) were recruited through social media groups, as well as dog behavior, training and veterinary organizations, including the Animal Behaviour Training Council (ABTC), Association of Pet Behavioural Counsellors (APBC), Association of Pet Dog Trainers (APDT), Fellowship of Animal Behaviour Clinicians (FABC) and LV. To be included in the study, participants must have been at least 18 years of age, a resident of the UK, and belong to one of the groups with the inclusion criteria listed in Table 2. The population of Dog Owners used in Phase Two was different from those participating with their dogs in Phase One.

Table 2. *Inclusion criteria for each participant group recruited for questionnaire completion.*

Participant (subject) group	**Specific inclusion criteria**
Dog owner	Own at least one dog at the time of the study
Veterinarian	Working in clinical practice within the UK
Dog trainer	Registered as an animal trainer, training instructor, or behavior technician with the ABTC, APDT, or APBC.
Dog behavior expert	Obtained a postgraduate degree in a behavior qualification or be registered as a Clinical Animal Behaviorist or Certified Clinical Animal Behaviorist with the ABTC, APBC, or CCAB.

Data analysis

Demographic data were analyzed to produce descriptive population summaries. To assess inter-rater reliability, three measures were calculated: the interclass correlation coefficient (ICC), proportion of incorrect scores and the proportion of error scores. ICC estimates and their 95% confidence intervals were calculated using a two-way mixed effects model, seeking absolute agreement (ICC 3,k) across all fourteen videos and raters. The proportion of incorrect scores was calculated by subtracting the participant's FAS score from the actual FAS score assigned by researchers (see Video

Selection and Expert Scoring). Scores were coded (correct = 0, incorrect = 1), with differences of zero considered correct and all other differences considered incorrect. The proportion of incorrect participant scores by video, FAS score and participant group was calculated manually and expressed as a proportion and percentage. Error scores were calculated to determine the size of the difference between participant scores and the assigned FAS (how incorrect a participant was, rather than just correct or incorrect). This was determined by subtracting the participant's FAS score from the actual FAS score. The absolute value of differences was taken. To statistically evaluate the differences between the participant group's error scores for each video, analyses were conducted using non-parametric tests (Kruskal-Wallis ANOVA by ranks with a Dunn's correction for multiple comparisons), as data were not normally distributed, as identified by the Shapiro-Wilk test (Behaviour experts; $W = 0.5026$ $p < 0.0001$, Dog trainers; $W = 0.7472$ $p = <0.0001$, Owners; $W = 0.8080$ $p = <0.0001$, Veterinarians; $W = 0.8030$ $p = p = <0.0001$).

To assess intra-rater reliability, the intraclass correlation coefficient (ICC) was calculated using a two-way mixed model (ICC 3,k), seeking absolute agreement across all seven repeated videos and all raters. Significant differences between first and second scoring by participant group were evaluated using Wilcoxon matched-pairs signed-rank tests to determine whether participants assigned significantly different FAS scores when viewing videos ($n = 7$) for a second time, compared to their scoring during the first viewing. For this study, intra-rater reliability was defined as no significant difference between the first and second scorings of the same video (scores from questionnaire sections 3 and 4 compared) across the seven repeated videos and all participants.

To assess concurrent validity, the analysis of the relationship between participants' FAS Scores and their cumulative LCAS scores for each video was undertaken via Spearman correlation coefficients for each participant group. Actual and participant cumulative LCAS and FAS scoring data were not normally distributed. Therefore, Spearman rank correlations were used instead of Pearson's correlation coefficients in the following analyses. Correlations are described according to a weak correlation 0 to 0.3, moderate correlation 0.3 to 0.7 and strong correlation 0.7 to 1.0 (Dancey and

Reidy, 2007). Scatter graphs were plotted to visualize the relationships within these data.

To assess ease-of-use, thematic analysis was performed on participant comments from the questionnaire, in response to the prompts “What was good about using the FAS scale?” and “What was challenging about using the FAS scale?” by generating codes to participant answers and evaluating for the themes presented (See Supplementary Material S6). The themes were then reviewed and refined by the primary author (EG). Ease of use of the FAS spectrum, as answered by participants, was also reported.

Results

Due to the requirement to complete all questions, no data were missing.

Demographic data

Seventy-nine participants completed the questionnaire, comprising forty-five dog owners (56.95%), twenty-one veterinarians (26.58%), eight behavior experts (10.13%), and five trainers (6.33%). The sample was predominantly female (n = 72, 91.14%) and ranged in age from 21 - 75 years old (M = 40; SD ± 13). Most participants (n = 48, 60.76%) had no prior knowledge of the FAS Spectrum.

Inter-rater reliability

Inter-rater reliability was excellent (Intraclass correlation coefficient (ICC 3,k), 0.99 , (95% CI 0.99 – 1.00) for all videos and participants’ scores. ICC by participant group for all scores were excellent, as follows; dog trainers ICC 0.97 (95% CI 0.93 – 0.99), veterinarians ICC 0.97 (95% CI 0.95 – 0.99), owners ICC 0.99 (95% CI 0.98 – 1.00) and behavior experts ICC 0.97 (95% CI 0.93 – 0.99).

The average percentage of incorrect scores across all videos and FAS scores was 64.92% for the owners, 60.36% for the dog trainers, 57.14% for the veterinarians and 50.89% for the behavior experts (Table 3).

Table 3: Proportion and percentage of incorrect scores for each video, overall and per participant group. The following acronyms indicate an FAS category, as noted in Table 1: PIA 0-1 indicates “Perked/Interested/Anxious,” and AEA 0-1 indicates “Alert/Excited/Anxious.” 4a and 4b indicate FAS 4 Flight and Freeze/ Fret, respectively.

FAS Score	Video	Proportion of Incorrect	% Incorrect Overall	% Incorrect Owner	% Incorrect Trainer	% Incorrect Behaviour Expert	% Incorrect Vet
0	5	0.53	53.16	57.78	60	37.5	47.62

0	12	0.51	50.63	55.56	80	12.5	47.62
PIA 0-1	14	0.86	86.08	88.89	80	87.5	80.95
PIA 0-1	9	0.59	59.49	62.22	60	37.5	61.9
AEA 0-1	8	0.72	72.15	73.33	60	75	71.43
AEA 0-1	3	0.46	45.57	51.11	60	37.5	33.33
1	2	0.68	68.35	66.67	100	62.5	66.67
1	7	0.59	59.49	68.89	60	25	52.38
2	6	0.76	75.95	80	100	87.5	57.14
2	11	0.68	68.35	71.11	80	75	57.14
3	1	0.73	73.42	86.67	20	37.5	71.43
3	4	0.56	55.7	57.78	60	75	42.86
4a	10	0.49	49.37	48.89	20	12.5	71.43
4b	13	0.39	39.24	40	20	50	38.1
			Average % Incorrect:	64.92	61.43	50.89	57.14

Across all participant categories, video 14 (FAS score PIA 0-1) had the highest percentage of incorrect scores (86.08%) and video 13 (FAS score 4b Freeze/Fret) had the lowest (39.24%) (Figure 4). The percentage of incorrect scores for all participant groups, categorized by FAS score, is shown in Figure 5. FAS score 0 (51.90%) and 4 (44.30%) had the lowest percentage of incorrect answers, with FAS score PIA 0-1 (72.78%) and FAS score 2 (72.15%) being the highest (Figure 5).

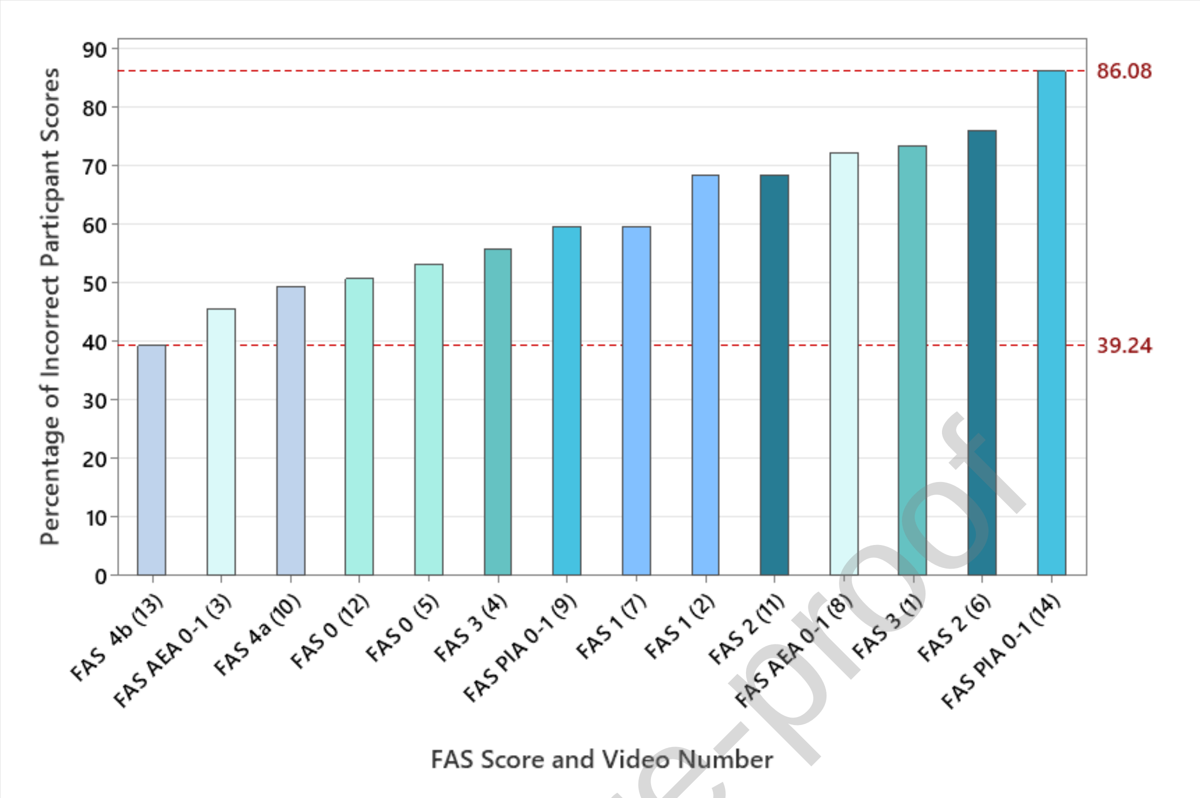

Figure 4: Percentage of incorrect scores (those not matching the actual assigned FAS score) per video and associated FAS score. Red reference lines illustrate the FAS scores with the highest (FAS PIA 0-1, Video 14, 86.08%) and lowest (FAS 4b, Video 13, 39.24%) percentages of incorrect scores.

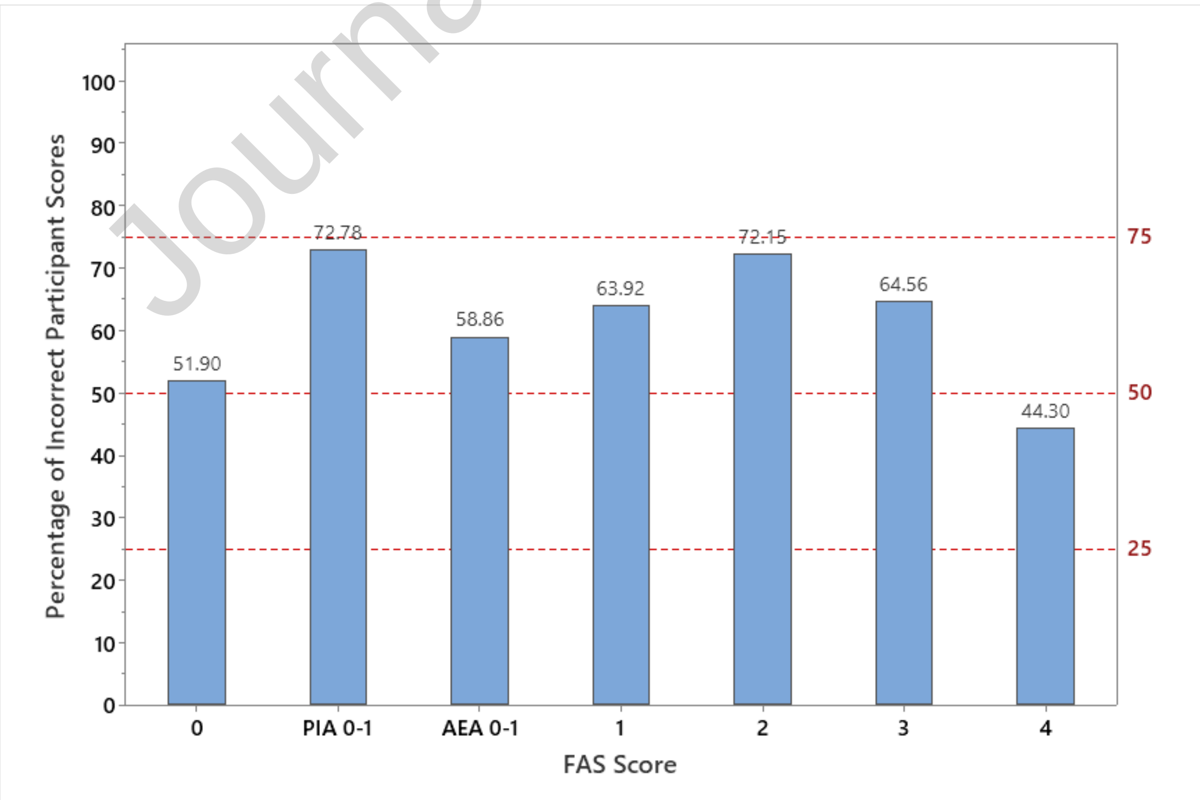

Figure 5: Percentage of incorrect scores for each FAS score, across all participant groups.

No significant difference in error score (how incorrect an assigned score was) between any of the participant groups was identified for any videos other than videos 10 (FAS 4a; $p = 0.0355$) and 1 (FAS 3; $p = 0.0025$) which occurred in the owner vs behavior expert groups, indicating that owners scored further away on the FAS spectrum to the actual FAS score, compared to behavior experts.

Intra-rater reliability

Intra-rater reliability was very good (Intraclass correlation coefficient (ICC 3,k), 0.83 , 95% CI 0.802-0.855) for all videos and participant's scores.

Wilcoxon matched-pairs signed-rank tests used to identify significantly different FAS scores between the first and second scorings, found no significant differences within Behaviour expert ($W = 10$, $p = 0.6895$), Owner ($W = 208$, $p = 0.8037$) or Dog Trainer ($W = 31$, $p = 0.1982$) groups. The veterinarian group's scores were found to have a statistically significant difference between the first and second scores ($W = 1069$, $p = 0.0001$). A significant difference between first and second scores was identified when examining the videos by FAS score; PIA 0-1 ($W = 300$, $p = 0.0337$). When the FAS score PIA 0-1 was further analyzed by the participant group, a significant difference was identified in the veterinarian participant group ($W = 91$, $p = 0.002$). No significant differences were found for the remaining FAS scores 0 ($W = 53$, $p = 0.5064$), AEA 0-1 ($W = 22$, $p = 0.8128$), 1 ($W = -33$, $p = 0.8007$), 2 ($W = 145$, $p = 0.1138$), 3 ($W = 147$, $p = 0.1749$) or 4 ($W = 71$, $p = 0.1859$).

Validity

Spearman's correlation identified a significant, moderate positive correlation between participants' and actual LCAS scores ($\rho = 0.618$, $p <0.001$, $n = 79$). When assessed by participant group there was a significant, strong positive correlation between the participant and the actual LCAS score in the Dog trainer group ($\rho = 0.728$, $p<0.001$, $n = 5$) and a significant, moderate positive correlation in the Behaviour expert ($\rho = 0.666$, $p<0.001$, $n = 8$), Dog owner ($\rho = 0.728$, $p<0.001$, $n = 45$) and Veterinarian ($\rho = 0.546$, $p<0.001$, $n = 21$) groups.

There was a significant, strong positive correlation between participants' FAS scores and cumulative LCAS scores (Spearman rank correlation, $\rho = 0.811$, $p<0.001$, $n = 79$). When assessed by participant group (Figure 6), a significant and strong positive correlation was found between participant FAS scores and cumulative LCAS scores for all participant groups. Behaviour expert (ρ = 0.754, p<0.001, n = 8), Dog trainer (ρ = 0.810, p<0.001, n = 5), Dog owner (ρ = 0.822, p<0.001, n = 45) and Veterinarian (ρ = 0.812, p<0.001, n = 21). Scatter plots were created to visualize the relationships for this data (Figure 6).

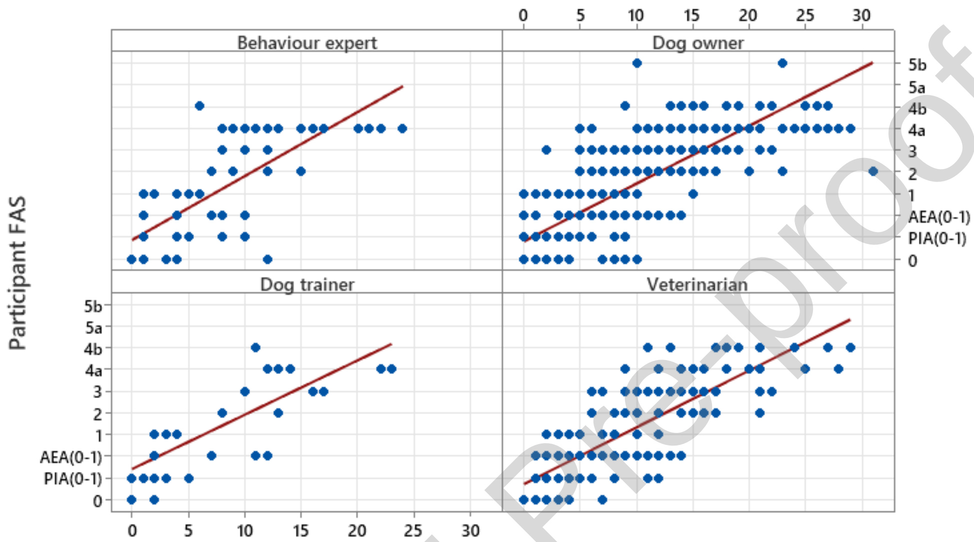

Figure 6: Relationship between participant cumulative LCAS (x axis) and FAS scores assigned by each participant group (y axis). The red line represents the regression fit, demonstrating a strong positive relationship. Multiple data points may overlie each other.

Ease of use and thematic analysis

Most participants experienced some difficulty using the FAS Spectrum (60.76%; n = 48), while the remainder found it easy (30.38%; n = 24) or difficult (8.86%; n = 7) to use. Yet most participants said they would use the spectrum again (77.22%; n = 61).

Participants were asked "What was good about using the FAS scale?" and "What was challenging about using the FAS scale?". A randomly selected series of participant comments are provided in Table 4. A detailed list of the questions and the participants' answers is provided as supplementary information (S6).

Table 4: Themes, sub themes and response count of comments made by questionnaire participants when asked to provide free-text answers to what was "good" and "challenging" about the Spectrum of Fear, Anxiety and Stress.

Themes	Response count	Sub themes	Negative feedback	Positive feedback
Structure	53	Categories Numbering	"Dog may show signs that potentially fit in more than one category" "Some dogs didn't seem to fit one bracket" "Lots of behavior is covered by multiple options" "Some dogs move between stages or do not fit nicely into a set stage" "Didn't like the numbering" "Numbering of bands is odd."	"Liked scale and how it broke certain elements into freeze/flight" "Different behaviors considered and broken down"
Design	44	Descriptions Images Color coding	"The Descriptions s were a bit long. More pictures to illustrate behavior would help"	"Descriptions were clear" "The inclusion of images was particularly helpful" "very detailed, good color system" "Color coding helps with severity measure"
Behavioral signs	**27**	Behaviors observed Valence of behaviour	"Not all behaviors covered i.e., some dogs had slight signs but not covered in scale" "It does not cover all behaviors"	"May point out subtle signs to owners" "Good to list some specific

		Breed specific descriptors	“Some behaviors missing, such as shaking off, hard to separate some of the lower bands.”	behaviors to look for“ “Good for the more subtle signs that we don't always notice “
		Subtle signs	“Different breeds exhibit differently e.g., no tail” “What’s anxiety or excitement differences are on the scale” “Not all dogs are the same, they don’t climb the stress ladder in the same order. Sone were more stressed than scale suggested, others less so.”	
Video footage	19	Viewing video Sound Length of video	“Difficult to evaluate subtler signs in some videos, pupil size, ear position, commissures, mild trembling etc.” “Some of the criteria are difficult to assess (e.g., tightness of the face from video)” “No sound” “Clip were very short, no sound, on a small screen difficult to judge what’s what.”	
Observer	10	Subjectivity Scoring accuracy	“I struggled with some as I thought they seemed anxious but didn’t fit in the higher categories” “Some of the categories, particularly lower down, more difficult to differentiate and are subjective”	

			"Sometimes not fully congruent with what I observed."	
Use	4	Ease of use	"Middle of FAS spectrum harder to grade especially as difficult to evaluate subtler signs in some videos" "A lot of the bottom categories are very similar and this makes distinguishing between them tricky."	"Easy to pinpoint some behavior" "It seemed straightforward to use initially" "Fairly easy scoring system"

Discussion

This study aimed to assess inter-rater and intra-rater reliability, concurrent validity and accessibility of an established behavioral assessment tool (Fear Free® Spectrum of Fear, Anxiety, and Stress) for measuring fear, anxiety, and acute stress in dogs visiting the veterinary practice. While assessment is crucial for ensuring patient and staff welfare and safety, a reliable and accurate assessment remains challenging. The results of this study suggest that while the FAS spectrum is indeed accessible and valid, it may require further refinement to enhance its reliability and usability.

The demographics of the study participants may have influenced results, as groups were not equal in number. Over half of the participants were dog owners (45/79), which was likely an artefact of recruitment techniques used within the study. Future studies may benefit from wider recruitment practices to recruit more participants to non-owner groups. Participants were 91% female, similar to other online dog behavior studies (86.81% [Daniels et al., 2023]), which reflects the prevailing gender distribution among dog owners and veterinarians (Jelinski et al., 2022; Anderson et al., 2023), and those most likely to complete online surveys (Smith and Smith, 2008).

The interclass correlation coefficient was excellent across all participants scoring all videos, indicating that the FAS spectrum appears reliable overall. The inter-rater reliability was further assessed by analyzing the percentage of incorrect scores for each video and FAS score for each participant group. If the FAS spectrum is a reliable assessment tool, we should expect high inter-rater reliability between participant groups along with a low percentage of incorrect and error scores. The percentage of incorrect scores was expected to be lowest for extreme ends of the FAS spectrum (FAS 0 and FAS 4; FAS 5 not available due to video recording limitations), in line with findings that fearful behaviors are easier to identify (Diesel et al., 2008), and the more ambiguous, subtle changes of the mid-range FAS scores would be more difficult to assess, resulting in higher error scores. A high percentage of incorrect scores for a specific video or FAS score would indicate participants' reduced ability to correctly assign the actual FAS score. This assumption was found to be largely correct, with FAS 4 (44.30%) and FAS 0 (51.90%) achieving the lowest overall percentage of incorrect scores (Figure 5). This, and the higher reported percentage of incorrect scores for owners scoring FAS 1 in our study (66.67% and 68.89%, Figure 4), contradicted the findings of previous research, suggesting that owners recognize FAS level 1 most correctly (Raileanu, 2021) and supported the findings of King et al. (2022) with the reliability of assessment the highest in stressed, anxious dogs (higher FAS scores).

When individually assessed, video 14 (PIA 0-1; 86.08%) had the highest overall percentage of incorrect scores, likely due to subtle descriptor differences between adjacent FAS scores. Despite this, the percentage of incorrect scores for FAS PIA 0-1 was similar across participant groups (Owners: 88.89%, Trainers: 80%, Behavior experts: 87.50%, Vets: 80.95%), suggesting reasonable inter-rater reliability. As expected, video 13 (FAS 4 Freeze/Fret; 39.24%) had the lowest percentage of incorrect scores. The dog in the video demonstrated freezing (behavioral inhibition), commonly mistaken for compliance in a veterinary setting, which is problematic as misinterpretation increases the risk of injury (Steimer, 2002), as well as reflecting poor awareness of emotional distress. The results of this study suggest that the FAS spectrum may aid interpretation of freeze type behaviors, reducing risk of injury and enhancing patient welfare, as on the whole participants of the study coded FAS score 4 accurately. However, the overall percentage of incorrect scores for videos demonstrating FAS 4 remained relatively high (Video 10; 49.37% and Video 13; 39.24%), raising concerns about a failure of some participants to accurately identify signs of fear, anxiety, and stress without delay, preventing safe intervention, which could lead to injury, emotional distress, misdiagnosis and poor client retention. Early recognition is crucial for timely treatment adjustments and effective patient care and yet veterinarians correctly identified fear and anxiety signs in only 42.86% of videos, a result similar to the 49.39% reported by Catalán et al (2020), suggesting that early-stage alterations to examination processes may not be put in place in nearly half of the example videos used within the study.

There was a lower-than-expected percentage (45.57%) of incorrect scores for video 3 (FAS AEA 0-1, Figure 4), despite its mid-spectrum FAS score, contradicting the hypothesis that AEA 0-1 would be harder to interpret. The dog in Video 3 was an enthusiastic Labrador and the term “excited” in the “Alert, Excited, Anxious” category (AEA 0-1 FAS) may have helped participants correctly classify this dog. Care should be taken as erratic behavior can represent a ‘fidget’ type response to increasing fear and anxiety, rather than excitement, which is commonly misinterpreted. While the FAS spectrum does include this behavior under “FAS 4 Freeze/Fret”, its limited description may have made correct classification difficult for participants with minimal behavioral knowledge.

Video 1 (FAS 3) had a higher-than-expected incorrect score percentage (73.42%), likely due to conflict-related behaviors (e.g., tail wagging while barking [no sound provided], approaching the researcher while mouthing and tense), which may have been mistaken as excitement rather than an escalation of FAS. Participants unfamiliar with anxiety-related behavioral display variations may have scored FAS lower, as they may have seen the tail wagging as an indicator of valence, as tail movements are commonly used to provide a holistic description of dog behavior (Tami and Gallagher, 2009). This was supported by a lower percentage of incorrect scores among behavior experts (37.5%) and dog trainers (20%) compared to veterinarians (71.43%) and owners (86.67%) for this video, consistent with research showing that formal training improves inter-rater reliability (Diesel et al., 2008) and understanding of canine body language.

On average, even the behavior experts in this study had more than 50% incorrect scores, while the small sample size provides an opportunity for more scoping research, the high prevalence of incorrect scores highlights significant concerns around the potential inaccuracy of expert scoring or the validity of the tool. However, the role of education should be considered, given that most participants had no prior knowledge of, or exposure to, the FAS spectrum prior to study participation and that limited training was provided within the study. While extensive online certification programs exist (Fear Free® Certified Professional) very limited training resources are provided when downloading the FAS spectrum or freely accessible on the Fear Free website public domain and there is an opportunity for instructional information on completing the spectrum to be developed for dissemination alongside the spectrum.

However, this highlights a limitation in the perceived self-explanatory nature of the spectrum. Providing additional training alongside the spectrum may improve reliability, as even benchmarked ordinal scales require extensive training (Jokela et al., 2023) and training observers before conducting assessments helps maximize the reliability of observations (Taylor and Mills, 2006; Mullen et al., 2008), as shown to be effective with other behavioral instruments such as the AnimalFACS tool (Zhang et al., 2019)

Despite the high percentage of incorrect scores (Figure 4), inter-rater reliability was good for 12 out of the 14 videos, as the difference between actual and participant FAS scores when incorrect was generally small and there was no significant difference in error scores, in keeping with preliminary studies with similar formats (Tami and Gallagher, 2009). However, two videos, videos 10 (FAS 4a Flight) and 1 (FAS 3), demonstrated poor inter-rater reliability between owners and behavioral experts due to high error scores in the owner group. This was unexpected, given that these videos represented FAS 4 and 3 respectively, and we hypothesized that all groups would score higher FAS scores more correctly. This further supports the argument that enhancing owner's ability to recognize body language through education is an essential step in improving canine welfare (Philpotts et al., 2019).

Error scores provided insight into the degree of incorrectness across participant groups, offering an advantage over the percentage of incorrect scores or a general mixed model where only the presence of correct or incorrect could be assessed. Error score calculations relied on accurate identification of FAS score by researchers to assign a score to the videos. Given that Mercier et al. (2023) identified fair to moderate inter-rater reliability within a small sample of veterinary behavior residents and specialists, the current study deemed assessment by a single resident and behavior expert appropriate, but thought should be given to whether this may potentially account for lower reliability of scores. Error scores were assessed per video rather than per FAS Score due to highlighted differences between videos within the same FAS score, as discussed above. For example, while Video 14 (FAS PIA 0-1) had the highest percentage (86.08%) of incorrect participant scores, error scores showed that most participants' scores were within one score of the assigned FAS score. However, video 13 (FAS 4b Freeze/Fret), which had the lowest percentage of incorrect scores (39.24%), displayed a larger range of participant scores when assessed by error score, highlighting the importance of considering both accuracy and degree of incorrectness, despite excellent agreement (ICC) overall. This is particularly important when the degree of incorrectness could mean assigning a dog as significantly less fearful and anxious than true, which poses an inherent safety risk (Meints et al., 2018). Future studies should consider assessing how closely participants' scores are to the "correct" score when examining ordinal or ranked data and focus training around defining these differences.

Error scores could have been influenced by question design, as participants had the ability to score videos up to a score of FAS 5, despite no videos representing this score being included. The inclusion of FAS score 5 videos may have provided lower error scores by offering a clear benchmark; however, similar to other studies with standardized veterinary examination protocols (Kim et al., 2022), aggression scores were low during data collection, likely explained by the study's inclusion criteria, which restricted the recruitment of dogs with behavioral warnings and the systematic approach to clinical examination, which facilitated individual dog's refusal to participate in specific sections. While ensuring researcher safety, this limited the assessment of the entire FAS spectrum, which could be further evaluated using pre-existing nonclinical video examples of FAS score 5 in future research. Nevertheless, given that there is a higher inter-rater agreement when identifying more extreme behaviors (Mariti et al., 2015) this is less of a concern than the ability to identify escalating behavioral signs (FAS score 2-4). Behavioral assessment studies often exclude dogs showing aggression during clinical examination (Stollar et al., 2022), which should be adhered to, as it would be unethical and unsafe to provoke severe behaviors. Owners were asked not to provide food or toys during the mock consultation in line with other standardization protocols in behavioral assessment studies (Hauser et al., 2020) as conflict-related behaviors may arise when treats are used to encourage a dog's interaction with a researcher (Kuhne et al., 2014). However, this did limit the assessment of a changing willingness to take food, which is an aspect of the FAS spectrum, and may have improved participant ability to assess escalating FAS score.

Intra-rater reliability was good across all videos and participants, indicating that participants could use the spectrum in sequential trials with a good degree of correlation and agreement between their first and second scores. The limited time between the completion of survey sections may have introduced recall bias rather than representing true high intra-rater agreement. To better assess intra-rater reliability, future research could benefit from issuing two surveys with an increased test-retest interval between them. Furthermore, this study aimed to investigate whether significant differences existed between participant groups. While the Wilcoxon paired sample test is not an ideal measure of reliability (Koo and Li, 2016), it can be used to analyze agreement or significant differences between groups. Significant differences were identified in the veterinarian group with veterinarians showing a significant difference between the first and second scores across all videos ($p = 0.0001$) and specifically for FAS PIA 0-1. This could be explained by an increase in the assigned

FAS score resulting from learning during the survey process. However, this would be an expected effect across all participant groups if this was the case. As veterinarians may only have one opportunity for observation (a single consultation appointment), it is essential that intra-rater reliability is sufficient. Behavior has historically been taught poorly in veterinary curriculums (Demaline, 2018); however, given the progressive inclusion in more programs, this result was unexpected. The need for ongoing behavioral education among all groups is evident, but especially veterinarians, as bastions of animal health and welfare and experts to whom owners often turn for advice. It is essential that veterinarians recognize early signs of FAS in order to adapt protocols, as there may be a risk of normalization of these signs due to their high prevalence in the veterinary environment.

Similarly to previous validation studies (Mills et al., 2020), criterion validity could not be fully assessed, as there is no comparable gold standard scale. Mills et al. (2020) suggested that the LCAS could be used to assess anxiety in a wider range of contexts, therefore, this study extrapolated the scale to the veterinary environment as a source of acute, episodic stress due to its validity and crossover between the descriptors within the LCAS and FAS spectrum. Validity assessments were performed using seven repeat videos in the latter half of the survey, rather than the full initial 14 videos to reduce time constraints. Participants' cumulative LCAS was assessed against the researchers' cumulative LCAS to ensure that their measure was a useful validation against FAS score, which was found to have a moderate positive association ($r_s = 0.618$, $p<0.001$). While participant FAS scores strongly correlated with cumulative LCAS ($r_s = 0.811$, $p<0.001$), demonstrating good concurrent validity, there may have been a learning effect through the repeat use of the FAS spectrum. Future studies would benefit from assessing concurrent FAS and LCAS scores on first attempts.

Participants were encouraged to identify both “good” and “challenging” aspects of the FAS spectrum, with the aim of formulating practical recommendations for improvement should they be deemed necessary based on the study results. Participants appreciated the use of the color-coded traffic light style system, and the use of multiple images was a clear advantage over previous scales. However, cartoon-style images may limit interpretation for some participants, especially given concerns about “Different breeds exhibit(ing) differently.” Future research should focus on recruiting different dog breeds to ensure diverse representation, acknowledging the existence of variation in behavioral signaling between breeds (Goodwin et al., 1997; Vas et al., 2005).

Participants highlighted that sensitive yet subtle indicators of fear and anxiety, such as lip-licking and yawning, were absent from the spectrum. While the inclusion of these subtle signs would be beneficial, this is similar to other cumulative fear, anxiety and stress scores (Mandese et al., 2021). These signs may be difficult to interpret via video, relying on real-time observation and can indicate merely arousal, not valence (Posner et al., 2005). Video recording may have posed some difficulties in interpreting FAS scores, with comments such as “some criteria difficult to assess from video.”, as not all behavioral signs exhibited by a dog may be readily observable on video footage (Diesel et al., 2008). However, this should not affect the assessment of inter-rater reliability, as each participant encountered the same limitations. Whilst the use of video footage within the study could pose some risk to generalization of the FAS spectrum for use in live observation studies, the authors chose to utilize video recordings due to the significant correlation between video and live field observations (Curby et al., 2016) and the high inter-rater reliability demonstrated when assessing dogs using video footage (Arena et al., 2017) . It is recommended that intra-rater reliability be assessed via video recordings to ensure any difference in results arises from variation in observer, rather than dog's behaviour (Bateson and Martin, 2021).Live observer scoring may be an opportunity within future work on this topic, as suggested by Diesel et al (2008), increasing participants' ability to observe more subtle body language signs. Context must also be considered, as veterinarians in general practice typically observe their patients whilst taking history and .this may be when they are most likely to make an FAS assessment. However, whilst this is preferable to habituate the dog to the environment, due to the restricted length of consultations some veterinarians may take a history, whilst examining and observing their patient simultaneously . Video clips from each aspect (during habituation and examination) were used within the participant questionnaire, to assess participant ability across these situations and account for varying emotional states across a consultation. This reinforces that FAS assessment should be an ongoing, continuous process throughout an animal's veterinary experience to ensure that changes in FAS score are detected, and appropriate action is taken.

Despite participants' comments that "clips were very short", the average length (17 seconds) was similar to the video length (20 seconds) in other studies (Tami and Gallagher, 2009), replicating the fleeting nature of real-life observation where decision making in a short time frame is paramount to maintaining safety. The lack of sound on videos was also noted by participants. However, this was an intentional decision given the lack of vocalizations described on the FAS spectrum. It should be noted that the absence of vocalizations in the FAS spectrum itself is a criticism of the scale rather than a methodological error.

Based on participants' feedback, assigning the correct FAS score was challenging when observed behaviors spanned multiple categories and with variation in individual displays of stress between dogs. Multiple participants commented that “overlap between the descriptions meant I was unsure,” “Overlap between categories,” and “Some overlap of behaviors.” Retrospectively, the questionnaire could have provided more instruction on what to do when a fluctuation between scores occurred, e.g., to assign the highest of two sequential scores. The overlap of behaviors is a significant concern in other recent studies (Jokela et al., 2023; Mercier et al., 2023) and cannot be avoided due to the escalating and grading nature of Likert scales, as there will always be a continuous escalation to behaviour and therefore a point of crossover. Effective use of the FAS spectrum needs to enable accurate assessment of not only FAS score but how the measure changes over time with subsequent assessments, to aid clinicians recognizing break or stop points in their patient examinations. Overlapping constructs in the FAS spectrum may lead to inaccurate scoring of individuals, likely contributing to the identified error rates. Given that collapsing scales have been shown to increase inter-rater reliability in similar studies (Mercier et al., 2023), adapting the FAS spectrum into more behaviorally distinct categories, with benchmarks for intervention, with minimal overlap would be beneficial. Condensing categories, e.g., PIA and AEA 0-1, would likely be equally valuable, as participants identified that “the middle of FAS spectrum harder to grade” and “a lot of categories are very similar.”

Recommendations

Collapsing some elements of the spectrum, with less subtle variation between items, would increase inter-rater reliability. The numbering of the FAS bands should also be reconsidered, as there are multiple FAS 0-1 scores and removal of the term “excited” from the AEA category wording may help to avoid misinterpretation. More weight should also be given to individualized stress responses, with greater emphasis on displacement behaviors, conflicts and fidget/high-arousal type behaviors as is currently given to those behaviors typical of fight-or-flight responses. Due to the complexity and individual variation of behavioral responses to stressors, it is unreasonable to expect a rapid, on the spot, patient side assessment instrument to include every nuanced behavior associated with fear and anxiety. Yet, researchers developing these instruments must ensure that such a tool offers sufficient breadth and detail to inform clinical decisions with minimal risk. Despite most participants experiencing some difficulty using the FAS Spectrum (60.76%; n = 48), participants (61/79) were open to the continued use of the FAS scale. With the recommendation of this study to reduce the likelihood of incorrect scoring, it holds promise as a reliable and validated assessment tool for acute canine stress in the veterinary environment.

It is relevant to note that this study used the version of the FAS spectrum available freely at the time of the study (2022), and while an updated version of the FAS spectrum has since been released, the recommendations of this study remain valid.

Conclusion

The Fear Free® spectrum of Fear, Anxiety, and Stress demonstrates good concurrent validity and intra-rater reliability for most participants, although lower agreement was demonstrated for more subtle behavioral signs (FAS 0-1 PIA) and those most likely to utilize it (i.e., veterinarians). While inter-rater reliability was reasonable, all participants struggled to score subtler categories. The percentage of incorrect scores across participant groups, including behavior experts, suggests that even experienced professionals cannot currently consistently rate behaviors correctly using this tool. Very few studies have looked to validate an existing behavioral scale for use within the veterinary environment; this study fulfils this niche and offers an accessible tool for in-practice use. Given the risks associated with misinterpreting and mismanaging fearful and anxious dogs in practice, the implications of this study are extensive. While some owners and veterinarians may not be amenable to altering veterinary protocols to reduce animal stress, the spectrum offers the opportunity to engage animal caregivers and advise them about escalating stress responses and how they can be mitigated, thereby increasing access to those in need of behavioral support.

The findings of the current study provide sufficient evidence for expanding owner and veterinarian education on identifying fear and anxiety in practice. While there has been recent significant progress in bringing these concerns to the attention of the wider population, this study highlights an error in the dissemination of accurate information (due to average higher incorrect scores in owner versus animal professional groups) and underscores the need for more widely published, accessible, reliable and validated tools. Overall, this study suggests that the FAS spectrum offers an opportunity for education and, with some adjustments, could be promoted for wider use within clinical settings to aid identification of signs of stress, thereby optimizing safety and welfare.

Although this study identifies areas for improvement within the Fear Free® spectrum, it is worth noting that it still provides one of the few viable options for assessing behavior within a practice setting. Overall, the FAS spectrum provides a solid foundation for the further development of qualitative behavioral assessment tools within the field. This suggests that it is in the best interest of the broader clinical animal behavior community to support its development into a behavioral assessment tool that can be successfully used in non-research settings. If the FAS spectrum can be refined, it will provide a global audience of professionals and animal caregivers with the means to advocate for the animals in their care with greater accuracy, making meaningful steps toward improved psychological care.

Acknowledgements:

Thank you to all the participants who took part in the questionnaire, as well as the caregivers of the dogs who supported us in obtaining video footage to use within the study. Thanks to Sarah Liversage and Catherine Tindale for helping to pilot the study. Special thanks to Laura James for proofreading the thesis on which this paper is based.

Authorship:

The idea for the paper was conceived by Emma Gatehouse with support from Sagi Denenberg, Loni Loftus and Annika Bremhorst. The experiments were designed by Emma Gatehouse with support from Sagi Denenberg, Loni Loftus and Annika Bremhorst. The experiments were performed by Emma Gatehouse. The data were analyzed by Emma Gatehouse and Loni Loftus. The paper was written by Emma Gatehouse and reviewed by Sagi Denenberg, Loni Loftus and Annika Bremhorst. All authors have approved the final article prior to publication.

CRediT authorship contribution statement

Emma Gatehouse: Conceptualization, methodology, formal analysis, investigation, visualization, resources, data curation, writing – original draft. **Annika Bremhorst:** Conceptualization, methodology, supervision, writing – review and editing. **Sagi Denenberg**: Conceptualization, methodology, supervision, writing – review and editing. **Loni Loftus**: conceptualization, methodology, formal analysis, data curation, supervision, writing – review and editing.

Conflict of interest statement:

Emma Gatehouse received a grant from the Fear Free® organization to aid completion of the research. The funders had no role in study design, data collection and analysis, decisions to publish or preparation

Ethical approval

All procedures were performed in compliance with the relevant laws and institutional guidelines and were approved by the appropriate institutional committees

The study received approval from the Royal (Dick) School of Veterinary Studies Human Ethical Review Committee (HERC); 18/01/2023 HERC_2022_141, Royal (Dick) School of Veterinary Studies Veterinary Ethical Review Committee (VERC); 03/02/2023 VERC/138.22, and the University of Bristol Animal Welfare and Ethical Review Body (AWERB) 01/03/2023 VIN/23/017, prior to commencing data collection.

Study information and written or electronic informed consent documents were provided to owners and questionnaire participants prior to participation. Caregivers of dogs used for video recordings gave permission for collected data (including video footage) to appear in relevant publications and reports.

References

Anderson, K.L., Casey, R.A., Cooper, B., Upjohn, M.M., Christley, R.M., 2023. National Dog Survey: Describing UK dog and ownership demographics. Animals 13. https://doi.org/10.3390/ANI13061072/S1

Arena, L., Wemelsfelder, F., Messori, S., Ferri, N., Barnard, S., 2017. Application of free choice profiling to assess the emotional state of dogs housed in shelter environments. Appl. Anim. Behav. Sci. 195, 72–79. https://doi.org/10.1016/j.applanim.2017.06.005

Bateson, M., Martin, P., 2021. Measuring Behaviour. Cambridge University Press. https://doi.org/10.1017/9781108776462

Beerda, B., Schilder, M.B.H., Van Hooff, J.A.R.A.M., De Vries, H.W., Mol, J.A., 2000. Behavioural and hormonal indicators of enduring environmental stress in dogs. Anim. Welf. 9, 49–62. https://doi.org/10.1017/S0962728600022247

Bellamy, N., 2015. Principles of clinical outcome assessment, in: Rheumatology: Sixth Edition. Mosby, pp. 9–19. https://doi.org/10.1016/B978-0-323-09138-1.00002-4

Campbell, W.E., 1975. Behaviour problems in dogs. American Veterinary Publications, Santa Barbara.

Catalán, A.I., Rojas, C.A., Chávez, G.A., 2020. Recognition of aggressive and anxious behaviors in canines by a group of Chilean veterinarians. J. Vet. Behav. 38, 8–13. https://doi.org/10.1016/j.jveb.2020.05.001

Csoltova, E., Martineau, M., Boissy, A., Gilbert, C., 2017. Behavioral and physiological reactions in dogs to a veterinary examination: Owner-dog interactions improve canine well-being. Physiol. Behav. 177, 270–281. https://doi.org/10.1016/J.PHYSBEH.2017.05.013

Curby, T.W., Johnson, P., Mashburn, A.J., Carlis, L., 2016. Live versus video observations. J Psychoeduc. Assess. 34, 765–781. https://doi.org/10.1177/0734282915627115.

Dancey, C. P. & Reidy, J., 2007. *Statistics Without Maths for Psychology*. Illustrated ed. Harlow: Pearson Education. ISBN 0132051605; 9780132051606.

Daniels, J.T., Busby, D., Chase-Topping, M., Brown, S.M., 2023. A survey of dog behavior modification practice in the UK: Who is offering it, what methods are they using and how effective do their clients perceive practitioners to be? J. Vet. Behav. 59, 1–7. https://doi.org/10.1016/J.JVEB.2022.11.011

Demaline, B., 2018. Fear in the veterinary clinic: History and development of the Fear Free[SM] initiative. Conspectus Borealis 4.

Diederich, C., Giffroy, J.M., 2006. Behavioural testing in dogs: A review of methodology in search for standardization. Appl. Anim. Behav. Sci. pp. 51–72. https://doi.org/10.1016/j.applanim.2005.11.018

Diesel, G., Brodbelt, D., Pfeiffer, D.U., 2008. Reliability of assessment of dogs' behavioural responses by staff working at a welfare charity in the UK. Appl. Anim. Behav. Sci. 115, 171–181. https://doi.org/10.1016/J.APPLANIM.2008.05.005

Döring, D., Roscher, A., Scheipl, F., Küchenhoff, H., Erhard, M.H., 2009. Fear-related behaviour of dogs in veterinary practice. Vet. J. 182, 38–43. https://doi.org/10.1016/j.tvjl.2008.05.006

Edwards, P.T., Smith, B.P., McArthur, M.L., Hazel, S.J., 2019. Fearful Fido: Investigating dog experience in the veterinary context in an effort to reduce distress. Appl. Anim. Behav. Sci. https://doi.org/10.1016/j.applanim.2019.02.009

Fear Free, 2023. Fear Free Veterinary Professional Certification Programme [https://www.fearfree.com/certifications/fear-free-individual-veterinary-professional-certification-programs/].

Fear Free, 2022. The Spectrum of Fear, Anxiety, and Stress.

Godbout, M., Palestrini, C., Beauchamp, G., Frank, D., 2007. Puppy behavior at the veterinary clinic: A pilot study. J. Vet. Behav. 2, 126–135. https://doi.org/10.1016/J.JVEB.2007.06.002

Goodwin, D., Bradshaw, J. W., Wickens, S. M.,1997. Paedomorphosis affects agonistic visual signals of domestic dogs. Anim. Behav. 53(2), 297-304. https://doi.org/10.1006/anbe.1996.0370

Hauser, H., Campbell, S., Korpivaara, M., Stefanovski, D., Quinlan, M., Siracusa, C., 2020. In-hospital administration of dexmedetomidine oromucosal gel for stress reduction in dogs during veterinary visits: A randomized, double-blinded, placebo-controlled study. J. Vet. Behav. 39, 77–85. https://doi.org/10.1016/J.JVEB.2020.05.002

Hekman, J.P., Karas, A.Z., Dreschel, N.A., 2012. Salivary cortisol concentrations and behavior in a population of healthy dogs hospitalized for elective procedures. Appl. Anim. Behav. Sci. 141, 149–157. https://doi.org/10.1016/J.APPLANIM.2012.08.007

Hekman, J.P., Karas, A.Z., Sharp, C.R., 2014. Psychogenic stress in hospitalized dogs: Cross species comparisons, implications for health care, and the challenges of evaluation. Animals (Basel) 4, 331. https://doi.org/10.3390/ANI4020331

Horwitz, D., Mills, D.S., 2012. BSAVA Manual of Canine and Feline Behavioural Medicine, Second edition. ed. British Small Animal Veterinary Association, Quedgeley, Gloucester.

Jelinski, M.D., Schreiner, B., Neale, A., Townsend, H.G.G., 2022. Demographic survey of private veterinary practices in western Canada. Can. Vet. J. 63, 27.

Jokela, F., Mercier, P., Honeckman, L., Dunham, A.E. & Overall, K.L., 2023. Sensitivity of benchmarked behavioral assays for distress: Should we attend to certain behaviors during exams? J. Vet. Behav. 66, 36–42. https://doi.org/10.1016/j.jveb.2023.06.012

Kim, Sun A, Borchardt, M.R., Lee, K., Stelow, E.A., Bain, M.J., 2022. Effects of trazodone on behavioral and physiological signs of stress in dogs during veterinary visits: a randomized double-blind placebo-controlled crossover clinical trial. J. Am. Vet. Med. Assoc. 260, 876–883. https://doi.org/10.2460/JAVMA.20.10.0547

King, T., Flint, H.E., Hunt, A.B.G., Werzowa, W.T. & Logan, D.W., 2022. Effect of music on stress parameters in dogs during a mock veterinary visit. Animals (Basel), 12(2), 187. https://doi.org/10.3390/ani12020187

Koo, T.K. & Li, M.Y., 2016. Cracking the code: Providing insight into the fundamentals of research and evidence-based practice—a guideline of selecting and reporting intraclass correlation coefficients for reliability research. J. Chiropr. Med., 15(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Koolhaas, J.M., Bartolomucci, A., Buwalda, B., Boer, S.F. De, Flügge, G., Korte, S.M., Meerlo, P., Murison, R., Olivier, B., Palanza, P., Richter-levin, G., Sgoifo, A., Steimer, T., Stiedl, O., Dijk, G. Van, Wöhr, M., Fuchs, E., 2011. Neuroscience and biobehavioral reviews stress revisited : A critical evaluation of the stress concept. Neurosci. Biobehav. Rev. 35, 1291–1301. https://doi.org/10.1016/j.neubiorev.2011.02.003

Korpivaara, M., Huhtinen, M., Pohjanjousi, P., Overall, K., 2022. Tasipimidine, a novel orally administered alpha-2 adrenoceptor agonist, alleviates canine acute anxiety associated with owner departure—A pilot study. J. Vet. Behav. 58, 54–61. https://doi.org/10.1016/j.jveb.2022.11.003

Kuhne, F., Hößler, J.C., Struwe, R., 2014. Emotions in dogs being petted by a familiar or unfamiliar person: Validating behavioural indicators of emotional states using heart rate variability. Appl. Anim. Behav. Sci. 161, 113–120. https://doi.org/10.1016/J.APPLANIM.2014.09.020

Lefman, S.H., Prittie, J.E., 2019. Psychogenic stress in hospitalized veterinary patients: Causation, implications, and therapies. J. Vet. Emerg. Crit. Care. 29, 107–120. https://doi.org/10.1111/VEC.12821

Lind, A.K., Hydbring-Sandberg, E., Forkman, B., Keeling, L.J., 2017. Assessing stress in dogs during a visit to the veterinary clinic: Correlations between dog behavior in standardized tests and assessments by veterinary staff and owners. J. Vet. Behav. 17, 24–31. https://doi.org/10.1016/j.jveb.2016.10.003

Mandese, W.W., Griffin, F.C., Reynolds, P.S., Blew, A.C., Deriberprey, A.S., Estrada, A.H., 2021. Stress in client-owned dogs related to clinical exam location: A randomised crossover trial. J. Small. Anim. Pract. 62, 82–88. https://doi.org/10.1111/jsap.13248

Mariti, C., Raspanti, E., Zilocchi, M., Carlone, B., Gazzano, A., 2015. The assessment of dog welfare in the waiting room of a veterinary clinic. Anim. Welf. 24, 299–305. https://doi.org/10.7120/09627286.24.3.299

McAuliffe, L.R., Koch, C.S., Serpell, J. & Campbell, K.L., 2022. Associations between atopic dermatitis and anxiety, aggression, and fear-based behaviors in dogs. J. Am. Anim. Hosp. Assoc. 58(4), pp.161–167. https://doi.org/10.5326/JAAHA-MS-7210

Meints, K., Brelsford, V., Keuster, T. De, 2018. Teaching children and parents to understand dog signaling. Front. Vet. Sci. 5, 257. https://doi.org/10.3389/FVETS.2018.00257

Mercier, P., Honeckman, L., Jokela, F., Dunham, A.E., Overall, K.L., 2023. Using standardized scales to assess fear at veterinary visits: Intra- and inter-rater reliability. J. Vet. Behav. 62, 12–17. https://doi.org/10.1016/J.JVEB.2023.02.004

Mills, D., Mueller, H.W., McPeake, K., Engel, O., 2020. Development and psychometric validation of the Lincoln Canine Anxiety Scale. Front. Vet. Sci. 7, 171. https://doi.org/10.3389/fvets.2020.00171

Moberg, G.P., Mench, J.A., 2000. The Biology of Animal Stress: Basic Principles and Implications for Animal Welfare. CABI Publishing, New York. https://doi.org/10.1079/9780851993591.0001

Mullen, B., Champagne, T., Krishnamurty, S., Dickson, D., Gao, R.X., 2008. Exploring the safety and therapeutic effects of deep pressure stimulation using a weighted blanket. Occup. Ther. Ment. Health. 24, 65–89. https://doi.org/10.1300/J004v24n01_05

National Research Council (US), 2008. Stress and Distress: Definitions, in: Recognition and Alleviation of Distress in Laboratory Animals. National Academies Press (US), Washington D.C.

Overall, K.L., 2013. Manual of Clinical Behavioral Medicine for Dogs and Cats. Mosby, St Louis. ISBN 9780323008907.

Pfaff, D.W., Martin, E., Ribeiro, A.C., 2007. CNS arousal necessary for supporting stress responses: Reply to “More appraisal please”: A day and walker commentary on Pfaff et al. “Relations between mechanisms of CNS arousal and mechanisms of stress.” Stress. 10, 314–315. https://doi.org/10.1080/10253890701638279

Philpotts, I., Dillon, J., Rooney, N., 2019. Improving the welfare of companion dogs—Is owner education the solution? Animals (Basel) 9. https://doi.org/10.3390/ANI9090662

Posner, J., Russell, J.A., Peterson, B.S., 2005. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. Dev. Psychopathol. 17, 715. https://doi.org/10.1017/S0954579405050340

Raileanu, Y.A., 2021. A pilot study for Fear Free client education based on the Fear Free scale, in: The 43rd Veterinary Symposium of ANIMAL WELFARE. Rishon Le-Zion.

Riemer, S., Heritier, C., Windschnurer, I., Pratsch, L., Arhant, C., Affenzeller, N., 2021. A review on mitigating fear and aggression in dogs and cats in a veterinary setting. Animals (Basel) 11, 1–27. https://doi.org/10.3390/ANI11010158

Siracusa, C., Manteca, X., Cerón, J., Martínez-Subiela, S., Cuenca, R., Lavin, S., Garcia, F., Pastor, J., 2008. Perioperative stress response in dogs undergoing elective surgery: variations in behavioural, neuroendocrine, immune and acute phase responses. Anim. Welf. 17, 259–273. https://doi.org/10.1017/S0962728600032188

Smith, W.G., 2008. Does gender influence online survey participation? A record-linkage analysis of university faculty online survey response behavior. Online Submission, 21 pp. Available at: https://eric.ed.gov/?id=ED501717

Souza, P., Guo, K., Mills, D.S., Resende, B., Albuquerque, N., 2023. How do dogs behave when presented with situations of different emotional valences? Animals. 13, 1027. https://doi.org/10.3390/ANI13061027/S1

Stanford, T.L., 1981. Behavior of dogs entering a veterinary clinic. Appl. Anim. Ethol. 7, 271–279. https://doi.org/10.1016/0304-3762(81)90083-3

Steimer, T., 2002. The biology of fear- and anxiety-related behaviors. Dialogues. Clin. Neurosci. 4, 231. https://doi.org/10.31887/DCNS.2002.4.3/TSTEIMER

Stephen, J.M., Ledger, R.A., 2005. An audit of behavioral indicators of poor welfare in kenneled dogs in the United Kingdom. J. Appl. Anim. Welf. Sci. 8, 79–95. https://doi.org/10.1207/S15327604JAWS0802_1

Stollar, O.O., Moore, G.E., Mukhopadhyay, A., Gwin, W., Ogata, N., 2022. Effects of a single dose of orally administered gabapentin in dogs during a veterinary visit: A double-blinded, placebo-controlled study. J. Am. Vet. Med. Assoc. 260, 1031–1040. https://doi.org/10.2460/JAVMA.21.03.0167

Tami, G., Gallagher, A., 2009. Description of the behaviour of domestic dog (Canis familiaris) by experienced and inexperienced people. Appl. Anim. Behav. Sci. 120, 159–169. https://doi.org/10.1016/J.APPLANIM.2009.06.009

Taylor, K.D., Mills, D.S., 2006. The development and assessment of temperament tests for adult companion dogs. J. Vet. Behav., 94–108. https://doi.org/10.1016/j.jveb.2006.09.002

Travain, T., Colombo, E., Heinzl, E., Bellucci, D., Prato-Previde, E., Valsecchi, P., 2014. Hot dogs: Thermography in the assessment of stress in dogs (Canis familiaris)—A pilot study. J. Vet. Behav. 10. https://doi.org/10.1016/j.jveb.2014.11.003

Uccheddu, S., Miklósi, Á., Gintner, S., Gácsi, M., 2022. Comparing pears to apples: Unlike dogs, cats need habituation before lab tests. Animals, 12, 3046. https://doi.org/10.3390/ANI12213046

Vas, J., Topál, J., Gácsi, M., Miklósi, Á., Csányi, V., 2005. A friend or an enemy? Dogs' reaction to an unfamiliar person showing behavioural cues of threat and friendliness at different times. Appl. Anim. Behav. Sci. 94, 99–115. https://doi.org/10.1016/J.APPLANIM.2005.02.001

Zhang, E.Q., Leung, V.S.Y., Pang, D.S.J., 2019. Influence of rater training on inter-And intrarater reliability when using the rat grimace scale. J. Am. Assoc. Lab. Anim. Sci. 58, 178–183. https://doi.org/10.30802/AALAS-JAALAS-18-000044

Declaration of interests

☐ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☒ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Emma Gatehouse reports financial support was provided by Fear Free LLC. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Highlights

- Extreme fear, anxiety, and stress behaviors are easier to identify correctly.
- Mild signs of escalating fear, anxiety, and stress are harder to identify correctly.
- Veterinarians are less reliable when rescoring videos of mild escalating FAS signs.
- Behavior experts assess moderate FAS scores more accurately than dog owners.
- The FAS spectrum is reliable but could improve with further adaptation.