Applied Animal Behaviour Science 238 (2021) 105302

Contents lists available at ScienceDirect

Applied Animal Behaviour Science

journal homepage: www.elsevier.com/locate/applanim

The hierarchical structure of dog personality in a new behavioural assessment: A validation approach

Kenth Svartberg *

Swedish University of Agricultural Sciences, Department of Anatomy, Physiology and Biochemistry, Box 7011, SE-750 07, Uppsala, Sweden

ARTICLE INFO

Keywords:
Dogs
Personality
Temperament
Experimental assessment
Hierarchical factor analysis
Construct validity

ABSTRACT

Experimental assessments can be useful in the study of individual differences among dogs. One example of such assessment is the Dog Mentality Assessment (DMA), in which stable traits, referred to as personality traits, have been detected. Due to limited access to the DMA for dogs of non-working breeds, a new experimental assessment named Behaviour and Personality Assessment in Dogs (BPH) was developed in 2012 with the DMA as a model. In this study, behavioural ratings from 12,117 dogs assessed with the BPH were analysed in two steps: first, a hierarchical factor analysis procedure was carried out, and second, the construct validity of the extracted factors was studied. Two measures of validity were used: correlations with subjective ratings during the assessment (internal construct validity (ICV)) and correlations with data from a web-based questionnaire regarding everyday behaviour (external construct validity (ECV)). The ECV was also used to investigate on which level of the factor hierarchy everyday behaviour was best predicted. The approach revealed a hierarchy of factors, from one general factor at the top to 28 specific factors at the bottom, with generally high ICV. The first factor, *Boldness*, is related to six of the eight subtests and is associated with a positive attitude towards unfamiliar persons, interest for object play, low fear, and high degree of exploration. Most of the specific factors stem from the factors *Sociability*, *Playfulness* and *Non-social fearfulness* at the third level in the hierarchy, factors with high or at least moderate ECV. *Sociability* seems to be the best predictor for attitude towards unknown persons and dogs outside the assessment situation, including positive interest, fear, and aggression. The broader factors at the first levels correspond well to a range of everyday behaviours but for some behavioural tendencies more specific factors appearing at lower levels in the hierarchy were of greater importance. For example, noise-related fear was predicted first by a factor from the 12th factor level. The results from the ECV analysis indicate consistency between contexts and suggest that the BPH can reveal dog personality traits. The information from the assessment may give indications regarding welfare as well as potential problem-causing and preferred behaviour. Given a genetic basis for the traits, the most promising application is in dog breeding, where a combination of broad and narrow factors, relevant for the breed in question, may be used as measures in breeding objectives.

1. Introduction

In recent decades, there has been a substantial increase in scientific interest in the concept of personality in animals (Whitham and Washburn, 2017). Personality traits, which in animals can be defined as behavioural tendencies consistent over time and across situations (Réale et al., 2007), can be considered as causal dispositions: stable properties within the individual that are manifested under certain circumstances, most often of a novel or challenging character (Réale et al., 2010; Kaiser and Müller, 2021). Traits may differ in generality and be hierarchically structured, from narrow traits that relate to a specific behaviour tendency, to broad personality domains based on sets of correlated behaviours that affect behaviour in a range of contexts (Réale et al., 2007; Uher, 2008; Araya-Ajoy and Dingemanse, 2014).

One of the most studied species within animal personality is the domestic dog (Fratkin, 2017). Knowledge of dog personality may be utilized in many different areas. One central aspect is behavioural predictions of individual dogs. For example, identifying dogs that are at a particular risk of problem-causing behaviour, such as aggression, enables preventative action to be taken, and may increase the probability for good owner-matching (Bennett et al., 2012). To what extent adult behaviour is possible to predict from puppy tests is an issue that has

* Correspondence to Present address: Voxome 102, SE-747 91, Alunda, Sweden.
E-mail address: kenth.svartberg@slu.se.

https://doi.org/10.1016/j.applanim.2021.105302
Received 12 June 2020; Received in revised form 12 March 2021; Accepted 14 March 2021
Available online 24 March 2021
0168-1591/© 2021 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

attracted considerable interest due to possible applications for both working and companion dogs (McGarrity et al., 2015). Information about young dog's characteristic behaviour can be used to improve selection efficiency in potential working dogs (Sinn et al., 2010; Foyer et al., 2014). Dog personality is also relevant from a welfare point-of-view. By focusing on how individual dogs react to given situations, preventive steps may be taken in adjusting routines and training to minimize stress and negative emotions (Protopopova, 2016; Lush and Ijichi, 2018). Many of the identified personality traits in dogs have been shown to be influenced by genetic variation (van den Berg, 2016), which provides opportunities for behavioural-based dog breeding. This applies both in breeding programmes for working dogs (e.g., Wilsson and Sinn, 2012) and in companion dogs with the aim to increase welfare and prevalence of preferred traits (King et al., 2012; Arvelius et al., 2014). Studies on genetically influenced personality variation in dogs have also contributed to insights in the domestication process (Marshall-Pescini et al., 2015; Smith et al., 2017) and in the mechanisms behind human behaviour and personality (vonHoldt et al., 2017; Sarviaho et al., 2019).

The approach to personality traits as dispositions manifested in novel and challenging situations requires methodological considerations. Due to the possibility of presenting stimuli of both novel and challenging character to the animal, behavioural assessments in experimental set-ups (experimental assessments) have advantages in animal personality research (Réale et al., 2010; Kaiser and Müller, 2021). Experimental assessments played a central role in the early study of individual differences in dogs (e.g., Pavlov, 1927; Scott and Fuller, 1965; Murphree et al., 1969; Goddard and Beilharz, 1984). More recently, experimental assessment has proven useful even in large-sample dog personality studies, suggested by studies based on data from the Dog Mentality Assessment (DMA). The DMA is organised by the Swedish Working Dog Association (SWDA). The DMA is an experimental set-up with 10 subtests where the dog's behaviour is assessed using rating scales with predefined steps based on the dog's behavioural reaction in the specific subtest (behavioural rating scales, described as a 'middle-way' between behaviour coding and subjective rating; Svartberg, 2007). Even though the DMA was developed for working breeds, it became popular among owners and breeders of dogs from other breeds. As of 2020, approximately 120,000 dogs of more than 280 breeds have been assessed with DMA. Results from studies based on DMA data suggest that five or six narrow traits as well as one higher-order trait (Shyness-boldness, or Boldness) are detected by the assessment, which are rank-order consistent over repeated assessments with approximately one month in between (Svartberg and Forkman, 2002; Svartberg et al., 2005). The traits have been validated against data from owner assessments and working dog trials, which indicates that they may be used as indicators for working dog aptitude and as predictors of everyday life behaviour, such as attitudes towards unknown persons and non-social fear (Svartberg, 2002, 2005). The traits are genetically influenced (Strandberg et al., 2005; Saetre et al., 2006; Arvelius et al., 2014) and have been used in the development of a BLUP-index to select against fearfulness (Arvelius and Grandinson, 2012). Variation in the DMA traits among as well as within breeds seem to be related to recent breeding goals, which suggest that the domestication is an ongoing process and that the DMA can assist with relevant indicators (Svartberg, 2006; Sundman et al., 2016; Wheat et al., 2019; Kolm et al., 2020).

The capacity of the DMA is limited, and dogs of working breeds have priority. The restricted access for dogs of other breeds, and requests for a more general assessment from breeders as well as breed organizations of non-working breeds, especially companion dog breeds, were responded to by the Swedish Kennel Club (SKC). In 2009, a development group was created with the aim to develop such an assessment, with a main use as a breeding tool in behavioural-based dog breeding. With the procedure suggested by Taylor and Mills (2006) as a guideline, an assessment was developed termed *Behaviour and Personality Assessment in Dogs* (Beteende- och personlighetsbeskrivning hund (BPH) in Swedish), and a final version was presented in 2012 (Arvelius et al., 2012). Many of the subtests are similar to those of the DMA, but efforts were made to increase the accessibility for dogs of different sizes (e.g., play objects of different size, and a low-placed dummy in the visual appearance subtest for enhanced visibility for dogs of all sizes). Efforts were also made to increase the degree of standardisation, for example by using automated equipment and a more comprehensive procedure description. One major difference is the more detailed assessment, with 241 behavioural ratings made in the BPH, compared to 33 in the DMA.

Rayment et al. (2016) have proposed a hierarchical perspective when investigating dog personality, where narrow and broad traits, and the relationships among them, can be identified. An appropriate factor analysis approach has been suggested by Goldberg (2006), and applied to dog personality (Jones, 2008), where factors are extracted from the first factor and on down. This technique permits the development of a hierarchical representation of factors that can contribute to the understanding of how different aspects of personality relate. As a next step, a set of factors relevant for the purpose can be retained based on their validity—how well the factors represent what they are supposed to (Hair et al., 1998)—for example by investigating the correspondence to relevant behaviour outside the assessment situation.

In this study, the procedure suggested by Goldberg (2006) was applied to a data set containing behavioural ratings from 12,117 dogs of 267 breeds that have been assessed using the BPH. As a second step, the construct validity of the factors was studied in two ways by comparing the factor scores with, in the first instance, subjective ratings during the assessment (internal construct validity (ICV)) and, in the second, data from a web-based questionnaire completed by owners with items related to everyday behaviour (external construct validity (ECV)). The ECV was also used to investigate on which level in the factor hierarchy different aspects of everyday behaviour was best predicted.

2. Materials and methods

2.1. General information about the BPH

The BPH is open to all dogs that have reached an age of 12 months, are registered in the SKC, and owned by a Swedish citizen. Dogs of mixed breed can enter the assessment with a competition license issued by the SKC. Bitches in heat are assessed after other dogs. The BPH includes eight subtests: Unfamiliar person (S1), Object play (S2), Food interest (S3), Visual surprise (S4), Metallic noise (S5), Approaching person (S6), Unstable surface (S7) and Gunshots (S8, optional). Each subtest has one to three phases, which are divided in sequences, described briefly in Table 1 and in depth in supplementary material. The assessment, which takes approximately 45 min to complete, including a verbal summary, is carried out in a flat, open area of approximately 100 square meters, sited to limit disturbance. The equipment for the subtests is located along a pathway and carried out in the same order for each dog (the equipment used in the BPH is developed or compiled centrally by the SKC; authorized organisers, such as dog schools, can purchase equipment from the SKC). Three individuals are involved in the assessment: an assessment leader (AL), a rater, and a figurant, all trained and authorized by the SKC. The AL is the handler's guide during the assessment (encountered first in S1, where the AL acts as the unfamiliar person), the rater makes all behavioural assessments, and the figurant is acting in S6 (as the approaching person) and S8 (the shooter of the gunshots). Before the assessment, the handler of the dog, usually the owner, is informed about the assessment, and the dog's identification is checked. In general, the handler should remain quiet during the assessment to ensure that the dog's reactions are due to its own initiative. Outside S2, the handler is not allowed to play with the dog, and no treats should be given to the dog except in S3. A long tether can be attached to the dog's collar during the assessment to ensure that the dog remains in the assessment area. The dog should wear an ordinary collar or harness (no choke chains or any anti-pulling devices) and is walked on a 1.8 m long leash during the assessment (supplied by the organiser). The rater is obliged to interrupt

Table 1
The subtests that are included in the BPH. (AL = assessment leader; S = subtest; see supplementary material for more details).

Subtest	Phase 1	Phase 2	Phase 3
Unfamiliar person (S1)	The AL approaches the handler and the dog, and invites the dog to interact (3 sequences)	The AL invites the dog for a short walk with interaction (6 sequences)	The dog is physically touched (handled) by the AL (4 sequences)
Object play (S2)	The handler invites the dog to play with an play object familiar to the dog (2 sequences)	The handler invites the dog to play with a standardized play object (2 sequences)	The AL invites the dog to play tug-of-war with the standardized play object (4 sequences)
Food interest (S3)	The dog is released and free to eat open treats and try to reach hidden treats during 60 sec (1 sequence)		
Visual surprise (S4)	A dummy is rapidly raised in front of the dog (1 sequence)	The dog is free to approach the dummy and initiate contact with it (3 sequences)	After contact with the dummy, the handler takes the dog for a walk passing the dummy four times (4 sequences)
Metallic noise (S5)	A metal barrel with chains inside starts to rotate (and creates a rattling noise) in front of the dog (1 sequence)	The dog is free to approach the barrel and initiate contact with it (3 sequences)	After contact with the barrel, the handler takes the dog for a walk passing the barrel four times (4 sequences)
Approaching person (S6)	An oddly dressed person approaches slowly towards the dog (3 sequences)	The dog is released and free to approach the person (6 sequences)	
Unstable surface (S7)	The dog is encouraged by the handler to step up on a unstable surface and pass it back and forth (4 sequences)	The dog is encouraged by the handler to step up on a second unstable surface and pass it back and forth (4 sequences)	
Gunshots (S8)	A shot from a starting gun is fired when the dog is walked by it's handler (2 sequences)	A second shot is fired while the handler and the dog are standing still (3 sequences)	

the assessment if the dog shows such signs of fear as prevent recovery before the next subtest, or due to aggressive behaviour or health issues. The assessment can be stopped at any time by the handler. A subtest, or any part of a subtest, can be interrupted by the rater without stopping the assessment. Spectators are instructed by the AL to not interfere with the assessment procedure. No other dogs are allowed in the assessment area. Significant places, such as the starting point for the handler, are marked on the ground. The AL uses a stopwatch (Casio Module 3425) programmed to beep every 10 s to time the assessment. The rater makes records of the dog's behaviour using a handhold tablet (Samsung Galaxy Tab P6800/P7500, OS Android) with an application specifically developed for this purpose by the SKC. The AL also uses the application to communicate with the SKC database after the assessment to upload the results. During the assessment, the rater makes records using the scales for each subtest; unscored scales are interpreted by the application as non-observed reactions and are set to zero (the first step on the scale, in general defined as "no signs"). The BPH is a non-competitive assessment within the SKC organisation. The dog's behaviour is not valuated in BPH; no points are awarded, and the dogs are not ranked in any way. However, an accomplished BPH, entailing that the dog has carried out subtests 1–7, is used as a criterion in some regards: the owner of the dog can buy a BPH rosette, and a breeder with five or more BPH-accomplished dogs can receive a BPH-breeder diploma. Accomplished BPH is used as a criterion for dogs of non-working breeds to start in working-dog trials, organised by the SWDA. For some breeds, both parents must have accomplished BPH to register offspring in the SKC (as of March 2021, these include Czechoslovakian Wolfdog, Saarlos Wolfhound, Neapolitan Mastiff, Rafeiro of Alentejo and Italian Cane Corso). One new assessment attempt is given to a dog with interrupted BPH. This second assessment may be carried out not earlier than six weeks later and may not be carried out at the same assessment station.

2.2. Sample description

The data were collected in the BPH over 1,855 days between March 2012 and November 2017. Thirty-three sets of assessment equipment were used at 52 assessment stations. At each assessment occasion, 1–15 dogs were assessed (mean = 6.5 dogs). In total, 66 raters scored the dogs' reactions in the assessment (range = 27–1021 dogs/rater; mean = 183.6 dogs/rater) and the number of AL used was 128 (range = 1–811 dogs/AL; mean = 94.7 dogs/AL). A majority of the raters and ALs were of female gender (76 % and 83 %, respectively). In total, the data set contained data from 12,117 dogs that carried out all subtests in the assessment, including 6,601 females (54,5 %) and 5,516 males (45,5 %). Of the females, 934 dogs (14.1 %) were in heat. Of all dogs, 1,106 (9.1 %) were neutered or castrated and 343 dogs (2.8 %) in the data set had a tether attached during the assessment. The data set included dogs from 267 breeds (see supplementary material). Labrador Retrievers were most represented in the sample with 752 (6.2 %) dogs; 30 breeds had more than 100 assessed dogs, and the median number of dogs per breed was 10. All 10 breed groups according to the nomenclature of the Fédération Cynologique Internationale were represented in the sample (Fig. 1). The age of the dogs when assessed ranged from 365 to 4,731 days (mean = 879.6 days), where 6,617 dogs (54.6 %) were between 1 and 2 years old. Fourteen of the dogs in the sample had made the assessment twice. The assessment result used from these dogs was chosen according to the largest number of subtests accomplished (N = 12). If the same number of subtests were carried out in both assessments, data from the first assessment were used (N = 2).

2.3. The behavioural rating scales in the BPH

Eight broader behaviour categories are defined in the BPH: aggressive behaviour, fear-related behaviour, exploratory behaviour, greeting behaviour, submissive behaviour, play-related behaviour, food-related behaviour, and activity-related behaviour. Based on these categories, 32 behaviour types are defined with associated rating scales. Twelve of the behaviour types have more than one associated scale, either of a different type (e.g., intensity and duration) or with different steps. In total, 47 behavioural rating scales are used (Table 2). The scales are based on the behavioural reaction's intensity, duration, latency, frequency, pace, or quality. Each step on the scales, which vary from three to eight, is pre-defined and has a written behavioural description in the protocol. The first step (0) on each scale is in general defined as a non-observed reaction. The majority of the scales are used in the assessment more than once. In total, the protocol includes 241 behavioural ratings (per subtest: 71, 21, 7, 40, 24, 54, 12 and 12, respectively), which in this study were used in the EFA procedure. The categories, types, and scales are presented in Table 2, and described in detail in the supplemental material.

2.4. Analysis of the behavioural rating scales

The EFA of the behavioural rating variables was based on a matrix of

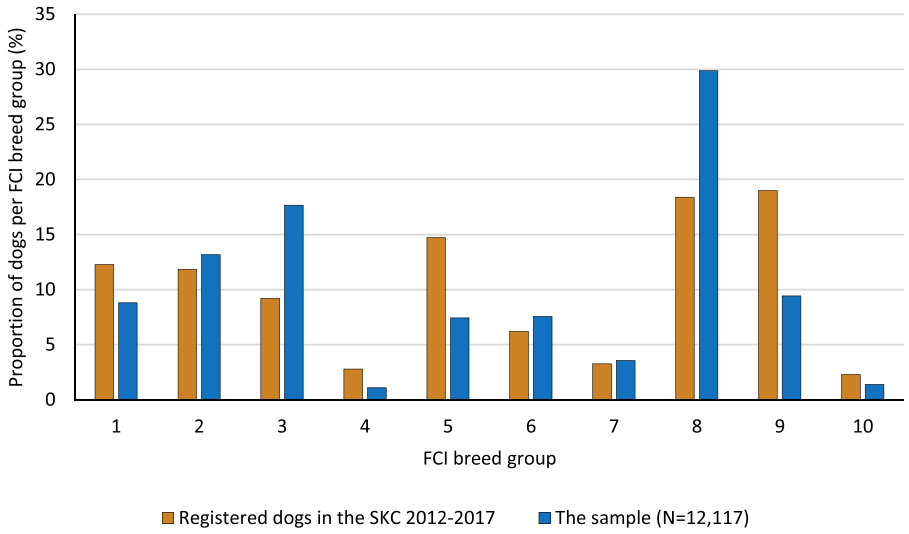

Fig. 1. The proportion of dogs per breed group according to the Fédération Cynologique Internationale (FCI) breed group nomenclature in the sample (blue bars) and in the population of registered dogs in the Swedish Kennel Club (SKC) during the equivalent period of data sampling (2012–2017; orange bars) (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

polychoric correlations, which is relatively insensitive to deviations from normal distribution (Olsson, 1979). However, a number of the variables had a high proportion of zero values (i.e., not observed behaviour). To minimise the risk of poor reliability of the polychoric correlations, a criterion was used where at least 1 % of the dogs should have a record other than zero, which resulted in exclusion of 47 variables. In addition, the variable immediacy of offensive reaction was excluded, since it is dependent on the score for the variable offensive reaction. In S4 (phase 1 & 2), S5 (phase 1 & 2) and S6 (phase 2, seq. 1–4) the dogs are carrying out different numbers of sequences depending on when the dog initiates contact with the stimulus or figurant. To be able to compare the dogs in the data set, aggregated measures of the behaviour recorded in these sequences were calculated, which substituted the original variables: 1) *duration*, number of sequences in which the behaviour was observed, and 2) *intensity*; the maximum score of the behaviour. These were calculated for avoidance and passive avoidance in all three subtests, for threat and silent threat in S4 and S6, and for exploration in S6. Exploration in S4 and S5 are recorded both in phase 1 (for all dogs) and in the first sequence in phase 2 (for dogs that did not initiate contact in phase 1); these two variables were merged into one, where the highest score was used. These behavioural measures substituted the original variables, which left 179 variables to be included in the analyses (see supplementary material). In the data set, there were 12,270 missing values (mv), corresponding to 0.6 % of the total data points (range = 0–48 mv/dog, mean = 1.0 mv/dog). A majority of the mv were due to interrupted phases or sequences in S1, S7, and S8 (9,215 mv from 38 scales), and 2,635 mv were from the scale Neutral state latency in S8, which was added to the protocol first in 2014. The mv were replaced using median imputation. In the EFA, principal axis was used as extraction method, and the factors were obliquely rotated using the oblimin method. I followed a top-down approach developed by Goldberg (2006), in which factors are extracted from the first factor down, until the additional factor has no variable with its highest factor loading on it. Factor scores were calculated for each extracted factor by using the loadings as weights on standardized variables. Correlations between factors from one level with factors from adjoining levels where calculated using Spearman rank order correlation.

2.5. Construct validity analysis

2.5.1. Internal construct validity

In addition to the behavioural ratings, the rater makes subjective ratings as a basis for the verbal summary after the assessment. The protocol included 19 subjective rating scales with five steps where only the lowest (0=“not at all”) and highest (4=“very”) steps are defined. In this study, 18 of these scales were expected to correspond to the content of at least one of the extracted factors and therefore used in the analysis of the extracted factors’ ICV (Table 3). There were few missing values (0–6 mv/scale) with the exceptions of Energetic and Angry overall, which had 2,376 mv each, due to the addition of these scale first in 2014.

2.5.2. External construct validity

For the ECV analysis, data from an online questionnaire were used (Asp et al., 2015). Information and link to the questionnaire were posted on a SKC website, and dog owners planning to take the BPH were encouraged to respond. In addition, information regarding the questionnaire was sent to owners of dogs registered for the assessment. Questionnaires received within 6 months prior to assessment, for dogs at least 365 days old, were included in the analyses. These criteria were met by 1,956 dogs (1,011 females, 945 males) of 190 breeds (range = 1–129 dogs/breed; mean = 10.3 dogs/breed, see supplementary material). The median age was 687.5 days (range = 365–4,574 days), and the questionnaire was received in median 8 days (range = 1–180 days) before the BPH. The questionnaire included questions regarding the respondent and the identity of the dog, and 132 behaviour-related items, describing situations dogs might encounter in everyday life. Of these, 99 were used in the ECV analysis; 72 originated from a Swedish translation of the Canine Behavioural Assessment and Research Questionnaire (C-BARQ; Hsu and Serpell, 2003; Duffy and Serpell, 2012) with 15 added items (Svartberg, 2005), with additional 12 items for this study (see supplementary material). The items were grouped in sections related to behavioural categories, with a brief explanation of typical signs. The respondent was requested to describe the dog’s typical behaviour in the recent past as objectively and neutrally as possible. For 50 of the items (related to trainability, separation-related behaviour, attachment and attention-seeking, play behaviour, social contact and miscellaneous), a rating scale was used to indicate the frequency of the behaviour (“never”, “seldom”, “sometimes”, “usually”, or “always”). For the remaining 39 items, the scale indicated the intensity of the behavioural expression, ranging from no sign of the behaviour (e.g., “no

Table 2
The behaviour categories and types that are assessed in the BPH using behavioural rating scales (AL = assessment leader; S = subtest; see supplementary material for more details).

Behaviour category	Behaviour type	Type of scale	Description/example of signs	Scored in subtest	No of scales	No of steps
Aggressive behaviour	Active threat (threat)	Intensity	Growling, low frequency barking, threatening body posture, ears pricked forward, corners of the mouth pulled forward, raised tail, direct stare and lunging towards the person/test stimulus	S1, S4, S6	2	3/4
	Offensive reaction	Intensity	An immediate threat reaction in response to the test stimulus	S4	1	6
		latency (not analysed)	How immediate the offensive reaction (if noted) is in relation to the start of the test stimulus	S4	1	3
	Silent threat	Intensity	Stiff and up-right body posture, erect and stiff tail, direct stare and low intensity growling	S1, S4, S6	1	4
	Biting (not analyzed)	Intensity	The occurence of bites, bite intentions and/or snarls (seen as a precursor for biting)	S1, S4, S6	2	3/4
Fear-related behaviour	Active avoidance (avoidance)	Intensity	Evasive maneuver, shrinking back, backing off, shelter seeking, flight intention and flight	S1, S4, S5, S6, S8	3	3/4/4
	Defensive reaction	Intensity	An immediate active avoidance reaction in response to the start of the test stimulus	S4, S5	1	6
	Passive avoidance	Intensity	Low body posture with low tail and ears pulled back, pacing, changing glance, panting, salivating, licking, yawning and scratching	S1, S4, S5, S6, S8	1	4
	Separation reaction	Intensity	How unwilling the dog is to be separated from the handler, often expressed as straining in the leash towards the handler	S1	1	4
	Recovering latency	Latency	Latency until the dog has recovered from a fearful event; when the dog acts in an open way without any signs of fear the dog is defined as recovered	S4, S5, S8	1	6
	Approach latency	Latency	Latency, measured as number of sequences, until the dog initiated contact with the test stimulus	S4, S5, S6	2	5/8
	Change of pace	Pace	A change of pace when the dog passes the test stimulus; either hesitation - slowing down while passing the test stimulus (seemingly to avoid it) - or acceleration - increasing the pace while passing the test stimulus (seemingly to avoid it)	S4, S5, S7	3	2/3
	Startle reaction	Intensity	An immediate reaction in response to the gunshot that may be limited to the head but can also be seen as a short flight	S8	1	4
Exploratory behaviour	Exploration	Intensity	Olfactory and/or visual inspection together with stretched neck or steps towards or laterally in relation to the test stimulus	S1, S4, S5, S6	3	3/4/4
	Disinterest	Duration/ frequency	General exploration of other than the test stimulus, for example ground inspection or inspection of the surroundings	S2, S3, S4, S5, S6	1	4
Greeting behaviour	Greeting	Intensity	The intensity with which the dog greets the person; common signs are tail waving, pacing, jumping against the person, whining, barking and other social vocalization	S1, S6	1	4
		Duration	The duration of the greeting event	S1, S6	1	4
Submissive behaviour	Submissive greeting	Intensity	Ears held back or flattened, head held low, body lowered, tail low or tucked and/or licking or licking intentions during greeting	S1, S6	1	4
Play-related behaviour	Toy approach	Pace	The pace of the dog's approach towards the object after it has been thrown by the handler	S2	1	4
	AL/toy approach	Pace	The pace of the dog's approach towards the object and/or the AL when the AL holds the object and invites the dog to play	S2	1	4
	Grabbing	Intensity	The intensity with which the dog grabs the object	S2	1	4
		Quality	The type of grabbing; from grabbing with the front teeth to grabbing with "full mouth"	S2	1	3
	Holding	Quality	How steady the dog holds the object after grabbing it	S2	1	3
	Play interest	Intensity	The intensity with which the dog plays with the object; common signs are intense grabbing, holding, shaking and tossing of the object while running around	S2	1	4
		Duration	The duration of the play event	S2	1	4
	Tug-of-war	Intensity	The intensity with which the dog is holding and pulling the object while the AL is holding it	S2	1	4
		Duration	The duration of the tug-of-war event	S2	1	4
	Handler-directed play	Duration	Play invitations to the handler; common signs are looks and approaches towards the handler, the dog may bring the object to the handler	S2	1	4
	AL-directed play	Duration	Play invitations to the AL; common signs are looks and approaches towards the AL, the dog may bring the object to the AL	S2	1	4
Food-related behaviour	Food approach	Pace	The pace of the dog's approach towards the treats	S3	1	5
	Appetite	Intensity	The intensity of the dog's eating of the open treats	S3	1	5
	Persistence	Duration	The dog's interest of the food station (eating the open and attempting to reach the hidden treats)	S3	1	5
	Physical manipulation	Duration	The dog's physical manipulation (scratching, digging, biting, chewing, pulling) of the food station while attempting to reach the treats	S3	1	5
	Handler/AL contact	Frequency	The number of the dog's contact initiations (looks, approaches) made by the dog towards the handler and/or the AL (positioned besides each other)	S3	1	5
		Duration	The duration of the dog's contact initiations (looks, approaches) towards the handler and/or the AL	S3	1	5
Activity-related behaviour	Activity increment	Intensity	Increased visual attention, high held ears, raised head and body, pacing, whining and/or barking in response to the gunshot	S8	1	4
	Vocalization	Duration	The duration of the dog's vocalization (for example, whining and/or barking)	S8	1	4
	Neutral state latency	Latency	Latency until the dog has returned to it's neutral state of activity after the gunshot; the neutral state is characterized with relaxation and that the dog initiates low-arousal activities	S8	1	6

Table 3
Definitions of and basic statistics for the subjective rating scales that were used in the analysis of the extracted factors' ICV (S = subtest). The scales range from 0 (defined as "not at all") to 4 (defined as "very").

Subtest	Scale	Definition	N	Min.	Max.	Median	Mean	Stand dev.
S1	People friendly	Enjoys the company of unfamiliar persons, responds to invitations from strangers and often invites to interactions (1)	12117	0	4	2	1.95	1.03
S1	Confident	Has confidence and acts rationally in fearful situations (2)	12116	0	4	3	3.06	1.02
S2	Energetic	Spirited and alert, reactive, always on the move (3)	12117	0	4	3	2.54	0.98
S2	Playful	Responds to play invitations and invites others to play (4)	12117	0	4	2	2.21	1.19
S3	Energetic	See (3)	12116	0	4	2	2.23	0.93
S4	Curious	Interested in events that occur and explores them in a positive spirit (5)	12117	0	4	2	2.34	1.15
S4	Angry	Gets easily angry and shows it with threatening behaviour (6)	12117	0	4	0	0.68	0.85
S4	Confident	See (2)	12117	0	4	3	2.65	1.00
S5	Curious	See (5)	12117	0	4	3	2.50	1.22
S5	Confident	See (2)	12117	0	4	3	3.18	0.96
S6	People friendly	See (1)	12114	0	4	2	1.87	1.08
S6	Angry	See (6)	12114	0	4	0	0.70	0.92
S6	Confident	See (2)	12114	0	4	3	2.87	0.96
S7	Confident	See (2)	12115	0	4	4	3.32	0.99
S8	Confident	See (2)	12111	0	4	4	3.60	0.87
Overall	Positive	Has a generally positive attitude	12117	0	4	3	2.84	0.86
Overall	Energetic	See (3)	9741	0	4	3	2.56	0.81
Overall	Angry	See (6)	9741	0	4	0	0.09	0.38

Table 4
Basic statistics, internal consistency reliability estimates and origins for the subscales from the web-based questionnaire used in the ECV analysis. The subscales are based on items with scales ranging from 1, defined as never occurring (frequency)/no signs (intensity), to 5, defined as always occurring/intense expression.

Questionnaire subscale	Valid N	Minimum	Maximum	Median	Mean	Standard deviation	No of items	Average inter-item correlation	Cronbach's alpha	Type of scale	Origin
Stranger-directed aggression	1931	1.00	3.88	1.20	1.28	0.32	10	0.54	0.92	intensity	Hsu and Serpell (2003)
Dog-directed fear or aggression	1903	1.00	4.63	1.25	1.43	0.45	8	0.41	0.85	intensity	Hsu and Serpell (2003)
Separation-related behaviour	1872	1.00	5.00	1.13	1.30	0.45	8	0.44	0.88	frequency	Hsu and Serpell (2003)
Trainability	1954	1.88	5.00	3.75	3.68	0.52	8	0.29	0.77	frequency	Hsu and Serpell (2003)
Owner-directed aggression	1935	1.00	3.75	1.00	1.03	0.14	8	0.54	0.90	intensity	Hsu and Serpell (2003)
Attachment or attention seeking behaviour	1876	1.00	5.00	2.83	2.92	0.64	6	0.30	0.72	frequency	Hsu and Serpell (2003)
Excitability	1879	1.00	5.00	3.00	2.92	0.79	6	0.41	0.81	intensity	Hsu and Serpell (2003)
Nonsocial fear	1905	1.00	4.50	1.33	1.39	0.42	6	0.37	0.78	intensity	Hsu and Serpell (2003)
Dog-directed interest	1862	1.00	5.00	4.00	3.84	0.80	5	0.59	0.88	frequency	Svartberg (2005)
Human-directed play interest	1873	1.00	5.00	3.80	3.73	0.84	5	0.56	0.86	frequency	Svartberg (2005)
Stranger-directed interest	1871	1.00	5.00	4.00	3.84	0.86	5	0.61	0.89	frequency	Svartberg (2005)
Stranger-directed fear	1904	1.00	5.00	1.00	1.26	0.51	4	0.77	0.93	intensity	Hsu and Serpell (2003)
Stairs/surface fear	1876	1.00	5.00	1.25	1.33	0.49	4	0.60	0.86	intensity	Present study
Dog rivalry	1247	1.00	5.00	1.25	1.33	0.51	4	0.62	0.87	intensity	Duffy and Serpell (2012)
Noise-related fear (1)	1840	1.00	4.67	1.33	1.45	0.67	3	0.62	0.83	intensity	Present study
Energy level	1851	1.00	5.00	3.50	3.47	0.87	2	0.65	–	frequency	Duffy and Serpell (2012)
Exploration tendency	1832	1.50	5.00	4.00	4.02	0.66	2	0.54	–	frequency	Present study
Fear recovery latency	1820	1.00	5.00	2.00	1.75	0.65	2	0.65	–	frequency	Present study
Appetite	1865	1.00	5.00	5.00	4.74	0.62	1	–	–	frequency	Present study
Contact seeking tendency	1835	1.00	5.00	3.00	3.44	0.95	1	–	–	frequency	Present study
Handling fear (2)	1898	1.00	5.00	1.00	1.41	0.76	1	–	–	intensity	Hsu and Serpell (2003)
Submissive tendency	1856	1.00	5.00	2.00	2.31	1.19	1	–	–	frequency	Present study
Tendency to impose	1861	1.00	5.00	3.00	2.54	1.20	1	–	–	frequency	Present study

(1) two items in 'noise-related fear' are also used in the subscale 'nonsocial fear'; (2) one item from the C-BARQ subscale 'touch sensitivity' is used in 'handling fear' (see supplementary material).

visible signs of aggression") to intense expression (e.g., "serious aggression: snaps, bites, or attempts to bite"). The answers in the questionnaire were transformed to scales from 1 to 5, and condensed to 23 subscales used in the ECV analysis (Table 4). Subscale scores were calculated according to Duffy and Serpell (2012) and Svartberg (2005). For subscales with only one item, the original scores served as subscale scores. For testing the internal consistency reliability of the subscales, the average inter-item correlation and Cronbach's alpha (for subscales with > 2 items) were calculated for each subscale based on polychoric correlations.

2.5.3. Validation criteria

The content of each extracted factor was interpreted based on the variables loading > 0.4 or < -0.4 on the factor. For the ICV analysis, subjective rating scales from the same subtest as the factor's loading variables that could be expected to correspond to the factors content were used as ICV criteria. Expected correlations are presented in Table 5. For the ECV analysis, the expected correlations were based on similarity in content between each factor and questionnaire subscale (Table 6). Even though the stimuli in S1 and S6 are unfamiliar persons, subscales related to reactions to unfamiliar dogs were used as ECV criteria, as well when validating the factors related to *Sociability* and *Aggressiveness*. This was due to a possible shared social component. For factors that were extracted at several levels, the highest obtained correlation in absolute terms in the expected direction was used as a measure of ECV. When investigating the correlation to subscales at different levels in the hierarchical solution, all correlations were of interest, expected or not. Since even small correlation coefficients reached statistical significance with the large sample sizes (for $p = 0.05$ approx. $r_s = 0.02$ in the ICV analysis and $r_s = 0.05$ in the ECV analysis) the effect size, measured by the magnitude of the correlation coefficient (Spearman rank order correlation), was used as a validity index (Bosco et al., 2015). Based on previous comparisons between subjective rating and behaviour coding or rating (Capitanio, 1999; Konečná et al., 2008; Wilsson and Sinn, 2012), $r_s \geq$

Table 5
Correlations (Spearman rank order correlation coefficients) between extracted factors and subjective rating scales from the BPH data set used as ICV index (S = subtest). Green fields indicate expected positive correlations (underlined coefficients); red fields indicate expected negative correlations (italic coefficients). $r_s > 0.30$ and < -0.30 are presented; $r_s > 0.60$ and < -0.60 are in bold. (For interpretation of the references to colour, the reader is referred to the web version of this article).

		Subjective rating scale																	
Factor	Factor relationship	People friendly S1	Confident S1	Energetic S2	Playful S2	Energetic S3	Curious S4	Angry S4	Confident S4	Curious S5	Confident S5	People friendly S6	Angry S6	Confident S6	Confident S7	Confident S8	Positive overall	Energetic overall	Angry overall
Boldness		0.53	0.46	0.50	0.58	0.31	0.51		0.52	0.49	0.54	0.56		0.47	0.41		0.58	0.46	
Sociability/playfulness		**0.64**	0.43	**0.62**	**0.74**	0.32						**0.61**		0.34			0.53	0.47	
Sociability	Person interactions	**0.77**	**0.62**	0.37	0.36				0.32			**0.71**		0.54			0.46	0.35	
Greeting	Person interactions	**0.78**	0.48	0.39	0.38							**0.73**		0.31			0.48	0.38	
Approaching person sociability	Person interactions	**0.62**	0.40	0.32	0.33		0.30					**0.77**		0.41			0.45	0.31	
Handling greeting	Person interactions	**0.77**	0.51	0.33	0.31							**0.61**					0.42	0.32	
Initial greeting	Person interactions	**0.72**	0.36	0.41	0.40							**0.62**					0.46	0.39	
Handling anxiety	Person interactions	-0.54	***-0.63***									-0.46		-0.35		-0.32	-0.35		
Handling avoidance	Person interactions	-0.49	***-0.65***						-0.33		-0.30	-0.44		-0.37			-0.31		
Separation reaction	Person interactions		*-0.58*																
Playfulness	Play	0.39		**0.68**	**0.87**	0.35						0.36					0.42	0.47	
Standardized object play	Play	0.36		**0.67**	**0.87**							0.34					0.41	0.44	
Familiar object play	Play	0.33		**0.61**	**0.74**												0.38	0.44	
Fetch play	Play	0.30		**0.61**	**0.79**												0.36	0.41	
Tug-of-war play	Play	0.40		**0.65**	**0.82**							0.39					0.43	0.44	
Non-social fearfulness	Non-social fear		-0.33				***-0.64***		***-0.69***	***-0.60***	***-0.70***			*-0.45*	*-0.59*	-0.32	*-0.44*	-0.32	
Surface fear	Non-social fear						-0.41		-0.45	-0.45	-0.54				***-0.82***		-0.35		
Surprise & noise fear	Non-social fear						***-0.70***		***-0.78***	***-0.64***	***-0.71***			-0.39	-0.40		*-0.42*		
Gunshot fear	Non-social fear						-0.37		-0.41	-0.42	-0.54			-0.33		***-0.61***	-0.37		
Remaining fear	Non-social fear						-0.53		*-0.52*	-0.58	***-0.60***				-0.33		-0.37		
Noise fear	Non-social fear						-0.54		-0.57	**-0.71**	***-0.77***			-0.32	-0.41		-0.36		
Remaining avoidance	Non-social fear						**-0.61**		***-0.69***	-0.51	*-0.58*			-0.35	-0.37		-0.38		
Remaining pace change	Non-social fear						-0.55		***-0.63***	-0.60	***-0.70***			-0.35	-0.39		-0.34		
Surface hesitation	Non-social fear						-0.46		-0.48	-0.44	-0.53				***-0.77***		-0.35		
Surface acceleration	Non-social fear						-0.30		-0.38	-0.36	-0.48				***-0.74***				
Anxiety	Anxiety	-0.31	*-0.45*				-0.37		*-0.45*	-0.34	*-0.49*	-0.36		*-0.50*		*-0.44*	*-0.40*		
Surprise & appr. person anxiety	Anxiety	-0.31	-0.33				-0.40		*-0.47*	-0.32	-0.45	-0.50		*-0.49*			*-0.42*	-0.30	
Approaching person anxiety	Anxiety						-0.30		-0.37		-0.36	-0.33		*-0.45*			-0.32		
Suprise & noise anxiety	Anxiety						-0.54		*-0.58*	-0.45	*-0.55*			-0.36	-0.30		-0.34		
Aggressiveness	Aggression							0.43	-0.39				**0.72**	*-0.57*					
Approaching person aggression	Aggression							0.35	-0.32				**0.77**	***-0.60***					
Surprise aggression	Aggression						**-0.66**	**0.62**	***-0.78***	-0.37	-0.54			-0.37			-0.32		
Imposing	Aggression							0.34				0.31	0.34						
Approaching person exploration	Exploration									0.32		0.44							
Noise exploration	Exploration																		
Disinterest	Exploration						*-0.43*		*-0.38*	*-0.34*									
Handler play	Handler				0.34				0.33										
Handler contact	Handler			0.30															
Food interest				0.34	0.40	**0.72**				0.33							0.31	0.31	
Gunshot activity				0.31	0.32		0.33			0.33							0.30	0.35	
Submissiveness																			

Table 6
Correlations (Spearman rank order correlation coefficients) between extracted factors from the BPH and 16 of the subscales from the questionnaire (green indicates expected positive correlations (underlined coefficients); red indicates expected negative correlations (italic coefficients)). $r_s > 0.10$ and < -0.10 are presented (no correlations of that magnitude were found for the seven subscales that are not included); $r_s > 0.30$ and < -0.30 are in bold. (For interpretation of the references to colour, the reader is referred to the web version of this article).

		Questionnaire subscale															
Factor	Factor relationship	Nonsocial fear	Stairs/surface fear	Noise-related fear	Fear recovery latency	Stranger-directed aggression	Stranger-directed fear	Handling fear	Stranger-directed interest	Dog-directed fear or aggression	Dog rivalry	Dog-directed interest	Human-directed play interest	Appetite	Energy level	Exploration tendency	Trainability
Boldness		*-0.14*			*-0.19*	-0.24	***-0.32***	*-0.11*	**0.31**	*-0.18*		0.18	0.29		0.17	0.23	
Sociability/playfulness					*-0.16*	-0.28	***-0.33***	*-0.14*	**0.37**	*-0.23*	-0.10	0.24	**0.39**	0.12	0.20	0.19	0.16
Sociability	Person interactions				*-0.19*	*-0.30*	***-0.40***	*-0.19*	**0.43**	*-0.23*		0.23	0.21		0.16	0.17	
Greeting					-0.17	-0.27	**-0.38**	-0.17	**0.42**	-0.24		0.21	0.23		0.17	0.15	
Approaching person sociability					*-0.18*	*-0.27*	***-0.37***	-0.16	**0.40**	*-0.23*		0.19	0.19		0.12	0.16	
Handling greeting					-0.16	-0.24	**-0.35**	-0.17	**0.38**	-0.20		0.18	0.19		0.14	0.13	
Initial greeting					-0.14	-0.27	**-0.34**	-0.16	**0.39**	-0.22		0.21	0.24	0.10	0.17	0.15	
Handling anxiety					0.16	0.25	**0.33**	0.17	**-0.33**	0.18		-0.19	-0.17		-0.13	-0.13	0.11
Handling avoidance		0.11		0.11	0.19	0.22	0.30	0.16	-0.30	0.16		-0.14				-0.11	
Separation reaction					0.12	0.15	0.21	0.11	-0.23	0.13		-0.13					
Playfulness	Play					-0.17	-0.15		0.18	-0.14	-0.11	0.19	**0.44**	0.12	0.20	0.16	0.25
Standardized object play					0.10	-0.19	-0.17		0.19	-0.15	-0.11	0.21	**0.43**	0.10	0.20	0.17	0.22
Familiar object play						-0.12	-0.12		0.13			0.13	**0.43**	0.11	0.17	0.12	**0.31**
Fetch play						-0.13	-0.11		0.12	-0.10		0.14	**0.38**		0.18	0.14	0.22
Tug-of-war play					-0.13	-0.23	-0.22		0.24	-0.18		0.23	**0.41**		0.20	0.17	0.15
Non-social fearfulness	Non-social fear	0.18	0.12	0.14	0.16	0.11	0.20		-0.13				-0.10			*-0.19*	
Surface fear		0.13	0.13		0.13	0.11	0.16		-0.11							-0.14	
Surprise & noise fear		0.17	0.10	0.13	0.13		0.16		-0.12							*-0.17*	
Gunshot fear		0.24		0.27	0.19	0.14	0.20		-0.17	0.10						-0.18	
Remaining fear		0.12							-0.10							-0.16	
Noise fear		0.17	0.11	0.13	0.14		0.14		-0.11							*-0.18*	
Remaining avoidance		0.13			0.10		0.14		-0.11							-0.15	
Remaining pace change		0.13			0.10		0.12									-0.16	
Surface hesitation		0.14	0.13		0.12		0.15									-0.14	
Surface acceleration		0.14	0.13		0.16		0.17		-0.11							-0.13	
Anxiety	Anxiety	0.16		0.16	0.18	0.13	0.23	0.10	-0.23	0.12		-0.15	-0.11		-0.12	-0.19	0.10
Surprise & appr. person anxiety		0.10			0.15	0.12	0.23		-0.23	0.12		-0.14	-0.13		-0.11	-0.20	
Approaching person anxiety					0.10		0.17		-0.16							-0.15	
Suprise & noise anxiety		0.12			0.12		0.14									-0.17	
Aggressiveness	Aggression				0.11	0.21	0.21		-0.22	0.12							
Approaching person aggression						0.22	0.23	0.11	-0.25	0.14							
Surprise aggression		0.14			0.12	0.13	0.17		-0.12								
Imposing							-0.12		0.12							0.11	
Approaching person exploration	Exploration						-0.14		0.20	-0.15		0.15				0.16	
Noise exploration							0.11										
Disinterest		0.10															
Handler play	Handler	-0.14		-0.12		0.13	0.13		-0.18	0.12		-0.12	0.19				0.25
Handler contact													0.21				0.20
Food interest													0.19	0.16		0.13	0.11
Gunshot activity				0.11			-0.13		0.14				0.21		0.15	0.16	
Submissiveness							0.13										

0.40 was used as a limit for ICV (indicating moderate validity), and $r_s \geq 0.60$ was treated as high validity. Previous studies indicate that considerably lower correlations are to be expected when experimental assessment data and questionnaire data are compared (Svartberg, 2005; Jones, 2008; Mirkó et al., 2013). Based on this, $r_s \geq 0.15$ and $r_s \geq 0.30$ were used as limits of moderate and high ECV, respectively.

2.6. Interrater reliability

An interrater reliability study was carried out in 2016 at the request of the SKC, where 47 raters independently of each other assessed 15 dogs each in the BPH (Svartberg, 2016). Since the results of that study are relevant for the present study, a summary of the interrater reliability study is presented in the supplementary material, and the results are briefly presented in the results section in this study. Besides percentage

Table 7
Label, description, number and level/-s in the hierarchy for each factor (in order of appearance). Positive loadings (PL) refers to loadings > 0.4; negative loadings (NL) refers to loadings < -0.4. The number of the column refers to on what factor level/-s in the hierarchy the factor emerged on; the number in the cell refers to the factor's number on each level (AL = assessment leader; S = subtest).

		Factor level/factor number																							
Factor label	Description	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	28
Boldness	PL from greeting (S1 & S6), play variables (S2) and exploration (S4 & S5); NL from fear-related variables (S1, S4, S5, S6 & S7)	1																								
Sociability/ playfulness	PL from greeting (S1 & S6) and play variables (S2); NL from fear-related variables in S1		1																							
Non-social fearfulness	PL from fear-related variables in S4 and S5 (all levels), S6 (levels 2–4 & 6–7), and S7 (levels 2–5); NL from exploration in S4 and S5		2	2	2	2	2	2																		
Sociability	PL from greeting in S1 & S6; NL from fear-related variables in S1 and S6			1	1	1	1	1	1	1																
Playfulness	PL from play variables in S2 (and at some levels from food approach in S3); NL from disinterest in S2			3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3					
Aggressiveness	PL mainly from threat and avoidance variables in S6, but also from threat variables in S4				4	4	4	4																		
Submissiveness	PL from submissive greeting variables in S1 and S6					5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
Surface fear	PL from fear-related variables in S7						6	6	6	6	6	6	6	6	6	6	6	6	6	6						
Handling anxiety	PL from passive avoidance variables during handling by the AL in S1 (phase 2 & 3)							7			7			7	7	1	7	7	18	18	18	18	7	23	24	24
Surprise & noise fear	PL from fear-related variables in S4 and S5; NL from exploration in S4 and S5								2	2	2	2	2	2	2	2										
Approaching person aggression	PL from threat and avoidance variables in S6								4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
Anxiety	PL from passive avoidance variables in S4 and S6 (all levels), S1 (levels 8–9 & 11–12), S5 (levels 8–11), and S8 (levels 8–11)								8	8	8	8	8													
Handling avoidance	PL from avoidance variables during handling by the AL in S1 (phase 2 & 3)									7	10	7	7	13	11	11	11	13	13	13	13	13	13	13	13	13
Imposing	PL from silent threat variables in S4 and S6									9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
Greeting	PL from greeting variables in S1 and S6										1	1	1	1	1	7	1	1	1	1	1	1	1			
Gunshot activity	PL from activity increment, vocalization and neutral state latency in S8											10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
Surprise aggression	PL from threat and avoidance variables in S4											11	11	11	8	8	8	11	11	11	11	11	11	11	11	11
Gunshot fear	PL from starle reaction (except levels 12 & 13), avoidance, passive avoidance, and recovery latency in S8												12	12	12	12	12	12	12	12	12	12	12	12	12	12
Surprise & appr. person anxiety	PL from passive avoidance variables in S4 and S6													8	13	13	13	8	8	8	8	8	8	8	8	
Handler play	PL from handler play variables in S2														14	14	14	14	14	14	14					14
Approaching person sociability	PL from greeting variables, and NL from threat, avoidance and approach latency in S6 (phase 2, seq. 5–6)															15	15	15	15	15	15	15	15	15	15	15
Remaining fear	PL from avoidance and change of pace variables and NL from exploration variables in phase 3 of S4 and S5																2	2	2							
Noise fear	PL from fear-related variables and NL from exploration in phase 1 and 2 of S5																16	17	17	17	17	17	2	2	2	23
Food interest	PL from food-related behaviour (appetite, persistence and physical manipulation) and NL from disinterest in S2																	16	16	16	16	16	16	16	16	16
Approaching person exploration	PL from exploration variables and NL from approach latency in S6																		7	7	7	7	18	7	7	17
Remaining avoidance	PL from avoidance variables in phase 3 of S4 and S5																			2	2	2	17	17	17	2
Remaining pace change	PL from change of pace variables in phase 3 of S4 and S5																			19	19	19	19	18	18	19
Surface hesitation	PL from hesitation variables in S7																				6	6	6	6	6	6
Surface acceleration	PL from acceleration variables in S7																				20	20	20	20	20	20
Standardized object play	PL from play variables in phase 2 (levels 21–24) and phase 3 (levels 21–24) of S2; NL from disinterest in S2 (levels 21–24)																					3	3	3	3	

(*continued on next page*)

Table 7 (*continued*)

Factor label	Description	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	28
		Factor level/factor number																								
Familiar object play	PL from play variables in phase 1 (levels 21–28) and from handler play variables in phase 2 and 3 (levels 21–24) of S2																					14	14	14	21	21
Disinterest	PL from disinterest variables in S4 and S5 (levels 21–24) and S6 (level 28)																					21	21	21	23	18
Noise exploration	PL from exploration variables in phase 3 of S5																						22	22	14	25
Handling greeting	PL from greeting variables during handling by the AL in S1 (phase 2)																							1	1	1
Initial greeting	PL from greeting variables in phase 1 of S1 and S6																							19	19	7
Handler contact	PL from the handler/AL contact variables in S3																								22	22
Approaching person anxiety	PL from passive avoidance variables S6																									8
Surprise & noise anxiety	PL from passive avoidance variables in S4 and S5																									26
Fetch play	PL from play variables in phase 2 of S2																									3
Tug-of-war play	PL from play variables in phase 3 of S2; NL from disinterest in S2																									27
Separation reaction	PL from separation reaction variables in S1																									28
Proportion of total variation explained (%)		13.3	20.4	26.3	30.2	33.5	36.4	38.7	40.9	42.8	44.5	46.2	47.9	49.4	50.8	52.1	53.3	54.5	55.6	56.7	57.7	58.7	59.7	60.6	61.6	64.9

agreement with the reference rater, agreement was estimated using the agreement coefficient suggested by Gwet (2014), both unweighted (AC1) and weighted (AC2).

2.7. Software

The software R was used for most of the analyses (R Development Core Team, 2014). The psych package was used for calculating the polychoric correlations, the EFA, and the appropriateness of the data set, and the reliability estimates for the questionnaire subscales (Revelle, 2018). For basic data analyses and Spearman rank order correlation analyses, the software STATISTICA was used (StatSoft, Inc., 2013).

3. Results

3.1. The hierarchical factor analysis

The Bartlett's test of sphericity (χ^2 = 3029556, df = 15931, p < 0.001) as well as the measure of sampling adequacy (MSA = 0.85) indicated that the data set with 179 behavioural rating scales from 12,117 dogs was appropriate for EFA. The correlation matrix was not positive definite; smoothing was done by adjusting the eigen values. The top-down EFA procedure reached the stopping criterion at 24 factors. Analyses of the solutions with more factors showed also that the 28-factor solution was within the criterion, i.e., that all factors had variables with its highest loading on them, and that all factors were interpretable. Therefore, all factors down to the 24-factor level were analysed and, in addition, the 28-factor solution, explaining 64.9 % of the total variance. In total, 41 factors were identified and labelled (one additional factor at the 8th factor level was undefined, and not included in the analysis). A brief description of the factors is presented in Table 7. The factor loadings and the inter-correlations between factors are presented in supplementary material. The hierarchical relationships among the factors are graphically presented in Fig. 2.

The first factor was loaded (defined as loadings > 0.4 or < -0.4) by variables from S1, S2, S4, S5, S6 and S7. Variables related to greeting, play, and exploration were loading positively, whereas fear-related behaviour, such as avoidance, passive avoidance, recovering latency and approach latency, loaded negatively on the factor. Due to the similarity in loading pattern of the higher-order factor found in the DMA (Svartberg and Forkman, 2002) the factor was labelled *Boldness*. This first factor split off at the second level into *Sociability/playfulness*, which had positive loadings from greeting (S1 & S6) and play variables (S2), and negative loadings from variables related to fear in the first subtest, and *Non-social fearfulness*, which had positive loadings from fear-related variables in S4, S5 and S7. At the third level, *Playfulness* separated from *Sociability*, and at the fourth factor level a new factor emerged, *Aggressiveness*, which had positive loadings from threat and avoidance variables from S6, and from threat variables from S4. At the fifth level, a *Submissiveness* factor emerged, with positive loadings from submissive greeting in S1 and S6.

Additional factors appeared at lower levels in the factor hierarchy due to split offs of broader factors, especially from *Sociability*, *Non-social fearfulness* and *Playfulness*; at the 28-factor level, 15 factors could be considered as split-offs from these three factors (Fig. 2). Relatively unrelated factors emerged at the forthcoming factor levels, such as *Imposing* (loadings from silent threat in S6), *Anxiety* (loadings from passive avoidance variables), *Gunshot activity* (loadings from activity-related variables in S8) and *Gunshot fear* (loadings from fear-related variables in S8). More specific factors emerged further down in the hierarchy, for example, *Handler contact* and *Food interest* with loadings from variables recorded in subtest 3; *Disinterest* (loadings from disinterest variables from S4, S5 and S6); *Noise exploration* (loadings from exploration variables in S5); and *Approaching person exploration* (loadings from exploration variables from S6) were evident at the 28-factor level.

Fig. 2. A graphic presentation of the factor hierarchy from the top-down factor analysis procedure. Thin lines represent $r_s > 0.6$ and thick lines $r_s > 0.8$; dotted lines indicate negative correlations. The width of each box corresponds to the amount of variance accounted for by that factor.

3.2. Internal construct validity

Based on the correlation magnitudes, 27 factors reached high ICV and an additional eight factors reached moderate ICV against at least one expected subjective rating scales (Table 5). All factors with high ICV were related to interaction with person (S1 or S6), non-social fear (S4, S5 or S7), play (S2) or aggression (S4 or S6). The factors *Imposing*, *Submissiveness*, *Gunshot activity*, *Noise exploration*, *Handler play* and *Handler contact* did not reach the moderate ICV criterion of $r_s \geq 0.40$ with any of the subjective rating scales.

3.3. External construct validity

Sufficient average inter-item correlations (close to or above 0.3) were found for all of the questionnaire subscales, and the alpha values where all > 0.7 (Table 4). The criterion for moderate ECV was reached by 30 factors (Table 6). High validity was found for 13 of these factors, which all were related to interaction with person or play, the *Boldness* factor included. *Submissiveness*, *Imposing*, *Surprise & noise anxiety*, six of the factors related to *Non-social fearfulness* and two exploration-related factors had no expected correlations of $r_s \geq 0.15$ in absolute terms.

3.4. External construct validity at different factor levels

A correlation with any of the factors higher than 0.15 or lower than -0.15 was found for 14 of the questionnaire subscales (Table 8). Four reached the highest correlation at any of the first three factor levels ('stranger-directed aggression', 'dog-directed interest', 'exploration tendency' and 'energy level'), and for six additionally subscales a correlation magnitude within 0.01 from the maximum correlation was found, with a factor on any of the three first factor levels ('fear recovery latency', 'stranger-directed fear', 'handling fear', 'stranger-directed interest', 'human-directed play interest', and 'dog-directed fear or aggression'). In six of these cases, the highest correlations were with the *Sociability* factor. For the remaining four subscales, the highest correlations were attained with factors from level 16 or below in the factor hierarchy. For the subscales 'non-social fear' and 'noise-related fear' this was due to the extraction of *Gunshot fear* (first extracted at the 12th factor level). For the subscale 'trainability', a considerable higher correlation was reached at the 21st factor level due to the appearance of *Familiar object play* factor. Validity for the subscale 'appetite' was reached at the 17th level with the extraction of *Food interest*.

3.5. Interrater reliability

The mean of AC1 for the 179 behavioural rating variables used in the hierarchical EFA procedure was 0.80 (SD = 0.17; range = 0.37–1.00), whereas the mean of AC2 was 0.92 (SD = 0.09; range = 0.49–1.00). The mean of AC1 for the 18 subjective rating variables used in the ICV analysis was 0.49 (SD = 0.17; range = 0.32–0.91), and the mean of AC2 was 0.86 (SD = 0.06; range = 0.76–0.98). Both agreement coefficients were higher for the behavioural ratings compared with the subjective ratings (Mann-Whitney U Test; AC1: $Z = 5.49$; $p < 0.001$; AC2: $Z = 3.55$; $p < 0.001$). Using loadings > 0.45, the mean agreement coefficients were calculated for the 41 extracted factors (AC1: mean = 0.79 SD = 0.10; range = 0.53–0.97; AC2: mean = 0.91; SD = 0.06; range = 0.78–0.99). See supplementary material for detailed results.

4. Discussion

The EFA procedure resulted in a hierarchy of factors from one to 28 factors, in total 41 factors, which could predict 14 of 23 categories of everyday behaviour. The correlations could be found with factors from all levels in the hierarchy. The broader factors at the first levels seem to

Table 8
The maximum correlations (Spearman rank order correlation coefficients) in absolute terms for each factor level for 14 of the questionnaire subscales used in the ECV analysis that reached $r_s > 0.15$ or $r_s < -0.15$. The highest correlation for each subscale is in bold. The number after the correlation coefficient indicates the factor at the actual level with which the subscale reached the highest correlation, and + or - within brackets indicate the sign of the correlation.

Factor level	Non-social fear		Noise-related fear		Fear recovery latency		Stranger-directed fear		Handling fear		Stranger-directed aggression		Stranger-directed interest		Human-directed play interest		Trainability		Dog-directed fear or aggression		Dog-directed interest		Energy level		Exploration tendency		Appetite	
1	0.14	1 (-)	<0.10		0.19	1 (-)	0.32	1 (-)	0.11	1 (-)	0.24	1 (-)	0.31	1 (+)	0.29	1 (+)	<0.10		0.18	1 (-)	0.18	1 (+)	0.17	1 (+)	**0.23**	1 (+)	<0.10	
2	0.18	2 (+)	0.12	2 (+)	0.16	2 (+)	0.33	1 (-)	0.14	1 (-)	0.28	1 (-)	0.37	1 (+)	0.39	1 (+)	0.16	1 (+)	0.23	1 (-)	**0.24**	1 (+)	0.20	1 (+)	0.19	1 (+)	0.12	1 (+)
3	0.18	2 (+)	0.13	2 (+)	0.19	1 (-)	0.40	1 (-)	0.19	1 (-)	**0.30**	1 (-)	0.43	1 (+)	0.43	3 (+)	0.23	3 (+)	0.23	1 (-)	0.20	1 (+)	**0.20**	3 (+)	0.18	2 (-)	0.11	3 (+)
4	0.18	2 (+)	0.13	2 (+)	**0.19**	1 (-)	**0.40**	1 (-)	**0.19**	1 (-)	0.28	1 (-)	0.42	1 (+)	0.43	3 (+)	0.25	3 (+)	0.23	1 (-)	0.23	1 (+)	0.19	3 (+)	0.17	2 (-)	0.12	3 (+)
5	0.18	2 (+)	0.13	2 (+)	0.19	1 (-)	0.40	1 (-)	0.19	1 (-)	0.28	1 (-)	**0.43**	1 (+)	0.43	3 (+)	0.25	3 (+)	0.23	1 (-)	0.22	1 (+)	0.19	3 (+)	0.18	2 (-)	0.12	3 (+)
6	0.18	2 (+)	0.14	2 (+)	0.19	1 (-)	0.40	1 (-)	0.19	1 (-)	0.28	1 (-)	0.43	1 (+)	0.43	3 (+)	0.25	3 (+)	0.23	1 (-)	0.22	1 (+)	0.19	3 (+)	0.19	2 (-)	0.12	3 (+)
7	0.17	2 (+)	0.14	2 (+)	0.18	1 (-)	0.39	1 (-)	0.18	1 (-)	0.27	1 (-)	0.43	1 (+)	0.44	3 (+)	0.24	3 (+)	0.23	1 (-)	0.22	1 (+)	0.19	3 (+)	0.18	2 (-)	0.11	3 (+)
8	0.17	2 (+)	0.13	2 (+)	0.18	1 (-)	0.39	1 (-)	0.19	1 (-)	0.28	1 (-)	0.43	1 (+)	0.44	3 (+)	0.25	3 (+)	0.23	1 (-)	0.22	1 (+)	0.19	3 (+)	0.17	8 (-)	0.11	3 (+)
9	0.17	2 (+)	0.13	2 (+)	0.16	1 (-)	0.38	1 (-)	0.18	1 (-)	0.27	1 (-)	0.42	1 (+)	0.44	3 (+)	0.25	3 (+)	0.23	1 (-)	0.22	1 (+)	0.19	3 (+)	0.18	8 (-)	0.11	3 (+)
10	0.16	2 (+)	0.16	8 (+)	0.17	10 (+)	0.37	1 (-)	0.17	1 (-)	0.26	1 (-)	0.42	1 (+)	0.44	3 (+)	0.24	3 (+)	0.23	1 (-)	0.21	1 (+)	0.19	3 (+)	0.19	8 (-)	0.11	3 (+)
11	0.16	2 (+)	0.15	8 (+)	0.19	7 (+)	0.37	1 (-)	0.17	1 (-)	0.26	1 (-)	0.42	1 (+)	0.44	3 (+)	0.25	3 (+)	0.23	1 (-)	0.21	1 (+)	0.19	3 (+)	0.19	8 (-)	0.11	3 (+)
12	0.23	12 (+)	0.25	12 (+)	0.17	1 (-)	0.38	1 (-)	0.17	1 (-)	0.27	1 (-)	0.42	1 (+)	0.44	3 (+)	0.24	3 (+)	0.23	1 (-)	0.21	1 (+)	0.19	3 (+)	0.17	2 (-)	0.11	3 (+)
13	0.23	12 (+)	0.25	12 (+)	0.19	13 (+)	0.37	1 (-)	0.17	1 (-)	0.26	1 (-)	0.41	1 (+)	0.44	3 (+)	0.24	3 (+)	0.23	1 (-)	0.20	1 (+)	0.19	3 (+)	0.20	8 (-)	0.11	3 (+)
14	0.24	12 (+)	0.26	12 (+)	0.18	12 (+)	0.37	1 (-)	0.16	1 (-)	0.26	1 (-)	0.42	1 (+)	0.44	3 (+)	0.24	3 (+)	**0.24**	1 (-)	0.21	1 (+)	0.19	3 (+)	0.19	13 (-)	0.11	3 (+)
15	0.24	12 (+)	0.26	12 (+)	0.18	12 (+)	0.36	7 (-)	0.17	7 (-)	0.26	7 (-)	0.41	7 (+)	0.44	3 (+)	0.24	14 (+)	0.23	7 (-)	0.21	7 (+)	0.19	3 (+)	0.17	2 (-)	0.11	3 (+)
16	0.24	12 (+)	**0.27**	12 (+)	0.19	12 (+)	0.36	1 (-)	0.17	1 (-)	0.26	1 (-)	0.40	1 (+)	0.44	3 (+)	0.25	14 (+)	0.23	1 (-)	0.21	1 (+)	0.19	3 (+)	0.17	13 (-)	0.11	3 (+)
17	**0.24**	12 (+)	0.26	12 (+)	0.18	12 (+)	0.36	1 (-)	0.17	1 (-)	0.26	1 (-)	0.41	1 (+)	0.44	3 (+)	0.24	3 (+)	0.23	1 (-)	0.21	1 (+)	0.20	3 (+)	0.18	12 (-)	0.15	16 (+)
18	0.24	12 (+)	0.26	12 (+)	0.19	12 (+)	0.36	1 (-)	0.17	1 (-)	0.26	1 (-)	0.40	1 (+)	**0.44**	3 (+)	0.24	3 (+)	0.23	1 (-)	0.21	1 (+)	0.20	3 (+)	0.18	12 (-)	0.16	16 (+)
19	0.24	12 (+)	0.26	12 (+)	0.18	12 (+)	0.36	1 (-)	0.17	1 (-)	0.26	1 (-)	0.40	1 (+)	0.44	3 (+)	0.25	14 (+)	0.23	1 (-)	0.21	1 (+)	0.20	3 (+)	0.17	12 (-)	0.16	16 (+)
20	0.24	12 (+)	0.25	12 (+)	0.19	12 (+)	0.36	1 (-)	0.17	1 (-)	0.26	15 (-)	0.40	1 (+)	0.44	3 (+)	0.25	14 (+)	0.23	1 (-)	0.21	1 (+)	0.20	3 (+)	0.17	12 (-)	0.16	16 (+)
21	0.24	12 (+)	0.25	12 (+)	0.18	12 (+)	0.37	15 (-)	0.17	1 (-)	0.27	15 (-)	0.40	1 (+)	0.43	3 (+)	0.31	14 (+)	0.23	1 (-)	0.21	1 (+)	0.20	3 (+)	0.17	12 (-)	**0.16**	16 (+)
22	0.24	12 (+)	0.26	12 (+)	0.18	12 (+)	0.37	15 (-)	0.17	1 (-)	0.27	15 (-)	0.40	1 (+)	0.42	3 (+)	0.31	14 (+)	0.23	15 (-)	0.20	1 (+)	0.20	3 (+)	0.18	2 (-)	0.16	16 (+)
23	0.23	12 (+)	0.25	12 (+)	0.18	15 (-)	0.37	15 (-)	0.16	1 (-)	0.27	15 (-)	0.40	15 (+)	0.42	3 (+)	**0.31**	14 (+)	0.23	15 (-)	0.20	19 (+)	0.20	3 (+)	0.18	2 (-)	0.16	16 (+)
24	0.23	12 (+)	0.25	12 (+)	0.18	15 (-)	0.37	15 (-)	0.16	1 (-)	0.27	15 (-)	0.40	15 (+)	0.42	3 (+)	0.30	21 (+)	0.23	15 (-)	0.21	19 (+)	0.20	3 (+)	0.18	2 (-)	0.16	16 (+)
28	0.23	12 (+)	0.26	12 (+)	0.18	15 (-)	0.37	15 (-)	0.17	1 (-)	0.27	15 (-)	0.40	15 (+)	0.43	21 (+)	0.30	21 (+)	0.23	15 (-)	0.23	27 (+)	0.20	27 (+)	0.17	27 (+)	0.16	16 (+)

predict several categories of everyday behaviour well, such as attitude towards unknown persons. For some behaviour, however, more specific factors at lower levels in the factor hierarchy appear better predictors. For example, the correspondence to noise-related and non-social fear in everyday life increased substantially with a factor that appeared from the 12th-factor level. The correspondence to behaviour in the everyday context for the validated factors suggests consistency across situations, which is one main criterion for personality traits. Thus, the validated factors can be seen as personality traits in companion dogs.

As suggested by Rayment et al. (2016), a hierarchical perspective may be advantageous when investigating the structure of dog personality. Constraining the number of factors to retain by using a statistical criterion, such as Horn's parallel analysis (Horn, 1965), which only considers the correlation pattern within the actual data set, may give insufficient results. Relevant factors may be missed, and less important factors may be given too much attention. Using criteria outside the assessment, such as the relationship to questionnaire data in this study, may limit this risk and yield a set of valid factors. In general, the broader factors at the first factor levels seem to be good predictors for several aspects of everyday behaviour. However, *Submissiveness*, which appeared already at the 5th-factor level, was not validated in either the ICV or ECV analysis and seems to be a poor everyday behaviour predictor. In contrast, the prediction of non-social and noise-related fear increased substantially with the extraction of the factor *Gunshot fear*, which appeared from the 12th level. For two additional questionnaire subscales, 'appetite' and 'trainability', the correlations were substantially increased by factors that appeared at lower levels in the hierarchy, from the 17th and 21st factor levels, respectively. Thus, the method used in this study makes it possible to sort out factors of different specificity that are relevant for their application.

Of the 41 extracted factors, 27 reached the criterion of high ICV and an additional eight factors had moderate ICV. This result suggests that most of the extracted factors, especially factors related to human interactions, non-social fear, play, or aggression, reflect what they appear to reflect in the assessment. When it comes to ECV, the indicator of how well the factor predict similar behaviour outside the assessment, 30 factors reached the criterion of moderate validity, and 13 that of high validity. The factors with high ECV were all related to human interactions or play, which indicates that attitude towards unfamiliar persons, such as fear of, aggression towards and positive interest in, and interest to play person-oriented object games, are well-predicted by measures extracted from the assessment. High correlations were also found for the subscale 'trainability', which was best predicted by the dog's interest in object play, especially with the familiar object, and the dog's contact with the handler in the play and food subtests. Compared with the DMA, the BPH seems to be at least as good in predicting these aspects of everyday behaviour; the largest correlations for aggressiveness towards, fear of, and positive interest in unfamiliar persons and dogs, as well as for trainability and object play, were all higher in this study compared with what was found in Svartberg (2005). Especially the correlations with the dog-related subscales are notable. Factors that correspond well to attitude towards stranger outside the assessment, such as *Sociability* and *Greeting*, also correlate with subscales related to fear of, aggression towards, and interest in unfamiliar dogs. This suggests that the dog's social attitude towards unfamiliar persons and unfamiliar dogs have a common denominator, which is assessed in the *Sociability*-related factors in BPH. The relationship between social attitude towards persons and dogs has been indicated previously (Jones, 2008; Rayment et al., 2016), but only weakly in the DMA (Svartberg, 2005). It is possible that the assessment method used in the BPH, with many specific ratings, capture nuances in the dog's social behaviour better than the DMA's more summarizing scales.

Two factors with both low ICV and ECV were *Imposing* and *Submissiveness*. These factors are related to the well-studied concept of social dominance (Drews, 1993). The definitions of silent threat (e.g., stiff and up-right body posture, erect and stiff tail, direct stare and low intensity growling), which loads on *Imposing*, and submissive greeting (e.g., ears held back or flattened, head held low, body lowered, tail low or tucked, often slowly waving, and licking or licking intentions), loading on *Submissiveness*, are highly similar to the behavioural expressions in dogs used to define dominance and submission, respectively (Trisko and Smuts, 2015; Bonanni et al., 2017). Besides being a descriptor of a social organization within a group, dominance has also been proposed as a personality trait in animals in general (Baenninger, 1981; Gosling and John, 1999) as well as in dogs (Jones and Gosling, 2005). However, the concept of dominance in dogs, especially when it comes to dominance as a personality trait, has been questioned (e.g., Bradshaw et al., 2009). The low validity may be due to difficulties in perceiving behavioural expressions related to dominance and submissiveness for both dog owners and raters, which will lead to inaccurate assessments. However, the results from the interrater study suggest good rater agreement for *Submissiveness* and an agreement for *Imposing* that was only slightly lower than the average. Low reliability of the two questionnaire subscales 'tendency to impose' and 'submissive tendency' may have affected the result negatively; they are both one-item subscales, and the item related to 'tendency to impose' refers to behaviour in an encounter with a dog, not a person, as in the BPH. Another cause for the poor validity estimates may be of a more general nature. Besides the low correlations found with the subscales 'tendency to impose' and 'submissive tendency', similar results were found for 'excitability', 'attachment or attention seeking behaviour', 'separation-related behaviour', 'contact seeking tendency', 'dog rivalry', and 'owner-directed aggression'. These subscales are all related to interactions with family members in the home context, and it is possible that these behaviours are specific to the well-known home situation. The lack of consistency between the assessment situation and the home situation may call into doubt these behaviours' classification as personality traits. If they are aspects of dog personality, an experimental set-up like the BPH, which exposes the dog for novel and unexpected stimuli, does not seem to be appropriate to assess them.

The surface-related factors – *Surface fear*, *Surface acceleration* and *Surface hesitation* – had high ICV, but did not reach the ECV criterion of $r_s > 0.15$ with the subscale 'stairs/surface fear' that was intended specifically to validate these three factors. The subscale had sufficient internal consistency, and the surface-related factors had high rater agreement according to the interrater reliability study. A previous study on floor fearfulness indicates that this type of fear, compared with gunshot fearfulness, may be more difficult to identify in an experimental set-up and reliably assess by dog owners (Hydbring-Sandberg et al., 2004). It is possible that surface fearfulness may be highly specific to the surface type, which implies difficulties in designing standardised experimental assessments for this type of fearfulness.

The many expected correlations imply adequate convergent validity for the majority of the extracted factors. However, there are also a number of unexpected correlations, which indicate poor discriminant validity. For example, factors related to greeting in the assessment, which are based solely on measures of the dog's positive interest in the figurant, are negatively correlated with both fear of and aggression towards unfamiliar persons in the questionnaire. A similar negative correlation pattern is found between fear-related factors in the assessment and exploration in the home situation. Unexpected positive correlations were found between *Playfulness* and positive interest in both unfamiliar persons and dogs as well as to exploration tendency, and between *Non-social fearfulness* and stranger-directed fear in everyday life. Discriminant validity is desirable from an assessment development point-of-view. However, these results should be regarded as relevant since they seem to be caused by broader personality tendencies expressed both in and outside the assessment. One such is reflected by the first factor *Boldness*. This factor is in the assessment situation related to a high degree of greeting and a low degree of fear-related social behaviour, high interest to play object games as well as low fear and high exploration in the non-social subtests. The results from the ICV analysis are in line with

this, with substantial correlations to the subjective rating scales confident, curious, energetic, playful, people friendly, and positive. According to the correlations with questionnaire subscales, *Boldness* is related to similar tendencies outside the assessment situation. In most aspects, this factor seems equivalent to the same-labelled higher-order dimension found in analyses of DMA data (Svartberg and Forkman, 2002; Svartberg, 2005). This dimension could reflect fearfulness, a personality trait suggested in dogs (Jones and Gosling, 2005) and in animals in general (Boissy, 1995). However, absence of fear is not necessarily related to the aspects of high *Boldness*. It is unlikely that a high tendency to play, explore, be active, and greet strangers can be explained solely by low fear. Thus, the *Boldness* dimension seems to reflect a wider array of behavioural aspects than just general fearfulness. From a comparative perspective, *Boldness* seems to be related to the General Factor of Personality (GFP) in humans, which is correlated with Extraversion, Neuroticism (inverted), Agreeableness, Openness and Conscientiousness (Rushton and Irwing, 2008). Equivalents of these domains have been identified in dogs (Gosling et al., 2003; Ley et al., 2009). A bold dog could be described as outgoing, active, friendly, fearless, curious, and cooperative, which is in line with high-GFP personality. Thus, it is possible that the DMA and the BPH capture a dog variant of the human GFP. There are indications of a similar dimension in wolves (Fox, 1972; MacDonald, 1987) as well as other animal species (Wilson et al., 1994; Kaiser and Müller, 2021). Since boldness in animals and the GFP in humans are genetically based, and related to life-history strategy (Bell, 2007; Rushton et al., 2008), it is possible that they are analogous due to shared evolutionary roots.

Besides the contribution to our understanding of dog personality, the results of this study may have several applications. The validated factors composed by behavioural ratings, suggested to reflect personality traits in dogs, are in this regard of special interest. Since they have higher interrater reliability compared with the subjective ratings, these measures are probably the most reliable and predictable of measures from the BPH. The most promising application, given that the personality traits have a genetic base, is to breeding. Decisions in companion-dog breeding are mostly based on dog appearance, whereas behavioural traits are often overlooked (King et al., 2012). Some of the everyday behaviours that the factors correspond to, such as friendliness and trainability, are preferred traits in companion dogs and associated with positive welfare, whereas others are considered undesirable or even problematic, such as anxiety, fear, and aggression, and related to poor well-being (King et al., 2009; Mellor and Beausoleil, 2015; Cannas et al., 2018). Thus, the BPH may be an objective breeding tool, where suggested personality traits can be used for selection of well-adjusted companion dogs. The diversity of traits may make it possible to design breeding goals that are specifically appropriate for the breed in focus. This calls for future research on the genetic influence of the suggested dog personality traits, to ensure that effective selection is possible. Besides breeding, the personality traits may be used as a guide for prospective companion-dog owners when choosing breeds or genetic lines within a breed. Data from the BPH are already accessible via open-web resources provided by the SKC, but with the suggested personality traits from this study the possibility of utilising the information grows. Even though the dog's behaviour is likely known to the owner of the dog prior to the assessment, the unbiased rater's assessment may be an eye-opener for owners, and result in adequate training and preventive actions that can minimize potential problem-causing behaviour and behaviour associated with poor welfare. Even though the BPH is a non-scientific assessment, the validated personality traits suggested by the present results may be useful in several disciplines of research. One field of special interest is in genetic research, where the BPH seems to be a promising tool for dog-personality phenotyping and used in the search for links between genes and behaviour.

There are some possible limitations of this study. One relates to the representativeness of the sample. Even though all FCI breed groups and 267 breeds were represented in the data set, there might be categories of dogs that differ from the general pattern found in this study. In any future studies or applications, it is of importance to ensure the factors' internal consistency and validity for each sample. A second possible limitation is the large number of assessment stations, raters, and ALs involved in the BPH-system, which entails a risk of undesirable variation. This may have influenced the results of this study, and may endanger future usefulness of BPH data. Besides the need for continuous educational efforts for raters and ALs to ensure consensus, the organizers of the BPH must be encouraged to follow the established procedure in order to reach standardization. A third possible limitation is that 47 behavioural rating scales were not analysed due to a low frequency of scores other than zero (< 1 %). They may be aspects of dog personality and of great importance in some applications. For example, in predictions of potential problem-causing behaviour, rare behaviours such as biting and threat in the first subtest should be regarded in addition to the more common behaviours assessed by the factor scores. A fourth possible limitation regards the treatment of missing data, which were median imputed. The majority of the missing values were due to interrupted phases or sequences in S1, S7, and S8, which most likely were caused by the dog's aversive reaction. The imputation of medians may have influenced the correlations in the ICV and ECV analyses. However, since the frequency of missing data was low (0.6 %), this may not have had major impact on the results. A fifth possible limitation concerns the subscales that were extracted from only one or two items, which may result in low content validity and low capability to discriminate between responses. This, in turn, may have had impact on the results from the ECV analyses, where these subscales were used as criteria.

5. Conclusions

The results in this study indicate that the BPH can reveal relevant aspects of the dog's personality, tendencies both general and more specific, related to a range of behaviour outside the assessment. These traits may give indications of the dog's well-being and ability to cope with challenges in the home environment, especially challenges of novel and unexpected character. The most promising application is to breeding; given a genetic base for the personality traits, the results may be useful in behavioural-based breeding in companion dogs. Since there are relevant traits of different specificity and content, the BPH may be suitable for different dog types and breeds with different behavioural expressions. The traits from the BPH may also be highly useful for both owners of specific dogs, who may gain insights about theirs dog, and prospective dog owners, who can use traits as guidelines in making choices. The BPH and the suggested personality traits may also be useful in research, especially in genetic research, where the BPH seems a promising tool for dog-personality phenotyping.

Declaration of Competing Interest

The author was one of the developers of the BPH and acted during the development process as a consultant for the SKC.

Acknowledgements

I am very grateful to Dr. James Serpell for letting me use C-BARQ questions for this study; to the Swedish Kennel Club for providing me with the BPH data set and for the distribution of information regarding the questionnaire to dog-owners; and to all persons involved in the organization of the BPH. A special thanks to Björn Forkman for advice during the study and an anonymous reviewer who provided valuable comments that helped improve and clarify this manuscript. Finally, I am very grateful to all dog owners who let their dogs carry out the BPH and describe their everyday behaviour by answering the questionnaire. This work was supported by the Agria SKK Research Fund (grant number N2017-0006).

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.applanim.2021.105302.

References

Araya-Ajoy, Y.G., Dingemanse, N.J., 2014. Characterizing behavioural 'characters': an evolutionary framework. Proc. Royal Soc. B. 281, 20132645 https://doi.org/10.1098/rspb.2013.2645.

Arvelius, P., Grandinson, K., 2012. Creating a BLUP-index to select against fearfulness in Rough Collie. J. Vet. Behav. 7, 57–58. https://doi.org/10.1016/j.jveb.2011.12.009.

Arvelius, P., Blixt, C., Svartberg, K., Trenkle-Nyberg, S., 2012. A new behavior and personality test to be used as a tool in dog breeding. J. Vet. Behav. 1, 57. https://doi.org/10.1016/j.jveb.2011.12.008.

Arvelius, P., Asp, H.E., Fikse, W.F., Strandberg, E., Nilsson, K., 2014. Genetic analysis of a temperament test as a tool to select against everyday life fearfulness in Rough Collie. J. Anim. Sci. 92, 4843–4855. https://doi.org/10.2527/jas.2014-8169.

Asp, H.E., Fikse, W.F., Nilsson, K., Strandberg, E., 2015. Breed differences in everyday behaviour of dogs. Appl. Anim. Behav. Sci. 169, 69–77. https://doi.org/10.1016/j.applanim.2015.04.010.

Baenninger, R., 1981. Dominance: on distinguishing the baby from the bathwater. Behav. Brain Sci. 4, 431–432. https://doi.org/10.1017/S0140525X0000964X.

Bell, A.M., 2007. Evolutionary biology: animal personalities. Nature 447, 539. https://doi.org/10.1038/447539a.

Bennett, S.L., Litster, A., Weng, H.Y., Walker, S.L., Luescher, A.U., 2012. Investigating behavior assessment instruments to predict aggression in dogs. Appl. Anim. Behav. Sci. 141, 139–148. https://doi.org/10.1016/j.applanim.2012.08.005.

Boissy, A., 1995. Fear and fearfulness in animals. Q. Rev. Biol. 70, 165–191. https://doi.org/10.1086/418981.

Bonanni, R., Cafazzo, S., Abis, A., Barillari, E., Valsecchi, P., Natoli, E., 2017. Age-graded dominance hierarchies and social tolerance in packs of free-ranging dogs. Behav. Ecol. 28, 1004–1020. https://doi.org/10.1093/beheco/arx059.

Bosco, F.A., Aguinis, H., Singh, K., Field, J.G., Pierce, C.A., 2015. Correlational effect size benchmarks. J. Appl. Psychol. 100, 431. https://doi.org/10.1037/a0038047.

Bradshaw, J.W., Blackwell, E.J., Casey, R.A., 2009. Dominance in domestic dogs—useful construct or bad habit? J. Vet. Behav. 4, 135–144. https://doi.org/10.1016/j.jveb.2008.08.004.

Cannas, S., Talamonti, Z., Mazzola, S., Minero, M., Picciolini, A., Palestrini, C., 2018. Factors associated with dog behavioral problems referred to a behavior clinic. J. Vet. Behav. 24, 42–47. https://doi.org/10.1016/j.jveb.2017.12.004.

Capitanio, J.P., 1999. Personality dimensions in adult male rhesus macaques: prediction of behaviors across time and situation. Am. J. Primatol. 47, 299–320. https://doi.org/10.1002/(sici)1098-2345(1999)47:4%3C299::aid-ajp3%3E3.0.co;2-p.

Drews, C., 1993. The concept and definition of dominance in animal behaviour. Behaviour 125, 283–313. https://doi.org/10.1163/156853993x00290.

Duffy, D.L., Serpell, J.A., 2012. Predictive validity of a method for evaluating temperament in young guide and service dogs. Appl. Anim. Behav. Sci. 138, 99–109. https://doi.org/10.1016/j.applanim.2012.02.011.

Fox, M.W., 1972. Socio-ecological implications of individual differences in wolf litters: a developmental and evolutionary perspective. Behaviour 41, 298–313. https://doi.org/10.1163/156853972x00077.

Foyer, P., Bjällerhag, N., Wilsson, E., Jensen, P., 2014. Behaviour and experiences of dogs during the first year of life predict the outcome in a later temperament test. Appl. Anim. Behav. Sci. 155, 93–100. https://doi.org/10.1016/j.applanim.2014.03.006.

Fratkin, J.L., 2017. Personality in dogs. In: Vonk, J., Weiss, A., Kuczaj, S.A. (Eds.), Personality in Nonhuman Animals. Springer International Publishing, Cham, pp. 205–224. https://doi.org/10.1007/978-3-319-59300-5_10.

Goddard, M.E., Beilharz, R.G., 1984. A factor analysis of fearfulness in potential guide dogs. Appl. Anim. Behav. Sci. 12, 253–265. https://doi.org/10.1016/0168-1591(84)90118-7.

Goldberg, L.R., 2006. Doing it all bass-ackwards: the development of hierarchical factor structures from the top down. J. Res. Pers. 40, 347–358. https://doi.org/10.1016/j.jrp.2006.01.001.

Gosling, S.D., John, O.P., 1999. Personality dimensions in nonhuman animals: a cross-species review. Curr. Dir. Psychol. Sci. 8, 69–75. https://doi.org/10.1111/1467-8721.00017.

Gosling, S.D., Kwan, V.S., John, O.P., 2003. A dog's got personality: a cross-species comparative approach to personality judgments in dogs and humans. J. Pers. Soc. Psychol. 85, 1161–1169. https://doi.org/10.1037/0022-3514.85.6.1161.

Gwet, K.L., 2014. Handbook of Inter-Rater Reliability, 4th edition. Advanced Analytics, LLC, Gaithersburg.

Hair, J.F., Anderson, R.E., Tatham, R.L., Black, W.C., 1998. Multivariate Data Analysis. Prentice Hall, Upper Saddle River.

Horn, J.L., 1965. A rationale and test for the number of factors in factor analysis. Psychometrika 30, 179–185. https://doi.org/10.1007/bf02289447.

Hsu, Y., Serpell, J.A., 2003. Development and validation of a questionnaire for measuring behavior and temperament traits in pet dogs. J. Am. Vet. Med. Assoc. 223, 1293–1300. https://doi.org/10.2460/javma.2003.223.1293.

Hydbring-Sandberg, E., von Walter, L.W., Hoglund, K., Svartberg, K., Swenson, L., Forkman, B., 2004. Physiological reactions to fear provocation in dogs. J. Endocrinol. 180, 439–448. https://doi.org/10.1677/joe.0.1800439.

Jones, A.C., 2008. Development and Validation of a Dog Personality Questionnaire. PhD Dissertation. The University of Texas, Austin.

Jones, A.C., Gosling, S.D., 2005. Temperament and personality in dogs (*Canis familiaris*): a review and evaluation of past research. Appl. Anim. Behav. Sci. 95, 1–53. https://doi.org/10.1016/j.applanim.2005.04.008.

Kaiser, M.I., Müller, C., 2021. What is an animal personality? Biol. Philos. 36, 1–25. https://doi.org/10.1007/s10539-020-09776-w.

King, T., Marston, L.C., Bennett, P.C., 2009. Describing the ideal Australian companion dog. Appl. Anim. Behav. Sci. 120, 84–93. https://doi.org/10.1016/j.applanim.2009.04.011.

King, T., Marston, L.C., Bennett, P.C., 2012. Breeding dogs for beauty and behaviour: why scientists need to do more to develop valid and reliable behaviour assessments for dogs kept as companions. Appl. Anim. Behav. Sci. 137, 1–12. https://doi.org/10.1016/j.applanim.2011.11.016.

Kolm, N., Temrin, H., Miklósi, Á., Kubinyi, E., Garamszegi, L.Z., 2020. The link between selection for function and human-directed play behaviour in dogs. Biol. Lett. 16 (9), 20200366 https://doi.org/10.1098/rsbl.2020.0366.

Konečná, M., Lhota, S., Weiss, A., Urbánek, T., Adamová, T., Pluháček, J., 2008. Personality in free-ranging Hanuman langur (*Semnopithecus entellus*) males: subjective ratings and recorded behavior. J. Comp. Psychol. 122, 379–389. https://doi.org/10.1037/a0012625.

Ley, J.M., Bennett, P.C., Coleman, G.J., 2009. A refinement and validation of the Monash Canine Personality Questionnaire (MCPQ). Appl. Anim. Behav. Sci. 116, 220–227. https://doi.org/10.1016/j.applanim.2008.09.009.

Lush, J., Ijichi, C., 2018. A preliminary investigation into personality and pain in dogs. J. Vet. Behav. 24, 62–68. https://doi.org/10.1016/j.jveb.2018.01.005.

MacDonald, K.B., 1987. Development and stability of personality characteristics in pre-pubertal wolves: implications for pack organization and behavior. In: Frank, H. (Ed.), Man and Wolf. Dr W. Junk Publ., Dordrecht, pp. 293–312.

Marshall-Pescini, S., Virányi, Z., Range, F., 2015. The effect of domestication on inhibitory control: wolves and dogs compared. PLoS One 10, e0118469. https://doi.org/10.1371/journal.pone.0118469.

McGarrity, M.E., Sinn, D.L., Gosling, S.D., 2015. Which personality dimensions do puppy tests measure? A systematic procedure for categorizing behavioral assays. Behav. Process. 110, 117–124. https://doi.org/10.1016/j.beproc.2014.09.029.

Mellor, D.J., Beausoleil, N.J., 2015. Extending the 'Five Domains' model for animal welfare assessment to incorporate positive welfare states. Anim. Welfare 24, 241–253. https://doi.org/10.7120/09627286.24.3.241.

Mirkó, E., Dóka, A., Miklósi, Á., 2013. Association between subjective rating and behaviour coding and the role of experience in making video assessments on the personality of the domestic dog (*Canis familiaris*). Appl. Anim. Behav. Sci. 149, 45–54. https://doi.org/10.1016/j.applanim.2013.10.003.

Murphree, O.D., Peters, J.E., Dykman, R.A., 1969. Behavioral comparisons of nervous, stable, and crossbred pointers at ages 2, 3, 6, 9, and 12 months. Cond. Reflex 4, 20–23. https://doi.org/10.1007/BF03000074.

Olsson, U., 1979. On the robustness of factor analysis against crude classification of the observations. Multivar. Behav. Res. 14, 485–500. https://doi.org/10.1207/s15327906mbr1404_7.

Pavlov, I.P., 1927. In: Anrep, G.V. (Ed.), Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex. Oxford University Press, Oxford.

Protopopova, A., 2016. Effects of sheltering on physiology, immune function, behavior, and the welfare of dogs. Physiol. Behav. 159, 95–103. https://doi.org/10.1016/j.physbeh.2016.03.020.

R Development Core Team, 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. R Foundation for Statistical Computing, Vienna.

Rayment, D.J., Peters, R.A., Marston, L.C., De Groef, B., 2016. Investigating canine personality structure using owner questionnaires measuring pet dog behaviour and personality. Appl. Anim. Behav. Sci. 180, 100–106. https://doi.org/10.1016/j.applanim.2016.04.002.

Réale, D., Reader, S.M., Sol, D., McDougall, P.T., Dingemanse, N.J., 2007. Integrating animal temperament within ecology and evolution. Biol. Rev. 82, 291–318. https://doi.org/10.1111/j.1469-185X.2007.00010.x.

Réale, D., Dingemanse, N.J., Kazem, A.J., Wright, J., 2010. Evolutionary and ecological approaches to the study of personality. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 365, 3937–3946. https://doi.org/10.1098/rstb.2010.0222.

Revelle, W., 2018. psych: Procedures for Personality and Psychological Research. Northwestern University, Evanston, Illinois, USA. https://CRAN.R-project.org/package=psychVersion=1.8.4.

Rushton, J.P., Irwing, P., 2008. A General Factor of Personality (GFP) from two meta-analyses of the big five. Pers. Individ. Differ. 45, 679–683. https://doi.org/10.1016/j.paid.2008.07.015.

Rushton, J.P., Bons, T.A., Hur, Y.M., 2008. The genetics and evolution of the general factor of personality. J. Res. Pers. 42, 1173–1185. https://doi.org/10.1016/j.jrp.2008.03.002.

Saetre, P., Strandberg, E., Sundgren, P.E., Pettersson, U., Jazin, E., Bergström, T.F., 2006. The genetic contribution to canine personality. Genes Brain Behav. 5, 240–248. https://doi.org/10.1111/j.1601-183x.2005.00155.x.

Sarviaho, R., Hakosalo, O., Tiira, K., Sulkama, S., Salmela, E., Hytönen, M.K., Sillanpää, M.J., Lohi, H., 2019. Two novel genomic regions associated with fearfulness in dogs overlap human neuropsychiatric loci. Transl. Psychiatry 9, 1–11. https://doi.org/10.1038/s41398-018-0361-x.

Scott, J.P., Fuller, J.L., 1965. Dog Behaviour: The Genetic Basis. The University of Chicago Press, Chicago.

Sinn, D.L., Gosling, S.D., Hilliard, S., 2010. Personality and performance in military working dogs: reliability and predictive validity of behavioral tests. Appl. Anim. Behav. Sci. 127, 51–65. https://doi.org/10.1016/j.applanim.2010.08.007.

Smith, B.P., Browne, M., Serpell, J.A., 2017. Owner-reported behavioural characteristics of dingoes (*Canis dingo*) living as companion animals: A comparison to 'modern' and 'ancient' dog breeds. Appl. Anim. Behav. Sci. 187, 77–84. https://doi.org/10.1016/j.applanim.2016.11.010.

StatSoft, Inc, 2013. STATISTICA (Data Analysis Software System), Version 12. www.statsoft.com.

Strandberg, E., Jacobsson, J., Saetre, P., 2005. Direct genetic, maternal and litter effects on behaviour in German shepherd dogs in Sweden. Livest. Prod. Sci. 93, 33–42. https://doi.org/10.1016/j.livprodsci.2004.11.004.

Sundman, A.S., Johnsson, M., Wright, D., Jensen, P., 2016. Similar recent selection criteria associated with different behavioural effects in two dog breeds. Genes Brain Behav. 15, 750–756. https://doi.org/10.1111/gbb.12317.

Svartberg, K., 2002. Shyness–boldness predicts performance in working dogs. Appl. Anim. Behav. Sci. 79, 157–174. https://doi.org/10.1016/s0168-1591(02)00120-x.

Svartberg, K., 2005. A comparison of behaviour in test and in everyday life: evidence of three consistent boldness-related personality traits in dogs. Appl. Anim. Behav. Sci. 91, 103–128. https://doi.org/10.1016/j.applanim.2004.08.030.

Svartberg, K., 2006. Breed-typical behaviour in dogs—historical remnants or recent constructs? Appl. Anim. Behav. Sci. 96, 293–313. https://doi.org/10.1016/j.applanim.2005.06.014.

Svartberg, K., 2007. Individual differences in behaviour – dog personality. In: Jensen, P. (Ed.), The Behavioural Biology of Dogs. CAB International, Wallingsford, pp. 182–206.

Svartberg, 2016. Utvärdering av BPH – Beteende- och Personlighetsbeskrivning Hund – under åren 2012-2015: Beskrivaröverensstämmelser, kopplingar mellan BPH och vardag samt utvärdering av de sammanfattande" spindelvärdena". Report for the Swedish Kennel Club. https://www.skk.se/globalassets/dokument/om-skk/bph/bph-utvardering-2012-2015.pdf.

Svartberg, K., Forkman, B., 2002. Personality traits in the domestic dog (*Canis familiaris*). Appl. Anim. Behav. Sci. 79, 133–155. https://doi.org/10.1016/s0168-1591(02)00121-1.

Svartberg, K., Tapper, I., Temrin, H., Radesäter, T., Thorman, S., 2005. Consistency of personality traits in dogs. Anim. Behav. 69, 283–291. https://doi.org/10.1016/j.anbehav.2004.04.011.

Taylor, K.D., Mills, D.S., 2006. The development and assessment of temperament tests for adult companion dogs. J. Vet. Behav. 1 (3), 94–108. https://doi.org/10.1016/j.jveb.2006.09.002, 1, 94-108.

Trisko, R.K., Smuts, B.B., 2015. Dominance relationships in a group of domestic dogs (*Canis lupus familiaris*). Behaviour 152, 677–704. https://doi.org/10.1163/1568539x-00003249.

Uher, J., 2008. Comparative personality research: methodological approaches. Eur. J. Pers. 22, 427–455. https://doi.org/10.1002/per.680.

van den Berg, L., 2016. Genetics of dog behavior. In: Serpell, J. (Ed.), The Domestic Dog: Its Evolution, Behaviour and Interactions with People, 2nd ed. Cambridge University Press, Cambridge, pp. 69–92. https://doi.org/10.1017/9781139161800.005.

vonHoldt, B.M., Shuldiner, E., Koch, I.J., Kartzinel, R.Y., Hogan, A., Brubaker, L., Wanser, S., Stahler, D., Wynne, C.D.L., Ostrander, E.A., Sinsheimer, J.S., Udell, M.A., 2017. Structural variants in genes associated with human Williams-Beuren syndrome underlie stereotypical hypersociability in domestic dogs. Sci. Adv. 3, e1700398 https://doi.org/10.1126/sciadv.1700398.

Wheat, C.H., Fitzpatrick, J.L., Rogell, B., Temrin, H., 2019. Behavioural correlations of the domestication syndrome are decoupled in modern dog breeds. Nat. Commun. 10, 2422. https://doi.org/10.1038/s41467-019-10426-3.

Whitham, W., Washburn, D.A., 2017. A history of animal personality research. In: Vonk, J., Weiss, A., Kuczaj, S.A. (Eds.), Personality in Nonhuman Animals. Springer International Publishing, Cham, pp. 3–16. https://doi.org/10.1007/978-3-319-59300-5_1.

Wilson, D.S., Clark, A.B., Coleman, K., Dearstyne, T., 1994. Shyness and boldness in humans and other animals. Trends Ecol. Evol. 9, 442–446. https://doi.org/10.1016/0169-5347(94)90134-1.

Wilsson, E., Sinn, D.L., 2012. Are there differences between behavioral measurement methods? A comparison of the predictive validity of two ratings methods in a working dog program. Appl. Anim. Behav. Sci. 141, 158–172. https://doi.org/10.1016/j.applanim.2012.08.012.