Inter-rater and test–retest reliability of the Monash Canine Personality Questionnaire-Revised (MCPQ-R)

Jacqui M. Ley [a,*], Paul McGreevy [b], Pauleen C. Bennett [a]

[a] Animal Welfare Science Centre, Department of Psychology, Monash University, P.O. Box 197, Caulfield East, VIC 3145, Australia
[b] Faculty of Veterinary Science, The University of Sydney, Sydney, NSW 2006, Australia

ARTICLE INFO

ABSTRACT

In developing a valid test for measuring canine personality, it is important to test the reliability of the instrument. The Monash Canine Personality Questionnaire-Revised (MCPQ-R) is a 26-item, adjective-based, owner-administered, questionnaire that measures canine personality along five dimensions: Extraversion, Motivation, Training Focus, Amicability and Neuroticism. Its reliability was tested using inter-rater reliability measures and test–retest measures. Sixty-five couples rated their pet dog individually using the MCPQ-R, with one member of each couple completing the same questionnaire after a 6-month interval. Inter-rater reliability values were similar to or stronger than those found in similar research in dogs, with statistically significant ($P < 0.001$) positive values for all five dimensions ranging from 0.75 (Neuroticism) to 0.86 (Extraversion). Test–retest results were also positive and statistically significant ($P < 0.001$) for all dimensions ranging from 0.79 (Neuroticism) to 0.93 (Motivation). The results of this study suggest the MCPQ-R is reliable for assessing canine personality along the five identified dimensions.

© 2009 Elsevier B.V. All rights reserved.

To help in the matching of owners and pet dogs and handlers and working dogs, the personality of dogs needs to be described accurately and reliably. Personality, in this context, describes the stable and characteristic behavioural tendencies of adult individual dogs (Ley et al., 2007). Although recent research has suggested most companion dogs perform satisfactorily as pets (Bennett and Rohlf, 2007), the high number of animals surrendered to animal shelters annually (Marston et al., 2004) implies that owner-pet matching could be improved. Research investigating relinquishment of companion dogs to shelters (Patronek et al., 1997; Salman et al., 1998; Marston and Bennett, 2003) has shown that the behaviour of the dogs plays a major part in relinquishment to a shelter. Similarly, research investigating training of working dogs (e.g. guiding dogs or police dogs) (Goddard and

Beilharz, 1984; Wilsson and Sundgren, 1997) has shown that the behaviour of the dog plays a role in its success or failure as a working dog. A robust and strongly predictive suite of tests will aid in selection of both companion and working dogs.

The Monash Canine Personality Questionnaire-Revised (MCPQ-R) is a recently developed questionnaire for measuring canine personality along five dimensions: Extraversion, Motivation, Training Focus, Amicability and Neuroticism (Ley et al., 2007). The development process of the MCPQ-R employed a methodology established during the validation of the Big Five taxonomy for describing human personality (John, 1990; McCrae and John, 1992). While alternative tests for assessing canine personality have been reported, they have been limited by their use of only small populations of study dogs (Goddard and Beilharz, 1986; Wilsson and Sundgren, 1997; Murphy, 1998) or their forced adaptation of human models of personality to dogs (Gosling et al., 2003a,b). In developing the MCPQ-R, our goal was to produce a psychometrically

sound instrument for assessing personality in any and all dogs who are individually familiar to a person capable of completing the simple questionnaire.

The validity of the MCPQ-R was recently explored and found to be satisfactory (Ley et al., 2007). However, in addition to validity, the reliability of a measure such as the MCPQ-R affects its utility. Reliability is the consistency with which the tool assesses constructs of interest. Two aspects of reliability are considered important, internal consistency reliability and test reliability. Internal consistency reliability is a measure of how well all the items in a scale measure the concept of interest (Litwin, 1995). It is typically measured using Cronbach's coefficient alpha and is considered adequate if alpha is equal to or greater than 0.7. However, John and Soto (2007, pp. 469–470) suggest combining Cronbach's alpha with mean inter-item correlations for the scale as a more accurate method of assessing internal consistency. Previously, the authors reported alpha levels of between 0.74 and 0.87 and inter-item correlations of between 0.37 and 0.53, showing MCPQ-R to have reasonable internal reliability (Ley et al., 2007).

Test reliability describes the consistency of test results from multiple measurements over time (Hair et al., 1998, pp. 117–119). Perhaps the easiest method of testing the reliability of a scale is to use test–retest correlations, whereby a questionnaire is administered to the same set of participants under the same conditions on at least two separate occasions, and the results from the occasions are then correlated. Although popular, this method of assessing test reliability is based on the assumption that the variable under consideration is stable across time. In human and animal personality research, correlations from 0.48 and 0.90 have been reported (French, 1993; Gosling et al., 2003a,b), with some variation in results possibly reflecting maturation, senescence or learning in the lag between test administrations. Personality in adults is considered to be relatively stable, whereas in infants and juveniles development patterns of both inheritance and learning usually increase variability. An additional limitation of test–retest correlations is that participants may 'drop out' of the study before the second administration.

Another common method for assessing the reliability of personality scales is to correlate the results of questionnaire administration from two or more participants, who are equally familiar with the same subject, rating that individual. This is quick, easy and cost effective as it requires only one administration of the questionnaire. It is limited by the level of knowledge of the subject by each participant and possible association between item scores, in accord with implicit personality theory (IPT). IPT describes the theories people have regarding how different personality traits co-vary (Borkenau, 1992). For example, a person may consider that being quiet and being studious tend to occur together and would rate a quiet person more highly for studiousness regardless of whether the person displays the trait of studiousness to a high level. Increasing acquaintanceship between judge and subject tends to reduce the effect of IPT (Borkenau, 1992). Research in humans and other animals have reported a wide range of inter-rater reliability levels. McCrae and Costa (1987) reported intra-class correlations between raters of

between 0.30 and 0.65 for people rating a person they knew well. Convergent correlations of the aggregated peer ratings in this study with the ratees self-reports varied from 0.29 up to 0.62 (McCrae and Costa, 1987). Reliability research in animal personality is discussed in a detailed review by Gosling (2001). Martau et al. (1985) report inter-rater reliabilities calculated using Pearson-product moment correlations of between −0.005 and 0.92 across four species of new world monkeys. Gosling et al. (2003a,b) reported an average reliability of 0.62 in pet dogs rated by their owner and a person who knew them well.

This paper explores the inter-rater and test–retest reliability of the MCPQ-R for assessing canine personality.

1. Materials and method

1.1. Participants

1.1.1. Inter-rater reliability test

Participants were 65 pairs of people ($n = 130$) who responded to stories in suburban community and commercial radio and articles in local and state newspapers introducing this research. To be included in the study, participants were required to live in Australia and speak English as their first language. Each pair of participants owned a dog aged 18 months or older that they had lived with for at least 12 months.

Of the respondents recruited, 55 (42.3%) were male and 75 female (57.7%). Their mean age was 48.8 years (S.D. 14.2) and most (58.5%) participants had lived with five or more dogs (see Table 1).

The 65 dogs were aged between 18 months and 16 years with a mean age of 5.7 years (S.D. 3.4). There were 34 female (52.3%) and 31 males (47.7%). Most (90.3%) were desexed; only three males and four females being entire. Twenty-four different Australian National Kennel Council (ANKC) recognized breeds were represented ($n = 41$, 63.1%) along with two unrecognized breeds (spoodle and Swiss white shepherd, $n = 2$, 3%) and 22 (33.8%) mixed breeds. Most dogs (84%) were acquired before 6 months of age and 61 (63%) spent 80% or more time indoors.

1.1.2. Test–retest reliability

At the start of this study, all participants indicated they were happy to be contacted a second time to complete the questionnaire again for their dog for the test–retest reliability study. Due to participants not responding to the second contact, the number of participants was reduced from 130 participants to 100 participants; 59 (59%) female and 41 (41%) male. The mean age for this group was 50.2 years ±13.9. Questionnaire responses from one participant randomly chosen from each pair were used for the inter-rater reliability study. Table 1 shows the demographic information for the participants who remained in the study and those who dropped out before the test–retest study.

1.2. Methods

Participants contacted the researcher by phone or by email to indicate their interest in taking part in the study. The aim and data collection process was described to them

Table 1
Demographic information and independent samples *T*-tests and Chi-square tests for all participants and dogs and for those remaining in the study and those who dropped out after the inter-rater reliability (IRR) phase.

Characteristic	Total	Remained	Dropped out	Comparison of remained and dropped out
Number of owners	130	100	30	
Sex of owners				
Male (%)	55 (42.3)	41 (41)	14 (46.7)	Chi-Square
Female (%)	75 (57.7)	59 (59)	16 (53.3)	NS
Owner age (years)				
Mean	48.8	50.2	44.0	Independent *T*-test
S.D.	14.2	13.9	14.4	$t = -2.09$, df = 126, $P = 0.04$
Number of dogs lived with in lifetime				
1 *N* (%)	6 (4.6)	3 (3)	3 (10)	
2–3 *N* (%)	26 (20.0)	21 (21)	5 (16.7)	Chi-Square
4–5 *N* (%)	22 (16.9)	16 (16)	6 (20)	NS
>5 *N* (%)	76 (58.5)	60 (60)	16 (53.3)	
Knowledge of dog[a]				
Mean	5.8	5.8	5.8	Independent *T*-test
S.D.	0.6	0.6	0.5	NS
Time dog spent inside (%)				
Mean	65.7	65.7	66.7	Independent *T*-test
S.D.	29.7	31.8	21.2	NS
Number of dogs	65	50	15	
Sex of dogs				
Male (%)	31 (47.7)	24 (48)	8 (53.3)	Chi Square
Female (%)	34 (52.3)	26 (52)	7 (46.7)	NS
Dog age (years)				
Mean	5.7	6.3	5.1	Independent *T*-test
S.D.	3.4	3.6	2.5	NS

[a] As rated by the participant on a six-point scale with 1 = not very well and 6 = very well.

at this time, including the fact they would be contacted a second time in 6 months to repeat the questionnaire for their dog. Participants were asked to supply their name, address, a contact number and/or an email address which were retained on a separate data base from the personality data. This was so the participants could be contacted for the second data collection. Participants were sent two research packs and asked to complete one and have someone who lived with them and the dog, who was aged 18 years or older and spoke English as their first language to complete the second pack. The packs contained an information sheet about the research, a consent form, demographic questions about the participants (their first name, sex, age, number of dogs they had lived with, and a rating from 1 to 6 of how well the participant feels they know the subject dog), questions about the participants' dog (its name, age, sex, sexual status, breed, estimates of height, weight, build and coat type, age acquired, and an estimate of how much time it spends inside), the MCPQ-R and a reply paid envelope. Participants were instructed to complete the questionnaire alone and not to discuss their answers with the other member of their pair. Six months later, this pack, minus the consent form, was sent out to participants for the second data collection. When the questionnaires were returned to the researcher, each dog was assigned a number and the two owners were randomly designated as either Person A or Person B. They retained these designations throughout the study. The participants were asked to supply their dog's name and their first name only on the questionnaire to ensure at the

second data collection point that the same dogs were rate by the same people at both data collection times.

1.3. Data analysis

The data were analyzed using SPSS v 14.0 (SPSS Inc., SPSS Inc. Headquarters, 233 S. Wacker Drive, 11th floor Chicago, IL 60606). Scale scores for the five canine personality dimensions were calculated from the ratings for each dog from each person in each pair using the same method as described in earlier papers by two of the authors (Ley et al., 2008). That is, using the method of Percentage of Maximum Possible (POMP) (Cohen et al., 1999), the ratings for the words from each of the five scales were summed, divided by the maximum score possible for the scale and converted to a percentage.

1.3.1. Inter-rater reliability and test–retest reliability

Each canine personality dimension from each type of test administration was considered separately. Intra-class correlations (ICC) were used to assess the difference between scores for each personality dimension given by the raters. A level of 0.6 equates to 'good' reliability when utilizing the ICC (Cicchetti, 1994).

2. Results

2.1. Respondents

Table 1 shows the demographic information for the participants who remained in the study for the test–retest

Table 2
Mean, standard errors and intra-class correlations (ICC) for the five dimensions of the Monash Canine Personality Questionnaire-Revised (MCPQ-R) using inter-rater reliability and test–retest methodologies.

Method of testing reliability	Personality dimensions									
	Extraversion		Motivation		Training focus		Amicability		Neuroticism	
	1[a]	2[b]	1[a]	2[b]	1[a]	2[b]	1[a]	2[b]	1[a]	2[b]
Inter-rater reliability										
Mean	65.6	68.5	69.8	66.4	78.3	77.4	77.8	77.2	44.8	44.4
SE	2.4	2.6	2.4	2.5	1.9	1.9	2.4	2.3	2.7	2.7
ICC	0.86		0.82		0.73		0.81		0.75	
P<	0.001		0.001		0.001		0.001		0.001	
df	60		60		60		64		63	
Test–retest reliability										
Mean	65.2	67.8	67.1	67.1	78.8	77.9	79.5	82.8	46.3	44.4
SE	2.6	2.8	2.5	2.9	2.2	2.3	2.9	2.7	2.4	2.8
ICC	0.91		0.93		0.89		0.85		0.79	
P<	0.001		0.001		0.001		0.001		0.001	
df	46		47		47		49		47	

[a] First administration of MCPQ-R.
[b] Second administration of MCPQ-R.

phase and those who dropped out of the study at this point. Comparing the 30 people who dropped out with the 100 people who remained in the study using independent samples T-tests and Chi square tests, no significant differences were found for owner sex, the number of dogs they had previously lived with, their knowledge of their dog or the amount of time their dog spends inside between the two groups. Owner age was significantly different with the mean age of the people remaining in the study being older than those who dropped out. No significant differences were found between the dogs that remained in the study and those that dropped out for dog sex and dog age.

2.1.1. Inter-rater reliability

The results for the inter-rater reliability analyses are shown in Table 2. All correlations were positive and significant at $P < 0.001$. The strongest ICC was for the dimension Extraversion (ICC = 0.86, $P < 0.001$), while the weakest was for Neuroticism (ICC = 0.75, $P < 0.001$). All ICC values were higher than the 'good reliability' level of 0.6 reported by Cicchetti (1994).

2.1.2. Test–retest reliability

Table 2 shows the results of the correlations for the test–retest reliability analyses. All correlations were positive, as expected, and all were significant at $P < 0.001$. The strongest correlation was for the dimension Motivation (ICC = 0.93, $P < 0.001$) while the weakest correlation was for Neuroticism (ICC = 0.79, $P < 0.001$). Again all values were higher than 0.6 suggested as demonstrating good reliability for the results.

3. Discussion

The aim of the current project was to test the reliability of the MCPQ-R, a recently developed instrument for measuring canine personality using five dimensions: Extraversion, Motivation, Training Focus, Amicability and Neuroticism. Two established methodologies, based on inter-rater and test–retest reliability tests, were used to measure reliability. Comparison of the participants who remained in the study with those who dropped out revealed a significant difference for owner age, with the mean age of those who remained in the study being older than those who dropped out. Otherwise there were no significant differences between the owners on the demographic information collected.

The inter-rater reliability test is widely used in human and animal personality research. Among the human literature, reported intra-class correlations have ranged between 0.3 and 0.65 (McCrae and Costa, 1987). The reported inter-rater correlations in animal personality research tend to be stronger than those of human research. For example, King and Figueredo (1997) reported a range of inter-rater reliabilities of 0.55–0.81 for descriptors of personality in Chimpanzee (Pan troglodytes) while inter-rater reliabilities of 0.6 were reported for vervet monkeys (Cercopithecus aethiops) (McGuire et al., 1994). A study in horses found an average inter-rater correlation of 0.4 (Morris et al., 2002). Most relevant to the current study, the mean inter-rater reliability correlation coefficient in dogs was 0.62, with the highest correlation being 0.76 for Extraversion dimension and the lowest correlation being 0.55 for the Openness and Agreeableness dimensions as measured by a modified version of the Big Five Inventory (BFI); a 44-item questionnaire for assessing human personality (Gosling and Vazire, 2002). The results of the current study show that the correlation between the raters for each dog all of the dimensions are strong, being greater than the 0.62 average reported by Gosling and Vazire (2002).

The two lowest correlations were found for Training Focus and Neuroticism in the inter-rater reliability test. Dogs vary in their responses to different people (Vas et al., 2005) and so the lower correlations for Neuroticism and Training Focus may be a function of the variability of the relationships that dog have with their carers. Relationships between carers and dogs may be affected by human attributes such as personality, dog-handling skills, timing, consistency and empathy. It is possible that the nature of

this relationship can influence the dog's apparent fearfulness or timidity. However, this is not to say that dogs do not have stable personality traits, rather the degree to which the expression of these traits may be affected by variable interactions with different people may give the appearance of variability in personality. In a sense this underlines the behavioural flexibility of dogs as they adapt to humans. It highlights the need for any dog-matching process to be complemented by a robust and relevant measure of human personality. A confident person, for example, may rate a dog as relaxed and moderately fearful while a less confident handler may rate the dog as tense and very fearful. The extent to which interaction with different people affects how canine personality is scored has not been explored. It was not possible to collect human personality data from participants as the scope of this short study was to assess the reliability of the MCPQ-R. Future studies are planned to explore the relationships between canine and human personality and how they affect the human animal bond.

The test–retest format for assessing reliability is rarely reported in animal personality studies. This may be, in part, due to the nature of the research undertaken. Many studies have been less concerned with developing a robust test for assessing animal personality than with measuring how behavioural individual differences affect the chances of an individual animal reproducing (Huntingford, 1976) or surviving its environment, for example the effect of personality on foraging behaviour (Kieffer and Colgan, 1991; Oers et al., 2005). However, in human personality research, test–retest results are considered important for assessing the reliability of a personality inventory (Costa and McCrae, 1992). The results of the current study are important for assessing the reliability of the MCPQ-R.

It is reported that a test–retest correlation greater than or equal to 0.7 is desirable (Litwin, 1995, pp. 5–31). After a 3-month lag, test–retest correlations of between 0.75 and 0.83 have been reported for the NEO-FFI, a 60-item, statement based questionnaire for assessing human personality (Costa and McCrae, 1992, p. 45). Martau et al. (1985) reported averaged correlations of between 0.17 and 0.98 for carer ratings of Japanese Macaques made 12 months apart. The results of the current study show that the correlation between the two data collection points for all of the dimensions of the MCPQ-R are strong.

In several animal personality studies, human instruments have been modified and applied to a non-human species (e.g. Martau et al., 1985; e.g. Gosling and Vazire, 2002; Morris et al., 2002). The MCPQ-R is a test developed specifically for dogs. Finding the inter-rater reliability and test–retest reliability results for the MCPQ-R align with those found with other species and being similar to results found specifically in dogs supports the MCPQ-R as a reliable test for rating canine personality.

4. Conclusions

The results of the current study support the use of the MCPQ-R for measuring personality in dogs. The results of the inter-rater reliability tests revealed that dog owners show strong agreement in their ratings of familiar dogs for all five

personality dimensions measured by the MCPQ-R. Despite the correlations being strong and significant there was less agreement between owners' ratings of Training Focus and Neuroticism, a finding that may reflect claims that dogs behave differently when interacting with different people and should prompt further investigation into the effects of human behaviour on dog behaviour. The test–retest results also showed high agreement for all five dimensions. Overall the results support the MCPQ-R as reliable for assessing canine personality along the five identified dimensions.

References

Bennett, P.C., Rohlf, V.I., 2007. Owner-companion dog interactions: relationships between demographic variables, potentially problematic behaviours, training engagement and shared activities. Appl. Anim. Behav. Sci. 102, 65–84.

Borkenau, P., 1992. Implicit personality theory and the Five Factor Model. J. Pers. 60, 295–327.

Cicchetti, D.V., 1994. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol. Assess. 6, 284–290.

Cohen, P., Cohen, J., Aiken, L.S., West, S.G., 1999. The problem of units and the circumstance for POMP. Mult. Behav. Res. 34, 315.

Costa, P.T., McCrae, R.R., 1992. Manual for NEO-PI-R. Psychological Assessment Resources Inc., Florida.

French, J.M., 1993. Assessment of donkey temperament and the influence of home environment. Appl. Anim. Behav. Sci. 36, 249–257.

Goddard, M.E., Beilharz, R.G., 1984. A factor analysis of fearfulness in potential guide dogs. Appl. Anim. Behav. Sci. 12, 253–265.

Goddard, M.E., Beilharz, R.G., 1986. Early prediction of adult behaviour in potential guide dogs. Appl. Anim. Behav. Sci. 15, 247–260.

Gosling, S.D., 2001. From mice to men: what can we learn about personality from animal research? Psychol. Bull. 127, 45–86.

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

Gosling, S.D., Rentfrow, P.J., Swann, J., William, B., 2003b. A very brief measure of the Big-Five personality domains. J. Res. Pers. 37, 504–528.

Gosling, S.D., Vazire, S., 2002. Are we barking up the right tree? Evaluating a comparative approach to personality. J. Res. Pers. 36, 607–614.

Hair, J.A., Anderson, R.E., Tatham, R.L., Black, W.C., 1998. Multivariate Data Analysis, 5th ed. Prentice-Hall International, London, pp. 89, 114, 577–666.

Huntingford, F.A., 1976. The relationship between anti-predator behaviour and aggression among conspecifics in the three spined stickleback, Gasterosteus aculeatus. Anim. Behav. 24, 245–260.

John, O.P., 1990. The "Big Five" factor taxonomy: dimensions of personality in the natural language and in questionnaires. In: Pervin, L.A. (Ed.), Handbook of Personality: Theory and Research. The Guilford Press, New York, pp. 66–100.

John, O.P., Soto, C.J., 2007. The importance of being valid: reliability and the process of construct validation. In: Robins, R.W., Fraley, R.C., Krueger, R.F. (Eds.), Handbook of Research Methods in Personality Psychology. The Guilford Press, New York, pp. 461–494.

Kieffer, J.D., Colgan, P.W., 1991. Individual variation in learning by foraging pumpkinseed sunfish, Lepomis gibbosus: the influence of habitat. Anim. Behav. 41, 603–611.

King, J.E., Figueredo, A.J., 1997. The five-factor model plus dominance in chimpanzee personality. J. Res. Pers. 31, 257–271.

Ley, J., Bennett, P.C., Coleman, G.J., 2007. Personality dimensions that emerge in companion canines. Appl. Anim. Behav. Sci. 110, 305–317.

Ley, J., Bennett, P.C., Coleman, G.J., 2008. A refinement and validation of the Monash Canine Personality Questionnaire (MCPQ-R). Appl. Anim. Behav. Sci. 116, 220–227.

Litwin, M.S., 1995. How to Measure Survey Reliability and Validity, vol. 7. Sage Publications Inc., Thousand Oaks, CA, pp. 5–60.

Marston, L.C., Bennett, P.C., 2003. Reforging the bond: towards successful canine adoption. Appl. Anim. Behav. Sci. 83, 227–245.

Marston, L.C., Bennett, P.C., Coleman, G.J., 2004. What happens to shelter dogs? An analysis of data for one year from three Australian Shelters. J. Appl. Anim. Welf. Sci. 7, 27–47.

Martau, P., Caine, N., Candland, D., 1985. Reliability of the emotions profile index, primate form, with Papio hamadryas, Macaca fuscata, and two Saimiri species. Primates 26, 501–505.

McCrae, R.R., Costa, P.T., 1987. Validation of the Five-Factor Model of personality across instruments and observers. J. Per. Soc. Psychol. 52, 81–90.

McCrae, R.R., John, O.P., 1992. An introduction to the five-factor model and its applications. J. Pers. 60, 175–215.

McGuire, M.T., Raleigh, M.J., Pollack, D.B., 1994. Personality features in vervet monkeys: the effects of sex, age, social status, and group composition. Am. J. Primat. 33, 1–13.

Morris, P.H., Gale, A., Duffy, K., 2002. Can judges agree on the personality of horses? Pers. Individ. Diff. 33, 67–81.

Murphy, J.A., 1998. Describing categories of temperament in potential guide dogs for the blind. Appl. Anim. Behav. Sci. 58, 163–178.

Oers, K.v., Klunder, M., Drent, P.J., 2005. Context dependence of personalities: risk-taking behaviour in a social and a nonsocial situation. Behav. Ecol. 716–723.

Patronek, G.J., Beck, A.M., Glickman, L.T., 1997. Dynamics of dog and cat populations in a community. J. Am. Vet. Med. Assoc. 210, 367–642.

Salman, M., New, J., Scarlett, J., Kass, P., Ruch-Gallie, R., Hetts, S., 1998. Human and animal factors related to the relinquishment of dogs and cats in 12 selected animal shelters in the United States. J. Appl. Anim. Welf. Sci. 1, 207–226.

Vas, J., Topal, J., Gacsi, M., Miklosi, A., Csanyi, V., 2005. A friend or an enemy? Dogs' reaction to an unfamiliar person showing behavioural cues of threat and friendliness at different times. Appl. Anim. Behav. Sci. 94, 99–115.

Wilsson, E., Sundgren, P.-E., 1997. The use of a behaviour test for the selection of dogs for service and breeding: I. Method of testing and evaluating test results in the adult dog, demands on different kinds of service dogs, sex and breed differences. Appl. Anim. Behav. Sci. 53, 279–295.