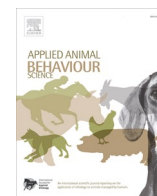


Title	Comparison between the Canine Behavioral Assessment and Research Questionnaire and Monash Canine Personality Questionnaire – Revised to predict training outcome in apprentice assistance dogs
Authors	Marcato, Marinara; Tedesco, Salvatore; O'Mahony, Conor; O'Flynn, Brendan; Galvin, Paul
Publication date	2023-10-18
Original Citation	Marcato, M., Tedesco, S., O'Mahony, C., O'Flynn, B. and Galvin, P. (2023) 'Comparison between the Canine Behavioral Assessment and Research Questionnaire and Monash Canine Personality Questionnaire – Revised to predict training outcome in apprentice assistance dogs', Applied Animal Behaviour Science, 269, 106086 (15 pp). Available at: https://doi.org/10.1016/j.applanim.2023.106086
Type of publication	Article (peer-reviewed)
Link to publisher's version	https://github.com/mmarcato/dog_questionnaire - 10.1016/j.applanim.2023.106086
Rights	© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/). - http://creativecommons.org/licenses/by/4.0/
Download date	2026-10-29 14:53:16
Item downloaded from	https://hdl.handle.net/10468/15664



UCC

University College Cork, Ireland
 Coláiste na hOllscoile Corcaigh



Comparison between the canine behavioral assessment and research questionnaire and monash canine personality questionnaire – Revised to predict training outcome in apprentice assistance dogs

Marinara Marcato^{*,1}, Salvatore Tedesco, Conor O'Mahony, Brendan O'Flynn, Paul Galvin

Tyndall National Institute, Lee Maltings Complex Dyke Parade, Cork T12 RSCP, Ireland

ARTICLE INFO

Keywords:

Assistance dogs
Training outcome prediction
C-BARQ
MCPQ-R
Dog questionnaires

ABSTRACT

Ratings of canine behaviour and personality are a convenient method used by dog training organisations to gather information about prospective and trainee dogs. The objective of this study was to compare the use of two rating questionnaires to predict training outcomes in assistance dogs. It was of interest to investigate the predictive power of a questionnaire answered by the dog trainer in case no puppy raiser questionnaire was available. Two standardised ratings were used, in particular, the Canine Behavioral Assessment and Research Questionnaire (C-BARQ) was completed by puppy raisers around the time the dogs started formal training and the Monash Canine Personality Questionnaire - Revised (MCPQ-R) was completed by dog trainers at ten weeks of training. Rating data were independently analysed to investigate their relationship with training outcomes. The results from the univariate logistic regression analysis were used to select the variables for the reduced feature sets that were used for modelling. The novel machine learning models built with data collected using the C-BARQ and MCPQ-R achieved similar performance in predicting training outcomes, an area under the ROC curve of 0.84 and 0.85, respectively. The novel models developed in this research were the most effective early prediction of suitability for assistance work compared to previously reported studies. The MCPQ-R was demonstrated for the first time to be a reliable canine behavioural assessment method for estimating future outcomes in trainee dogs. The dataset and code used are publicly available on [GitHub](https://github.com).

1. Introduction

Ratings of canine behaviour and personality have been designed and used to identify key behavioural patterns and personality traits in dogs. Ratings are essentially questionnaires consisting of either context-based or adjective-based items that are assessed through scales or codings (Barnard et al., 2016). They represent a convenient method for dog training organisations to gather information about prospective and trainee dogs in an objective and systemic way. A variety of studies have been carried out with the goal of assessing the suitability of dogs for assistance work using ratings (Marcato et al., 2022). This study aims to evaluate and compare the validity of traditional and novel canine behaviour and personality assessment ratings to predict training outcomes for trainee assistance dogs.

Two criteria were used in this research to identify possible ratings: (1) designed to assess all aspects of dog personality and behaviour, and

(2) standardised and tested for reliability and validity purposes in previous studies. The following aspects were considered to select the most appropriate rating for the two groups of participants: length, vocabulary and familiarity. The length of the questionnaire affects the time it would take for participants to complete it. The vocabulary influences their understanding of the concept that the item aims to evaluate. Finally, familiarity considers their ability to answer the items based on previous experiences with the dog.

The C-BARQ rating was deemed to be the most appropriate for the puppy raisers as they may not have a sufficient understanding of the technical vocabulary used in adjective-based ratings. Puppy raisers are still willing to complete this long questionnaire as they only answer it for one dog. The context-based items are not an issue in this instance, as puppy raisers are very familiar with the dog's responses in a range of situations.

The MCPQ-R rating was chosen for the dog trainers because they are

* Corresponding author.

E-mail address: marinara.marcato@tyndall.ie (M. Marcato).

¹ 0000-0002-7938-5227

not familiar with the dog's response to some particular situations referenced in context-based ratings like the PTSQ and they usually complete the questionnaire for multiple dogs at a time; therefore, shorter rating questionnaires are preferred to promote participation. Although the vocabulary used in this adjective-based questionnaire can be considered technical, dog trainers usually have a great understanding of dog behaviour and personality traits.

The validity and reliability of both the C-BARQ and MCPQ-R were comprehensively evaluated in the literature. The validity and internal consistency of the C-BARQ and MCPQ-R were demonstrated in [Hsu and Serpell \(2003\)](#); [Ley et al. \(2008\)](#) and [Ley et al. \(2009\)](#), respectively. The inter-rater and test-retest reliability of the C-BARQ and MCPQ-R were reported in [Duffy and Serpell \(2008\)](#) and [Ley et al. \(2009\)](#), respectively.

The main goal of this study was to apply various machine learning methods to improve the prediction of training outcomes in trainee assistance dogs using ratings. In order to achieve that goal, two machine learning models were built separately using data from the C-BARQ as rated by puppy raisers, and MCPQ-R as rated by dog trainers. This is the first study to explore the employment of the MCPQ-R rating for this purpose; hence, it was also of interest to compare model performance with the widely used C-BARQ. This study focuses on investigating the association between behaviour and personality (as assessed using standardised ratings) and training outcomes. More specifically, it was hypothesised that the MCPQ-R rating can be used to differentiate between dogs that successfully completed the assistance dog training programme and those that were withdrawn for behavioural reasons.

2. Methodology

The Irish Guide Dogs for the Blind (IGDB) invited puppy raisers to complete the rating C-BARQ when the dogs were due to be returned to their training facility for formal training at the age of 13–18 months. Subsequently, dog trainers were invited to complete the rating MCPQ-R when the dogs reached 10 weeks of formal training around the age of 15–21 months.

2.1. Ethical approval

The Social Research Experimentation Committee (SREC) at University College Cork (UCC) reviewed and approved the questionnaire data collection procedures under request number 2019–016. Puppy raisers, dog trainers and dog instructors related to the dogs participating in the research reviewed the information sheet and signed the consent form before completing the questionnaires.

2.2. Subjects

This study is part of a collaboration between Tyndall National Institute and the IGDB which are located in Cork, Ireland. All dogs starting formal training at the IGDB from April/2019 to April/2021 were recruited to participate in this study. Dogs who completed the training were classified as “success” after qualifying as service or guide dogs and being matched with a blind or disabled person, respectively. Dogs withdrawn from the training programme for behavioural reasons were classified as “fail”. Dogs who were withdrawn from the training programme due to health issues or selection as breeding stock were excluded.

2.3. Puppy raiser questionnaire - C-BARQ

The Canine Behavioral Assessment and Research Questionnaire (C-BARQ) comprises 100 items which are grouped into seven sections, namely trainability, aggression, fear and anxiety, separation-related behaviour, excitability, attachment and attention-seeking, and miscellaneous. Participants describe how their dog has been behaving in the recent past using a 5-point scale ([Hsu and Serpell, 2003](#)).

Following the latest recommendations of the questionnaire's creators ([Duffy and Serpell, 2012](#)), a total of 14 factors were calculated by averaging the original items as outlined in [Table A.1](#). Two strategies were adopted to handle missing values in the calculation of the factors similar to [Duffy and Serpell \(2012\)](#). Firstly, factors were only calculated if at least 75% of the items necessary for their calculation were available. Secondly, missing values were removed before taking their average for the calculation of the factors.

2.4. Dog trainer questionnaire - MCPQ-R

The Monash Canine Personality Questionnaire - Revised (MCPQ-R) is composed of 26 adjective-based items measuring canine personality along five dimensions: extraversion, motivation, training focus, amicability, and neuroticism. Participants describe their dog's personality using a 6-point scale from 1 (really does not describe my dog) to 6 (really describes my dog).

2.5. Data analysis

The data were analysed using R version 4.2.1 ([Team, 2022](#)). Logistic regression models were used for both feature selection and prediction of training outcome ([Hosmer et al., 2013](#)). For the purpose of the analysis, the C-BARQ and MCPQ-R variables were processed as numerical features representing interval variables. The original C-BARQ dataset comprised two demographic variables (sex and breed), 100 items and 14 factors, resulting in 116 features. The original MCPQ-R dataset contained two demographic variables (sex and breed) and 26 items, resulting in a total of 28 features. The reduced feature dataset was created from the original C-BARQ and MCPQ-R datasets, by selecting the features that met the following criteria:

1. a significant association ($p < 0.05$) with training outcomes as calculated by logistic regression;
2. a significant reduction ($p < 0.05$) of deviance in the logistic regression model as evaluated by the chi-squared test;
3. items not included in the calculation of factors that met criteria 1 and 2 - C-BARQ only;
4. less than 5% missing values - C-BARQ only;
5. a correlation value below 0.75 with another independent variable.

The reduced feature dataset contained only features that achieved a significant level of association with training outcome (criteria 1 and 2) as proposed in [Asher et al. \(2013\)](#). It removed items which were used for calculating significant factors while keeping only those not used for calculating the selected factors (criterion 3). Variables containing more than 5% missing data were removed, instead of the 10% as denoted in [Duffy2012](#) and [Bray2019](#) (criterion 4) due to the size of our dataset. If the correlation between two independent variables reached the threshold of 0.75, the variable which had the highest mean absolute correlation considering all other covariates was removed (criterion 5).

The backward feature selection algorithm was implemented using Recursive Feature Elimination (RFE) with a Leave One Out Cross Validation (LOOCV) sampling strategy. The reduced feature dataset was pre-processed in three steps as follows: imputation of missing values with the mean, normalisation of values using z-score, and finally dimensionality reduction with Principal Component Analysis (PCA). PCA retained the principal components that captured 95% of the variance in the reduced dataset.

A multivariate logistic regression model was used to predict the probability of withdrawal and optimised for maximising precision. The area under the Receiver Operating Characteristic (ROC) curve was used to validate and compare models. Sensitivity, specificity and area under the Precision-Recall (PR) curve were also reported. All metrics were calculated using the predictions on the validation examples provided by LOOCV.

The model was intended to aid assistance dog organisations when deciding whether to continue training a dog or release it from training by flagging only dogs that were very likely to be withdrawn; however, they were ultimately responsible for making a final decision. The flagging system utilised was similar to that presented in Harvey et al. (2017), Asher et al. (2013) and Bray et al. (2019). In particular, dogs were assigned a yellow flag if their probability of withdrawal was above an arbitrary threshold of 0.70, or a green flag if their probability of withdrawal was below an arbitrary threshold of 0.30.

The impact of the features from the reduced dataset on the principal components was investigated through the coefficient loadings on the principal components selected by RFE. The estimates from the optimal multivariate logistic regression model were analysed through the calculation of the Odds Ratio (OR) and 95% Confidence Interval (CI).

3. Results

3.1. Puppy raiser questionnaire - C-BARQ

The puppy raisers of 63 dogs rated their dog's behaviour using C-BARQ. In this sample, the breakdown of the dog's sex was female ($N = 30$) and male ($N = 33$), and the breed was Labrador Retriever ($N = 25$), Labrador Retriever crossed with Golden Retriever ($N = 36$), Golden Doodle ($N = 1$), German Shepherd ($N = 1$), and age at assessment in months was (Mean = 16.1, SD = 3.4). The Golden Doodle and German Shepherd were combined into one class called "Other" for further analysis. The failure rate in this sample was 18%, as 11 failed and 52 succeeded in their training. Dogs were withdrawn after 14 weeks of training on average. Only three out of the total 11 dogs that failed to graduate were withdrawn before 10 weeks.

3.1.1. Feature selection

The results from the univariate logistic regression models are shown in Table A.2 where the 29 variables satisfying the first two criteria are highlighted. Demographic information such as breed (Other vs. LR $p = 0.44$ and Other vs LRxGR $p = 0.16$) and sex ($p = 0.61$) were not significantly associated with training outcome, hence both variables were discarded (criterion 1). A total of 29 variables met criteria 1 and 2 which indicated a significant association with training outcome and reduction in deviance, respectively. Specifically, there were 23 original items (out of 100) and six derived factors (out of 14). Criterion 3 was applied to those 23 items to retain only items not included in the calculation of the six significant factors. Hence, nine original items and six derived factors satisfied the first three criteria. Criterion 4 was applied to those 15 variables and resulted in the removal of two variables with more than 5% missing data.

Finally, criterion 5 was applied to those 13 variables to remove highly correlated variables. The correlation heatmap is shown in Fig. A.1. Only one variable was removed for reaching a correlation level of more than 0.75, namely "F_07 Dog directed aggression", because of its high correlation with "F_08 Dog directed fear".

Consequently, the C-BARQ reduced feature set was created with 12 variables. There were seven items belonging to the following sections: Trainability ("I_01 When off the leash returns immediately when called"), Aggression ("I_27 Toward cats, squirrels or other small animals entering your yard"), Miscellaneous ("I_75 Chases or would chase birds given the opportunity", "I_82 Bogs persistently for food when people are eating", "I_85 Pulls excessively hard when on the leash", "I_90 Hyperactive restless has trouble settling down", "I_99 Licks people or objects excessively"); and five factors "F_02 Stranger directed aggression", "F_06 Nonsocial fear", "F_08 Dog directed fear", "F_10 Separation related problems", and "F_11 Excitability".

3.1.2. Model

Data pre-processing techniques were applied to the reduced feature set, including normalisation, imputation and PCA. PCA retained 10 first

Table 1

Best logistic regression model using selected by RFE.

	Estimate	Std. Error	z value	p-value	OR	2.5%	97.5%
(Intercept)	- 2.06	0.47	4.40	0.00	0.13	0.04	0.29
PC01	- 0.80	0.24	3.31	0.00	0.45	0.26	0.68

principal components as they captured 95% of the variance. The explained variance per principal component is shown in Fig. A.2. PC01, the first principal component, captured 34% of variance in the data, which was higher than the next three largest components combined.

The optimal logistic regression model selected by RFE contained a single principal component (PC01) as a predictor. The parameters estimated for the optimal model using PC01 are shown in Table 1. The OR and 95% CI were interpreted to understand the relationship between PC01 and training outcomes. The probability of withdrawal decreases by 55% ($0.45 - 1 = -0.55$) for every unit increase of PC01. We are 95% confident that a one unit increase in PC01 results in 32–74% ($0.68 - 1 = -0.32$, $0.26 - 1 = -0.74$) lower odds of being withdrawn from training.

Model fitness was evaluated considering the reduction in deviance resulting from the addition of one variable (PC01). The null deviance was 58.35 on 62 degrees of freedom, and the residual deviance was 39.33 on 61 degrees of freedom. Hence, the difference in deviance was $58.35 - 39.33 = 19.02$ on $62 - 61 = 1$ degrees of freedom. The model can be considered a good fit for the data as it fitted the data significantly better than the null model $P(\chi^2_1 \geq 19.02) = 1.30e^{-05}$. Thus, this significant reduction in deviance supported the hypothesis that PC01 was a significant variable in predicting training outcomes.

3.1.3. Performance

The ROC and PR curves are shown in Fig. A.3 and the area under the ROC curve was 0.84 and the PR curve was 0.64. The precision, specificity, and sensitivity estimated using LOOCV prediction results were 0.57, 0.94 and 0.36, respectively.

Fig. 1a and b show the probability of failure predicted by the best logistic regression model for all the 63 dogs in the dataset considering the 12 C-BARQ items and factors as rated by their puppy raisers. Two boxes show idealised arbitrary thresholds for a flagging system, where the green flag is assigned to dogs likely to succeed and the yellow flag to dogs likely to be withdrawn.

Fig. 1a displays the training outcome obtained by each dog; i.e., whether the dog failed or succeeded in their training. It demonstrates that the lowest estimates belonged to dogs who succeeded and the highest estimates to dogs who failed. Considering the flagging system in the cross-validated results, green flags were assigned to 53 dogs and yellow flags to four dogs. From the green-flagged group, a total of 48 dogs succeeded and only five failed their training resulting in 91% precision in predicting success. In the yellow-flagged group, all four dogs failed to yield 100% precision in the prediction of failure.

Fig. 1b shows the same probability curve highlighting the status of training outcome, whether the dog became a guide dog, service dog or was withdrawn from training. Guide dogs achieved a lower estimated probability of withdrawal than service dogs. In particular, guide dogs had an average probability of withdrawal of 0.10, service dogs 0.14 and withdrawn dogs 0.44. The six non-flagged dogs got a very mixed outcome as expected; two were withdrawn, two became assistance dogs, and two became guide dogs.

3.1.4. Feature importance

The coefficient loading on PC01 is shown in Fig. 2. The top five variables that most strongly contributed to PC01 were "F_11 Excitability", "F_08 Dog directed fear", "F_06 Nonsocial fear", "F_10 Separation related behaviour" and aggression "I_27 Toward cats squirrels or other small animals entering your yard". The only positive coefficient

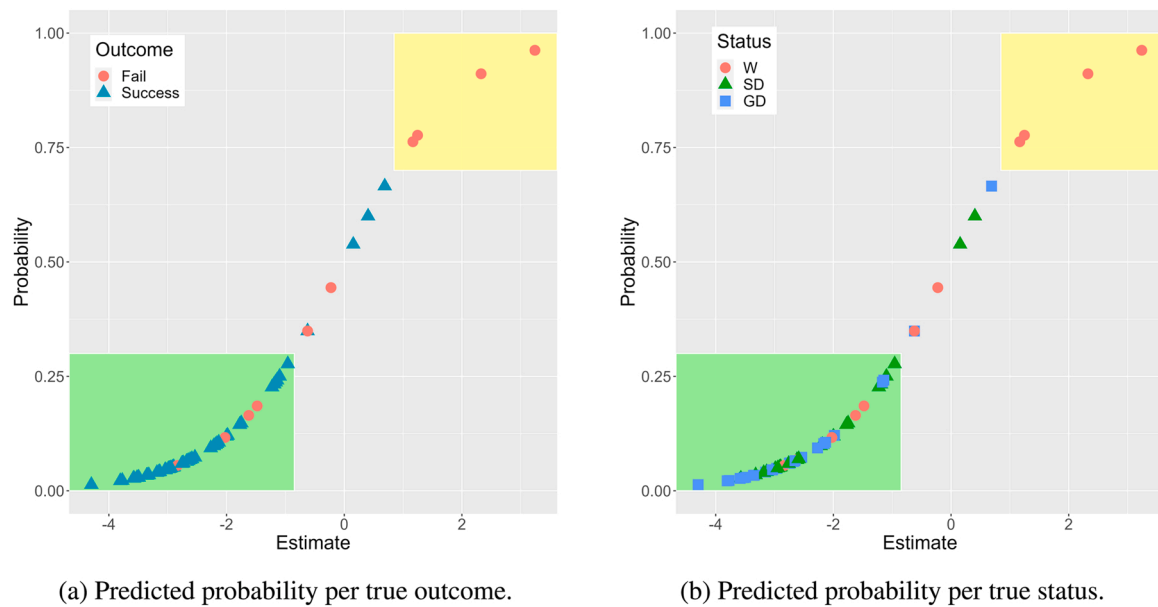


Fig. 1. Probability of withdrawal predicted by the best logistic regression model using C-BARQ data. Status: W = withdrawn, SD = service dog, GD = guide dog.

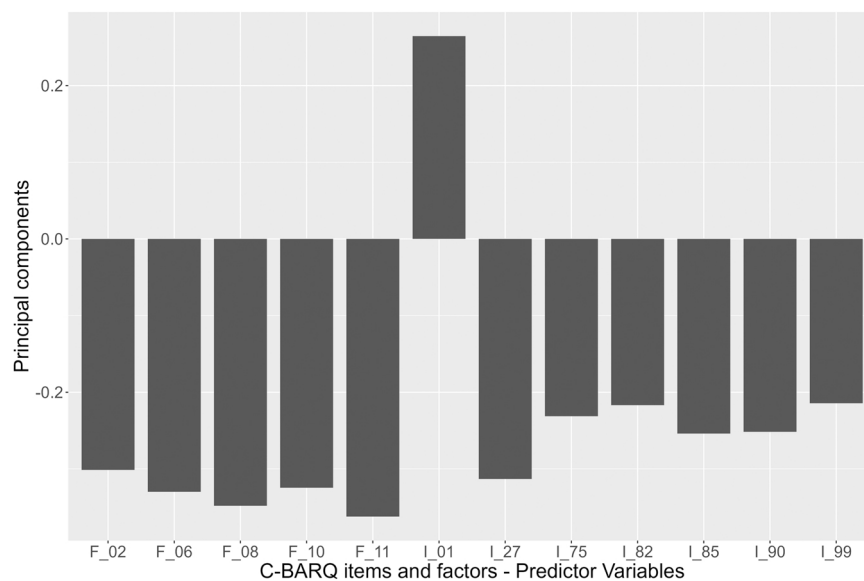


Fig. 2. PC01 coefficient loading from original items and factors.

was related to the variable “I_01 When off the leash returns immediately when called” while all others were negative. This result indicates that higher scores in I_01 produce higher values of PC01; however, higher scores in any of the other original variables cause lower values of PC01.

3.2. Dog trainer questionnaire - MCPQ-R

The dog trainers of 89 dogs rated their dog's behaviour using MCPQ-R. In this sample, the breakdown of the dogs' sex was female (N = 44) and male (N = 45), and the breed was Labrador Retriever (N = 37), Labrador Retriever crossed with Golden Retriever (N = 49), Golden Doodle (N = 2) and German Shepherd (N = 1) and age at assessment in months was (Mean = 18.34, SD = 2.6). A total of 71 dogs succeeded (70%) and 18 dogs failed (20%). Dogs were withdrawn after 16 weeks of training on average, only three out of the 18 dogs that failed to graduate as service dogs were withdrawn before 10 weeks.

3.2.1. Feature selection

The results from the univariate logistic regression are shown in Table A.4 where the 13 variables satisfying the first two criteria are highlighted. Demographic information such as breed (Other vs. LR p-value = 0.82 and Other vs. LRxGR p-value = 0.40) and sex (p-value = 0.32) was not significantly associated with training outcome. Criterion 1 and 2 were applied preserving only a total of 13 MCPQ-R variables. The correlation between the variables was evaluated and is shown in Fig. A.4. Criterion 5 was applied resulting in the removal of one highly correlated item, namely “Neuroticism Fearful” due to its high correlation with “Neuroticism Nervous”.

Consequently, the MCPQ-R reduced feature set was created with 12 variables. They belonged to the following sub-scales Motivation (Persevering), Training Focus (Attentive, Obedient, Reliable, Trainable), Amicability (Easy going, Friendly, Relaxed, Sociable), and Neuroticism sub-scale (Nervous, Submissive, Timid).

Table 2

Best logistic regression model using selected by RFE.

	Estimate	Std. Error	z value	p-value	OR	2.5%	97.5%
Intercept	-1.99	0.40	-4.97	0.00	0.14	0.06	0.28
PC1	0.77	0.20	3.95	0.00	2.17	1.54	3.37

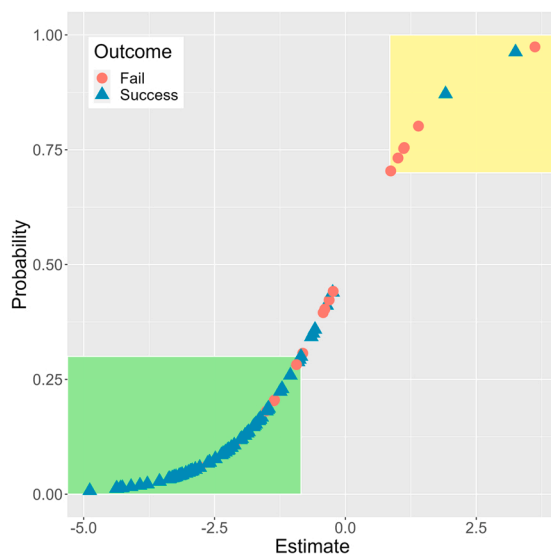
3.2.2. Model

Data pre-processing techniques were applied to the reduced feature set, including normalisation, imputation and PCA. PCA retained nine first principal components as they captured 95% of the variance. The explained variance per principal component is shown in Fig. A.5. PC1, the first principal component, captured 36% of variance in the data, which was higher than the next two largest components combined.

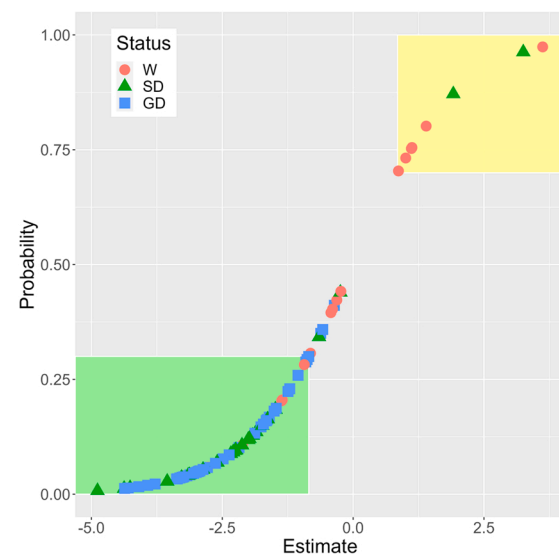
The optimal logistic regression model selected by RFE contained a single principal component (PC1) as a predictor. The parameters

estimated for the optimal model using PC01 are shown in Table 2. The OR and 95% CI were interpreted to understand the relationship between PC1 and training outcomes. The probability of withdrawal increases by $(2.17 - 1 = 1.17)$ 117% for every unit increase of PC1. We are 95% confident that a one unit increase in PC1 results in a 54–237% ($1.54 - 1 = 0.54$, $3.37 - 1 = 2.37$) higher odds of being withdrawn from training.

Model fitness was evaluated considering the difference in deviance resulting from the addition of one variable (PC1). The null deviance was 89.62 on 88 degrees of freedom and the residual deviance was 63.58 on 87 degrees of freedom. Thus, the difference in deviance was $89.62 - 63.58 = 26.04$ on $88 - 87 = 1$ degrees of freedom. The model can be considered a good fit for the data as it fitted the data significantly better than the null model $P(\chi^2_2 \geq 26.04) = 3.34e - 07$. Thus, this significant reduction in deviance supported the hypothesis that PC1 was a significant variable in predicting training outcomes.



(a) Predicted probability per true outcome.



(b) Predicted probability per true status.

Fig. 3. Probability of withdrawal predicted by the best logistic regression model using MCPQ-R data. Status: W = withdrawn, SD = service dog, GD = guide dog.

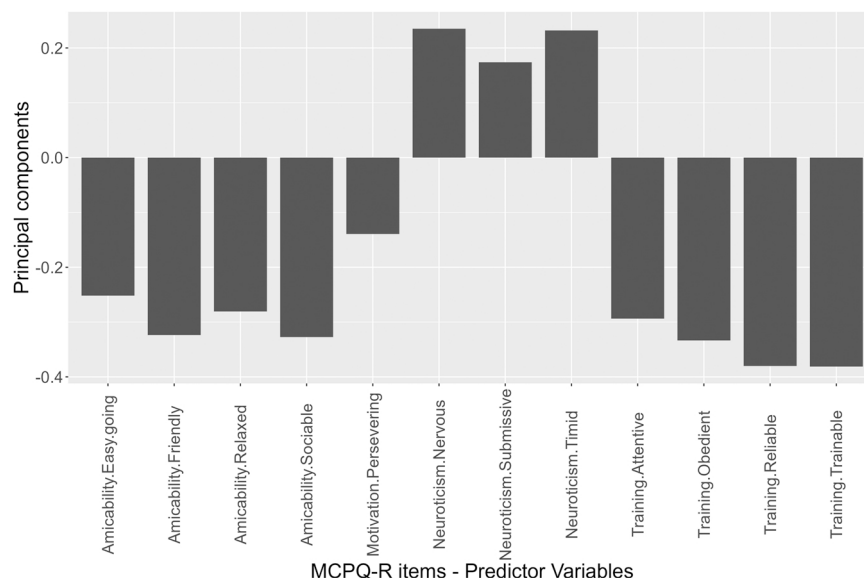


Fig. 4. PC1 coefficient loading from original items.

3.2.3. Performance

The ROC and PR curves are shown in Fig. A.6 and the area under the ROC curve was 0.85 and the PR curve was 0.58. The precision, specificity and sensitivity estimated using LOOCV results were 0.75, 0.97 and 0.33, respectively.

Fig. 3a and b show the probability of failure predicted by the best logistic regression model for all the 89 dogs in the dataset considering the 12 MCPQ-R items as rated by their dog trainers. Two boxes show idealised thresholds for a flagging system, where the green flag was assigned to dogs likely to succeed and the yellow flag indicates that dogs were highly likely to be withdrawn. The thresholds selected were 0.30 for training success and 0.70 for withdrawal.

Fig. 3a displays the training outcome obtained by each dog; i.e., whether the dog failed or succeeded in their training. It demonstrates that the lowest estimates belonged to dogs who succeeded and the highest estimate to dogs who failed. Considering the flagging system in this idealised demonstration, green flags were assigned to 71 dogs and yellow flags to six dogs. From the green-flagged group, a total of 64 dogs succeeded and only seven failed their training resulting in 90% precision in predicting success. In the yellow-flagged group, six dogs failed and two dogs succeeded yielding 75% precision in the prediction of failure.

Fig. 3b shows the same probability curve highlighting the status of training outcome; i.e., whether the dog became a guide dog, or service dog or was withdrawn from training. Guide dogs had a lower estimated probability of withdrawal than service dogs. In particular, guide dogs had an average estimated probability of withdrawal of 0.13, service dogs 0.17 and withdrawn dogs 0.43. The 10 non-flagged dogs got a mixed training outcome as expected; five were withdrawn from training, two graduated as service dogs, and three graduated as guide dogs.

3.2.4. Feature importance

The coefficient loading on PC1 is shown in Fig. 4. The top five variables that most strongly contribute to PC1 are related to the Training Focus (Trainable, Reliable, Obedient), and Amicability (Sociable, Friendly) sub-scales. The only positive coefficients were related to variables in the sub-scale Neuroticism, all others were negative. Therefore, higher scores in any of the items in the Neuroticism sub-scale, namely Nervous, Sensitive and Timid, produce higher values of PC1; however, higher scores in any of the other original variables will cause lower values of PC1.

4. Discussion

This study investigated the predictive power of two standardised behavioural ratings of prospective assistance dogs to estimate their training outcomes. It demonstrated the benefit of collecting data on trainee assistance dogs using canine behaviour and personality questionnaires as it allows the creation of statistical models for the prediction of training outcomes. Puppy raisers answered the C-BARQ around the time when dogs started their formal training (16 months) and dog trainers answered the MCPQ-R after 10 weeks of training (18 months). Machine learning algorithms were independently developed to predict training outcomes based on the data from each of the ratings separately.

The results demonstrated for the first time that the MCPQ-R can be used for estimating a dog's probability of withdrawal from training with comparable performance to the well-established C-BARQ. Moreover, the C-BARQ model significantly outperformed all models previously presented in the literature by applying various machine learning methods (feature selection and reduction techniques) in this field of research for the first time. This is the first study to validate the use of the MCPQ-R for the prediction of training outcomes in trainee assistance dogs. This finding allows dog training organisations to get an early prediction of training outcomes, through not only the puppy raiser's knowledge but also that of the dog trainer.

The models built on C-BARQ ratings completed by puppy raisers and MCPQ-R completed by dog trainers can greatly help assistance dog

training organisations, by classifying dogs into green and yellow flags before they start formal training, and at 10 weeks of training, respectively. The high predictive value of the yellow and green flags indicated that yellow-flagged dogs were very likely to fail, and green-flagged dogs were very likely to succeed. The application of these models in this industry allows for resources to be more optimally managed. In particular, dogs assigned a yellow flag can be more closely assessed, and problematic behaviours can be addressed at an early stage of training. Whenever training resources are limited (e.g. personnel, material, financial), the training of dogs assigned a green flag can be prioritised.

The methodology and results obtained from the C-BARQ and MCPQ-R are discussed in Sections 4.1 and 4.2. The relationship between each of the rating's original variables and the training outcome predictions are complex due to the data reduction technique employed (i.e., PCA). Hence, the interpretation of the optimal model was limited due to PCA. However, a high-level qualitative interpretation is possible considering the association between original variables and principal components, and the relationship between the principal components and probabilities estimated by the logistic regression model.

The results obtained from the C-BARQ and MCPQ-R models developed in this research are compared in Section 4.3. A comprehensive comparison with previous research is provided in Section 4.4. It examines the model created in this research with those from previous research on training outcome prediction using the C-BARQ rating. Then, the prediction performance of the C-BARQ and MCPQ-R models developed in this research is compared with previous research using diverse questionnaires for the prediction of training outcomes in prospective assistance dogs.

4.1. Puppy Raiser Questionnaire - C-BARQ

The original C-BARQ dataset contained 100 items, 14 factors and two demographic variables. Although Duffy and Serpell (2012) and Bray et al. (2019) used all factors and miscellaneous items for modelling, a preliminary feature selection was necessary due to the sample size in this study. Exploratory data analysis revealed that while some factors did not show a significant relationship with training outcomes, the items used to calculate them did show a significant relationship with training outcomes. This result suggested that while the specific behaviours captured by those items provided meaningful information for the prediction of training outcomes, the general behavioural categories derived by combining those items with other items were not as meaningful. Hence, we chose to remove factors not significantly associated with training outcomes from the reduced dataset and retained their items if they were significantly associated with training outcomes, as those items added important necessary information that was not captured by any other independent variables in the reduced dataset. Including both items and factors ensures that the model considered both specific and broad behaviours pertinent to the dependent variable.

After applying the five criteria for feature selection, the reduced C-BARQ dataset contained 12 features; specifically, they were seven items and five factors. The results from the PCA indicate that the only variable with a positive coefficient loading on PC01 was "I_01 When off the leash returns immediately when called", while all the other 11 variables - namely, "I_27 Toward cats squirrels or other small animals entering your yard", "I_75 Chases or would chase birds given the opportunity", "I_82 Begs persistently for food when people are eating", "I_85 Pulls excessively hard when on the leash", "I_90 Hyperactive restless has trouble settling down", "I_99 Licks people or objects excessively", "F_02 Stranger directed aggression", "F_06 Nonsocial fear", "F_08 Dog directed fear", "F_10 Separation related problems", and "F_11 Excitability" - had negative coefficient loadings, as shown in Fig. 2.

The top five variables with the highest coefficients on PC01 were "F_11 Excitability", "F_08 Dog directed fear", "F_06 Nonsocial fear", "F_10 Separation related behaviour", and aggression "I_27 Toward cats squirrels or other small animals entering your yard". Effectively, these

were the variables that most strongly influenced the value of PC01 and consequently the probability of withdrawal.

The optimal logistic regression model estimate associated with PC01 is negative (−0.80) as described in Table 1; therefore, positive values of PC01 result in negative point estimates. Accordingly, this indicates that the optimal logistic regression model learnt that dogs which scored low in “I_01” and high in any of the other 11 variables are more likely to be withdrawn from their training programme.

4.2. Dog trainer questionnaire - MCPQ-R

The original MCPQ-R dataset comprised 26 items and two demographic variables. Preliminary modelling experiments, utilising the five personality sub-scales proposed in Ley et al. (2009), yielded sub-optimal performance in predicting training outcomes. These initial results suggested that the five sub-scales might represent overly broad canine personality dimensions for the prediction of training outcomes; whereas, this particular classification task relies on complex behavioural patterns captured by more granular information. Consequently, the five sub-scales were not utilised in the final models presented in this study.

After applying three criteria for feature selection using the 26 items, the reduced MCPQ-R dataset contained 12 variables relating to four out of the five sub-scales. The results from the PCA indicate that all the variables with a positive coefficient loading on PC1 were associated with the sub-scale Neuroticism (Nervous, Submissive, Timid), while all other variables in the sub-scales Motivation (Persevering), Training Focus (Attentive, Obedient, Reliable, Trainable) and Amicability (Easy going, Friendly, Relaxed, Sociable) had negative coefficient loadings, as shown in Fig. 4.

The top five MCPQ-R items with the biggest impact on PC1 and, consequently, on training outcome estimations are related to the Training Focus (Trainable, Reliable, Obedient) and Amicability (Sociable, Friendly). Effectively, these are the variables that most strongly influence the value of PC1 and consequently the probability of withdrawal.

The optimal logistic regression model estimate associated with PC1 is positive (0.77) as described in Table 2. Accordingly, this indicates that the optimal logistic regression model learnt that dogs which scored high in items related to the Neuroticism sub-scale and low in items related to Motivation, Training Focus and Amicability sub-scales are more likely to be withdrawn from their training programme.

4.3. Comparison between C-BARQ and MCPQ-R models

This section compares the C-BARQ and MCPQ-R from both methodological and performance perspectives in the context of training outcome prediction of prospect assistance dogs. This was the first study to not only investigate the relationship between MCPQ-R and training outcomes but also develop a machine learning model to predict training outcomes using MCPQ-R data. This research demonstrated that the novel MCPQ-R model achieved similar results to the widely used C-BARQ.

From a practical perspective, collecting data with the C-BARQ is more challenging than with the MCPQ-R. The completion rate for the C-BARQ filled out by the puppy raisers (46%) was lower than the MCPQ-R filled out by the dog trainers (74%). Two main reasons are attributed to this outcome: firstly, the length of the questionnaire as the MCPQ-R is significantly shorter (26 vs. 100 items); and secondly, the role of participants, as dog trainers were more involved with the training organisation than puppy raisers.

The similar performance results obtained by the MCPQ-R compared to the widely-used C-BARQ provided a scientific basis for the use of the MCPQ-R for predicting training outcomes in assistance dogs. The C-BARQ model achieved an area under the ROC curve of 0.84 and PR curve of 0.64, a precision of 0.57, a specificity of 0.94, and a sensitivity of 0.36. The MCPQ-R model achieved an area under the ROC of 0.85 and

Table 3

Number of models for which C-BARQ factors were significantly related to training outcome in assistance dogs.

Factors	Duffy and Serpell (2008)	Duffy and Serpell (2012)	Bray et al. (2019)	Hunt et al. (2020)	This study
F_01 Trainability	0	4	1	0	0
F_02 Stranger-directed aggression	1	4	0	NA	1
F_03 Owner-directed aggression	0	3	0	NA	0
F_04 Dog rivalry	NA	3	0	NA	0
F_05 Stranger-directed fear	1	3	1	NA	0
F_06 Nonsocial fear	1	4	0	0	1
F_07 Dog-directed aggression	NA	4	1	NA	1
F_08 Dog-directed fear	NA	3	0	NA	1
F_09 Touch sensitivity	0	3	0	0	0
F_10 Separation related behaviour	0	2	1	1	1
F_11 Excitability	1	3	0	1	1
F_12 Attachment and attention-seeking	1	1	0	0	0
F_13 Chasing	0	2	NA	0	0
F_14 Energy level	1	5	0	0	0

PR curve of 0.58, a precision of 0.75, a specificity of 0.97, and a sensitivity of 0.33. In cases where C-BARQ data are not available at the beginning of training, organisations can obtain reliable predictions of training outcomes by using the MCPQ-R model at ten weeks of training.

The proposed flagging system with two probability thresholds (lower = 0.30 and upper = 0.70) produced more precise predictions compared to using only one probability threshold of 0.50. The predictive value for withdrawal increased from 85% using the 0.5 threshold to 100% when using the 0.70 threshold for C-BARQ and remained 75% for MCPQ-R. Similarly, the predictive value for success increased from 88% to 91% for C-BARQ and from 85% to 90% for MCPQ-R when using the two thresholds.

While the C-BARQ and MCPQ-R models had similar performance results, the specific variables used were seemingly not analogous. In particular, the only similar variables used were “F_02 Stranger directed aggression” in the C-BARQ model and Sociable and Friendly from the Amicability sub-scales in the MCPQ-R model. However, while the C-BARQ model used “F_11 Excitability” and “I_90 Hyperactive restless has trouble settling down”, the MCPQ-R model did not use Excitable or Hyperactive from the Extraversion sub-scale. While the C-BARQ model used “F_08 Dog directed fear” and “F_06 Nonsocial fear”, the MCPQ-R model did not use Fearful from the Neuroticism sub-scale. While the C-BARQ model used “I_27 Toward cats squirrels or other small animals entering your yard”, “I_75 Chases or would chase birds given the opportunity” and “F_02 Stranger directed aggression”, the MCPQ-R model did not include Non-aggressive from the Amicability sub-scale.

Conclusively, the use of MCPQ-R rating offers several advantages over C-BARQ for the prediction of training outcomes in terms of data collection and model complexity. The main advantage is the length as the MCPQ-R (26 items) is significantly shorter than the C-BARQ (100 items) which affects the completion rate. The MCPQ-R data are simpler to model because less pre-processing is required as it comprises fewer variables, contains no missing data and no factors need to be extracted. In this research, the MCPQ-R data were collected after 10 weeks of training, while the C-BARQ data were collected before the dogs started

training. This is a research design limitation which can be overcome in future research by validating the use of this rating either earlier in the training or with puppy raisers.

4.4. Comparison with previous research

In this section, previous studies on the prediction of training outcomes in assistance dogs are compared to this study. In particular, Section 4.4.1 examines studies which analysed the connection between C-BARQ and training outcomes in assistance dogs. Section 4.4.2 discusses the performance of different models based on various questionnaires for the prediction of training outcomes in assistance dogs.

4.4.1. C-BARQ and training outcome

Six studies explored the relationship between C-BARQ and training outcomes in assistance dogs. In particular, two aimed at evaluating the association between the C-BARQ rating and training outcome (Duffy and Serpell, 2008; Hunt et al., 2020), and four studies used the rating to predict training outcome (Batt et al., 2009; Duffy and Serpell, 2012; Cleghern et al., 2018; Bray et al., 2019). However, it is not clear whether an overlap exists in the datasets used in previous research; i.e., the datasets utilised in Duffy and Serpell (2008), (2012); Bray et al. (2019) share a subset of data.

Table 3 shows a qualitative analysis of the four studies that investigated an association between C-BARQ factors and training outcomes in assistance dogs. Duffy and Serpell (2012) trained five logistic regression models separately for five organisations using data collected when the dogs were 12 months old, therefore the number of organisations for which each factor was significant is presented. The other studies trained only one model; hence, significant associations between C-BARQ factors and training outcomes are signalled with one, and not significant associations are marked with zero.

Two studies were excluded because they did not analyse factors (Batt et al., 2009; Cleghern et al., 2018). The former modified or combined the C-BARQ items; hence, no comparison can be performed. The latter reported a significant correlation between items I_10, I_11 and I_12 and training outcomes in guide dogs. Considering that those items are used for calculating F_02, it could be reasonably estimated that F_02 would have yielded a significant relationship with training outcome if that analysis was performed in that study.

The quantitative comparative analysis of these studies allowed the identification of trends. In particular, some factors were more frequently reported to be predictive of training outcomes, namely “F_02 Stranger-directed aggression”, “F_05 Stranger-directed fear”, “F_06 Nonsocial fear”, “F_07 Dog-directed aggression”, “F_10 Separation related behaviour”, and “F_11 Excitability”. Other factors were only associated with training outcomes in very few studies, namely “F_03 Owner-directed aggression”, “F_04 Dog rivalry”, “F_09 Touch sensitivity”, “F_12 Attachment and attention-seeking”, and “F_13 Chasing”. This difference may be due to the low variability in the scores, missing values or factors associated with different organisations such as training methodology and graduation requirements as discussed in Duffy and Serpell (2012).

4.4.2. Prediction performance

This section compares the prediction performance of the two novel models developed in this study (i.e., the C-BARQ and MCPQ-R models) with previous research. In particular, Duffy and Serpell (2012), Harvey et al. (2017), Cleghern et al. (2018), Bray et al. (2019) also utilised ratings to predict training outcomes in assistance dogs. This comparison revealed that both the C-BARQ and the MCPQ-R models achieved better prediction performance than previous studies. Some studies did not provide important performance metrics which complicates the comparison. The following paragraphs provide a more in-depth comparative analysis of this and previous studies.

The C-BARQ model developed in this research achieved an area under the ROC curve of 0.84 on the cross-validated dataset, which was

significantly higher than 0.64–0.72 obtained by Duffy and Serpell (2012) and 0.71 obtained by Bray et al. (2019). Cleghern et al. (2018) reports the model's accuracy, precision and sensitivity using C-BARQ scores to predict training success. The accuracy achieved by Cleghern et al. (2018)'s C-BARQ model was 0.51 while the accuracy achieved by both our C-BARQ and MCPQ-R models was 0.84. Specificity in predicting training withdrawal is mathematically equivalent to sensitivity in predicting training success, as both evaluate the proportion of successful dogs correctly identified. Hence, the specificity of 0.36 achieved by the C-BARQ model presented here can be compared to the sensitivity of 0.46 attained by the C-BARQ model reported by Cleghern et al. (2018). Precision or Positive Predictive Value for successful (PPV) dogs is equivalent to Negative Predictive Value (NPV) for withdrawn dogs. Accordingly, the NPVs of 0.88 and 0.85 achieved by our C-BARQ and MCPQ-R models, respectively, can be compared to the precision of 0.55 achieved by Cleghern et al. (2018)'s C-BARQ model. All in all, this indicates that while Cleghern et al. (2018)'s model better identified successful dogs, our model provided more exact overall predictions of training outcomes.

The prediction performance obtained by our model suggests that the C-BARQ ratings are more appropriate for estimating failure than success, which is in line with findings from previous research (Bray et al., 2019). In particular, the predicted accuracy for the green flags was 91% and for the yellow flags was 100% in our study compared to 62% in the top 10% most likely to succeed and 92% in the top 10% most likely to fail in Bray et al. (2019), respectively. These results were consistent with the purpose of C-BARQ which was intended to capture potentially problematic behaviours (Hsu and Serpell, 2003; Duffy and Serpell, 2008).

Harvey et al. (2017) collected data on 12-month-old dogs utilising the refined Puppy Training Supervisor Questionnaire (PTSQ) to predict the training outcome of guide dogs. A flagging system based on a composite z-value was proposed for the prediction of training outcomes.

The predictive value for the green flag was 85%, while our C-BARQ and MCPQ-R models' green flags achieved 100% and 75%, respectively. The sensitivity for the green flag was 8.4% which was much smaller than the precision achieved by the green flag of our C-BARQ and MCPQ-R models of 94% and 90%, respectively. The predictive value for the red flag was 83%, while our C-BARQ and MCPQ-R models' yellow flags achieved 91% and 90%, respectively. The sensitivity for the red flag was 8.4% which was much smaller than the sensitivity achieved by the yellow flag of our C-BARQ and MCPQ-R models of 36% and 33%, respectively.

4.5. Recommendations

More research is needed to demonstrate the total replacement of C-BARQ in favour of MCPQ-R at an earlier stage than ten weeks. Hence, researchers are encouraged to collect MCPQ-R data at an earlier stage, create training outcome prediction models, and compare the performance of such a model with the C-BARQ model. Dog trainers could be invited to complete the MCPQ-R within the first month of training and puppy raisers the month before the dogs start training. It is advisable to provide puppy raisers with a description of the typical behaviours related to each of the adjectives in the MCPQ-R. Finally, researchers should also investigate the use of this personality rating for predicting training outcomes on other types of working dogs.

5. Conclusion

The most popular standardised dog behaviour rating reported in the literature for predicting training outcomes in assistance dogs is the C-BARQ. Its main downside is that it contains 100 context-based items, which makes it unsuitable for dog trainers to answer due to the number and nature of the questions. The goal of this study was to address this limitation by investigating the application of a shorter adjective-based rating; thus, the MCPQ-R was used for the first time for this same

purpose.

Two models were built separately using data collected from the same cohort using the C-BARQ (completed by puppy raisers at the start of formal training) and the MCPQ-R (completed by dog trainers at around ten weeks of training). Performance metrics of the two classifiers demonstrated that the MCPQ-R achieved similar results to the C-BARQ in the same cohort of trainee assistance dogs.

Hence, this is the first study to reveal a relationship between MCPQ-R scores and training outcomes. Furthermore, the MCPQ-R was validated for predicting training outcomes in assistance dogs as the performance results were comparable to the widely used C-BARQ. The novel C-BARQ model created outperformed all models presented in the literature by employing advanced machine learning techniques for the first time in this area. Conclusively, this study significantly contributes to the state-of-the-art by providing the most effective classifiers for predicting training outcomes in trainee assistance dogs using data from two standardised dog behaviour and personality questionnaires.

Funding

This publication has emanated from research supported in part by a grant from the European Regional Development Fund (ERDF) through the Ireland-Wales INTERREG Programme under the Celtic Advanced

Life Science Innovation Network (CALIN) project [grant number 80885]; from Science Foundation Ireland (SFI) [grant number 12/RC/2289-P2] which is co-funded under the European Regional Development Fund; from SFI and the Irish Department of Agriculture, Food and Marine [grant number 16/RC/3835].

CRediT authorship contribution statement

Marinara Marcato: Conceptualisation, Methodology, Writing – original draft, Writing – review & editing, Software, Formal analysis, Investigation, Data curation, Project Administration. **Salvatore Tedesco:** Formal analysis, Writing – Review & editing, Supervision. **Conor O'Mahony:** Writing – review & editing, Supervision. **Brendan O'Flynn:** Writing – review & editing, Supervision. **Paul Galvin:** Conceptualisation, Funding acquisition, Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: This research is related to the PhD of Marinara Marcato, which is part of collaborations with the Irish Guide Dogs for the Blind.

Appendix A. Appendix

A.1. Puppy raiser questionnaire (C-BARQ)

This questionnaire was developed, standardised and validated by [Serpell and Hsu \(2001\)](#), [Hsu and Serpell \(2003\)](#), [Duffy and Serpell \(2012\)](#). The rating scales are divided into sections and all items had an NA option which was used to indicate that the respondent had never observed the dog in the situation described.

Detailed description of the C-BARQ factors and the original items used to calculate them.

Factor	Description	Items
F_01 Trainability	Dog shows a willingness to attend to the owner and obey simple commands. Dog is not easily distracted, tends to be a fast learner, responds positively to correction, and will fetch or retrieve objects	1, 2, 3, 4, 5, 6, 7, 8
F_02 Stranger-directed aggression	Dog shows threatening or aggressive responses to strangers approaching or invading the dog's or the owner's personal space, territory, or home range	10, 11, 12, 15, 16, 18, 20, 21, 22, 28
F_03 Owner-directed aggression	Dog shows threatening or aggressive responses to the owner or other members of the household when challenged, manhandled, stared at, stepped over, or when approached while in possession of food or objects	9, 13, 14, 17, 19, 25, 30, 31
F_04 Dog rivalry	Dog shows aggressive or threatening responses to other familiar dogs in the same household	32, 33, 34, 35
F_05 Stranger-directed fear	Dog shows fearful or wary responses when approached directly by strangers	36, 37, 39, 40
F_06 Nonsocial fear	Dog shows fearful or wary responses to sudden or loud noises (e.g. thunder), traffic, and unfamiliar objects and situations	38, 41, 42, 44, 47, 48
F_07 Dog-directed aggression	Dog shows threatening or aggressive responses when approached directly by unfamiliar dogs	23, 24, 26, 29
F_08 Dog-directed fear	Dog shows fearful or wary responses when approached directly by unfamiliar dogs	45, 46, 52, 53
F_09 Touch sensitivity	Dog shows fearful or wary responses to potentially painful or uncomfortable procedures, including bathing, grooming, nail clipping, and veterinary examinations	43, 49, 50, 51
F_10 Separation-related behaviour	Dog vocalises and/or is destructive when separated from the owner, often accompanied or preceded by behavioural and autonomic signs of anxiety including restlessness, loss of appetite, trembling, and excessive salivation	54, 55, 56, 57, 58, 59, 60, 61
F_11 Excitability	Dog displays strong reaction to potentially exciting or arousing events, such as going for walks or car trips, doorbells, the arrival of visitors, and the owner arriving home; has difficulty calming down after such events	62, 63, 64, 65, 66, 67
F_12 Attachment and attention-seeking	Dog maintains close proximity to the owner or other members of the household, solicits affection or attention, and displays agitation when the owner gives attention to third parties	68, 69, 70, 71, 72, 73
F_13 Chasing	Dog chases cats, birds, and/or other small animals, given the opportunity	27, 74, 75, 76
F_14 Energy level	Dog is energetic, "always on the go", and/or playful	91, 92

Table A.2

Univariate logistic regression model using the C-BARQ to predict training outcomes (fail) including the estimate, p-value, Odd Ratio (OR) and Confidence Interval (CI).

Variable	Estimate	p-value	OR	2.5%	97.5%
* I_01 When off the leash returns immediately when called	− 1.12	0.01	0.33	0.12	0.76
I_02 Obeys the sit command immediately	− 0.39	0.45	0.68	0.24	1.95
I_03 Obeys the stay command immediately	0.11	0.81	1.11	0.46	2.91
I_04 Seems to attend or listen closely to everything you say or do	− 0.12	0.83	0.89	0.29	2.81
^b I_05 Slow to respond to correction or punishment thick-skinned	− 0.08	0.83	0.93	0.48	1.98
I_06 Slow to learn new tricks or tasks	0.17	0.66	1.18	0.60	2.79
I_07 Easily distracted by interesting sights sounds or smells	− 0.78	0.06	0.46	0.19	1.02

(continued on next page)

Table A.2 (continued)

Variable	Estimate	p-value	OR	2.5%	97.5%
I_08 Will fetch or attempt to fetch sticks balls or objects	− 0.06	0.85	0.94	0.52	1.69
^{a,b} I_09 When verbally corrected or punished scolded shouted at etc by you or a household member	0.35	− 4.44	1.42		
^a I_10 When approached directly by an unfamiliar adult while being walked on a leash	0.33	− 4.68	1.39		
I_11 When approached directly by an unfamiliar child while being walked on a leash	18.21	0.99			
I_12 Toward unfamiliar persons approaching the dog while she is in your car at the gas station for example	1.09	0.22	2.97	0.47	22.66
^a I_13 When toys bones or other objects are taken away by a household member	0.33	− 4.68	1.39		
I_14 When bathed or groomed by a household member	− 15.03	1.00			
I_15 When an unfamiliar person approaches you or another member of your family at home	9.11	0.99			
I_16 When unfamiliar persons approach you or another member of your family away from your home	18.21	0.99			
^a I_17 When approached directly by a household member while she the dog is eating	0.33	− 4.68	1.39		
I_18 When mailmen or other delivery workers approach your home	0.38	0.54	1.47	0.35	4.80
I_19 When his or her food is taken away by a household member	− 14.98	1.00			
I_20 When strangers walk past your home while your dog is outside or in the yard	0.03	0.97	1.03	0.16	4.12
I_21 When an unfamiliar person tries to touch or pet the dog	18.21	0.99			
I_22 When joggers cyclists rollerbladers or skateboarders pass your home	0.86	0.29	2.37	0.38	13.30
I_23 When approached directly by an unfamiliar male dog while being walked on a leash	1.01	0.12	2.75	0.82	12.94
I_24 When approached directly by an unfamiliar female dog while being walked on a leash	0.83	0.20	2.30	0.62	11.63
I_25 When stared at directly by a member of the household	6.07	0.99	433.36	0.00	
I_26 Toward unfamiliar dogs visiting your home	1.31	0.01	3.70	1.51	12.95
* I_27 Toward cats, squirrels or other small animals entering your yard	1.00	0.04	2.73	1.15	8.14
I_28 Toward unfamiliar persons visiting your home	1.63	0.26	5.10	0.19	136.44
I_29 When barked growled or lunged at by another unfamiliar dog	1.07	0.01	2.92	1.43	8.11
^a I_30 When stepped over by a member of the household	0.33	− 4.68	1.39		
I_31 When you or a household member retrieves food or objects stolen by the dog	− 14.98	0.99			
^b I_32 Towards another familiar dog in your household NA if no other dogs	1.93	0.13	6.90	1.21	172.19
^b I_33 When approached at a favourite resting or sleeping place by another familiar household dog NA if no other dogs	− 15.05	0.99			
^b I_34 When approached while eating by another familiar household dog NA if no other dogs	− 16.30	1.00			
^b I_35 When approached while playing with or chewing a favourite toy bone object etc by another familiar household dog NA if no other dogs	0.93	0.20	2.54	0.67	18.11
I_36 When approached directly by an unfamiliar adult while away from your home	0.18	0.88	1.20	0.06	9.26
^a I_37 When approached directly by an unfamiliar child while away from your home	0.33	− 4.61	1.39		
I_38 In response to sudden or loud noises eg vacuum cleaner car backfire road drills objects being dropped etc	1.03	0.03	2.80	1.25	8.20
I_39 When unfamiliar persons visit your home	1.61	0.27	5.00	0.19	133.78
I_40 When an unfamiliar person tries to touch or pet the dog	18.32	0.99			
I_41 In heavy traffic	1.12	0.06	3.08	0.94	10.64
I_42 In response to strange or unfamiliar objects on or near the sidewalk eg plastic trash bags leaves litter flags flapping etc	0.68	0.33	1.97	0.43	7.59
I_43 When examined/treated by a veterinarian	0.91	0.05	2.47	1.06	7.04
I_44 During thunderstorms firework displays or similar events	0.23	0.69	1.26	0.32	3.76
I_45 When approached directly by an unfamiliar dog of the same or larger size	0.83	0.03	2.30	1.16	5.70
I_46 When approached directly by an unfamiliar dog of a smaller size	0.70	0.08	2.01	0.93	5.01
I_47 When first exposed to unfamiliar situations eg first car trip first time in elevator first visit to veterinarian etc	0.92	0.08	2.52	0.88	7.33
I_48 In response to wind or windblown objects	1.06	0.08	2.88	1.07	11.19
^b I_49 When having nails clipped by a household member	1.29	0.10	3.62	0.63	19.43
I_50 When groomed or bathed by a household member	0.47	0.22	1.60	0.71	3.52
I_51 When having his or her feet towed by a member of the household	− 14.96	0.99			
^b I_52 When unfamiliar dogs visit your home	1.39	0.06	4.03	1.28	22.55
I_53 When barked growled or lunged at by an unfamiliar dog	0.70	0.02	2.02	1.13	4.00
I_54 Shaking shivering or trembling	1.51	0.05	4.53	1.14	30.34
I_55 Excessive salivation	1.14	0.12	3.12	0.67	14.31
I_56 Restlessness agitation pacing	0.81	0.03	2.25	1.05	4.91
I_57 Whining	0.79	0.03	2.20	1.07	4.74
I_58 Howling	0.84	0.10	2.33	0.82	6.52
I_59 Chewing or scratching at doors floor windows curtains etc	0.32	0.32	1.38	0.70	2.58
I_60 Loss of appetite	1.16	0.19	3.20	0.50	24.82
I_61 Barking	0.66	0.03	1.93	1.07	3.60
I_62 When you or other members of the household come home after a brief absence	1.04	0.00	2.84	1.49	6.20
I_63 When playing with you or other members of your household	0.88	0.01	2.40	1.27	5.11
I_64 When doorbell rings	0.75	0.02	2.12	1.18	4.11
I_65 Just before being taken for a walk	0.73	0.01	2.07	1.19	3.85
I_66 Just before being taken on a car trip	0.64	0.03	1.90	1.06	3.53
I_67 When visitors arrive at your home	0.46	0.09	1.58	0.92	2.76
I_68 Displays a strong attachment for one particular member of the household	0.40	0.16	1.50	0.88	2.79
I_69 Tends to follow you or other members of household about the house from room to room	0.54	0.16	1.72	0.85	3.87
I_70 Tends to sit close to or in contact with you or others when you are sitting down	0.24	0.52	1.27	0.61	2.80
I_71 Tends to nudge nuzzle or paw you or others for attention when you are sitting down	0.41	0.17	1.51	0.85	2.83
I_72 Becomes agitated whines jumps up tries to intervene when you or others show affection for another person	0.46	0.15	1.58	0.83	2.98
I_73 Becomes agitated whines jumps up tries to intervene when you show affection for another dog or animal	0.34	0.27	1.41	0.75	2.60
^b I_74 Chases or would chase cats given the opportunity	0.15	0.57	1.16	0.69	1.92
* I_75 Chases or would chase birds given the opportunity	0.61	0.03	1.84	1.09	3.24
^b I_76 Chases or would chase squirrels rabbits and other small animals given the opportunity	0.32	0.23	1.38	0.82	2.40
^{a,b} I_77 Escapes or would escape from home or yard given the chance	1.05	0.01	2.86	1.44	6.84
I_78 Rolls in animal droppings or other smelly substances	− 0.04	0.90	0.96	0.52	1.69
I_79 Eats own or other animals droppings or faeces	− 0.36	0.31	0.70	0.30	1.29
I_80 Chews inappropriate objects	0.37	0.24	1.44	0.78	2.72
I_81 Mounts objects furniture or people	0.30	0.57	1.35	0.42	3.64

(continued on next page)

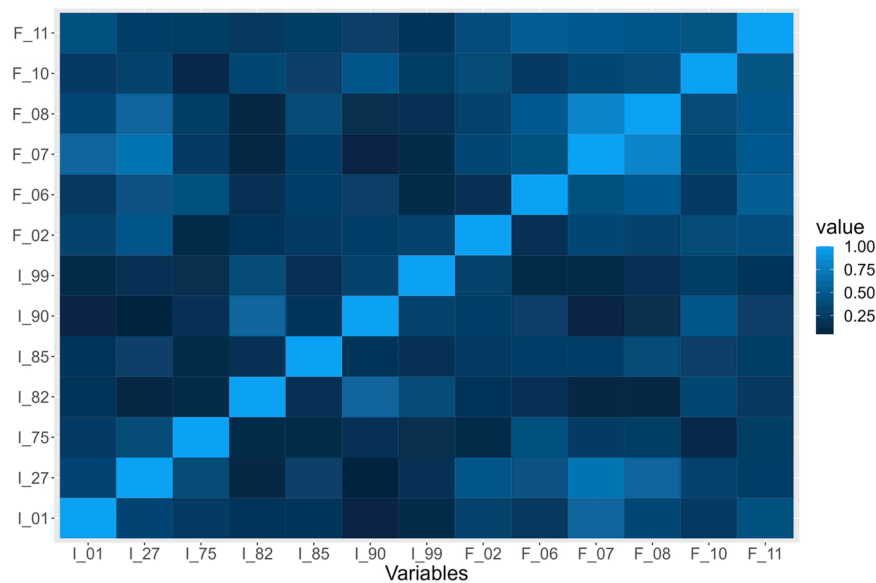


Fig. A.1. Correlation matrix of variables considered for the C-BARQ reduced feature set.

Table A.2 (continued)

Variable	Estimate	p-value	OR	2.5%	97.5%
* I_82 Bogs persistently for food when people are eating	0.68	0.03	1.98	1.10	3.80
I_83 Steals food	0.04	0.89	1.04	0.54	1.79
I_84 Nervous or frightened on stairs	0.07	0.90	1.07	0.23	2.69
* I_85 Pulls excessively hard when on the leash	0.82	0.04	2.27	1.07	5.31
I_86 Urinates against objects furnishings in your home	− 3.76	0.99			
I_87 Urinates when approached petted handled or picked up	− 3.76	0.99			
I_88 Urinates when left alone at night or during the daytime	− 14.88	0.99			
I_89 Defecates when left alone at night or during the daytime	− 14.88	0.99			
* I_90 Hyperactive restless has trouble settling down	0.87	0.04	2.39	1.03	5.86
I_91 Playful puppyish boisterous	0.66	0.07	1.93	0.98	4.10
I_92 Active energetic always on the go	0.57	0.08	1.77	0.95	3.51
* ^b I_93 Stares intently at nothing visible	1.03	0.05	2.80	1.14	9.28
^b I_94 Snaps at invisible flies	0.41	0.31	1.51	0.60	3.57
I_95 Chases own tail or hind end	0.38	0.29	1.46	0.72	3.04
^b I_96 Chases or follows shadows light spots etc	0.43	0.12	1.54	0.87	2.71
I_97 Barks persistently when alarmed or excited	0.26	0.50	1.30	0.55	2.75
I_98 Licks him or herself excessively	0.80	0.13	2.24	0.75	6.59
* I_99 Licks people or objects excessively	0.84	0.03	2.32	1.08	5.22
^b I_100 Displays other bizarre strange or repetitive behaviours	0.08	0.85	1.09	0.28	2.31
F_01 Trainability	− 1.03	0.21	0.36	0.07	1.80
* F_02 Stranger directed aggression	5.74	0.03	311.63	1.94	78300.96
F_03 Owner directed aggression	4.85	0.23	127.29	0.02	1356011.42
^b F_04 Dog rivalry	1.25	0.23	3.49	0.37	33.64
F_05 Stranger directed fear	3.35	0.15	28.60	0.27	5492.23
* F_06 Nonsocial fear	2.51	0.01	12.32	2.18	95.69
* F_07 Dog directed aggression	1.58	0.02	4.84	1.51	21.91
* F_08 Dog directed fear	0.96	0.03	2.61	1.19	7.54
F_09 Touch sensitivity	0.85	0.43	2.35	0.21	18.47
* F_10 Separation related problems	2.85	0.00	17.24	3.08	135.53
* F_11 Excitability	1.25	0.00	3.51	1.60	9.03
F_12 Attachment attention seeking	1.00	0.08	2.72	0.96	9.27
^b F_13 Chasing	0.62	0.07	1.87	0.96	3.82
F_14 Energy	0.75	0.05	2.11	1.03	4.73

95% Confidence Interval (CI): lower 2.5% and upper 97.5%. ^a No variance, constant values. ^b More than 5% missing data (4 data points). * Statistically significant with $p < 0.05$. Variables that satisfied criteria 1 and 2.

Table A.3

Variables considered for the reduced feature set and the number of missing values.

Variables	NAs
I_01 When off the leash returns immediately when called	0
I_27 Toward cats squirrels or other small animals entering your yard	2
I_75 Chases or would chase birds given the opportunity	1
I_77 Escapes or would escape from home or yard given the chance	6
I_82 Bogs persistently for food when people are eating	1
I_85 Pulls excessively hard when on the leash	1

(continued on next page)

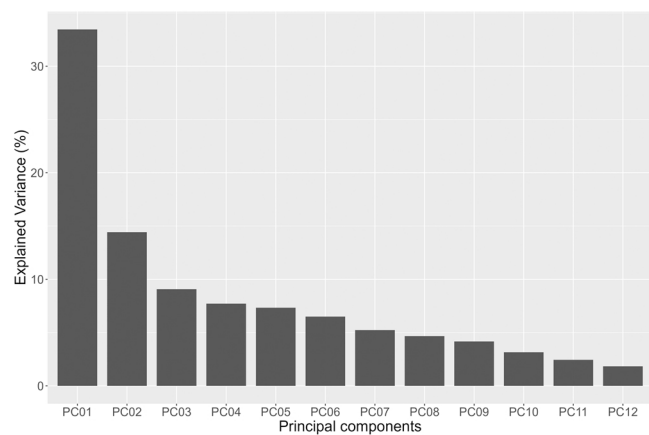


Fig. A.2. Explained variance per principal component (Scree Plot).

Table A.3 (continued)

Variables	NAs
I_90 Hyperactive restless has trouble settling down	1
I_93 Stares intently at nothing visible	7
I_99 Licks people or objects excessively	3
F_02 Stranger directed aggression	0
F_06 Nonsocial fear	0
F_07 Dog directed aggression	0
F_08 Dog directed fear	0
F_10 Separation related behaviour	1
F_11 Excitability	0
Outcome	0

Figures A.1, A.2, A.3.

A.2. Dog trainer questionnaire (MCPQ-R)

Table A.4

Univariate logistic regression model using the MCPQ-R to predict training outcomes (fail) including the estimate, p-value, Odd Ratio (OR) and Confidence Interval (CI).

Variables	Estimate	p-value	OR	2.5%	97.5%
Extraversion Active	− 0.22	0.38	0.80	0.48	1.31
Extraversion Energetic	− 0.14	0.54	0.87	0.55	1.35
Extraversion Excitable	0.17	0.45	1.18	0.77	1.82
Extraversion Hyperactive	0.37	0.07	1.45	0.97	2.18

(continued on next page)

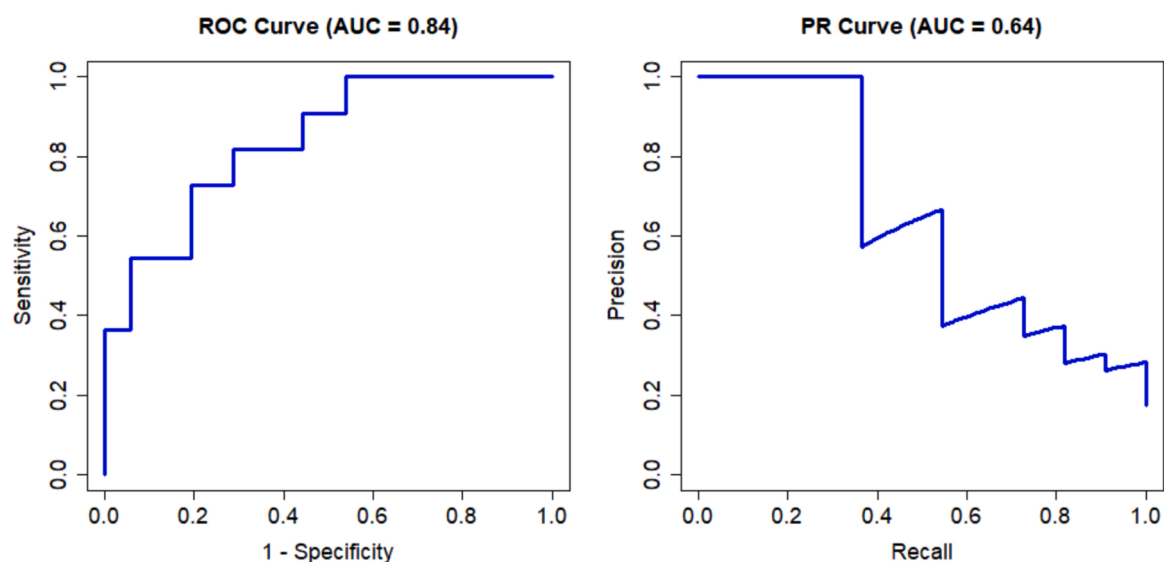


Fig. A.3. ROC and PR curves for the best logistic regression model using C-BARQ data selected by RFE.

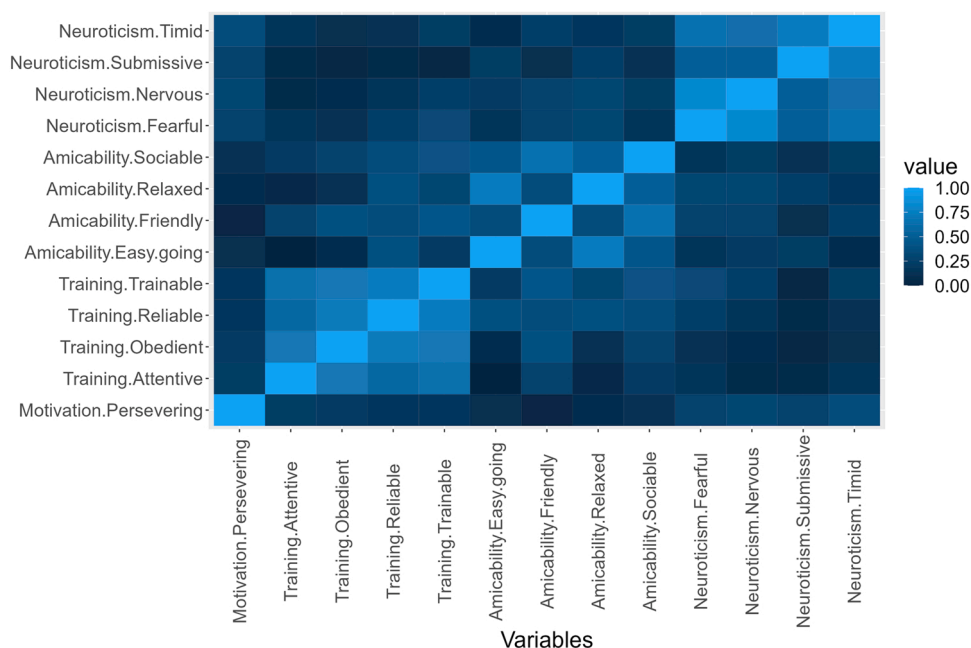


Fig. A.4. Correlation Matrix for variables considered for the MCPQ-R reduced feature set.

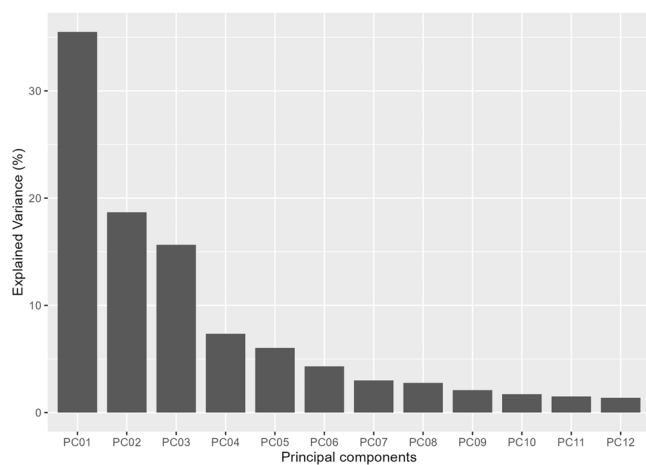


Fig. A.5. Explained variance per principal component (Scree Plot) for the MCPQ-R reduced feature set.

Table A.4 (continued)

Variables	Estimate	p-value	OR	2.5%	97.5%
Extraversion Lively	− 0.01	0.98	0.99	0.64	1.52
Extraversion Restless	0.18	0.35	1.19	0.82	1.74
Motivation Assertive	− 0.10	0.72	0.90	0.50	1.58
Motivation Determined	− 0.48	0.10	0.62	0.34	1.08
Motivation Independent	− 0.32	0.20	0.73	0.44	1.17
* Motivation Persevering	− 0.59	0.05	0.55	0.30	0.97
Motivation Tenacious	− 0.19	0.45	0.82	0.49	1.35
* Training Attentive	− 0.56	0.03	0.57	0.34	0.93
Training Biddable	− 0.49	0.07	0.62	0.36	1.02
Training Intelligent	− 0.39	0.13	0.68	0.41	1.12
* Training Obedient	− 0.63	0.02	0.53	0.31	0.89
* Training Reliable	− 1.09	0.00	0.34	0.17	0.60
* Training Trainable	− 1.09	0.00	0.34	0.18	0.57
* Amicability Easy going	− 0.84	0.00	0.43	0.24	0.70
* Amicability Friendly	− 0.76	0.00	0.47	0.27	0.76
Amicability Non aggressive	− 0.15	0.43	0.86	0.61	1.27
* Amicability Relaxed	− 0.73	0.00	0.48	0.29	0.74
* Amicability Sociable	− 0.64	0.01	0.53	0.31	0.84
* Neuroticism Fearful	0.74	0.00	2.10	1.37	3.39

(continued on next page)

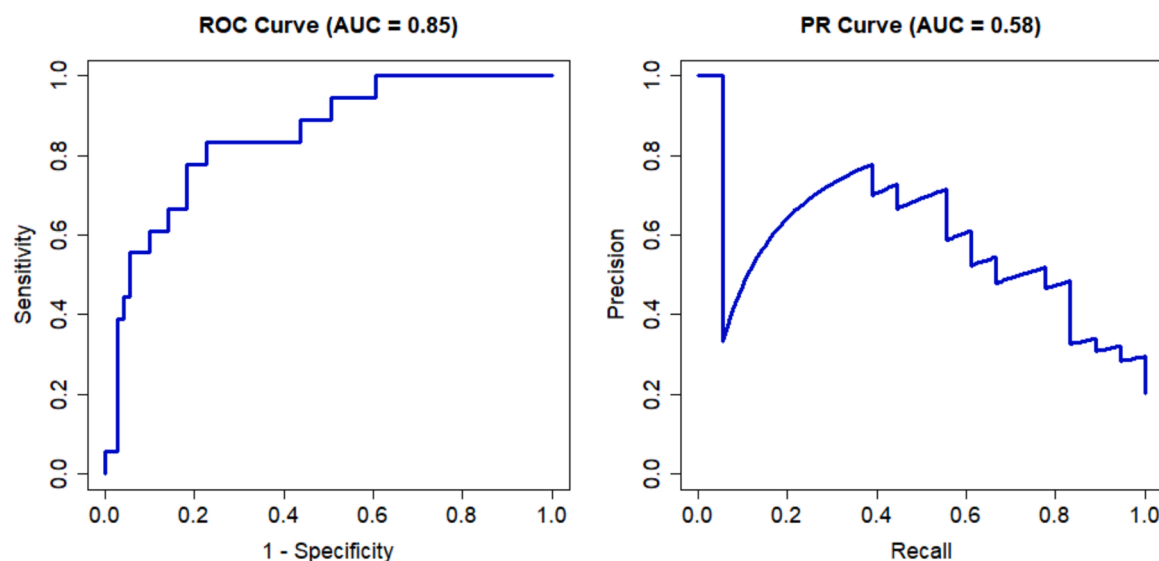


Fig. A.6. ROC and PR curves for the best logistic regression model using MCPQ-R data selected by RFE.

Table A.4 (continued)

Variables	Estimate	p-value	OR	2.5%	97.5%
* Neuroticism Nervous	0.65	0.00	1.91	1.28	2.98
* Neuroticism Submissive	0.64	0.03	1.90	1.09	3.44
* Neuroticism Timid	0.71	0.01	2.03	1.25	3.48

95% Confidence Interval (CI): lower 2.5% and upper 97.5%. * Statistically significant with $p < 0.05$. Variables that satisfied criteria 1 and 2.

Figures A.4, A.5, A.6.

REFERENCES

- Asher, L., Blythe, S., Roberts, R., Tothill, L., Craigon, P.J., Evans, K.M., Green, M.J., England, G.C., 2013. A standardized behavior test for potential guide dog puppies: methods and association with subsequent success in guide dog training. *J. Vet. Behav.* 8, 431–438. <https://doi.org/10.1016/j.jveb.2013.08.004>.
- Barnard, S., Marshall-Pescini, S., Passalacqua, C., Beghelli, V., Capra, A., Normando, S., Pelosi, A., Valsecchi, P., 2016. Does subjective rating reflect behavioural coding? personality in 2 month-old dog puppies: an open-field test and adjective-based questionnaire. *PLOS ONE* 11, e0149831. <https://doi.org/10.1371/journal.pone.0149831>.
- Batt, L.S., Batt, M.S., Baguley, J.A., McGreevy, P.D., 2009. The value of puppy raisers' assessments of potential guide dogs' behavioral tendencies and ability to graduate. *Anthrozoös* 22, 71–76. <https://doi.org/10.2752/175303708x390482>.
- Bray, E.E., Levy, K.M., Kennedy, B.S., Duffy, D.L., Serpell, J.A., MacLean, E.L., 2019. Predictive models of assistance dog training outcomes using the canine behavioral assessment and research questionnaire and a standardized temperament evaluation. *Front. Vet. Sci.* 6, 49. <https://doi.org/10.3389/fvets.2019.00049>.
- Cleghorn, Z., Gruen, G., Roberts, D., 2018. Using Decision Tree Learning as an Interpretable Model for Predicting Candidate Dog Success.
- Duffy, D.L., Serpell, J.A., 2008. Behavioral assessment of guide and service dogs. *J. Vet. Behav.* 3, 186–188. <https://doi.org/10.1016/j.jveb.2007.12.010>.
- 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>.
- Harvey, N.D., Craigon, P.J., Blythe, S.A., England, G.C.W., Asher, L., 2017. An evidence-based decision assistance model for predicting training outcome in juvenile guide dogs. *PLOS ONE* 12, e0174261. <https://doi.org/10.1371/journal.pone.0174261>.
- Hosmer, D., Lemeshow, S., Sturdivant, R., 2013. *Applied Logistic Regression*. Wiley Series in Probability and Statistics. Wiley.
- 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>.
- Hunt, R.L., England, G.C.W., Asher, L., Whiteside, H., Harvey, N.D., 2020. Concurrent and predictive criterion validity of a puppy behaviour questionnaire for predicting training outcome in juvenile guide dogs. *Animals* 10, 2382. <https://doi.org/10.3390/ani10122382>.
- Ley, J.M., Bennett, P., Coleman, G., 2008. Personality dimensions that emerge in companion canines. *Appl. Anim. Behav. Sci.* 110, 305–317. <https://doi.org/10.1016/j.applanim.2007.04.016>.
- 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>.
- Marcato, M., Kenny, J., O'Riordan, R., O'Mahony, C., O'Flynn, B., Galvin, P., 2022. Assistance dog selection and performance assessment methods using behavioural and physiological tools and devices. *Appl. Anim. Behav. Sci.* 254, 105691. <https://doi.org/10.1016/j.applanim.2022.105691>.
- Serpell, J.A., Hsu, Y., 2001. Development and validation of a novel method for evaluating behavior and temperament in guide dogs. *Appl. Anim. Behav. Sci.* 72, 347–364. [https://doi.org/10.1016/s0168-1591\(00\)00210-0](https://doi.org/10.1016/s0168-1591(00)00210-0).
- Team, R.C., 2022.R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.



Marinara Marcato is a PhD candidate at Tyndall National Institute. She has been working on a collaborative project entitled Guide Dog 4.0 with Swansea University and the Irish Guide Dogs for the Blind (IGDB). Her goal is to develop performance assessment techniques for guide and assistance dogs through the deployment of technology. Her research interests lie in the areas of artificial intelligence and wearable devices. Prior to joining Tyndall, Marinara earned her B.E. Hons Electrical Engineering degree from Londrina State University.



Dr. Salvatore Tedesco (Member, IEEE) is a Senior Researcher and a Team Leader at Tyndall National Institute, Cork, Ireland. He is responsible for leading a research team in the wearable and data analytics area focused on his main research interests on wearable technologies for healthcare and well-being, human motion analysis in sports and clinical populations, digital health, physiological monitoring, signal processing, edge analytics, and machine learning.



Dr. Conor O'Mahony (M. Eng. Sc '01, PhD '04) is a Principal Scientist with the Tyndall National Institute, Cork, Ireland. His team is focused on the development of Micro Transdermal Interface Platforms (MicroTIPs) - high-value, wearable systems that with applications in transdermal drug delivery, diagnostics and intelligent woundcare. He has published over sixty peer-reviewed journal articles and is a Senior Member of the IEEE.



Dr. Paul Galvin is Head of the ICT for Health Strategic Programmes, Head of the Life Sciences Interface Group and Head of the Bioelectronics Cluster at the Tyndall National Institute, where he drives an "ABC ecosystem" through collaborations with leading academic researchers, global leaders in the med-tech and pharmaceutical industries, and with clinical experts. He has published research articles in leading international peer-reviewed journals across a range of scientific disciplines.



Dr. Brendan O'Flynn is Head of Group for Wireless Sensor Networks with the Tyndall National Institute, Cork, Ireland. His group is focused on the development of smart, autonomous, and reconfigurable wireless sensing systems and the creation of information from data.