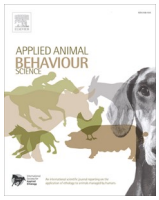
Applied Animal Behaviour Science 273 (2024) 106237

Contents lists available at ScienceDirect

Applied Animal Behaviour Science

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



Mind your language! Lessons from the application of an English published version of a Japanese horse personality instrument to a French population

Claire Ricci-Bonot [a,*], Teresa Romero [a,b], Christine Nicol [a,c], Daniel S. Mills [a]

[a] Animal Behaviour, Cognition and Welfare Group, Dept of Life Sciences, University of Lincoln, Lincolnshire LN6 7DL, United Kingdom
[b] Department of Psychology, University of Portsmouth, Hampshire PO1 2DY, United Kingdom
[c] The Royal Veterinary College, Hawkshead Lane North Mymms Hatfield, Hertfordshire AL9 7TA, United Kingdom

ARTICLE INFO

Keywords:
Character
Equus caballus
Personality
Questionnaire
Reliability
Temperament

ABSTRACT

Replicability is a fundamental tenet of the scientific method and scientific reporting, but there is a preponderance to publish scientific research in English to increase international recognition, regardless of the country of origin of the research. Questionnaires are widely used to assess personality in animals. These psychometric instruments are mainly published in English but can be used all over the world in other languages. However, without safeguards relating to the translation process, the replicability of the quality of the instrument may change from its originally reported value. This study focuses on the particular issue of cross-cultural reliability of psychometric instruments used for assessing animals that have been translated from their original context. We examined the replicability of the structure of a personality scale originally used in Japanese (but reported in the English literature) on an English population (n=100), and then the reliability of the structure of a French translation of the English version with additional translational safeguards (e.g. back-translation and sense checking) on a French population (n = 159 horses). Horses were rated by 3 evaluators to also allow calculation of inter-rater reliability. We found that there was greater reliability and similarity of structure between the adapted English translation and French version of the Japanese scale, than with the originally published structure of the instrument used in Japan. These results highlight the importance of never assuming the reliability and thus validity of semantic instruments used to assess animal behaviour which have been published in a different language to that in which they were originally developed.

1. Introduction

In the animal literature, the terms temperament and personality are often used interchangeably although it has been suggested the two should not be considered synonyms, but there appears to be little consensus over their distinction (Brady et al., 2018). One suggestion is that temperament should be used to refer to the behavioural tendencies resulting from individual biological differences present in early life, which are stable across time and situations (Bates, 1987; 1989; Lansade et al., 2008); by contrast personality may be considered what is evident as a result of the impact of environmental stimuli on the temperament (von Borstel, 2013; Finkemeier et al., 2018). However, for others (e.g. Brady et al., 2018), the latter is referred to as character, with personality referring to general biologically-based traits underpinning individual differences and temperament limited to a subset of these traits related to affect. Regardless of the terminology used, assessment of relatively stable behavioural traits, can help to improve the quality of life of captive animals such as zoo, farm, laboratory and working animals, as well as pets (see reviews by Mills, 1998; Powell and Gartner, 2011; Sebastian et al., 2011; Watters and Powell, 2012; Wilson et al., 2019).

Two methods are widely used to do this: behavioural tests and questionnaires (Manteca and Deag, 1993; Gosling, 2001). They both make it possible to assess individual differences but provide different information. Behavioural tests assess the reaction of individuals (behavioural and physiological responses) in various situations such as exposure to novelty, opening of an umbrella and animal-human encounter (cattle: Van Reenen et al., 2004; Graunke et al., 2013; goat: Lyons et al., 1988; Nawroth et al., 2017; horse: Lansade and Bouissou, 2008; Lansade et al., 2008; sheep: Beausoleil et al., 2012). However, this method only allows evaluation of the individual's reaction at a specific time, in situations that is often not checked for reliability and can sometimes be contextually quite artificial (Patronek and Bradley, 2016).

* Corresponding author.
E-mail address: cRicciBonot@lincoln.ac.uk (C. Ricci-Bonot).

https://doi.org/10.1016/j.applanim.2024.106237
Received 20 October 2023; Received in revised form 12 March 2024; Accepted 15 March 2024
Available online 18 March 2024

Accordingly, it is essential to examine the consistency of responses over time in order to assess the reliability of the behavioural response (Taylor and Mills, 2006) as well as consistency between assessors and institutions (Szabó et al., 2017). These issues of reliability also apply to questionnaires, but the latter have the advantage that they allow the rapid, relatively inexpensive gathering of data for a very large number of individuals. Questionnaires have the potential to use assessors who have been able to observe the individual over the long term and in various situations, to report on their observation of those aspects of behaviour which seem relatively stable, and so should be less affected by temporary states such as the current mental or physical state of the individual. Thus, they may be preferable to behaviour tests, if they can be shown to be valid. However, a potential problem with this methodology is that assessment can be more subjective as it might just reflect the impression of the evaluator; although the use of several assessors makes it possible to address this to a greater or lesser extent (Mills, 1998). It is also important when creating a psychometric instrument to ensure that terminology is standardised to avoid individual differences in interpretation, which may affect the validity of the instrument. To this end each term should be clearly and precisely defined to ensure that each assessor has a common understanding of the different terms (Mills, 2010). Another potential problem with psychometric instruments is that while they may be relatively simple to produce, good face validity is potentially used to imply good psychometric validity; this implies a high degree of reliability, without this being specifically assessed. Reliability refers to the consistency of the measurement procedure (John and Soto, 2007). In horses there are many psychometric instruments with a varying scientific basis used to assess individual traits, but few are associated with any form of reliability assessment (Visser et al., 2003; Momozawa et al., 2005; Lloyd et al., 2007). Where reliability has been assessed this is often in the form of the specific calculation of inter-rater reliability values (Visser et al., 2003; Lloyd et al., 2007); however, Momozawa and colleagues (2005) inferred construct reliability from the internal consistency of the factor structure between two groups of horses, without assessing the specific reliability of different raters. This normative approach implies that different items tend *on average,* to associate or not in a repeatable way and may, to some extent, mask problems with the reliability of individual items. It is also important to appreciate that both linguistic and cultural factors may influence the interpretation of translated items used in another country, and so cross-cultural validity including item reliability should always be evaluated and never assumed in such situations (Savalli et al., 2019).

In the current study we began by evaluating the personality of a group of French horses starting with a published English translation of an instrument developed in Japanese by Momozawa et al. (2005) as it had been shown to have a reliable structure by the original authors in Japan. However, it soon became apparent that this instrument may not be reliable in terms of factor structure when compared with the results obtained with our French population (the same items did not always load positively or negatively in both the Japanese and French versions). We realised we had made a number of assumptions concerning the reliability and validity of the published English version of the scale used before translating it into French. Accordingly, we decided to investigate the issue of cross-cultural reliability of psychometric instruments (for assessing animals) that have been translated from their original language. This appears to be an area that has been largely neglected (although see Savalli et al., 2019, for an example of best practice in this regard), but is particularly importance given the preponderance to publish scientific work in English regardless of the country of origin of the work. Specifically, we examined the reliability of the structure of a horse personality scale originally used in Japanese, (but reported in the English literature) on English and French populations. We then assessed both inter-rater reliability and the reliability of the structure of a French translation of the English version having applied additional safeguards relating to the translation process.

2. Material and methods

This research was approved by the University of Lincoln Research Ethics Committee (CosREC433).

2.1. Animals

This study used ponies and horses of different breeds from France and England. For the French population, 159 ponies and horses, aged 2–26 years (average age: 10.99 years, standard deviation = 4.86); 3 stallions, 71 geldings and 85 mares from 10 equestrian centres were evaluated (see the electronic supplementary material Table S1). For the English population, 100 equids came from Redwings horse sanctuary in the UK, 43 geldings and 57 females, aged 2–16 years (average age: 8.01 years, standard deviation = 3.13).

2.2. Psychometric instrument

Assessments were performed first on the French population from July 2018 to July 2019. The personality of the 159 horses was assessed by using a French adapted version of the originally published English translation of the Japanese psychometric instrument of Momozawa et al. (2005) with 20 items (see Table 1 and the electronic supplementary material Table S2). The text was first translated from English into French and then a bilingual post-doctoral researcher in ethology (French native) back-translated it from French to English to establish that its meaning had not been changed in translation. This adapted English version was

Table 1
Psychometric instrument used to assess horse personality (Adapted from Momozawa et al., 2005; Note: * only the definition of this item has been changed from the one used by Momozawa et al., 2005).

Items	Description (This horse tends to…)	1	9
Nervousness	become nervous about insects, noises, etc.	Calm	Nervous
Concentration	be trainable and undisturbed by the environment	Poor	Excellent
Self-reliance	be at ease if left alone away from the herd	Restless	At ease
Trainability	be trained easily and promptly	Poor	Excellent
Excitability	get excited easily	Not excitable	Excitable
Friendliness toward people	be never aggressive or fearful	Unfriendly	Friendly
Curiosity	be interested in novel objects and approach them	Rarely	Frequently
Memory	memorize what it learned or was trained	Poor	Excellent
Panic	get excited to an abnormal extent	Never	Frequently
Cooperation	be cooperative with a caretaker when handled	Never	Always
Inconsistent emotionality	be unpredictable from day to day	Consistent	Inconsistent
Stubbornness	be obstinate once it resists a command	Obedient	Stubborn
Docility	be docile in general	Active	Docile
Vigilance	be vigilant about surroundings	Never	Always
Perseverance	be patient with various stimuli	Impatient	Patient
Friendliness toward horses	interact with other horses in a friendly manner	Unfriendly	Friendly
Competitiveness	be dominant in antagonistic encounters with other horses	Subordinate	Dominant
Skittishness	get surprised easily	Not skittish	Skittish
Timidity	be timid in a novel environment	Audacious	Timid
Gate entrance*	easily enters in a riding arena, trailer, etc.	Rarely	Always

then compared to the original English published version of the instrument by another researcher (English native) for consistency of meaning.

The last item from the original instrument of Momozawa et al. (2005) referred to "gate entrance" (defined as "go easily through the starting gate") and because it related specifically to racehorses it was modified to match our population (as done previously by Dai et al. 2015 for similar reasons) by changing the definition ("easily enters a riding arena, trailer, etc."). The evaluators used a scale from 1 to 9 to score intensity of items on a semantic differential scale describing the horse. A score of 5 was indicated as the average value for defining a typical horse for that trait.

Given the issues outlined in the introduction, in order to address our questions concerning where sources of difference may lie, we needed to generate a dataset from our adapted English translation version of the instrument (see Table 1) after we had had undertaken the French study. The English dataset was generated by staff at the Redwings horse sanctuary from the evaluation of 100 horses between September 2021 and December 2021.

2.3. Inter-rater agreement

In order to explore inter-rater reliability, it was requested that each horse was evaluated by three people familiar with the animal (riding instructors, caretakers or horse owners) who should complete the same psychometric instrument independently. These 33 people were not the same for each equestrian centre (3 people per centre).

2.4. Data analysis

All the analyses were performed with R software (version 3.5.2).

The analyses were run on 98 horses for the English population and 157 equids for the French population. Incomplete responses were excluded.

2.4.1. Inter-rater agreement

In order to determine inter-rater reliability, the level of agreement across assessors was calculated for each item for each equestrian centre that had completed responses for more than 10 horses using Kendall's coefficients of concordance (W, with the R-package *irr*).

2.4.2. Assessment of reliability of the structure of the psychometric instrument, compared to its original source

In order to assess the reliability of the structure reported for the original psychometric instrument, a principal components analysis (PCA) was performed separately on the responses for both the English and French populations. For these analyses, the mean value of each item was used, based on the 3 scores given by each evaluator. The same mathematical procedure as used by Momozawa et al. (2005) was used, i.e. an orthogonal rotation (varimax) (within the R-package *psych*) and 5 principal components (PCs) extracted as per Momozawa et al. (2005). Components with loading values above +0.4 or below −0.4 were considered to belong to a given PC.

3. Results

3.1. Agreement between evaluators on specific items

For the English population, all items had at least fair agreement ($W > 0.3$) with some showing substantial concordance ($W > 0.6$) (see the electronic supplementary material Table S3). The item about inconsistent emotionality had least agreement between the evaluators, ($W = 0.373$), following by the items about curiosity ($W = 0.469$) and gate entrance ($W = 0.491$) The highest concordance among the evaluators was for the item about perseverance ($W = 0.682$).

For the 157 French horses, the evaluators were at least moderately/substantially concordant with one another for all personality traits ($W > 0.59$, in all instances and most > 0.7) (see the electronic supplementary material Table S4). The item about memory had least agreement between the evaluators, with a coefficient of concordance of 0.594, while the highest concordance among the evaluators was for the item about stubbornness ($W = 0.786$).

For the French equestrian centres with more than 10 horses, the concordance between the evaluators for the different traits was moderately high (see the electronic supplementary material Table S3). In Equestrian Centre A, the question about vigilance had least agreement between the evaluators ($W = 0.551$), with a similar pattern for Equestrian Centre B ($W = 0.454$); however in Equestrian Centre F, the evaluators were in less agreement for the items about skittishness ($W = 0.341$) and gate entrance ($W = 0.381$). For Equestrian Centre J, the evaluators were at least moderately concordant with one another for all personality traits (see the electronic supplementary material Table S3).

3.2. Assessment of reliability of the structure of the psychometric instrument, compared to its original source

For the English population, the first five factors of the PCA explained 82% of the total variance (see the electronic supplementary material Table S5). The first factor, accounting for 27% of the total variance, loaded positively for the following items: "concentration", "self-reliance", "trainability", "memory", "cooperation", "perseverance" and "gate entrance", and loaded negatively for the following items: "excitability", "panic", "inconsistent emotionality", "stubbornness" and "skittishness". This factor was labelled "Compliance". The second factor, explaining 24% of the total variance, loaded positively on the items "excitability", "panic", "inconsistent emotionality", "vigilance" and "skittishness", and negatively on the items "concentration", "self-reliance", "cooperation", "docility", "perseverance". This factor was labelled "Reactivity". The third factor, accounting for 17% of total variance, loaded positively on the items "friendliness toward people" and "curiosity", and negatively on the items "nervousness" and "timidity". This factor was labelled "Boldness". The fourth factor, explaining 10% of total variance, loaded positively on the item "friendliness toward horses" and negatively on the item "competitiveness", and was labelled "Sociability towards other horses". The fifth factor, accounting for 0.4% of variance, loaded negatively for the item "self-reliance" and was therefore labelled "Separation anxiety" (see Table 2).

For the French population, the first five factors from the PCA explained 69% of the total variance for the French population (see the electronic supplementary material Table S6). The first factor, explaining 23% of the total variance, loaded positively on the following items: "nervousness", "excitability", "panic", "inconsistent emotionality", "vigilance" and "timidity", and loaded negatively on the following items: "concentration", "self-reliance" and "perseverance". Given its similarity to the first factor extracted from the English population it was also labelled "Reactivity". The second factor, accounting for 18% of total variance, loaded positively on the items "concentration", "self-reliance", "trainability", "memory", "cooperation", "docility", "perseverance", "gate entrance" and negatively on the items "inconsistent emotionality" and "stubbornness". This factor was accordingly labelled "Compliance" in line with the English factor. The third factor, explaining 13% of total variance, loaded positively on the items "friendliness toward people", "curiosity", "cooperation", "docility" and "friendliness toward horses", and was therefore labelled "General sociability". The fourth factor, accounting for 8% of total variance, loaded positively on the items "curiosity" and "competitiveness", with a negative loading on "timidity". This factor was labelled "Competitiveness". The fifth factor, accounting for 6% of total variance, loaded positively for the item "skittishness" and was therefore labelled "Skittishness" (see Table 2).

These results were then compared with the published results of Momozawa et al., (2005) (Table 2).

For the first factor, the following items were common across studies:

Table 2
PCA factor loadings for each item for the different population after varimax rotation (Note: loadings with absolute value more than 0.4 are highlighted in bold; * = data from the article of Momozawa et al., 2005 where they distributed their Japanese instrument to a Japanese population in 2002 and 2003; English refers to our adapted English translation version of the instrument, distributed to an English population; French corresponds to the French version of the instrument distributed to a French population).

Items	PC1				PC2				PC3				PC4				PC5			
	Japanese (2002)*	Japanese (2003)*	French	English	Japanese (2002)*	Japanese (2003)*	French	English	Japanese (2002)*	Japanese (2003)*	French	English	Japanese (2002)*	Japanese (2003)*	French	English	Japanese (2002)*	Japanese (2003)*	French	English
Nervousness	**0.90**	**0.86**	**0.87**	- 0.23	- 0.06	- 0.26	- 0.17	0.23	- 0.03	- 0.03	- 0.14	**- 0.74**	- 0.05	0.15	- 0.14	0.03	- 0.11	0.25	0.00	0.36
Concentration	0.04	- 0.38	**- 0.58**	**0.70**	**0.82**	**0.77**	**0.58**	**- 0.48**	0.11	- 0.02	0.02	0.19	- 0.02	- 0.13	- 0.05	0.15	0.04	- 0.03	- 0.07	- 0.15
Self-reliance	- 0.29	**- 0.47**	**- 0.56**	**0.53**	**0.50**	0.25	**0.40**	**- 0.54**	0.25	0.01	0.01	0.16	0.02	0.25	- 0.08	0.13	0.30	**- 0.51**	- 0.17	**- 0.43**
Trainability	- 0.17	- 0.06	- 0.31	**0.82**	**0.81**	**0.82**	**0.80**	- 0.35	- 0.07	0.18	0.07	0.22	0.31	- 0.11	0.13	0.13	0.11	0.09	- 0.03	- 0.10
Excitability	**0.85**	**0.78**	**0.87**	**- 0.49**	- 0.09	- 0.29	- 0.08	**0.74**	0.28	- 0.09	- 0.15	0.11	- 0.02	0.24	0.08	- 0.18	0.00	0.18	0.04	0.19
Friendliness toward people	0.04	0.05	- 0.16	0.15	0.25	0.20	0.12	- 0.28	0.11	**0.86**	**0.85**	**0.78**	**0.82**	- 0.13	- 0.02	0.26	- 0.09	0.04	- 0.09	0.21
Curiosity	- 0.01	**- 0.70**	0.09	0.22	0.11	- 0.37	0.10	0.11	**0.81**	0.10	**0.54**	**0.83**	- 0.05	0.09	**0.59**	- 0.09	0.02	0.09	- 0.23	0.18
Memory	- 0.16	- 0.10	- 0.05	**0.78**	**0.75**	**0.75**	**0.67**	- 0.24	0.01	0.32	0.23	0.36	**0.40**	0.09	0.25	0.13	0.10	0.14	0.00	- 0.05
Panic	**0.87**	**0.86**	**0.83**	**- 0.53**	- 0.22	- 0.24	- 0.17	**0.54**	0.18	- 0.12	- 0.13	- 0.33	0.03	0.20	- 0.19	- 0.11	- 0.04	0.15	0.08	0.37
Cooperation	- 0.09	- 0.04	- 0.24	**0.78**	0.24	0.32	**0.46**	**- 0.51**	0.11	**0.79**	**0.60**	0.08	**0.71**	- 0.03	0.16	0.15	0.33	- 0.28	0.23	0.03
Inconsistent emotionality	**0.75**	**0.66**	**0.58**	**- 0.44**	- 0.16	- 0.11	**- 0.41**	**0.73**	0.17	- 0.22	- 0.18	- 0.23	- 0.17	**0.47**	0.02	0.02	0.01	- 0.14	0.01	- 0.25
Stubbornness	**0.50**	0.39	0.14	**- 0.84**	- 0.20	- 0.38	**- 0.76**	0.21	0.39	- 0.32	- 0.17	0.03	**- 0.45**	**0.51**	0.32	- 0.21	- 0.07	0.21	0.20	- 0.05
Docility	**- 0.53**	- 0.22	- 0.28	0.37	0.37	**0.67**	**0.60**	**- 0.71**	- 0.27	**0.43**	**0.48**	- 0.05	**0.41**	- 0.25	- 0.02	0.31	0.04	- 0.06	0.10	- 0.13
Vigilance	**0.87**	**0.86**	**0.73**	- 0.09	0.09	- 0.07	- 0.09	**0.88**	- 0.16	- 0.13	0.15	- 0.18	- 0.05	0.02	0.09	- 0.18	0.10	0.05	- 0.37	0.01
Perseverance	**- 0.42**	- 0.26	**- 0.60**	**0.53**	**0.65**	**0.63**	**0.44**	**- 0.63**	**- 0.46**	0.35	0.31	- 0.04	0.07	- 0.34	- 0.08	0.31	- 0.15	- 0.11	0.07	- 0.03
Friendliness toward horses	- 0.06	- 0.30	- 0.06	0.33	0.18	0.18	0.11	- 0.19	**0.54**	**0.60**	**0.77**	0.12	**0.48**	- 0.09	- 0.24	**0.86**	0.01	**0.47**	- 0.14	0.02
Competitiveness	0.16	- 0.07	0.10	- 0.11	0.02	- 0.15	0.01	0.25	0.39	- 0.07	- 0.32	0.21	**- 0.62**	**0.83**	**0.74**	**- 0.87**	- 0.08	0.01	0.12	0.06
Skittishness	0.82	**0.89**	0.04	**- 0.43**	- 0.04	- 0.28	- 0.06	**0.65**	- 0.05	0.10	- 0.11	- 0.37	- 0.09	- 0.08	0.00	- 0.16	- 0.04	0.05	**0.91**	0.18
Timidity	0.63	**0.80**	**0.49**	- 0.05	0.16	- 0.11	- 0.22	0.05	- 0.12	- 0.04	- 0.08	**- 0.86**	- 0.04	- 0.15	**- 0.57**	0.19	0.00	0.21	0.10	0.22
Gate entrance	- 0.14	- 0.34	- 0.24	**0.75**	0.08	- 0.20	**0.58**	- 0.14	- 0.02	0.11	- 0.02	0.37	0.16	- 0.19	0.33	0.08	**0.94**	**- 0.77**	0.06	- 0.09

"excitability", "panic", "inconsistent emotionality" and "vigilance". For the second factor, the common items were the following: "concentration", "trainability", "memory" and "perseverance". For the three remaining factors there is no consistency for more than one item in a factor.

4. Discussion

The aim of this research was to examine the reliability of the structure of personality in horses revealed using a semantic scale in different languages, and to consider the issues arising for future scientific reference. The original source instrument was created in Japan for a Japanese population but reported in an English language journal meaning that the original English translation version of the personality scale, presented in the article of Momozawa et al. (2005), had never undergone the quality control procedures associated with the translation of a psychometric instrument into another language. This raises a potential problem, since some people may assume that the published English version proposed in the journal is as valid as the Japanese one. This is clearly not the case. There was a greater similarity in the content of the PCs between the adapted English translation and French versions which underwent the important quality control procedures of comprehension analysis and back translation, than the Japanese version of the instrument reported in English in the article (original English version).

The Japanese version was used on two occasions (2002 and 2003), so it is possible to examine the reliability of the structure of personality revealed by the original instrument when used in its native language (Momozawa et al., 2005) and the English translation presented by them, but used by us in the UK with its very minor adaptation (see methods above).

In the research of Momozawa et al. (2005) the first PC, named "Anxiety", was very similar between the two versions (2002 and 2003) with seven common items loading positively in each of the PC analyses: 'nervousness', 'excitability', 'panic', 'inconsistent emotionality', 'vigilance', 'skittishness' and 'timidity'. The item 'stubbornness' loaded positively only in the 2002 version. However, the important items which loaded negatively are much more variable across the two surveys. For the one in 2002, 'docility' and 'perseverance' have important negative loadings, whereas in 2003 it was 'self-reliance' and 'curiosity', highlighting an inconsistency in the first PC even within the original Japanese versions. The first factor from the Japanese versions seems to most closely correspond to the second factor in the English version, labelled "Reactivity". The adapted English translation version, like the Japanese ones, had positive loadings for the following items: 'excitability', 'panic', 'inconsistent emotionality', 'vigilance' and 'skittishness'; it also had negative loadings for both 'docility' and 'perseverance' as in the 2002 Japanese version, and 'self-reliance' as in the 2003 Japanese version, but also 'concentration' and 'cooperation', which did not feature in the related PC for either Japanese version.

For the second PC in the Japanese versions, labelled "Trainability", there were positive loadings for the following items: 'concentration', 'trainability', 'memory' and 'perseverance', with the addition of 'self-reliance' in the 2002 version and 'docility' in the 2003 Japanese version. The corresponding English principal component appears to be PC1, which loaded positively for 'concentration', 'trainability', 'memory' and 'perseverance' as in the two Japanese versions, as well as 'self-reliance' as in the 2002 Japanese version, and 'cooperation' and 'gate entrance'; in addition the English component had negative loadings for 'excitability', 'panic', 'inconsistent emotionality', 'stubbornness' and 'skittishness', and so was labelled "Compliance", rather than the "Trainability" implied in the Japanese versions. This might suggest cultural differences in the perception of how behavioural predispositions cluster to form a recognizable personality trait. This is supported by the previous finding that individual differences in the definition of specific equine personality traits are common (Mills, 1998). Indeed, the social and cultural environment will affect the development of the personality trait and its perception in other people (Benet-Martínez and Oishi, 2008). People from the same culture are more likely to share the same beliefs, values and identities (Benet-Martínez and Oishi, 2008). For example, immorality is associated with disgust in the United States, New Zealand and Australia but not in Latin American countries (Scherer, 1997). Moreover, the same word may not have the same meaning in two different cultures. For example, depression (and *yuutsu* in Japanese) does not have the same meaning in terms of symptoms and feelings between Japanese-Americans and Japanese-nationals (Tanaka-Matsumi and Marsella, 1976).

For the three other PCs in the Japanese versions, there is little consistency between the 2002 and 2003 versions. In 2002, the third PC loaded positively for 'curiosity' and 'friendliness toward horses', and loaded negatively for 'perseverance', whereas in 2003, it loaded positively for 'friendliness toward people', 'cooperation', 'docility' and 'friendliness toward horses'. Thus, in 2002 this seems to correspond more to curiosity about horses and their sociability, whereas in 2003, this seems to describe compliance and sociability more broadly, i.e. toward horses and people. The third factor of the adapted English translation version was labelled "boldness" and loaded positively for the items 'curiosity' as in the 2002 Japanese version, and loaded negatively for the following items: 'nervousness' and 'timidity' and 'friendliness toward people' as in the 2003 Japanese version. The fourth PC, in the Japanese study of 2002, loaded positively on 'friendliness toward people', 'memory', 'cooperation', 'docility' and 'friendliness toward horses', and loaded negatively for the item's 'stubbornness' and 'competitiveness'. However, in 2003, it loaded positively for 'inconsistent emotionality', 'stubbornness' and 'competitiveness'. Thus, there seems to be a shift in the PC from describing compliant and sociable horses, to competitive and stubborn horses. The fourth PC of the adapted English translation version, termed "Sociability towards other horses", loaded positively for the item: 'friendliness toward horses', and loaded negatively for the item: 'competitiveness', as in the 2002 Japanese study.

For Momozawa and colleagues (2005), the factor four from the 2002 version corresponds to the third factor from 2003, with the following items in common: 'friendliness toward people', 'cooperation', 'docility', and 'friendliness toward horses'. For them, it corresponds to their third factor and they labelled it "Affability". They mentioned that the third factor from 2002 and the fourth factor from 2003 were not extracted in the other version. The adapted English translation version supports the idea of a sociability dimension, but its exact composition and construction from multiple items seems uncertain, and it might be argued that there is little advantage of using multiple over a single one asking about sociability with horses and a separate one for sociability towards people. We thus question the construct validity claim by Momozawa and colleagues (2005) for their third dimension (corresponding to factor 4 in the 2002 version and factor 3 of the 2003 version), which appears to be based largely on assessing the internal consistency of the common items for each factors using Cronbach's coefficients.

For the fifth factor, there is no consistency for more than one item between the Japanese versions and the adapted English translation one, and so further dimensions are not considered further.

If we now consider the results of the PCA for the French and adapted English translation versions which were subject to greater quality control procedures, to determine what effect these might have on the latent structure of personality revealed. The first two PCs are more similar between these two versions, than between the Japanese versions reported in English (first English translation). For the first PC, named "Reactivity", the following items loaded positively on both versions: 'excitability', 'panic', 'inconsistent emotionality' and 'vigilance' and the following negatively: 'concentration', 'self-reliance' and 'perseverance'. In the French version, the items 'nervousness' and 'timidity' also loaded positively. For the adapted English translation one, the item 'skittishness' also loaded positively, with 'cooperation' and 'docility' which loaded negatively. The items 'excitability', 'panic', 'inconsistent emotionality' and 'vigilance' group together in a single component in all

four versions of the instrument (Japanese 2002 and 2003, French, adapted English translation), suggesting a negative emotional lability seems to be a very consistent feature in horses. This is perhaps equivalent to the neuroticism dimension that has been reported previously in horses, e.g. Morris et al., (2002a); McGrogan et al., (2008); Ijichi et al., (2014) and many other species (Gosling and John, 1999).

Likewise there was extensive (but not complete) consistency between the French and adapted English translation versions for a second PC, labelled "Compliance", with 'concentration', 'self-reliance', 'trainability', 'memory', 'cooperation', 'perseverance' and 'gate entrance' loading positively, and 'inconsistent emotionality' and 'stubbornness' loading negatively. In the French version, the item 'docility' also loaded positively and for the adapted English translation version the following others items loaded negatively: 'excitability', 'panic' and 'skittishness'. The following items are common among all four versions within a single PC: 'concentration', 'trainability', 'memory' and 'perseverance'. This seems to reflect a collection of traits associated with effective learning, but whether this reflects a biological relatedness or a perceived association between these traits remains unknown, but deserves empirical experimental investigation. Although studies have shown the link between personality (specifically fear reactivity) and learning (Lansade et al., 2013); temperament seemed to be an important aspect in horse training (Heird et al., 1986; Christensen et al., 2005). Researchers found that emotivity is linked to trainability and learning ability (Hausberger et al., 2004; Christensen et al., 2012). However, there is no consensus about the definition of learning ability in horses. No study has shown a link between trained discipline and personality components (Sackman and Houpt, 2019).

As with the Japanese surveys, there was little agreement in the construction of further components, with only the items 'friendliness toward people' and 'curiosity' being common to the French and adapted English translation versions. For the two remaining PCs there was no consistency beyond a single item, suggesting that PCA was not useful beyond this point for identifying reliable latent structure, despite some of them having Eigen values >1.

There are several potential sources of inconsistency in structure for different language versions of a psychometric instrument, which should always be considered. First, it could relate to the actual meaning of the terms translated. In the current study, the original instrument was developed by a team of Japanese researchers and was distributed to a Japanese population. It was then translated into English (original English translation) for publication in their article (Momozawa et al., 2005). It is not mentioned in the article if there were any reliability checks between the Japanese and the original English translation of the instrument e.g. through back translation. It is therefore possible that in the process of translation from Japanese to English the meaning of certain items and their explanation have changed. This can be assumed to amplify inconsistencies in the structure which were already apparent in the two Japanese studies. By contrast, we adopted a robust procedure for the adaptation of the originally published English translation version of the instrument (albeit with a minor modification) into French, including back translation (Behr, 2018). This appears to have resulted in greater reliability and thus scientific replicability across countries (note we cannot comment on the validity of the instrument without further studies). This translation work was carried out by English and French native speakers, all researchers in the field of animal behaviour. It is worth appreciating that whilst we are focusing on the different languages used, and the risk of assuming a translation is valid, there is also a cultural dimension to the semantics used in a psychometric instrument that could result in variation in different populations even within a single language (Jackson et al., 2019). This may relate to differences in the use of language by different socio- or ethnic groups within a country (Harper and Jackson, 2018), by different geographic regions sharing a common primary language e.g. UK, North America and Australasia (Gilbert et al., 2019 – take the meaning of the word "pants" as an obvious example) as well as differences in language meaning across time (Kozlowski et al., 2019; Ducarme and Couvet, 2020).

From a further scientific perspective, it should be noted that the sample sizes used in relation to the Japanese instrument was relatively small (69 in 2002, and 70 in 2003). Whilst PCA can be used on any sized population, the risk of artefacts is increased with small populations and those with a poor ratio of items to subjects (Kline, 2014). As a rule of thumb, it is often recommended to have a minimum of 100 subjects when performing a PCA (Budaev, 2010). Accordingly it is not surprising that the robustness of the original PCs was demonstrably not strong.

The potential impact of variation in the types of horses used in the countries, needs to be considered. The two groups of horses used in Japan were 2 year old Thoroughbreds, belonging to the Hidaka Yearling Training Farm, associated with the Japan Racing Association. The English population of horses were 8 years old on average of different breeds and came from a horse rescue sanctuary in the UK whereas the French population came from 10 different institutions (equestrian centres, riding school, private yard, show jumping yard and driving horses), were of various breeds and 11 years old on average. Thus they might be expected to have very different personalities. However this should be revealed by differences in score not structure, since the aim of a psychometric instrument (such as that described here), should be to define a consistent structure to personality. As such reliability is an essential prerequisite to validity (Taylor and Mills, 2006). The structure should be common if the instrument defines something biologically meaningful rather than cultural traits (Mills, 1998). Although some researchers have recently suggested that personality structure could be affected by the diversity of social and ecological niches, i.e. the fit of a population to a specific environment (Smaldino et al., 2019), there is no research to validate this idea on domestic horses, to our knowledge. Moreover, the breeds of horses used in our research and the ones of Momozawa et al. (2005) are not specific to the country.

Another explanation could relate to the evaluation process employed. Each horse was evaluated by three caretakers in Momozawa's study (2005). This represents 34 caretakers in 2002 and 31 in 2003, of which 24 had already evaluated horses from the first group (2002). In their article, Momozawa et al. (2005) do not provide any information on the concordance of scores between evaluators, in particular, they do not report that they performed any test of inter-rater reliability. As the answers to an instrument correspond to the perception of an individual on a horse (subjectivity), the presence of 24 of the same evaluators in 2002 and 2003 could have introduced a bias in the apparent consistency of the instrument, on top of an already potentially very variable procedure. In order to ensure that this questionnaire was reliable they could have performed an inter-rater reliability analysis (Visser et al., 2003; Lloyd et al., 2007; Jolivald et al., 2022). This would have revealed whether the three assessors agreed on the different personality traits being assessed. In fact, the perception of a horse's personality could be affected by the familiarity of the rater with the horse, the context in which the rater observes the horse and the way the rater judges (e.g. they may be biased by previous interactions with the horse) (Funder et al., 1995; Funder, 1999; Gosling, 2001; Morris et al., 2002b). Lloyd and colleagues (2007) found that one of their three raters was not in accordance with the others and that individual only observed the horse when it was at pasture whereas the other two regularly manipulated the horse. In our research each horse from the English and French population was evaluated by 3 people (riding instructors, caretakers or owner of the horse). These individuals were not the same for each French equestrian centre (a total of 30 people). The inter-rater reliability assessment highlighted that the evaluators had difficulties in assessing some items, but this did not refer to the same items in the adapted English or French version. Indeed, the evaluators from the English population had difficulty in assessing the following items: 'curiosity', 'inconsistent emotionality' and 'gate entrance' whereas for the French population it was the following items: 'vigilance', 'skittishness' and 'gate entrance'. These items should therefore be removed from the instrument.

5. Conclusions

In conclusion, our study not only highlights the importance of quality checks when developing a psychometric instrument, but also the danger of assuming its scientific quality when it is sed in a different language and possibly culture. Quality control measures such as comprehension analysis and back translation quality control, are essential. The introduction of these in the current study may explain the greater similarity in the content of the PCs between the adapted English translation and French versions used by us than the original Japanese version.

CRediT authorship contribution statement

Claire Ricci-Bonot: Conceptualization, Data curation, Formal analysis, Methodology, Visualization, Writing – original draft, Writing – review & editing, Validation. **Teresa Romero:** Conceptualization, Formal analysis, Methodology, Visualization, Writing – review & editing, Validation. **Christine Nicol:** Conceptualization, Formal analysis, Methodology, Validation, Visualization, Writing – review & editing. **Daniel Simon Mills:** Conceptualization, Formal analysis, Methodology, Visualization, Writing – review & editing, Validation.

Declaration of Competing Interest

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

Acknowledgements

Thank you to the Redwings Horse Sanctuary and all the people who took the time to complete our questionnaires. I also would like to say thank you to Dr. Camille Troisi who help us in the translation of our questionnaire to French and Lucia Ripari for her help on the project.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.applanim.2024.106237.

References

Bates, J.E., 1987. Temperament in infancy. In: Osofsky, J.D. (Ed.), Handbook of Infant Development. John Wiley & Sons, Oxford, England, pp. 1101–1149.
Bates, J.E., 1989. Concepts and measures of temperament. In: Kohnstamm, G.A., Bates, J. E., Rothbart, M.K. (Eds.), Temperament in Childhood. Wiley, New York, pp. 3–26.
Beausoleil, N.J., Blache, D., Stafford, K.J., Mellor, D.J., Noble, A.D.L., 2012. Selection for temperament in sheep: Domain-general and context-specific traits. Appl. Anim. Behav. Sci. 139 (1-2), 74–85. https://doi.org/10.1016/j.applanim.2012.02.020.
Behr, D., 2018. Translating questionnaires for cross-national surveys: a description of a genre and its particularities based on the ISO 17100 categorization of translator competences. Transl. Interpret. 10 (2), 5–20. https://doi.org/10.12807/ti.110202.2018.a02.
Benet-Martínez, V., Oishi, S., 2008. Culture and personality. In: John, O.P., Robins, R.W., Pervin, L.A. (Eds.), Handbook of Personality: Theory and Research. The Guilford Press, New York, pp. 542–567.
von Borstel, U.K., 2013. Assessing and influencing personality for improvement of animal welfare: a review of equine studies. CAB Rev. 8, 1–27. https://doi.org/10.1079/PAVSNNR20138006.
Brady, K., Cracknell, N., Zulch, H., Mills, D.S., 2018. A Systematic review of the reliability and validity of behavioural tests used to assess behavioural characteristics important in working dogs. Front. Vet. Sci. 5, 103. https://doi.org/10.3389/fvets.2018.00103.
Budaev, S., 2010. Using principal components and factor analysis in animal behaviour research: caveats and guidelines. Ethology 116, 472–480. https://doi.org/10.1111/j.1439-0310.2010.01758.x.
Christensen, J.W., Keeling, L.J., Lindstrom, N.B., 2005. Response of horses to novel visual, olfactory and auditory stimuli. Appl. Anim. Behav. Sci. 93 (1-2), 53–65. https://doi.org/10.1016/j.applanim.2005.06.017.
Christensen, J.W., Ahrendt, L.P., Lintrup, R., Gaillard, C., Palme, R., Malmkvist, J., 2012. Does learning performance in horses relate to fearfulness, baseline stress hormone, and social rank? Appl. Anim. Behav. Sci. 140 (1-2), 44–52. https://doi.org/10.1016/j.applanim.2012.05.003.
Dai, F., Cogi, N.H., Heinzl, E.U.L., Dalla Costa, E., Canali, E., Minero, M., 2015. Validation of a fear test in sport horses using infrared thermography. J. Vet. Behav. 10 (2), 128–136. https://doi.org/10.1016/j.jveb.2014.12.001.
Ducarme, F., Couvet, D., 2020. What does 'nature'mean? Palgrave Commun. 6 (1), 1–8. https://doi.org/10.1057/s41599-020-0390-y.
Finkemeier, M., Langbein, J., Puppe, B., 2018. Personality research in mammalian farm animals: Concepts, measures, and relationship to welfare. Front. Vet. Sci. 5, 131. https://doi.org/10.3389/fvets.2018.00131.
Funder, D.C., 1999. Personality Judgment: A Realistic Approach to Person Perception. Academic Press, San Diego.
Funder, D.C., Kolar, D.C., Blackman, M.C., 1995. Agreement among judges of personality: interpersonal relations, similarity, and acquaintanceship. J. Personal. Soc. Psychol. 69 (4), 656–672. https://doi.org/10.1037/0022-3514.69.4.656.
Gilbert, P., Basran, J., MacArthur, M., Kirby, J.N., 2019. Differences in the semantics of prosocial words: an exploration of compassion and kindness. Mindfulness 10 (11), 2259–2271. https://doi.org/10.1007/s12671-019-01191-x.
Gosling, S.D., 2001. From mice to men - what can we learn about personality from animal research. Psychol. Bull. 127 (1), 45–86. https://doi.org/10.1037/0033-2909.127.1.45.
Gosling, S.D., John, O.P., 1999. Personality dimensions in nonhuman animals: a cross-species review. Curr. Dir. Psychol. Sci. 8 (3), 69–75. https://doi.org/10.1111/1467-8721.00017.
Graunke, K.L., Nürnberg, G., Repsilber, D., Puppe, B., Langbein, J., 2013. Describing temperament in an ungulate: a multidimensional approach. PLoS One 8, e74579. https://doi.org/10.1371/journal.pone.0074579.
Harper, K.C., Jackson, H., 2018. Dat'Niggas crazy: how hip-hop negotiates mental health. West J. Black Stud. 42 (3-4), 113–124.
Hausberger, M., Bruderer, C., Le Scolan, N., Pierre, J.S., 2004. Interplay between environmental and genetic factors in temperament/personality traits in horses (Equus caballus). J. Comp. Psychol. 118 (4), 434–446. https://doi.org/10.1037/0735-7036.118.4.434.
Heird, J.C., Whitaker, D.D., Bell, R.W., Ramsey, C.B., Lokey, C.E., 1986. The effect of handling at different ages on the subsequent learning ability of 2-year-old horses. Appl. Anim. Behav. Sci. 15 (1), 15–25. https://doi.org/10.1016/0168-1591(86)90018-3.
Ijichi, C., Collins, L.M., Elwood, R.W., 2014. Pain expression is linked to personality in horses. Appl. Anim. Behav. Sci. 152, 38–43. https://doi.org/10.1016/j.applanim.2013.12.007.
Jackson, J.C., Watts, J., Henry, T.R., List, J.M., Forkel, R., Mucha, P.J., Greenhill, S., Gray, R.D., Lindquist, K.A., 2019. Emotion semantics show both cultural variation and universal structure. Science 366 (6472), 1517–1522. https://doi.org/10.1126/science.aaw8160.
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.
Jolivald, A., Yarnell, K., Hall, C., Ijichi, C., 2022. Do you see what I see? Investigating the validity of an equine personality questionnaire. Appl. Anim. Behav. Sci. 248, 105567 https://doi.org/10.1016/j.applanim.2022.105567.
Kline, P., 2014. An easy guide to factor analysis. Routledge, New York.
Kozlowski, A.C., Taddy, M., Evans, J.A., 2019. The geometry of culture: Analyzing the meanings of class through word embeddings. Am. Sociol. Rev. 84 (5), 905–949. https://doi.org/10.1177/0003122419877135.
Lansade, L., Bouissou, M.F., 2008. Reactivity to humans: A temperament trait of horses which is stable across time and situations. Appl. Anim. Behav. Sci. 114 (3-4), 492–508. https://doi.org/10.1016/j.applanim.2008.04.012.
Lansade, L., Bouissou, M.F., Erhard, H.W., 2008. Fearfulness in horses: a temperament trait stable across time and situations. Appl. Anim. Behav. Sci. 115 (3-4), 182–200. https://doi.org/10.1016/j.applanim.2008.06.011.
Lansade, L., Coutureau, E., Marchand, A., Baranger, G., Valenchon, M., Calandreau, L., 2013. Dimensions of temperament modulate cue-controlled behavior: a study on Pavlovian to instrumental transfer in horses (Equus caballus). PLoS One 8, e64853. https://doi.org/10.1371/journal.pone.0064853.
Lloyd, A.S., Martin, J.E., Bornett-Gauci, H.L.I., Wilkinson, R.G., 2007. Evaluation of a novel method of horse personality assessment: Rater-agreement and links to behaviour. Appl. Anim. Behav. Sci. 105 (1-3), 205–222. https://doi.org/10.1016/j.applanim.2006.05.017.
Lyons, D.M., Price, E.O., Moberg, G.P., 1988. Individual differences in temperament of domestic dairy goats: constancy and change. Anim. Behav. 36 (5), 1323–1333. https://doi.org/10.1016/S0003-3472(88)80201-X.
Manteca, X., Deag, J.M., 1993. Individual Differences in Temperament of Domestic Animals: A Review of Methodology. Anim. Welf. 2 (3), 247–268. https://doi.org/10.1017/S0962728600015906.
McGrogan, C., Hutchison, M.D., King, J.E., 2008. Dimensions of horse personality based on owner and trainer supplied personality traits. Appl. Anim. Behav. Sci. 113 (1-3), 206–214. https://doi.org/10.1016/j.applanim.2007.10.006.
Mills, D.S., 1998. Personality and individual differences in the horse, their significance, use and measurement. Equine Vet. J. 30 (S27), 10–13. https://doi.org/10.1111/j.2042-3306.1998.tb05137.x.
Mills, D.S., 2010. Personality. In: Mills, D.S. (Ed.), Encyclopedia of Applied Animal Behaviour. CABI, Wallingford, pp. 463–465.
Momozawa, Y., Kusunose, R., Kikusui, T., Takeuchi, Y., Mori, Y., 2005. Assessment of equine temperament questionnaire by comparing factor structure between two separate surveys. Appl. Anim. Behav. Sci. 92 (1-2), 77–84. https://doi.org/10.1016/j.applanim.2004.11.006.

Morris, P.H., Gale, A., Duffy, K., 2002b. Can judges agree on the personality of horses? Personal. Individ. Differ. 33 (1), 67–81. https://doi.org/10.1016/S0191-8869(01)00136-2.

Morris, P.H., Gale, A., Howe, S., 2002a. The factor structure of horse personality. Anthrozoös 15 (4), 300–322. https://doi.org/10.2752/089279302786992414.

Nawroth, C., Prentice, P.M., McElligott, A.G., 2017. Individual personality differences in goats predict their performance in visual learning and non-associative cognitive tasks. Behav. Process. 134, 43–53. https://doi.org/10.1016/j.beproc.2016.08.001.

Patronek, G.J., Bradley, J., 2016. No better than flipping a coin: Reconsidering canine behavior evaluations in animal shelters. J. Vet. Behav. 15, 66–77. https://doi.org/10.1016/j.jveb.2016.08.001.

Powell, D.M., Gartner, M.C., 2011. Applications of personality to the management and conservation of nonhuman animals. In: From genes to animal behavior. Springer, Tokyo, pp. 185–199.

Savalli, C., Albuquerque, N., Vasconcellos, A.S., Ramos, D., de Mello, F.T., Mills, D.S., 2019. Assessment of emotional predisposition in dogs using PANAS (Positive and Negative Activation Scale) and associated relationships in a sample of dogs from Brazil. Sci. Rep. 9 (1), 1–9. https://doi.org/10.1038/s41598-019-54645-6.

Scherer, K.R., 1997. Profiles of emotion-antecedent appraisal: Testing theoretical predictions across cultures. Cogn. Emot. 11, 113–150. https://doi.org/10.1080/026999397379962.

Sebastian, T., Watts, J., Stookey, J., Buchanan, F., Waldner, C., 2011. Temperament in beef cattle: Methods of measurement and their relationship to production. Can. J. Anim. Sci. 91 (4), 557–565. https://doi.org/10.4141/CJAS2010-041.

Smaldino, P.E., Lukaszewski, A., von Rueden, C., Gurven, M., 2019. Niche diversity can explain cross-cultural differences in personality structure. Nat. Hum. Behav. 3, 1276–1283. https://doi.org/10.1038/s41562-019-0730-3.

Szabó, D., Mills, D.S., Range, F., Virányi, Z., Miklósi, Á., 2017. Is a local sample internationally representative? Reproducibility of four cognitive tests in family dogs across testing sites and breeds. Anim. Cogn. 20 (6), 1019–1033. https://doi.org/10.1007/s10071-017-1133-3.

Tanaka-Matsumi, J., Marsella, A.J., 1976. Cross-cultural variations in the phenomenological experience of depression: I. Word association studies. J. Cross Cult. Psychol. 7 (4), 379–396. https://doi.org/10.1177/002202217674001.

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

Van Reenen, C.G., Engel, B., Ruis-Heutinck, L.F.M., Van der Werf, J.T.N., Buist, W.G., Jones, R.B., Blokhuis, H.J., 2004. Behavioural reactivity of heifer calves in potentially alarming test situations: a multivariate and correlational analysis. Appl. Anim. Behav. Sci. 85 (1-2), 11–30. https://doi.org/10.1016/j.applanim.2003.09.007.

Visser, E.K., Van Reenen, C.G., Rundgren, M., Zetterqvist, M., Morgan, K., Blokhuis, H.J., 2003. Responses of horses in behavioural tests correlate with temperament assessed by riders. Equine Vet. J. 35 (2), 176–183. https://doi.org/10.2746/042516403776114108.

Watters, J.V., Powell, D.M., 2012. Measuring animal personality for use in population management in zoos: suggested methods and rationale. Zoo. Biol. 31 (1), 1–12. https://doi.org/10.1002/zoo.20379.

Wilson, V., Guenther, A., Øverli, Ø., Seltmann, M.W., Altschul, D., 2019. Future directions for personality research: contributing new insights to the understanding of animal behavior. Animals 9 (5), 240. https://doi.org/10.3390/ani9050240.