

How reliable are standardised behaviour tests and are they valid in predicting the suitability for use in police horses?

Rachel Flentje, BSc Hons

Dissertation submitted to the University of Chester for the degree of Master of Science (Research Methods) in part fulfilment of the Modular Program in Research Methods

September 2008

Acknowledgements

I would like to thank everybody who has contributed to the completion of this dissertation project. I would especially like to thank my supervisors Dr Emma Creighton and Catherine Perry for their guidance, support and encouragement during the course of the study. I would also like to thank the staff at Merseyside and Manchester police units for their kind permission in allowing their horses to be used in this study and for their time in completing the questionnaires. I am especially grateful to Kim Bailey and Lynn Maher for handling the horses and to Brian Seddon for sharing his knowledge on police horses, which aided in the questionnaire development. I also appreciate the generous assistance and time given by Sarah Dinning in assisting with the data collection and inter-observer reliability, and to Barbara Arch for her valuable comments on the statistical analysis. I would finally like to thank my family for their support, especially my father who was kind enough to proof read the thesis.

Declaration

The work is original and has not been submitted previously in support of any qualification or programme.

Signed

Word count: 24, 867

Table of contents

Contents	Page Number
Acknowledgements	i
Declaration	ii
Table of contents	iii-v
List of tables	vi-vii
List of figures	viii
Abstract	ix
 Chapter 1: Introduction	
1.1 Background and rationale	1-6
1.2 Aims and objectives	6-7
1.3 Structure of the dissertation	8-9
 Chapter 2: Literature review	
2.1 Police horses and their selection	10-12
2.2 Behaviour tests	12
2.2.1 Reaction to a novel object	12-18
2.2.2 Reaction to a loud/unexpected noise	19-21
2.2.3 Reaction to a sudden rapid and surprising object	21-23
2.2.4 Reaction to pain	23-24
2.2.5 Reaction to an unusual surface	24-27
2.2.6 Reaction to an unusual person	27-29
2.2.7 Reaction to social isolation	30-32
2.3 Predicting the suitability of animals	32-34
2.4 Summary	34-35
 Chapter 3: Methodology	
3.1 Research strategy	36-38
3.2 Ethics (animals)	38-44
3.3 Design of tests	44-45
3.3.1 Social isolation test	45-46
3.3.2 Stranger approach test	46-47
3.3.3 Unexpected noise test: Shaker	47-48
3.3.4 Novel object test: Lampshade	49-50
3.3.5 Sudden fast moving object test: Automatic umbrella opening	50-51
3.3.6 Pain/discomfort test: Skin pinch	51-52
3.3.7 Unusual surface: Large plastic sheet	52-53
 Chapter 4: Reliability assessment of the standardised behaviour tests	
4.1 Subjects and location	54-55
4.2 Materials	55
4.3 Test procedure	55-56
4.3.1 Stage one – social isolation tests	56-57
4.3.2 Stage two – handling tests	57-59

4.4	Analysis methods	59-61
4.5	Results	62-69
4.5.1	Test-retest reliability	62-64
4.5.2	Test for differences between trials	65-67
4.5.3	Inter-observer reliability	68-69
4.6	Discussion of reliability assessment	70-79
4.7	Summary	79-80
Chapter 5: Development of the ‘suitability of use questionnaire’		
5.1	Questionnaire design	81-83
5.1.1	Rating scales	83-84
5.2	Ethics (human)	84
5.3	Piloting	85-86
5.4	Target population/sampling strategy	86
5.5	Procedure	86-87
5.6	Analysis methods	87
5.7	Results of the reliability assessment	87-88
5.8	Development of the suitability of use indices	89-91
Chapter 6: Comparison of police horses’ behavioural reactions and suitability scores		
6.1	Aim of the chapter	92
6.2	Subjects and locations	92-93
6.3	Test procedure	93-94
6.4	Analysis methods	94-95
6.5	Results	95-98
6.5.1	Effects of location on the horses’ behaviour responses	99-100
Chapter 7: Discussion		
7.0	discussion	101-113
7.1	Implications of the study	113
7.2	Future research	113-114
Chapter 8: Conclusion		115-116
References		117-123
Appendices		
Appendix 1	Literature search strategy	124-129
Appendix 2	Evidence of ethical approval for use of animal subjects	130
Appendix 3	Informed consent for the use of police horses in this study	131-132
Appendix 4	Poster and abstract presented at the International Society for Equitation Science and the Universities Federation for Animal Welfare 2008.	133-134
Appendix 5	Coding framework used to record the horses’ behavioural reactions to the behaviour tests	135-137
Appendix 6	Test procedure for equine behaviour tests	138-139

Appendix 7	Suitability of use questionnaire	140-142
Appendix 8	Evidence of ethical approval for the use of human subjects in the study	143-145
Appendix 9	Informed consent for the use of human subjects in the study	146-147
Appendix 10	Covering letter	148
Appendix 11	Participant information sheet	149-151
Appendix 12	Project timetable	152-153
Appendix 13	Rationale of questions used in the suitability for use questionnaire	154-159

List of tables

Table number	Table legend	Page number
2.2.1.1	Summary of 'novel object tests', methods used and the claims authors made as to what their tests measured.	13-14
2.2.3.1	Summary of 'unexpected noise tests', methods used and the claims authors made as to what their tests measured.	20
2.2.3.1	Summary of 'surprising objects tests', methods used and claims authors made as to what their tests measured.	22
2.2.5.1	Summary of 'unusual surfaces tests', methods used and claims the authors made as to what their tests measured.	25
2.2.6.1	Summary of 'stranger approach tests', methods used and claims the authors made as to what their tests measured.	28
2.2.7.1	Summary of 'social isolation tests', methods used and claims the authors made as to what their tests measured.	31
4.2.1	List of equipment required to conduct the standardised behaviour tests on the horses.	55
4.4.1	Interpretation of Kappa values, adapted by Altman (1991).	60
4.4.2	Summary of the number of measures used for each standardised behaviour test and the alpha level achieved from using Bonferroni correction.	61
4.5.1.1	Reliability of responses to standardised behavioural tests in riding horses. Data showing moderate reliability or better are indicated in bold	63-64
4.5.2.1	Shows the median (IQR) and Wilcoxon's signed-rank test results for the test-retest data of the standardised behaviour tests, the results in bold indicate a significant difference between trials one and two	66-67
4.5.3.1	Inter-observer reliability of responses to standardised behaviour tests in riding horses. Data showing good reliability are shown in bold	69
5.7.1	Summary of the Weighted Kappa test (K_w) to assess the reliability of the suitability of use questionnaire. Bold values of K_w indicate moderate to good reliability between two independent raters.	88

5.8.1	Summary of police horse roles and situations relating to them, the mean (SD) for the horses overall scores for each role and the number of horses used in the analysis are also presented	90
6.4.1	Table to show informal interpretations of correlation coefficients taken from Martin & Bateson (1993).	94
6.5.1	Summary of Spearman's rank order (r_s) correlations between police horses' responses to the standardised behavioural tests and the scores of indices of police horse roles. Results in bold indicate a moderate or better correlation.	97-98
6.5.1.1	Summary of Mann Whitney U test results to explore the differences in horses' behavioural reactions between Merseyside and Manchester mounted police units. The means and standard deviations for each location are also shown	100

List of figures

Figure number	Figure legend	Page number
4.3.1.1	Layout of the arena for the social isolation tests	57
4.3.2.1	Layout of the arena for the handling tests	58

Abstract

Standardised behaviour tests have the potential to improve equine welfare by enabling the selection of horses suitable for a particular role. A range of tests are found in scientific literature which make such claims. However, with notable exceptions, authors have failed to establish the reliability of their tests or the validity of their claims.

This study aimed to determine a set of reliable behaviour tests based on those reported in literature; and to test the validity of these as predictors of horses' suitability as police horses.

Reliability of response to a novel object, social isolation and handling tests was established by tests repeated at three week intervals on 33 horses. Behavioural responses were recorded *post hoc* to enable the full range of responses to be measured. There was some evidence that the horses habituated to static stimuli used in the tests, and the majority of the tests were at best moderately reliable, with only reactivity tests showing good or better reliability. Tests with moderately reliable results or results which could be explained by habituation, were included in the validity testing. This compared the horses' responses to the behavioural tests with handlers' ratings of key aspects of police horse performance.

These subjective scales of suitability of police horses were developed from interviews with instructors and yard managers. The items were combined to develop a set of specific indices of police horse roles. Spearman's rank order correlations were used to explore whether any relationships existed between the horses' behavioural responses to the tests and subjective ratings of their suitability for their use. The results revealed eight moderate associations which were on the whole between horses' responses to the reactivity tests (unexpected noise and sudden moving object) and the suitability scores.

These findings suggest that standardised behavioural tests are not strongly reliable. The validity analysis indicated that either the tests used in this study did not serve as predictors of suitability for use, or that behavioural responses to standardised tests cannot be generalised to the wider challenges faced by these horses in their roles. However the reactivity tests did reveal stronger reliability and they were the only measures to show any association with the suitability scores, suggesting that reactivity tests may have some validity in measuring a horse's suitability. Overall, the findings of this study cannot support claims made for these (or similar) tests to be valid predictors of a horse's suitability for use.