

## **Chapter 2**

### **Literature review**

This review firstly describes the history and current selection of police horses. It then explores the different standardised behaviour tests for horses reported in the research literature and the extent to which the authors have addressed issues of feasibility, reliability and validity for each test. Methods used to assess the suitability of animals at a particular role are also explored. The search strategy used to obtain the literature can be seen in Appendix 1.

#### **2.1 Police horses and their selection**

The mounted police section dates back to the Eighteenth Century and was originally developed to deal with the plague of highwaymen (Metropolitan Police Service, n.d.). The modern day mounted branch continues to serve in remote parts of the country and in metropolitan areas where it conducts many roles including crowd control, crime prevention and community events (Wikipedia, 2008).

Very little literature is available on the selection of police horses. West Yorkshire Police (n.d.) choose their horses from various sources such as dealers, studs and private owners. The horses are already broken-in when purchased and are normally between four to seven years old.

In view of the limited available literature in this field, the mounted units at which the present study was conducted were asked about their methods of

selection. There is currently no objective assessment employed and the selectors/trainers are guided by their experience only, leading to a subjective assessment of the horses' personality and its suitability as a potential police horse. The horse is normally ridden at the owners' premises by the trainer/selector prior to recruitment and is further assessed during an initial trial period of about six weeks at the police unit.

The subjectivity of this method remains a problem, as it does not allow the trainers to accurately evaluate how the horse may behave in different environments or in regards to challenges related to police horse work, where an array of unusual sights and sounds are likely to be encountered. If it was possible to objectively assess the horses' personality and therefore its suitability prior to commencing training, the mounted units would be able to select horses with the appropriate personalities for the job. Such tests may also be useful for the selection of horses in other roles such as riding school horses and therapeutic riding horses.

This method of selection has already been adapted for the use of selecting dogs for a particular role, such as police dogs (Slabbert & Odendaal, 1999) and guide dogs (Serpell & Hsu, 2001). Potential young dogs are commonly subjected to a 'puppy test' to predict whether they would be suitable for a particular job. These tests involve exposing the pups to a series of controlled test situations that have been found to be useful for evaluating the relevant behaviour traits (e.g. aggression, fear, trainability etc) (Serpell & Hsu, 2001).

If a similar method can be found that is reliable and valid in predicting the suitability of horses for particular roles, it would be a valuable tool for the equine industry. To assess the concurrent validity of standardised behaviour tests in predicting a horse's suitability, the tests need to be validated by comparing them with another measure of suitability.

## **2.2 Behaviour tests**

The use of standardised behaviour tests facilitates the comparison of individuals in a standardised way (Manteca & Deag, 1993). The situation is highly controlled and the behaviour responses seen should reflect individual differences in horses' tendencies to respond in particular ways to environmental challenges. A variety of techniques have been used to develop such tests.

### **2.2.1 Reaction to a novel object**

Table 2.2.1.1 summarises the published studies on novel object tests in horses. In these tests, objects unknown to the horses are presented and both contact with the novel object and attention directed towards the novel object without physical contact were measured.

Table 2.2.1.1. Summary of ‘novel object tests’, methods used and the claims authors made as to what their tests measured.

<b>Authors</b>	<b>Novel object</b>	<b>Method</b>	<b>Conclusions drawn</b>
Seaman et al. (2002)	Bright blue metal saddle stand with a red plastic sledge in the centre.	The object was placed in the centre of a 10m circle and the times the horse entered the circle and approached the object were recorded.	Responses to behaviour tests were not predictors of horses’ general behavioural tendencies as assessed by questionnaires given to carers.
Le Scolan et al. (1997)	Six metal rails that formed a cage in which they placed a colourful helium inflated ball which had ‘arms and legs’ 50cm x 20cm.	A 10m circle was drawn around the object to evaluate the distance kept by the horse. They also measured behaviour using instantaneous scan sampling and all occurrence methods.	The authors concluded that this test measured a degree of nervousness.
Wolff et al. (1997)	Four green glass fibre poles placed on a circle of 1m diameter, two grey plastic tubes and a red-white striped plastic band around them.	A 10m circle was drawn around the object to evaluate the distance kept by the horse. They also measured the behaviour of the horse using instantaneous scan sampling and all occurrence methods.	They considered the responses to the tests to reflect an inherent degree of fearfulness.
Visser (2002)	A blue and white umbrella was lowered from the ceiling, to the floor.	The horse entered the arena and after two minutes an umbrella was lowered from the ceiling and the horse’s behaviour was recorded for five minutes once it hit the ground.	It was concluded that this test measured both horses’ temperament (flightiness and sensitivity) and suitability (show-jumping performance).

<b>Authors</b>	<b>Novel object</b>	<b>Method</b>	<b>Conclusions drawn</b>
McCall, Hall and McElhenney (2006)	A bright blue and orange plastic child's tricycle.	Measured the behaviour and heart rate of horses for 15 minutes.	Correlations were found between the horses' behaviour and heart rate and it was concluded that behaviour responses measure reactivity.
(Momozawa et al., 2003)	Balloons suspended from the ceiling	Measured the behaviour and heart rate of the horses.	A moderate correlation was found between the horse's behavioural response and heart rate, indicating this test measures a horse's reactivity. They also compared the tests with ratings of temperament given by caretakers and found the test invalid in measuring temperament.

### *Feasibility*

All the novel object tests summarised above can be conducted by a lay person, but some are much more practical than others. Methods which would be difficult to apply in an industrial setting include the methods used by Visser (2002) and Momozowa et al. (2003), in which objects were suspended from a ceiling. This would involve the experimenter reaching the ceiling, which is impractical in an indoor school, where the roof is usually very high. The objects designed by Wolff, Hausberger & Le scolan. (1997) and Le Scolan et al. (1997) are also less feasible, as they would be difficult to reproduce. Objects which could be used in an industrial setting include the ones used by both Seaman et al. (2002) and McCall et al. (2006) who used objects which could be quite easily obtained and quickly set up.

### *Reliability*

Very few authors have attempted to assess the reliability of their methods and where they do, their results vary. Visser (2002) assessed the reliability of her methods by testing the reactions of young horses on four occasions (aged 9 months and 10 months) and (21 months and 22 months), and explored whether behaviour was constant both within and between years. Within years she found behaviour to be mainly moderately correlated with a few high correlations. Between years the correlations were much lower. The author suggests this could be due to the same behavioural elements reflecting different underlying motivations at different ages, such as sexual maturing.

Seaman et al. (2002) used test-retest reliability and found the responses of horses to be inconsistent over three trials, with the time to enter the circle and touch the object increasing over trials. A possible reason for this is that horses were not restrained and could therefore explore the whole arena and by the third trial horses were possibly less inquisitive to investigate the object, leading to an increase latency to enter the circle and approach the object. McCall et al. (2006) found their test to be repeatable between days one and two and two and three, but found a decrease in rank correlation coefficients between days one and three. They did however conduct these tests on consecutive days and there is therefore a risk that the horses became habituated to the stimuli over time, which is something the authors comment on in their discussion. The other authors failed to assess the reliability of their tests, making their findings invalid.

Overall the data appear to be conflicting, with some authors finding horses' responses to novel object tests to be inconsistent and others finding some reliability. It also appears that more consistent results have been found for younger animals.

### *Validity*

Validity was assessed by many of the authors, usually by comparing the behavioural responses to the tests with data from a subjective questionnaire assessing some aspect of the animal's behaviour. The extent to which the test measures what the authors claim varies between studies. Seaman et al. (2002) validated their test by analysing the objective scores against a set of

subjective scores from questionnaires that asked questions about the horses' general behaviour and that were completed by farm leaders. No correlations were found between the two, indicating that their test is invalid as a measure of general behavioural tendencies.

Le Scolan et al. (1997) made some attempt to assess the validity of their test in measuring horses' temperament by comparing the results with ratings given by riding teachers. Although they claim to obtain significant correlations between the novel object test with the ratings of nervousness when ridden, the Spearman's rho value is 0.239, indicating a low correlation. An interesting comment by Le Scolan et al. (1997) (who used trained horses) compared their results with that of Wolff et al. (1997) (who used untrained horses) and found the behaviour patterns between these two types of horses differing, with untrained horses showing more fearful reactions towards the object, indicating this test may be affected by the horses' past experiences.

Visser et al. (2002) assessed the validity of their test in measuring temperament by comparing it with ratings given by riding teachers. Very few and only moderate correlations were found between the two. In another study (Visser et al., 2003) they assessed whether the test could predict show-jumping performance, by assessing the horse in training and jumping a novel course. They found the tests of use in predicting show-jumpers, but quite complicated multiple stepwise regression was used to obtain the results. Despite this the findings do show some promise of this test in predicting horses' suitability for use.

Momozowa et al. (2003) examined whether their test could assess a horse's temperament and behaviour tendencies. This was assessed by using a questionnaire to rate the horse's temperament and behaviour. No correlations were found between the temperament scores and behaviour tests, and they claimed to find correlations between the questionnaire about the horses general behavioural tendencies and standardised behaviour tests but the Spearman's rho values were low ( $<0.4$ ) indicating little relationship between the two. Both Momozowa et al. (2003) and McCall et al. (2006) use the tests to assess whether they measure horses' reactivity, by comparing the reactions of the horses with heart rate measures. They both found heart rates to increase during the test indicating this test is accurate in measuring reactivity.

Overall it appears that authors have made claims in regards to the validity of their tests as measures of temperament, behavioural tendencies and suitability. Scrutiny of the findings indicate low to moderate correlations at best, suggesting validity may not be as clear cut as claimed.

#### *Implications for this study*

There was some evidence to suggest behaviour tests can predict a horse's temperament/behaviour tendencies. As temperament can be related to performance (Visser, 2002) and could therefore be valid predictors of suitability, the novel object test was adopted in this current study to explore a horses' suitability for its role as a police horse.

### **2.2.2 Reaction to a loud/unexpected noise**

Examining the literature revealed two authors who have used this test. The table below summarises the different noises, methods used and what authors claimed the tests measured.

#### *Feasibility*

Both these tests are feasible, but both studies involve the horse being restrained, which does not allow the horse to express a full range of behaviour. The latter of the two methods is easier to replicate than the first, as the pots and pans need to be set up each time at a uniform height, which could be time consuming.

#### *Reliability*

The reliability of the horses' responses to the tests was not assessed in either of these studies.

#### *Validity*

The validity was not assessed by MacKenzie and Thiaboutot (1997). Anderson et al. (1999) concluded behaviour tests could not be used to assess a therapeutic riding horse's temperament, but they failed to examine this particular test independently, making it impossible to assess the validity of this test *per se*.

Table 2.2.2.1. Summary of 'unexpected noise tests', methods used and the claims authors made as to what their tests measured.

<b>Authors</b>	<b>Noise</b>	<b>Method</b>	<b>Conclusions drawn</b>
MacKenzie and Thiaboutot (1997)	Pots and pans were dropped from a uniformed height behind a visual barrier.	Horses were led to the barrier and the pots and pans were released, the time taken to lead the horse back to the centre of the barrier was measured.	The authors claimed that this test measured horses' reactivity, although this was not validated.
Anderson et al. (1999)	Popped a balloon.	One person held the horse whilst another stood at its flank with the balloon behind their back and popped it.	This test was combined with two other reactivity tests to gain a total score. It was concluded that these tests didn't measure suitability as a therapeutic riding horse.

### *Implications for this study*

As neither of these studies assessed the reliability or validity of this test individually, it was decided to include a more accurate examination of unexpected noise test in this study.

### **2.2.3 Reaction to a sudden rapid and surprising object**

Three authors report tests using sudden rapid and surprising objects. These are summarised in Table 2.2.3.1.

#### *Feasibility*

All three tests are feasible and could easily be replicated by a lay person. Using an umbrella is quick as the horses reaction can be seen straight away; umbrellas are easily accessible and cheap. The holiday garland used by Minero et al. (2006) would be more difficult to obtain and is not as sudden and quick moving as the umbrella. The length of the test was also excessive; horses are likely to have become habituated to the movement. Horses also were confined to a stable, which meant they would not have had the opportunity to flee, and would therefore spend more time immobile, which is what they found - only 3% of the time was spent moving.

#### *Reliability*

None of these studies attempted to establish the reliability of the horses' reactions to these tests.

Table 2.2.3.1. Summary of ‘surprising objects tests’, methods used and claims authors made as to what their tests measured.

<b>Authors</b>	<b>Surprising object</b>	<b>Method</b>	<b>Conclusions drawn</b>
MacKenzie and Thiaboutot (1997)	The opening of an automatic umbrella.	Horse was led towards a person sat on chair (a perpendicular line was drawn at their feet) and holding the umbrella, at an agreed distance the umbrella was released (opened), the time then taken by the horse to reach the perpendicular line was measured.	They concluded that this test measured reactivity, but this claim was not validated.
Anderson et al. (1999)	The opening of an automatic umbrella.	One person held the horse whilst an assistant released an umbrella while standing approximately 2m in front of the horse. The handler held the horse still for 20sec and its behaviour was recorded.	This test was combined with two other reactivity tests to gain a total score. It was concluded that these tests could not predict a horse’s suitability as a therapeutic riding horse.
Minero, Zucca, and Canali (2006)	Shook a 40cm long red and white synthetic holiday garland.	The garland was shaken within a stable containing the horse. The test lasted 23 minutes and the garland was shaken whenever the horse explored it. Behaviour, heart-rate and blood cortisol levels were measured.	They concluded that this test measured reactivity.

### *Validity*

Anderson et al. (1999) assess whether their reactivity tests measure a horse's temperament by comparing the results with a questionnaire designed to assess a horse's temperament, however they do not assess this particular test individually. Minero et al. (2006) assessed the validity of their test in measuring reactivity by comparing the results of the test with heart rate and blood cortisol values and found some relationship between the two. MacKenzie and Thiabouot (1997) failed to establish the validity of their test.

### *Implications for this study*

The method used in the majority of past studies was the opening of an automatic umbrella. It is unclear whether this test would be of use in predicting suitability due to the lack of reliability and validity established used in past studies. As this test would be simple to replicate, it was adapted in this current study.

#### **2.2.4 Reaction to pain**

One paper was found that attempted to investigate horses' reactions to a painful stimulus (MacKenzie & Thiabouot, 1997). They used a McCormick fruit pressure tester, which pressed on the horse's mandible to assess a horse's tolerance of pain. Pressure was applied until the horse raised its head and the numbers of pounds per square inch of pressure applied was recorded.

### *Feasibility*

This test is not very practical as the equipment used is not readily available to members of the general public.

### *Reliability*

The reliability of the horses' responses to this test was not established.

### *Validity*

The authors claim this test could measure a horses' reactivity, but did not compare their test with another measure to support this claim.

### *Implications for this study*

Unfortunately it is not clear from this study whether this test accurately measures a horse's reaction to pain, as it does not explore the reliability or validity of their measures. The authors' claimed their tests measured individual differences in reactivity, and therefore this test had the potential to measure suitability, so was included in this study.

## **2.2.5 Reaction to an unusual surface**

A number of studies have been conducted, using a variety of unusual surfaces and measures (see Table 2.2.5.1), that assess the performance of a horse led over an unusual surface.

Table 2.2.5.1. Summary of ‘unusual surfaces tests’, methods used and claims the authors made as to what their tests measured.

<b>Authors</b>	<b>Unusual surface</b>	<b>Method</b>	<b>What the test measured</b>
Mackenzie and Thiabouot (1997).	A clear plastic sheet on top of straw to raise it off the ground.	Firstly the horse’s walking speed was measured between two lines 10m apart (without sheet), then the horse’s speed was measured again with the sheet placed on the ground, the difference in walking speeds was used as a measure of reactivity.	They concluded that their test measured reactivity, but do not validate this claim.
Visser (2002)	Designed a bridge consisting of four concrete plates (2x1m)	The horse was led towards the bridge until the rope went tight, if the horse was resistant the handler started the trial again, three attempts were allowed to cross the bridge.	They interpreted the test to measure temperament, including a horses’ ‘willingness to perform’, ‘patience’ and ‘flightiness’
Wolff et al. (1997)	Made a bridge consisting of five wooden planks laid on the ground	The horse was led towards the bridge until the rope went tight, if the horse was resistant the handler started the trial again, three attempts were allowed to cross the bridge.	Measured a degree of fearfulness in the horse.
Le Scolan et al. (1997)	A close mossy mattress (200 x 100 x 10) was covered with a brown and white-patterned oil cloth.	Experimenter led the horse to the bridge and tried to make it cross, if it refused they started again. The test stopped when the horse crossed the bridge.	The authors indicated that this test measured a horse’s general fearfulness.

### *Feasibility*

The most practical of these tests is the use of a plastic sheet, which is easy to obtain, cheap and safe for horses to walk across. The other surfaces are not so easy to replicate and would take more time to set up, which is impractical in an industrial situation.

### *Reliability*

Only Visser (2002) assessed the reliability of her method. She measured reactions of young horses and found them to be consistent within years (at 9 & 10 months and 21 & 22 months) but not between years, indicating only short term consistency, possibly due to horses maturing over years.

### *Validity*

Visser (2002) assessed the validity of her tests as predictors of show-jumping suitability by comparing the results with riders' ratings and show-jumping performance in two separate studies. She found some moderate correlations between behavioural responses and riders' ratings, and a correlation between the variable 'horse stands still on bridge' and 'show-jumping performance' in a horse's second year of training. No correlations were found for any of the other variables measured or for the first year of training.

Le Scolan et al. (1997) assessed the validity of their test in measuring a riding horse's temperament by comparing the results with ratings by riding

instructors. One moderate correlation was found between the time taken to cross the bridge and the horse's fearfulness when ridden.

#### *Implications for this study*

The findings from above suggested that this test may be of use but needed further investigation. It was demonstrated that this test may have some validity and was therefore of interest to this study.

### **2.2.6 Reaction to an unknown person**

This method has been commonly used in farm animals and is employed to assess an animal's fearfulness towards humans (Manteca & Deag, 1993). When testing the reaction of horses, some authors have adapted/modified these methods, which are summarised in Table 2.2.6.1 below.

#### *Feasibility*

Both methods are practical. Jezierski et al. (1999) measured the horse's response whilst they were being restrained by another person, which may not be appropriate for trained horses as they may have been trained to stand still near the handler. It is therefore more appropriate to approach the horse whilst it is loose. In Hausberger and Muller's (2002) study horses were confined to a box and the behaviour displayed was therefore restricted. However, the authors of this study were only concerned with the horses' initial reaction, not their interaction with the humans.

Table 2.2.6.1. Summary of 'stranger approach tests', methods used and claims the authors made as to what their tests measured.

<b>Authors</b>	<b>Test used</b>	<b>Method</b>	<b>Conclusions drawn</b>
Seaman et al. (2002)	Used a 'person test'	A person unfamiliar to the horse stood in the centre of a 10m circle and it was recorded how often the horse entered the circle and approached the person.	Concluded that this test could not predict horses' general behavioural tendencies.
Jeziarski, Jaworski, and Gorecka (1999)	Conducted an 'unfamiliar person test' and a 'familiar person test'	They used both handled and unhandled horses in their study and recorded the horses' reaction when being held by a familiar person and approached by a stranger. They also recorded the horses' reaction when loose in the paddock and being approached by a familiar person.	They claimed that this test measured a horses reactivity.
Hausberger and Muller (2002)	Human approach test	The experimenter would suddenly appear at the door of the horse's box and note the reaction of the horse.	Measured a horses' temperament (friendliness).

### *Reliability*

Seaman et al. (2002) found this test to be unreliable in respect of the horses' reactions not being consistent over three trials. Jazierski et al. (1999) also found the tests to be inconsistent, as horses approached humans sooner with increased age, indicating horses possibly habituated to the humans. Hausberger and Muller (2002) repeated their test on five occasions - but on the same day. This would have increased the risk of habituation.

### *Validity*

The validity of this test in predicting a horse's general behavioural tendencies was assessed by Seaman et al. (2002). They found no relationship between the ratings of general behavioural tendencies given by farm leaders and reactions of the horses. Jazerski et al. (1999) assessed the validity of their test in measuring reactivity by comparing the scores with the horses heart rate, which they found increased when they were approached by the human, indicating this test does measure reactivity/fear. The validity of their test as a measure of temperament was not assessed by Hausberger and Muller (2002)

### *Implications for this study*

As Hausberger and Muller (2002) concluded that their test measured temperament and therefore the potential to measure suitability, their method of approaching the horses was adapted for this study.

### **2.2.7 Reaction to social isolation**

Four studies are reported using variations on social isolation tests, which involved confining a horse within a walled area and measuring the animal's behaviour, such as the frequency of defecation and the animal's movements (Manteca & Deag, 1993). The methods used by previous researchers are summarised in Table 2.2.7.1.

#### *Feasibility*

These three methods are very similar, with the difference being the test area. Seaman et al. (1997) assessed horses in a concrete paddock and Le Scolan et al. (1997) and Wolff et al. (1997) used an indoor school. The most appropriate of these to use was an indoor school as the ground surface is much safer for the horses. These arenas are also readily available at many equestrian establishments. The main criterion is that horses are prevented from seeing other animals.

#### *Reliability*

The reliability was assessed by Seaman et al. (2002), who conducted the test over three trials. They found consistency over trial one and two only, but inconsistency in a third trial. The other authors did not assess the reliability of their tests.

Table 2.2.7.1. Summary of 'social isolation tests', methods used and claims the authors made as to what their tests measured.

<b>Authors</b>	<b>Test used</b>	<b>Method</b>	<b>What the test measured</b>
Seaman et al. (2002)	Social isolation test	Individual horses were released into a familiar concrete paddock and their behaviour and movements were recorded.	They found this method invalid in predicting a horse's general behavioural tendencies.
Le Scolan et al. (1997)	Social isolation test	Horses were released individually into a familiar indoor school and their behaviour recorded.	They claimed this test measures a degree of gregariousness.
Wolff et al. (1997)	Social isolation test	Horses were released in a familiar indoor school and their behaviour recorded.	They claimed this test measures a degree of gregariousness.

### *Validity*

Seaman et al. (2002), found no relationship between the horses' behavioural responses and ratings given by farm leaders on the horses' general behavioural tendencies, indicating that this tests results cannot be generalised to a horse's normal behaviour. Le Scolan et al. (1997) assessed the test's validity in predicting a riding horse's temperament and they found moderate correlations between the indices of the social isolation test and subjective ratings of horses' suitability.

### *Implications for this study*

As this method showed some evidence of reliability and validity in measuring behavioural tendencies and suitability, it was employed in this study.

## **2.3 Predicting the suitability of animals**

This concept was introduced in Chapter 1, which described past studies assessing the suitability of animals for particular roles. Two main methods have been used which include objective measures of performance at a particular role (e.g. Slabbert & Odendaal, 1999; Visser et al., 2003) and subjective assessment of performance at a particular role by a human handler or trainer (e.g. Serpell & Hsu, 2001; Anderson et al., 1999).

### 1) Objective measures of suitability

The methodology used in these studies usually involved assessing the animals' reactions to a set of standardised behavioural tests prior to commencing training. Then, after some training, the animals were measured

on their suitability at the particular job. For example Visser (2002), conducted a series of behaviour tests on young horses and then, after training, assessed their performance at jumping a novel course. The authors looked for correlations between the two and found some evidence to suggest that the standardised behaviour tests may be of use in predicting suitability.

This approach would not be practical, as the methodology involves conducting a series of behaviour tests on youngsters to predict how good they will be in a particular role as an adult. This is not possible for use at the police units as horses are recruited as adults from the general riding horse population and they do not breed and raise foals specifically for the job.

It would be difficult to adopt this method to assess the performance of police horses, as they conduct multiple roles in various environments; i.e. one day they may be used for teaching officers to ride and on another day they may be controlling a crowd at a football match. It would therefore be difficult to objectively measure the performance of these animals.

## 2) Subjective measure of suitability

The other method uses a subjective questionnaire completed by a handler to assess an animal's potential suitability for a particular role. This method has been employed successfully in assessing canines, for example Serpell and Hsu (2001) used this method in guide dogs. They developed a subjective questionnaire to rate puppies' temperament at one year of age. These ratings were compared by an independent measure of guide dog

performance. The Seeing Eye Inc (TSE) guide dog organisation has an existing system, which is applied at 14-24 months of age, in which dogs are assessed by vets, trainers and supervisors. The questionnaire was found to be both reliable and valid in predicting guide dog suitability.

The use of a questionnaire was more appropriate for assessing police horses, as their behaviour could be recorded for a variety of situations relating to their job. However this method needed to be modified, as the methodology used above is for selecting suitable youngsters. As the police recruit their horses as adults, this study focused on developing reliable standardised behaviour tests for adult horses and determining which could be used to identify good police horses. This was achieved by comparing the behavioural responses of current police horses with their trainer's subjective ratings of relative suitability as a police horse. This method identified how good police horses performed at the tests, these tests could then be used to look for potential recruits who perform similarly to good police horses.

## **2.4 Summary**

Conducting this review has revealed that on the whole authors have failed to establish the reliability of their tests, which in turn could make them invalid. This problem has been recorded by Gosling (2001), who noticed from his review of personality methodology, that with few exceptions, researchers have largely ignored issues of reliability when using behavioural coding methods, perhaps assuming that such codings are reliable. A number of authors assessed test-retest reliability but none of the above used inter-

observer reliability to assess whether two people would code the horses reactions in the same way.

The literature also demonstrated that very few authors have examined whether these tests can actually predict a horse's personality/temperament and to its suitability to a particular role. The standardised behaviour tests varied in their feasibility and the modifications of these tests, to ensure their practicality for use in the equine industry, are described in Chapter 3.

The literature review also revealed that horses are currently selected using subjective assessment of the animals' personality. Very little research has been conducted so far to design and validate a method of selecting horses suitable for a particular role. The most appropriate method to assess whether the objective tests are valid in predicting suitability is by the use of a questionnaire (see Chapter 5).