

## **Chapter 4**

### **Reliability assessment of the standardised behaviour tests**

The reliability assessments were concerned with both test-retest and inter-observer reliability. The first task was to establish which of the behavioural tests generated reliable responses from the horses. The responses of horses to the behavioural tests outlined above were recorded on two occasions, with a three week interval to minimise habituation effects. This time period was chosen as Seaman et al. (2002) found horses' habituated to a novel object and person test after a nine day period. Visser (2002) used one month and reported no evidence of habituation over one month. The responses of the horses to the tests were videoed to allow a true visual record of the complex behaviours to be analysed (Martin & Bateson, 1993). Reliability was established by looking for the amount of agreement between the two trials and two observers. The tests which showed a reliable response then went on to be analysed for inter-observer reliability.

#### **4.1 Subjects and location**

The data were collected by Creighton (Personal communication, 2006) on 35 riding horses and ponies. These horses were housed at two different establishments, 20 at the Welsh College of Horticulture and 15 at the International League for the Protection of Horses, Penny Farm, Blackpool. Both establishments stabled a mixed selection of horses, giving a cross section of age, sex and breed to assess. The sample consisted of 24 (69%) geldings and 11 (31%) mares. The median (IQR) age of the horses was 10 (6-17) years. Breeds included ponies, cobs and horses.

At both establishments horses were stabled in an American barn system and each location offered a large indoor ménage for conducting the tests.

## 4.2. Materials

The standardised behaviour tests required the following equipment:

Table 4.2.1 List of equipment required to conduct the standardised behaviour tests on the horses

---

|                   |                                               |
|-------------------|-----------------------------------------------|
| General equipment | Video camera, tapes, tripod                   |
|                   | Stopwatch                                     |
|                   | Field journal                                 |
|                   | Subject detail sheets                         |
| Safety equipment  | BHS approved riding hat                       |
|                   | Sturdy boots                                  |
|                   | Gloves with grip                              |
|                   | Head collar and lead rope for horse           |
| Test equipment    | Large lampshade - novel object                |
|                   | Automatic umbrella – fast moving stimulus     |
|                   | Plastic sheet (5x3m approx) – unusual surface |
|                   | Loud percussion shaker – unexpected noise     |

---

## 4.3. Test procedure

Prior to commencing the tests, it was important to have the required equipment ready and to ensure the video camera was charged. All free moving tests were conducted first, including social isolation, stranger

approach and unexpected noise tests, followed by the handling tests, including novel object, surprising object, pain and unusual surface tests. A step by step procedure can be seen in Appendix 6.

#### **4.3.1. Stage one – social isolation tests (social isolation, stranger approach and unexpected noise)**

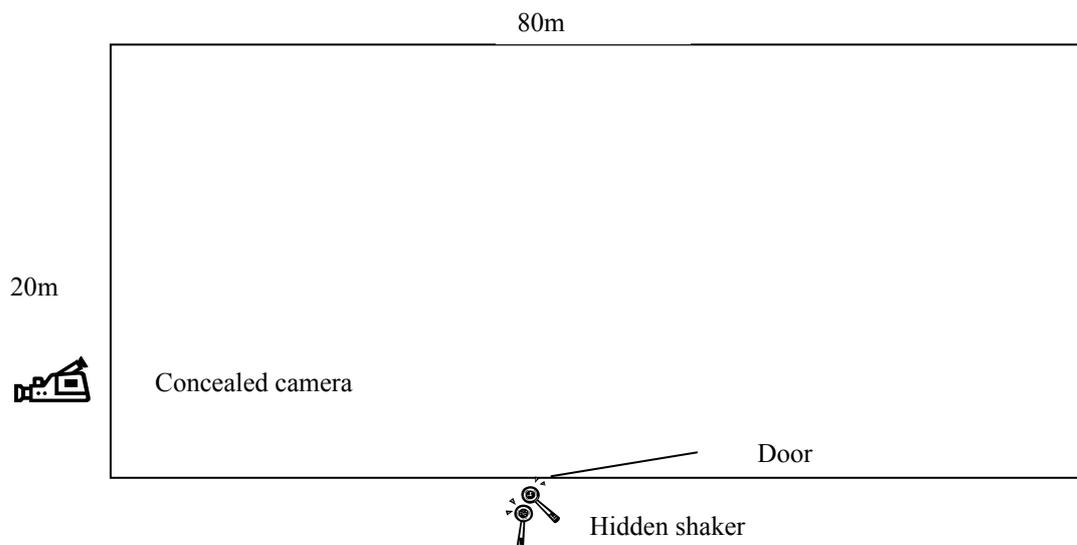
##### *Test arena*

Firstly the arena was prepared for the tests. The camera operator was positioned at a point which was concealed from the horse, with a wide camera angle to cover the ménage. The shaker was placed outside the ménage in preparation of the shaker test (see Figure 4.3.1.1).

##### *Procedure*

The horse was released into the arena and its behaviour was recorded for three minutes, during this time the tester remained out of sight and hearing of the horse. After three minutes the tester re-entered the arena and walked steadily and purposefully towards the horse. The tester then stopped 1m in front of the horse and stood still for one minute without interacting with the horse. The horse was then caught and led back to the door. The tester released the horse and left the arena and when out of sight the shaker was rattled for 10s. The tester re-entered the arena and caught the horse, giving the horse sufficient time to calm down before it was taken back to the stable

. Figure 4.3.1.1: Layout of the arena for the social isolation tests



This procedure was repeated for all the horses before conducting the handling tests.

#### **4.3.2. Stage two - handling tests (novel object, fast moving object, pinch and unusual surface)**

##### *Test Arena*

Arena was prepared by placing the equipment as shown in Figure 4.3.2.1

##### *Procedure*

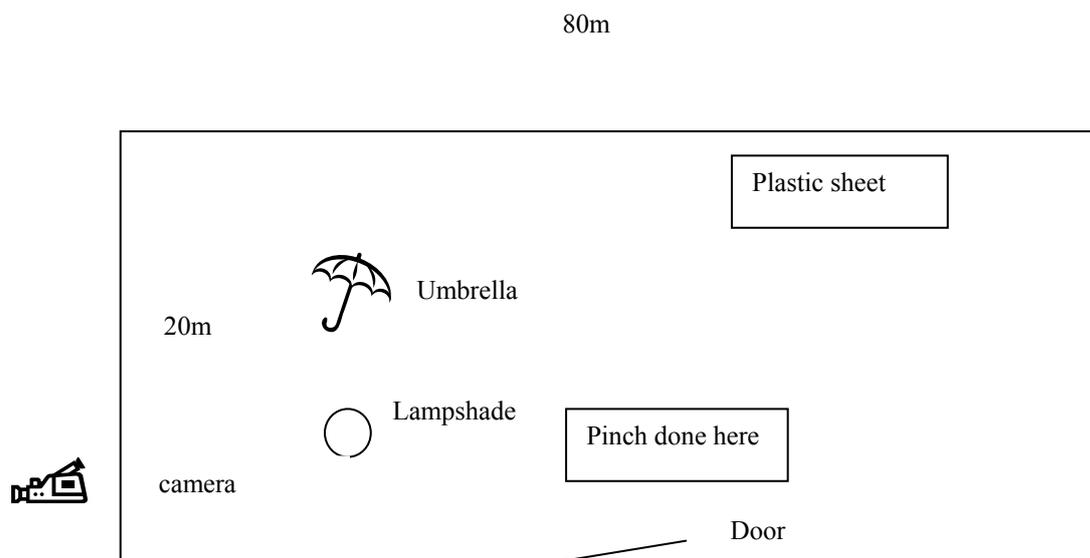
The horse was led on a loose rope towards the lampshade. One metre in front of it the handler stopped and stood still for one minute without interacting with the horse. The horse was free to move within the range of the 1.5m leadrope, the horse was then led away from the equipment before the handler

then pulled the umbrella out of their pocket, where it had been concealed.

The horse was released from the lead rope before the handler stepped back and opened the umbrella 1m in front of the horse's head. The handler stood still for one minute allowing the horse time to touch the umbrella.

The horse was caught and led close to the camera. When it stood still it was pinched on its chest, using their thumb and forefinger (no nails) for 10 seconds or until the horse moved away. Finally the horse was then led round to the plastic sheet, allowing a minimum of 20m approach. The handler led the horse on a loose rope and at this point did not interact with the horse and kept walking until the horse stopped or crossed the sheet. If the horse did stop, the handler gently encouraged the horse by talking to it and gently tugging the rope. The test stopped when either the horse crossed the sheet or refused to go any further.

Figure 4.3.2.1. Layout of the arena for the handling tests



This procedure was repeated for all horses. Both sets of tests were repeated on all the horses' after a three week period. Details of the horses' e.g. colour, height etc were recorded on field sheets, so that they could be re-identified when observing the videos.

#### **4.4. Analysis Methods**

Data were analysed using SPSS v.14. In order to establish whether the data were normally distributed, all numerical data were tested by the use of histograms and if there was uncertainty from the graphical evidence the Kolmogorov-Smirnov test for goodness of fit was used. The results indicated that the majority of the data were not normally distributed.

Categorical data were analysed for agreement between the horses' responses on their two sets of tests using Kappa (K) (for binary variables) and Weighted Kappa (Kw) (more than two ordinal categories). The interpretation of Kappa values adapted by Altman (1991) is shown in Table 4.4.1, where a Kappa value  $> 0.40$  indicates at least a moderate agreement. This value was used as a cut off point when deciding which standardised behaviour tests were reliable.

Table 4.4.1. Interpretation of Kappa values, adapted by Altman (1991).

| <b>Value of K</b> | <b>Strength of agreement</b> |
|-------------------|------------------------------|
| < 0.20            | Poor                         |
| 0.21-0.40         | Fair                         |
| 0.41-0.60         | Moderate                     |
| 0.61-0.80         | Good                         |
| 0.80-1.00         | Very good                    |

Continuous variables were analysed for limits of agreement, which were interpreted to judge whether agreement was acceptable (Altman, 1991). He states that for reasonably symmetrical distributions it would be expected that the range mean  $\pm$  2SD to include about 95% of observations. Therefore the mean difference (MD)  $\pm$  2 standard deviations (SD) can be taken as a 95% range of agreement; this range defines the 95% limits of agreement.

Secondary analysis was conducted on the standardised behaviour tests as findings from the literature suggest that there is a tendency for horses to habituate to certain stimuli. Seaman et al. (2002) found horses to habituate to both the 'novel object' and 'person test' in their study. Christensen et al. (2008) also found that horses habituate to non-threatening stimuli such as a non-moving object quicker than reactivity stimuli. To assess whether the horses in this study possibly habituated to the standardised behaviour tests, despite a three week gap in-between, the horses' behavioural responses between trials one and two were compared. Wilcoxon's signed-rank test was used to evaluate whether the distribution of the two paired variables in the two related samples was the same (Field, 2000). This non-parametric test was chosen as the majority of the data were not normally distributed and were mainly ordinal.

Bonferroni calculator (Simple Interactive Statistical Analysis, n.d.), was used to adjust the P value to account for the number of measures used for each of the behavioural tests, this was to decrease the chance of a type one error occurring (Field, 2000) from using multiple tests. The significance value was decreased from 0.05, for all the tests, the adjusted values are summarised in Table 4.4.2.

Table 4.4.2. Summary of the number of measures used for each standardised behaviour test and the alpha level achieved from using Bonferroni correction.

| <b>Standardised behaviour Test</b>                           | <b>Number of tests</b> | <b>P value before</b> | <b>P value after</b> |
|--------------------------------------------------------------|------------------------|-----------------------|----------------------|
| Social Isolation                                             | 5                      | 0.05                  | 0.01                 |
| Stranger approach, novel object and Surprising object tests  | 3                      | 0.05                  | 0.0166               |
| Unexpected noise and pinch tests                             | 2                      | 0.05                  | 0.025                |
| Unusual surface and encouragement over unusual surface tests | 1                      | 0.05                  | 0.05                 |

## **4.5. Results**

### **4.5.1 Test-retest reliability**

Table 4.5.1.1, shows the reliability of responses to standardised behavioural tests. The reliability analysis revealed low agreement for the majority of the measures taken for each of the standardised behaviour tests of social isolation, novel object, stranger approach and pinch tests.

Moderate agreement was shown between test-retest trials for the horses' tendency to vocalise during the social isolation test, in their initial response to a stranger, their flight reaction during the unexpected noise and sudden moving object test, in their tendency to touch the object during the sudden moving object test, and in the horses' initial reaction to an unusual surface and encouragement over an unusual surface.

Good agreement was found between trials for the horses' postural expression to an unexpected noise and sudden moving object.

Table 4.5.1.1. Reliability of responses to standardised behavioural tests in riding horses. Data showing moderate reliability or better are indicated in bold.

| Standardised behaviour test | Behavioural variable                                       | N  | Weighted Kappa | Kappa   | Limits of agreement                                         |
|-----------------------------|------------------------------------------------------------|----|----------------|---------|-------------------------------------------------------------|
| Social isolation test       | Duration of time stood within 2 horses length of door      | 32 | -              | -       | MD= 11.75 SD= 55.10<br>- 98.45 to 121.95<br>Range too wide  |
|                             | Duration of time exploratory walking                       | 32 | -              | -       | MD= 3.47 SD= 23.13<br>- 42.79 to 49.73<br>Range too wide    |
|                             | Duration of time alert walking                             | 32 | -              | -       | MD= 5.65 SD =17.78<br>- 30 to 41.21<br>Range too wide       |
|                             | Duration of time exploring anything else                   | 32 | -              | -       | MD= 0.14 SD= 21.81<br>- 43.48 to 43.76<br>Range too wide    |
| Stranger approach test      | Horse's initial reaction to the human                      | 33 | <b>0.475</b>   | -       | -                                                           |
|                             | Horse touches or doesn't touch human                       | 33 | -              | - 0.086 | -                                                           |
|                             | Amount of time horse spent in close proximity of the human | 33 | -              | -       | MD= 7.54 SD= 32.83<br>- 58.12 to 65.66 –<br>very wide range |

| Standardised Behaviour test             | Behavioural variable                                     | N  | Weighted Kappa | Kappa       | Limits of agreement                                  |
|-----------------------------------------|----------------------------------------------------------|----|----------------|-------------|------------------------------------------------------|
| Unexpected noise test                   | Horse's postural expression in reaction to the noise     | 33 | <b>0.856</b>   | -           | -                                                    |
|                                         | The horse's flight reaction                              | 33 | <b>0.442</b>   | -           | -                                                    |
| Novel object test                       | Duration of time exploring the object                    | 32 | -              | -           | MD= 9.43 SD= 11.38<br>-13.33s to 32.19s – wide range |
|                                         | Duration of time touching object                         | 32 | -              | -           | MD= 6.7 SD = 10.65<br>-14.26s to 28s – wide range    |
|                                         | The latency of time for the horse to touch the lampshade | 33 | 0.266          | -           | -                                                    |
| Sudden fast moving object test          | Horse's postural expression in relation to the umbrella  | 33 | <b>0.614</b>   | -           | -                                                    |
|                                         | The horse's flight reaction                              | 33 | <b>0.514</b>   | -           | -                                                    |
|                                         | Horse touches or doesn't touch umbrella                  | 33 | -              | <b>0.41</b> | -                                                    |
| Pain test                               | Horse's postural expressions                             | 33 | -              | 0.008       | -                                                    |
|                                         | Horse's movement in response to pressure                 | 33 | -              | 0.389       | -                                                    |
| Unusual surface test                    | Horse's initial reaction to the sheet                    | 33 | <b>0.431</b>   | -           | -                                                    |
| Encouragement over unusual surface test | Horse's reaction to encouragement                        | 33 | <b>0.545</b>   | -           | -                                                    |

#### **4.5.2 Test for differences between trials**

The exploration for evidence of habituation between trials revealed some statistically significant results after Bonferonni correction (see Table 4.5.2.1). These included a decrease in the time spent exploring and touching a novel object and an increase in the time to touch a novel object between trials. An increase was also found in the latency of time for the horse to touch the human and a decrease was found in the horses' flight reaction to an unexpected noise.

Table 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.

| Name of test                                                                   | Test median (IQR)  | Retest median (IQR) | W      | P value      | N  |
|--------------------------------------------------------------------------------|--------------------|---------------------|--------|--------------|----|
| Social isolation -time spent at door                                           | 57.90 (8.37-101.0) | 56.66 (24.7-106.5)  | - 1.14 | 0.252        | 32 |
| Social isolation -time spent alert walking                                     | 15.17 (4.8-38.49)  | 14.57 (1.9-22.87)   | - 1.52 | 0.126        | 32 |
| Social isolation -time spent exploratory walking                               | 22.32 (9.57-41.87) | 17.84 (5.76-42.50)  | - .645 | 0.519        | 32 |
| Social isolation - time spent exploring other                                  | 18.23 (6.5-32.86)  | 19 (9.14-40.47)     | - .274 | 0.784        | 32 |
| Stranger approach - time human stands still till horse moves away              | 42.96 (12.1-60.0)  | 60.0 (35.5 – 60.0)  | -1.51  | 0.130        | 33 |
| Stranger approach - tendency to touch human                                    | 1 (1-2)            | 1 (1-1)             | -1.38  | 0.166        | 33 |
| Stranger approach - time to touch human (only including horse's which touched) | 4.06 (2.15-16.13)  | 3.78 (2.05-5.09)    | -2.77  | <b>0.005</b> | 19 |
| Stranger approach – horse's initial response                                   | 3 (2-3)            | 3 (2-3)             | .000   | 1.000        | 33 |
| Unexpected noise test – initial reaction                                       | 3 (2-3)            | 3 (2.5-3)           | -1.73  | 0.083        | 33 |
| Unexpected noise test – flight reaction                                        | 3 (1-3)            | 1 (1-2.5)           | -3.22  | <b>0.001</b> | 33 |

| <b>Name of test</b>                                                 | <b>Test median (IQR)</b> | <b>Retest median (IQR)</b> | <b>W</b> | <b>P value</b> | <b>N</b> |
|---------------------------------------------------------------------|--------------------------|----------------------------|----------|----------------|----------|
| Novel object test - (time touching)                                 | 8.9 (4.7-18.5)           | 4.1 (0.127-7.4)            | -3.362   | <b>0.001</b>   | 32       |
| Novel object test - (time exploring)                                | 15.06 (7.2-26.7)         | 6.65 (1.5-10.29)           | -3.743   | <b>0.000</b>   | 32       |
| Novel object test – Time to touch object (<10s, >10s or not at all) | 1 (1-1)                  | 2 (1-3)                    | -3.34    | <b>0.001</b>   | 33       |
| Sudden moving object test – postural expression                     | 3 (2-4)                  | 3 (2-3)                    | -1.387   | 0.166          | 33       |
| Sudden moving object test – horse's flight reaction                 | 1 (1-2)                  | 1 (1-2)                    | -.587    | 0.557          | 33       |
| Sudden moving object test – horse's tendency to touch umbrella      | 2 (1-2)                  | 1 (1-2)                    | -1.26    | 0.206          | 33       |
| Pain test – postural expression                                     | 1 (1-2)                  | 1 (1-2)                    | -.775    | 0.44           | 33       |
| Pain test – tendency to move away                                   | 2 (1-2)                  | 2 (1-2)                    | -1.26    | 0.206          | 33       |
| Unusual surface test – initial reaction                             | 3 (2-4)                  | 3 (2-4)                    | -1.75    | 0.079          | 33       |
| Unusual surface test – response to encouragement                    | 2 (1-4)                  | 2 (1-4)                    | -.717    | 0.473          | 33       |

### **4.5.3 Inter observer reliability**

The standardised behavioural tests which generated moderately or better reliable data, went on to be assessed for inter-observer reliability to see if two observers agreed on the behavioural reaction of the horse using the behavioural coding scheme summarised in Appendix 5. The results of this are illustrated in table 4.5.3.1. All the tests were shown to have good agreement and were used to explore whether they could be used to predict suitability in police horses.

Table 4.5.3.1. Inter-observer reliability of responses to standardised behaviour tests in riding horses. Data showing good reliability are shown in bold.

| <b>Behaviour test</b>                 | <b>Behavioural variable</b>                                                     | <b>N</b> | <b>Reliability Kw</b> | <b>Reliability K</b> |
|---------------------------------------|---------------------------------------------------------------------------------|----------|-----------------------|----------------------|
| Stranger approach                     | Horse's initial reaction to the human                                           | 33       | <b>0.821</b>          | -                    |
| Unusual noise                         | Horse's postural expression, i.e. what its body did in relation to the noise    | 33       | <b>0.855</b>          | -                    |
|                                       | The horse's flight reaction                                                     | 33       | <b>0.838</b>          | -                    |
| Sudden fast moving object             | Horse's postural expression, i.e. what its body did in relation to the umbrella | 33       | <b>0.849</b>          | -                    |
|                                       | The horse's flight reaction                                                     | 33       | <b>0.848</b>          | -                    |
|                                       | Horse touches or doesn't touch umbrella                                         | 33       | -                     | <b>0.898</b>         |
| Unusual surface                       | Horse's initial reaction to the sheet                                           | 33       | <b>0.606</b>          | -                    |
| Encouragement over an unusual surface | Horse's reaction to encouragement                                               | 33       | <b>0.766</b>          | -                    |

#### **4.6. Discussion of reliability assessment**

##### *Standardised behaviour test measures showing low test-retest reliability*

Test-retest reliability revealed low agreement for the majority of the scores of the standardised behaviour tests. This included the novel object test, social isolation test, pain test and the majority of the scores for the stranger approach test. The lack of agreement for the novel object test is supported by Seaman et al. (2002), who found an inconsistency of horses' reactions to a novel object over three trials, separated by nine day gaps. They suggest that despite the time gaps between the trials, the horses may have habituated to the object.

Post hoc analysis of the data from this present study suggested habituation to the novel object and found a reduction in the amount of time the horse touched and explored the object on retest three weeks after the initial test. The time taken to touch the object was also greater. These results suggest that the horses in the study were less reactive to the novel object on retest, indicating habituation. A similar habituation to coloured mats over a one hour time period was found by Hall and Cassaday (2006) and Christensen et al. (2008) report rapid habituation to same coloured objects, with horses being presented with the objects every day for six days and their responses decreasing by the second day. As the colour of the novel object in this study remained the same in both trials, the decrease in exploration behaviour seen between test and retest is likely to be due to habituation and explains the lack of reliability of response to the novel object test.

These findings of habituation to the novel object test contrast with those of Visser (2002), who found good consistency of their test within years. Visser (2002) left a time gap of one month between each of their trials and it is possible that this slightly longer period between trials helped to reduce habituation. Visser (2002) used in her study young horses, which would not have experienced the same environmental influences as riding horses. Riding horses face various challenges on a daily basis it was more likely that these horses would habituate to the novel object quicker than less experienced horses. Visser's (2002) results may also have been more reliable due to the lowering of the umbrella; this moving stimulus may have caused horses to be more inquisitive and alert.

In summary this test of responses to a stationary novel object was shown to be unreliable in mature horses. However, as there was an ethological explanation for the data, in that horses habituated to the object, it was decided to consider the validity of the initial responses to the novel object test in assessing a horse's suitability for use in a given role.

The horses' responses to the social isolation test were also inconsistent. The measures showed a wide variety of responses between test and retest. This result is contrary to the findings of Seaman et al. (2002), who found the responses to social isolation to be consistent between trials separated by nine days, possibly suggesting a degree of short term consistency. Post-hoc analysis also revealed no evidence of habituation in this current study.

Possible suggestions for the unreliable social isolation test in this current study include external factors during the test. Although horses could not see other horses it is likely they could still hear them whilst isolated in the arena. Other factors such as yard routine e.g. horses being fed or turned out, and the amount of exercise the horse had received that day may have influenced the horses' behaviour. Controls were not implemented for these factors during the tests. However, for the tests to be useful to the horse industry they must be robust enough to provide reliable data in an industry setting where such behavioural factors will not be controlled. So, although the lack of reliability of responses to this test may be due to external influences, this very fact makes the tests unsuitable for use in selecting horses in an industrial setting.

The behavioural responses to the pinch test were also inconsistent. It was not possible to compare this finding with the literature as its reliability has not been assessed previously. A possible reason for its inconsistency is that the amount of pressure applied to the horses' chests may have differed between trials, or the horses may have had areas on their chests which differed in sensitivity. Post hoc analysis revealed no significant difference in the distribution of the two trials. As the pain test was not reliable it was excluded from the rest of the study.

The majority of the measures for the stranger approach test were also unreliable, including the latency to touch the human and the time spent with

the human. Wilcoxon's 'signed-rank test' was used to see whether there was a significant difference between measures in test and retest and it was found that in the second trial the latency to touch humans was shorter. This result is supported by Seaman et al. (2002), who found a decrease in the time it took horses to approach the human between test and retest nine days later. A possible reason could be that the horses are social, prey animals (Worthington, 1987) and after not being harmed by the human during the first set of test trials, may have been more trusting of the human on the retest, therefore approaching more quickly. There was no significant difference in the test retest in the time spent with the human.

As the latency to touch the human could be explained by ethological reasons, in that horses possibly became desensitised to the human, this measure was considered in the validity analysis. The later measure 'time spent with the human' was not included in the rest of the study as it was unreliable and could not be explained by ethological reasons.

#### *Behavioural test measures showing moderate test retest reliability*

The behavioural tests that showed moderate test retest reliability included the stranger approach test (initial response), the unusual surface test, flight reaction to an unusual noise, flight reaction and tendency to touch a sudden moving object and the encouragement over an unusual surface test.

It was not possible to compare the measure of the stranger approach (initial response) with past studies as most involve a motionless human.

Hausberger and Muller (2002) examined the horses postural expressions when being approached by a human but did not assess the reliability of their measures, the horses in their study were also confined to a box.

It could be that horses in this study acted similarly on the human's approach, as they were a moving stimuli, which horses find more difficult to habituate (Christensen et al, 2008) This may explain why the rest of the stranger approach measures were not reliable as they considered the horses' reactions once the human was stationary. As this test was found to be moderately consistent it was included in the validity analysis.

The horses' initial reaction to an unusual surface and its response to encouragement over the unusual surface also revealed moderate agreement. This is similar to Visser (2002), who found the horses behaviour on the unusual surface test to be moderately correlated with a one month gap between trials. Visser's (2002) results also show evidence of habituation, with the horses approaching the bridge quicker, standing in front of the bridge for less time and requiring fewer trials to cross the bridge. Wilcoxon's signed-rank test was used to assess whether the horses in this study possibly habituated to the tests, but the data revealed no significant difference in responses between the test retest for the horses' initial response on the approach and their response to encouragement.

As both the horses' initial response to an unusual surface and its response to encouragement over an unusual surface produced moderate agreement they were included in the rest of this study.

The horses' flight reaction to an unusual noise also showed moderate agreement. The reliability of this measure had not been assessed previously in the literature so it was not possible to compare this result with other findings. A surprising result came from the secondary analysis, exploring differences in distribution for each of the tests. A significant decrease was found in the reactions of horses between trials, which indicated possible habituation. This result was unexpected as horses could not predict the sound and therefore it would be expected that habituation would take much longer. As the results revealed moderate agreement for this test it was used in the rest of the study.

The horses' flight reaction to a fast moving object also revealed moderate agreement between trials. Again the reliability of this test has not been assessed in previous studies and no evidence of habituation was found during this study. This is supported by Christensen, Rundgren, & Olsson (2006), who investigated the use of desensitisation as a method of training horses and using a moving object found that it took five training sessions to habituate the horses to the object. The horses' tendency to touch the moving object in this study also revealed a moderate agreement and no evidence of habituation, both these measures were therefore included in the rest of this study.

*Behavioural test measures showing good reliability*

The more promising results came from the postural expression measures of response to the reactivity tests (unexpected noise and sudden moving stimulus), in which highly moderate to good agreement was shown between test and retest. Unfortunately it is not possible to compare these results with other studies as the literature revealed that past studies using similar tests have failed to assess their reliability. It was also found that horses did not habituate to these measures.

It is possible that horses' responses were less liable to habituation to this type of stimulus than to the static objects considered earlier. The possible reason for this lies in the unpredictability of the occurrence of noise and sudden movement, which makes it more difficult for the horses to habituate to this type of stimulus. Ethologically It is beneficial for prey animals not to habituate to these stimuli (Mills & McDonnell, 2005).

These types of tests are also less likely to be influenced by training and may be more suited for adult horses, as their initial response behaviour is not affected by the handler. During the unexpected noise test the handler was out of sight and during the fast moving object test the horse was loose, as not to influence their behavioural reactions. It is also unlikely that prior to the trials the horses would have encountered situations similar to the reactivity test, where the horses' behaviour was not influenced by a rider.

These tests were included in the study as they showed good reliability and did not reveal habituation.

All moderately reliable or better tests or those which could be explained by ethological reasons were assessed for inter-observer reliability, which was found to be good for all these tests. This indicated that the coding used, designed to measure the horses' responses could be repeated by another person.

#### *Limitations of the reliability analysis*

The first limitation is concerned with the method of video recording the horses' responses and their coding from the recording. Behaviour is usually easier to observe and to analyse live and in context, rather than by watching it later on a small screen which only allows restricted detail (Martin & Bateson, 1993). Unfortunately, the tests were poorly developed in the literature, which meant that behaviour measures could not be predicted. More than one measure needed to be conducted for each test, making the use of video unavoidable.

This method was also exceedingly time consuming, as measures had to be designed to account for any behavioural outcome. This meant modifying the categories at the design stage to ensure each horse could be categorised. When viewing the footage, it became evident that it was at times difficult to observe the horses' behaviour accurately, because of bad camera angles or the horse being positioned too near or too far from the camera. It was

sometimes necessary to watch the same footage repeatedly in order to identify different aspects of the horses' postural expression, which allowed them to be categorised appropriately. Another limitation linked to videoing related to the social isolation tests, during which it was necessary to conceal the camera operator in an area next to the arena and although best efforts were made to conceal the camera operator, some horses did find them, possibly due to olfactory clue and this may have influenced their behaviour. This may have potentially influenced the results, although it was evident from the videos, that only three horses appeared to find the operator and after a quick investigation continued to move around the arena.

The use of video recording did have its benefits however, in that it allowed the researcher to record the horses' behavioural reactions accurately, as it was possible to view the tapes repeatedly. This made it unlikely that changes in the horses' body language were overlooked, something that can happen when using direct observation (live recording) only. However, in an industrial setting it would be advisable to use pre-defined coding sheets, a method being less time consuming and avoiding problems associated with video recording as described above.

Leaving a three week gap between trials also proved to be a slight constraint, as it meant that a few horses were unable to undergo the second trial because of health reasons. This led to the reliability analysis having a reduced sample size, which was unavoidable as it would have been unethical

to test these horses. However, even if a shorter time period had been used, it may still have been possible for these horses to become injured.

The three week break was chosen as a period long enough to avoid habituation effects and the results indicate that it possibly wasn't long enough, as horses were shown to habituate to a few of the tests. Visser's (2002) results showed less evidence of habituation when leaving a four week gap, for the novel object test. In an attempt to overcome this methodological flaw the standardised behaviour tests were not rejected where any unreliability between test and retest was due to habituation or desensitisation.

#### **4.7 Summary**

In summary, test retest reliability was measured for all of the tests. This met the second objectives, which were to establish the reliability of the feasible tests. The tests and measures showing reliable results or results which could be explained by ethological reasons were used in the rest of the study. These tests and measures included:

- 1) Social isolation test – Horses' tendency to vocalise.
  
- 2) Stranger approach test
  - a) Horses' initial response to the stranger approaching.
  - b) Latency of time taken for the horse to touch the human.

3) Unexpected noise test

a) Horses' postural expression.

b) Horses' flight reaction.

4) Novel object test

a) Time spent touching the object.

b) Time spent exploring the object.

c) Latency of time to touch the object.

5) Sudden moving object test.

a) Horses' postural expression.

b) Horses' flight reaction.

c) Horses' tendency to touch the object.

6) Plastic sheet test – Horses' initial response.

7) Plastic sheet encouragement test – Horses' response to encouragement.