Animal Welfare 2009, 18: 381-390
ISSN 0962-7286

Inter-observer and test-retest reliability of on-farm behavioural observations in veal calves

EAM Bokkers*†, H Leruste§, LFM Heutinck‡, M Wolthuis-Fillerup‡, JTN van der Werf‡, BJ Lensink§ and CG van Reenen‡

† Animal Production Systems Group, Department of Animal Sciences, Wageningen University and Research Centre, PO Box 338, 6700 AH Wageningen, The Netherlands
‡ Animal Sciences Group, Wageningen University and Research Centre, PO Box 65, 8200 AB Lelystad, The Netherlands
§ Institut Supérieur d'Agriculture, 48 Boulevard Vauban, 59046 Lille, France
* Contact for correspondence and requests for reprints eddie.bokkers@wur.nl

Abstract

The objective of this study was to investigate inter-observer and test-retest reliability of different behavioural observations to be used in an on-farm, animal welfare monitoring system for veal calves. Twenty-three veal calf farms, varying in size, housing system, feeding regime and age of the calves were visited twice with two observers, simultaneously. Behavioural tests were conducted in eight pens per farm, measuring the response of calves to: a human entering the barn; a novel object; a passive, unfamiliar person; disturbance in the pen and an active approach by an unfamiliar and a familiar person. Furthermore, behaviour was recorded 20 min before and 20 min after feeding in eight other pens per farm. For all behavioural tests, inter-observer reliability was very high. Farm effects and test-retest reliabilities were high and significant for all behavioural tests, except for the test measuring response to disturbance in the pen. Although the active approach test with the familiar person was reliable, it was not feasible in practice due to the availability of the farmer. Since the active approach test with the unfamiliar person gave similar results, this test was recommended for an on-farm animal welfare monitoring system. For most behavioural elements recorded around feeding, farms differed significantly and inter-observer and test-retest reliabilities were high as well as being significant. The behavioural tests with entering the barn, novel object and unfamiliar person, and the behavioural observations before and after feeding were feasible and distinctive and reliable enough to be performed on-farm. These methods are promising tools to use as a monitor of animal welfare in veal calves.

Keywords: *animal welfare, behaviour, behavioural response test, monitoring, reliability, veal calves*

Introduction

It is generally accepted that an animal welfare monitoring system should be built upon animal-based parameters (EFSA 2006; Smulders *et al* 2006). An animal welfare monitoring system will be developed for veal calves as part of the Welfare Quality® EU programme. A veal calf tends to be a bull calf coming from a dairy farm. The diet for veal calves ranges from largely milk replacer and some solid feed (eg maize silage, grain, pellets) to milk replacer and solid feed during the first 12 weeks and thereafter only solid feed. Calves are at least two weeks old on arrival at the farm and the fattening period takes 21–35 weeks. Group housing is compulsory in the European Union after eight weeks of age. Group size may vary from 4 to 90 individuals per pen. Veal calves have a minimum space of 1.8 m^2 per calf and are generally kept on a (hardwood) slatted floor.

Fear in farm animals is recognised as an important indicator for welfare (Boissy & Bouissou 1995). Fear, however, cannot be assessed by a single measurement (Forkman *et al* 2007). Different measurements that gain insight into the level of fear should, therefore, be included in an animal welfare monitoring system. Several behavioural tests and observation methods in, eg pigs, dairy cattle, laying hens, and also in (veal) calves have been developed and validated to study the level of fear and its effect on animal welfare (Rushen *et al* 1999). Examples of such fear tests are measurement of the behavioural response of a calf to an active approach by a familiar and an unfamiliar person, and measurement of the latency to touch when a calf can voluntarily approach a person (Lensink *et al* 2000, 2001, 2003). Another example of investigating fear in calves is looking at the behavioural response to a novel object. Although a number of studies in (veal) calves and cattle suggest that fear of a human may be dissociated from fear of a novel object, novel object tests are widely used to study fear in farm animals (Hemsworth *et al* 1996; Jago *et al* 1999; Waiblinger *et al* 2003; van Reenen *et al* 2004, 2005; Graml *et al* 2008).

Next to fear, abnormal behaviour is widely accepted as an indicator of poor welfare (Fraser & Broom 1997; Anonymous 2001) while play behaviour is an indicator of good welfare (Fagen 1981; Newberry *et al* 1988). Play behaviour of calves has been studied extensively (eg Jensen *et al* 1998; Jensen & Kyhn 2000) as well as abnormal behaviour. Veal calves typically develop abnormal oral behaviour, comprising the following four behavioural elements: tongue playing; tongue rolling; sham ruminating and persistent biting/sucking on substrates such as bars and troughs (Bokkers & Koene 2001). Although housing system may affect the frequency of abnormal oral behaviour in veal calves (Bokkers & Koene 2001), it has been clearly demonstrated that abnormal oral behaviour evolves to a large extent as a result of a lack of appropriate roughage in the diet (Heeres *et al* 2000; van Vuuren *et al* 2004). Other abnormal (oral) behaviours in veal calves include cross-sucking and excessive self-licking. Cross sucking, defined as one calf sucking the ear, mouth, scrotum, prepuce, tail, udder area or navel of another calf (Lidfors 1993), is seen most often in young calves that have been separated from their mother. Persistent preputial sucking may adversely affect the prepuce (swelling, irritation, inflammation) of the calf being sucked and the calves that suck may risk poor health and reduced growth due to drinking urine (de Wilt 1985). Self licking is normal behaviour for a calf, but this can develop to an abnormal, excessive level, especially when a calf is kept in social isolation (Terosky *et al* 1997; Bokkers & Koene 2001).

Until now, relatively simple and feasible behavioural measures in veal calves have been sufficiently validated under experimental conditions. Before behavioural tests and observation methods, can be considered for inclusion into an animal welfare monitoring system, however, they have to be studied for feasibility and reliability under commercial conditions. The aim of this study was to investigate the feasibility of different behavioural responsiveness tests and of observation methods to study spontaneous behaviour in veal calves kept under commercial conditions, and to analyse inter-observer reliability and test-retest reliability for different variables. Variables that appear feasible and reliable may be suitable for an on-farm animal welfare monitoring system.

Materials and methods

Twenty-three veal farms were included in this study. These farms varied, intentionally, according to type and origin of calves, group size, size of farm, diet (amount of milk replacer and amount and type of solid feed), climate control, light intensity, and management. They were assumed to represent a cross-section of veal farms in The Netherlands.

Farm visits

Each farm was visited twice to collect data, with a 1–3 day interval between two visits in order to be able to study the test-retest reliability of the observations. Farms were visited either when the calves had been at the farm for 13–15 weeks (11 farms) or two weeks prior to slaughter (12 farms) to be able to study feasibility at different ages. These two age categories were chosen to observe calves at an age whereby abnormal behaviour might have been developed, but also to observe calves that vary in terms of ease of handling. During all visits, data were collected by two observers simultaneously in order to study inter-observer reliability. Four observers (two men and two women) visited the farms in different combinations, but always the same combination of observers within a farm. The observers wore the same overalls as the farmer. Although observers were experienced in behavioural research, they completed a training assessment with videos and photographs of calf behaviour and practised together at a farm beforehand.

Behavioural tests

With the exception of the Unit Entry Test (see below), eight pens per farm were selected for the behavioural responsiveness tests in such a way that observations could be done without mutual disturbance.

In the Unit Entry Test (UET), the first reaction of calves to the appearance of an unfamiliar person was measured. The test was performed approximately two hours after morning feeding. No one was allowed in the barn one hour prior to testing. A unit of the barn (defined as a part of the farm complex where groups of calves are kept in pens that are oriented to a central feeding corridor; units within a farm building are separated by a solid wall) was entered quietly and the first two pens on the left and right side were observed. The number of calves standing in each pen was recorded. The artificial light was then switched on and the observers entered the unit with a few steps inwards. After one minute, the number of calves standing per pen was registered again. In addition, the total number of calves per pen was recorded. Depending on the number of units, the UET was conducted 8–20 times at a farm.

In the Novel Object Test (NOT), the reaction to an unknown object was measured. The novel object (NO) was a plastic football covered with a grey, plastic bag that was tied up to a stand. The stand was placed in front of a pen. Observers remained outside the pen. The NO was turned into position, which meant dangling just above calf head level and, following this, a three-minute period of observation began. Latency of every first touch of the NO was recorded for each individual calf. A maximum time of three minutes was recorded for those calves not touching the NO during the test.

In the Human Approach Test (HAT), the reaction to a passive, unfamiliar person was tested. One observer leant against the front of the pen, allowing calves to approach voluntarily and touch the observer. Eye contact was avoided. The other observer was standing in front of the pen but could not be touched. The test began when the first observer with a clear but normal voice said "hello, you are looking good today" to catch the attention of the calves. Latency of every first touch was recorded for each individual calf. The maximum time of three minutes was recorded for calves that did not touch the observer.

The Rest Recovery Test (RRT) was conducted to study the effect of a deliberate disturbance at a moment (around 1300h) when calves are normally resting (Bokkers & Koene 2001). First, the number of calves standing was recorded per pen.

Then, the observers forced the calves to stand by entering the pen. The observers left the unit and returned after 30 min to record the number of calves standing in each pen again.

In the Calf Escape Test with an Unfamiliar person (CET_U), the behavioural response of a calf to an active approach by an unfamiliar person was measured. One of the observers entered the pen and waited for one minute to allow the calves to habituate. Next, the observer in the pen chose a calf standing with its head oriented towards the observer at a distance of approximately 1.5 m. In the four-stage test the person: i) made eye contact; ii) took one step towards the calf with one arm outstretched and stood still with two feet next to each other for one second; iii) took a second step and stood still again for one second and iv) touched the calf's snout. The test was ended whenever the calf moved one of its forelegs backwards. For each successful stage, one point was awarded (0- to 4-point scale), with 0 points for calves unable to make eye contact (maximally three attempts per calf). In large groups, (> 20 calves, three farms) a maximum of 10 randomly selected calves per pen were included in this test.

The Calf Escape Test with a Familiar person (CET_F) was similar to the CET_U except that the test was conducted by a familiar person (the farmer).

Observations of spontaneous behaviour

Behaviour in the home pen was observed in eight randomly selected pens, 20 min before and 20 min after evening feeding. Choosing a fixed moment of the day to conduct these observations meant that results between farms and days could be compared. These observations were conducted only in calves two weeks prior to slaughter (12 farms). Every two minutes, posture and activity of each calf were recorded (instantaneous scan sampling). The ethogram is shown in Table 1. In cases of small groups (< 20 calves per pen), four pens were observed simultaneously. The observers stood next to each other in the feeding corridor. Prior to starting the observation, a five-minute adaptation period was maintained. The observers had 30 s of observation time per pen and they switched to the next pen at the same moment (clockwise). In total, ten scans per pen per observation were recorded. After 20 min, the observers moved to the next four pens to repeat this procedure. After the calves had been fed, behavioural observations began, once again, at the first location. On farms with large groups, two pens were observed (2 × 20 min for each pen at approximately the same time as for small groups).

Testing schedule

The tests were all performed in the same order on each farm to a tight time schedule. A day would start with the UET, followed by the NOT and the HAT, before a one-hour break to allow the calves to lie down. After the break, observations resumed with the RRT followed by the CET_U and the CET_F. At the 12 appropriate farms, the observation day ended with the behavioural observations around feeding.

Statistical analysis

Data were analysed at pen level with the statistical software package, Genstat (2005). Spearman's rank correlations were calculated as a measure for reliability between days and between observers. *P*-values are given when correlations differed significantly from 0. Correlation coefficients were considered low when below 0.4, moderate when 0.4 to 0.7, high when 0.7 to 0.9, and very high when 0.9 and above (Martin & Bateson 1993). Inter-observer reliabilities were analysed for the three pairs of observers. Overall, inter-observer reliabilities were analysed with Kendall's tau. Test-retest reliability was analysed at pen level and at farm level.

In large groups, not all the calves could reach the novel object or the human within the three-minute duration of the test. As a result of this, a relatively large number of calves showed a maximum latency to touch of three minutes. In order to gain a robust variable for the NOT and the HAT, which is not affected by the unwanted effects of a large group or of a single calf (many pens have at least one curious calf), the average latency to touch of the first five calves was analysed. For the CET_U and the CET_F, the average score per pen was analysed. For the analysis of the UET, the number of calves standing at the start of the test was subtracted from the number of calves standing after one minute. A similar correction was performed for the number of calves standing after 30 min in the RRT. Spearman's rank correlations were also calculated between the different behavioural tests. Furthermore, Spearman's rank correlation of behavioural test data were calculated using residuals of an analysis of variance model with age (at the farm for 13–15 weeks) as a fixed effect. With the exception of lying idle, standing idle and walking, behaviour is expressed without distinguishing between lying and standing. Play behaviour is the sum of running, jumping, mounting, and butting behaviour. Comfort behaviour is the sum of stretching, scratching, nose licking and self-licking behaviour. Farm effects for all variables were analysed with the Kruskal-Wallis test.

Results

At all 23 farms, the UET, NOT, RRT and CET_U were conducted. At one farm, the HAT was not conducted due to practical constraints. Twenty out of 23 farmers were willing to perform the CET_F.

With the exception of the RRT on the second day of observation, a significant farm effect was found for all behavioural tests (Table 2). Although calves seemed to respond differently on the second day, no day effects were found for the behavioural tests. Inter-observer correlations for the behavioural tests were found to be very high and significant (Table 3). Test-retest correlations within pen and within farm were significant for all behavioural tests, except for the RRT (Table 4). Test-retest correlation coefficients were higher at farm level than at pen level, except for the RRT.

On both days, a highly significant positive correlation was found between NOT and HAT, and between CET_U and CET_F (Table 5). A significantly positive moderate correla-

Table 1 Ethogram for the behavioural observations around feeding (based on de Wilt 1985; Jensen *et al* 1998).

Observation	Definition
Posture	
Standing	Standing on three or four straightened legs, without locomotion.
Lying	Lying on brisket with either forelegs or hind legs bent or stretched.
Activity	
Walking	The left foreleg and right hindleg are placed forward simultaneously, followed by the right foreleg and left hindleg etc.
Running	The left foreleg and right hindleg are placed forward simultaneously, followed by the right foreleg and left hindleg etc. Performed with great vigour and velocity (trotting). Sometimes both fore and hindlegs are placed forward alternately (galloping).
Jumping	Pushing the forequarters upwards with a sudden movement of the forelegs and head, often followed by the kicking of both hindlegs backwards. The tail may be lifted and the ears set close against the neck.
Grooming (self-licking)	After extending the tongue, the upper side is shifted along the body while retracting it into the mouth and at the same time moving the head upwards. Sequence is usually repeated.
Nose licking	Moving the tongue over the muzzle, inserting the tip of the tongue in one or alternately in both nostrils and back into the mouth again.
Scratching	After bending the head and neck sideways the claw of the hindleg on the same side is rubbed over head, neck or shoulder repeatedly.
Mounting	Laying the head on the loins, back or withers of a pen mate, jumping upwards with both forelegs, putting them on either side of the head which is similtaneously raised.
Butting	Lowering the head with the muzzle pointing upwards, the forehead is pushed against the front of the head, neck, shoulder or other parts odf a pen mate and twisted or moved up and down.
Stretching	Bending the back upwards and then downwards (back stretch) with the tail slightly lifted or bent sideways to the seat bones (tail stretch), the hindlegs placed backwards and some-timnes lifted separately (leg stretch). The neck may be lifted while the muzzle downwards (neck stretch). The stretches may occur simultaneously.
(Sham) ruminating	After eructation, visible from the stretching of the head forwards and the widening of the throat, the bolus is chewed in a relatively slow and regular fashion and then swallowed. When chewing movements are performed in a relatively fast, irregular fashion and with only a limited range it is assumed to be sham ruminating.
Idle	No visible activity, calf is just looking ahead.
Tongue playing/rolling	Extending the tongue and swaying it sideways, turning and partly rolling and unrolling it. The tongue may also be repeatedly rolled and unrolled inside the open mouth.
Oral manipulation of prepuce or testicles	Sucking at the prepuce or testicles of a pen mate, sometimes resulting in drinking urine. Urine drinking may also occur spontaneously when a pen mate urinates.
Oral manipulation of fence/wall	Licking, nibbling, sucking or biting a fence or wall.
Oral manipulation of bucket/trough	Licking, nibbling, sucking or biting the bucket or the feeding trough.
Oral manipulation of pen mate	Licking, nibbling, sucking or biting a pen mate.
Oral manipulatin of floor	Licking, nibbling, sucking or biting the floor.
Other activity	Not performing one of the activities described in this ethogram.
Not visible	Activity is not visible for the observer.

tion was found between the NOT and the HAT and the UET for the first, but not for the second observation day. Both NOT and HAT were negatively correlated with the CET_U and the CET_F on both days. On farms where calves were reluctant to make contact with the human or the novel object (high average contact latencies), these calves were also difficult to approach by a human (low average response scores).

Correlations of behavioural test measures were not affected or only marginally so, when residuals after correction for age effects were used (results not shown). This means that test-retest reliabilities obtained in the present study were not as a result of consistent differences between ages.

For several behavioural elements, a significant farm effect was found before and after feeding (Table 6). Significance and strength of test-retest correlations differed per behavioural element and also occasionally before and after feeding (Table 7). Inter-observer correlations were generally high and significant for the different behavioural elements, with

https://doi.org/10.1017/S0962728600000786 Published online by Cambridge University Press

Table 2 Mean (± SD) results of behavioural tests and farm effect per behavioural test per observation day.

Test	Variable	Day	Mean (± SD)	df	Farm effect
UET	% standing after 1 min	1	36.2 (± 29.5)	22	$P < 0.001$
		2	31.6 (± 29.2)	22	$P < 0.001$
NOT	Latency to touch NO (s)	1	135.0 (± 33.0)	22	$P < 0.001$
		2	108.5 (± 38.2)	22	$P < 0.001$
HAT	Latency to touch humans (s)	1	110.2 (± 39.6)	21	$P < 0.001$
		2	96.1 (± 45.5)	21	$P < 0.001$
RRT	% standing after 30 min	1	5.9 (± 20.1)	21	$P < 0.05$
		2	3.4 (± 22.6)	21	ns
CET_U	Average pen score	1	1.7 (± 0.58)	22	$P < 0.001$
		2	1.8 (± 0.60)	22	$P < 0.001$
CET_F	Average pen score	1	1.6 (± 0.58)	19	$P < 0.001$
		2	1.7 (± 0.64)	19	$P < 0.001$

Table 3 Inter-observer reliability per behavioural test over two observation days. Spearman rank correlation coefficients are given for the pair-wise comparison and Kendall's coefficient of concordance for the overall comparison.

Test	Variable	Observer			
		1 and 2***	1 and 3***	1 and 4***	Kendall***
UET	% standing after 1 min	0.93 (n = 255)	0.98 (n = 506)	0.96 (n = 266)	0.80
NOT	Latency to touch NO (s)	0.99 (n = 104)	0.99 (n = 128)	0.99 (n = 112)	0.80
HAT	Latency to touch humans (s)	0.99 (n = 96)	0.99 (n = 128)	0.99 (n = 112)	0.81
RRT	% standing after 30 min	0.92 (n = 104)	0.96 (n = 128)	0.96 (n = 102)	0.76
CET_U	Average pen score	0.97 (n = 104)	0.96 (n = 128)	0.97 (n = 112)	0.82
CET_F	Average pen score	0.93 (n = 104)	0.91 (n = 120)	0.89 (n = 68)	0.82

*** $P < 0.001$ for all correlations.

Table 4 Reliability between days within pen and within farm per behavioural test.

Test	Variable	Days within pen	Days within farm
UET	% standing after 1 min	0.26*** (n = 516)	0.63*** (n = 23)
NOT	Latency to touch NO (s)	0.70*** (n = 180)	0.88*** (n = 23)
HAT	Latency to touch humans (s)	0.74*** (n = 176)	0.83*** (n = 22)
RRT	% standing after 30 min	0.14 (n = 172)	0.05 (n = 22)
CET_U	Average pen score	0.69*** (n = 180)	0.94*** (n = 23)
CET_F	Average pen score	0.68*** (n = 150)	0.93*** (n = 20)

*** $P < 0.001$ for all correlations.

the exception of the correlation between observer 1 and 4 for the behaviour, manipulating floor (Table 8).

Discussion

The results of this study showed that inter-observer reliability was high and significant for all behavioural tests. Thus, the observation methods were clear and feasible, although there is a need for experienced, well-trained individuals to conduct observations. Test-retest reliability was high and significant for all behavioural tests with the exception of the RRT. For the majority of behavioural tests, there were significant farm effects, which indicate that the behavioural observations used could detect differences between farms. This is important when these tests are utilised in order to monitor animal welfare. In this study, however, we did not aim for determining factors causing farm differences for different variables. Test-retest reliability of the behavioural tests was higher at farm level than at pen level which again is relevant for a monitoring system that should assess the level of welfare at farm level. The second visit to a farm, to conduct the retest observations, was within a few days in order to exclude age effects. This

Table 5 Spearman correlation coefficients between behavioural tests at farm level.

Test	NOT	HAT	RRT	CET_U	CET_F
Day 1					
UET	0.54*	0.52*	0.02	–0.26	–0.07
NOT		0.92**	–0.03	–0.81**	–0.64**
HAT			–0.20	–0.74**	–0.54**
RRT				–0.11	0.09
CET_U					0.84**
Day 2					
UET	0.29	0.07	–0.25	–0.12	0.11
NOT		0.73**	–0.34	–0.68**	–0.50*
HAT			–0.38	–0.66**	–0.48*
RRT				0.25	–0.13
CET_U					0.80**

* $P < 0.05$; $P < 0.01$.

increased the risk of habituation to the behavioural tests or calves lowering their threshold to express fear at the second observation day (Forkman *et al* 2007). But no such effect was found in our study.

In general, it was feasible to perform all behavioural tests at commercial farms. Nevertheless, some practical constraints need to be discussed.

The UET could be performed several times in barns that were divided into a number of separate units. In some barns, however, units within a barn were not separated by a solid wall. Therefore, calves can easily disturb one another. Furthermore, it was difficult to control whether anyone had been in the barn one hour prior to the test. This may be a reason for the low reliability for the UET between days within pen.

Occasional, minor practical constraints were seen in the NOT. Sometimes it was difficult to move the novel object between pens, eg when the ceiling was low, or when milk, water or gas pipes were suspended over the feeding corridor just above head level.

One practical constraint for the HAT was that on occasion the front of a pen differed not only between farms but also within them. It was easy for calves to touch a part of the observer when the front of a pen was made up only of a feeding trough and one bar. Touching the observer became more difficult when the front of a pen was a fence with grating. The effect of this constraint could not be tested in the present study.

The correlations found between NOT and HAT suggest that they evoke similar behavioural responses in calves. It would be attractive therefore to choose one of the tests to include in an animal welfare monitoring system. However, it remains unclear whether the response to a human and to a novel object has the same underlying cause (Andersen *et al* 2000; Grignard *et al* 2001; van Reenen *et al* 2004). For both NOT and HAT we analysed the latency to touch of the first five calves to have a variable with a low within-farm bias. This variable, however, is subjected to different dynamics in small and large groups. Assume, for example, 10% of the calves are not fearful of a novel object or of the human. In a pen of ten calves, it may be expected that one calf will touch the novel object or the human, while in a group of 40 calves, four individuals may be expected to touch the novel object or the human. When taking the first five calves for analysis, this may affect the average latency. In large groups, however, distance to the novel object or the human may be relatively large due to pen size. Calves may ignore the novel object or the human because they are busy with other activities in the rear of the pen, or they are neither attracted nor fearful of the novel object or the human at such a distance. This would introduce bias when analysing latencies of all calves. Furthermore, in small pens it may be that a fearful calf touches the novel object or the human just because it happens to be standing close to it, which may also affect average latency. As a contrast to the 10% assumption, we can assume that 50% of the calves are not fearful. Here, there is a risk of overestimating fearfulness in a large group. It is highly unlikely that twenty calves will touch the novel object or the human due to limited space around the novel object and in front of the human or as a result of calves simply not being attracted to the novel object or the human. In a group of ten calves, however, it is quite easy for five calves to approach the novel object or the human. Thus, although the analysis of latencies of the first five calves for the NOT and the HAT has advantages and disadvantages, it is a robust and useful variable to measure reliability of observations.

Both the CET_F and the CET_U had significant inter-observer and test-retest reliability which corresponds with the results of Lensink *et al* (2003). Although the CET_F had

Table 6 Mean (± SD) percentages of behaviour around feeding and farm effect per behavioural element.

Behaviour	Day	Mean (± SD)	Farm effect	Mean (± SD)	Farm effect
		Before feeding		After feeding	
Lying idle	1	6.2 (± 8.2)	0.001	7.2 (± 12.3)	0.001
	2	7.4 (± 9.9)	0.007	11.4 (± 15.3)	0.001
Standing idle	1	39.3 (± 13.0)	0.001	43.7 (± 21.5)	0.001
	2	29.6(± 12.4)	0.001	39.0 (± 20.4)	0.001
Tongue rolling/playing	1	0.6 (± 1.0)	0.034	0.6 (± 1.2)	0.746
	2	0.9 (± 1.9)	0.001	0.5 (± 0.9)	0.602
Manipulating fence	1	4.9 (± 4.1)	0.002	2.8 (± 3.0)	0.213
	2	6.4 (± 5.5)	0.001	3.6 (± 4.3)	0.027
Manipulating feeder	1	10.1 (± 9.6)	0.001	5.9 (± 7.7)	0.001
	2	12.1 (± 10.0)	0.005	6.6 (± 7.9)	0.002
Manipulating pen mate	1	3.6 (± 3.5)	0.001	2.5 (± 2.7)	0.001
	2	4.3 (± 3.9)	0.001	2.3 (± 2.1)	0.105
Manipulating floor	1	0.6 (± 1.1)	0.069	0.7 (± 1.4)	0.865
	2	1.1 (± 2.0)	0.001	0.7 (± 1.0)	0.243
Play behaviour	1	1.5 (± 2.4)	0.268	0.6 (± 0.9)	0.097
	2	1.8 (± 2.3)	0.138	0.7 (± 1.2)	0.431
(Sham) ruminating	1	4.3 (± 6.0)	0.001	2.9 (± 3.9)	0.001
	2	6.1 (± 8.6)	0.001	3.0 (± 4.2)	0.002
Comfort behaviour	1	8.1(± 2.8)	0.174	5.8 (± 3.2)	0.016
	2	8.2 (± 4.0)	0.004	5.6 (± 3.4)	0.154
Self licking*	1	3.0 (± 2.6)	0.153	2.8 (± 2.3)	0.114
	2	3.7 (± 2.7)	0.109	2.7 (± 2.4)	0.009
Walking	1	2.8 (± 2.7)	0.011	2.5 (± 2.2)	0.737
	2	2.6 (± 2.2)	0.090	2.0 (± 1.8)	0.485

* Self licking is also included in comfort behaviour.

a high reliability, the test had some practical disadvantages, which made this test less feasible. In particular, the farmer had to be persuaded to co-operate. Moreover, although the farmers were provided with instructions beforehand, each performed the test slightly differently. Such variation may attenuate the level of standardisation of a test. Nevertheless, in our study, the scores obtained by the farmer (CET_F) were significantly correlated with those obtained by a trained and experienced observer (CET_U), as found earlier for calves housed individually or in pairs (Lensink *et al* 2000, 2001). This indicates that our test was robust. These results also support the idea that veal calves generalised their response across a familiar and an unfamiliar person. This does not necessarily mean that calves are unable to discriminate between people, ie under certain conditions, calves were capable of this (Arave *et al* 1992; de Passillé *et al* 1996). In veal production, a generalised response across different people may be the result of the low intensity of contact between calves and humans under commercial conditions. The farmer is present in the barn twice a day to feed and inspect the calves and there are occasional extra visits to move or to treat calves. Human-animal interactions around feeding may be associated with positive experiences by calves, while other types of interaction may well be associated with negative experiences (eg certain veterinary procedures). It remains clear that unravelling discrimination and generalisation is complex (Rushen *et al* 1999).

Both the CET_U and the CET_F had practical constraints when calves were two weeks from slaughter. At this stage, the calves were heavy (250–350 kg) and little space remains for moving around in pens. The safety of the individual entering the pen and interacting with calves could not be guaranteed: as such, this test should not be performed with calves at this age and weight.

On the one hand, the negative correlations between the average latencies to contact with the human (HAT) and the novel object (NOT) and, on the other, the average score during the CET_U and CET _F are suggestive that the responsiveness of veal calves to these tests is mediated partly by the same underlying motivational state, possibly fear.

https://doi.org/10.1017/S0962728600000786 Published online by Cambridge University Press

Table 7 Reliability of behavioural observations around feeding between days at pen level and at farm level.

Behaviour	Days within pen (n = 90)	Days within farm (n = 12)	Days within pen (n = 90)	Days within farm (n = 12)
	Before feeding		After feeding	
Lying idle	0.63***	0.34	0.36**	0.67*
Standing idle	0.37***	0.60*	0.57***	0.76**
Tongue rolling/playing	0.41*	0.73**	0.41	0.36
Manipulating fence	0.56***	0.92***	0.26*	–0.02
Manipulating feeder	0.55***	0.73**	0.66***	0.79**
Manipulating pen mate	0.43***	0.64*	0.37***	0.36
Manipulating floor	0.32	0.63*	0.33	0.36
Play behaviour	0.29*	0.33	0.34	0.22
(Sham) ruminating	0.33**	0.60*	0.39***	0.75**
Comfort behaviour	0.30*	0.50	0.45***	0.82**
Self licking[1]	0.36***	0.67*	0.35**	0.86***
Walking	0.19	0.55	0.26*	0.64*

[1] Self licking is also included in comfort behaviour.
* $P < 0.05$; ** $P < 0.01$; $P < 0.001$.

Table 8 Inter-observer reliability (observers 1, 2, 3 and 4) for the behaviour around feeding (n = 64). Spearman rank correlation coefficients are given for the pair-wise comparisons and Kendall's coefficient of concordance for the overall comparison.

Behaviour	Combination of observers							
	Before feeding				After feeding			
	1 and 2	1 and 3	1 and 4	Kendall	1 and 2	1 and 3	1 and 4	Kendall
Lying idle	0.99*	0.98*	0.96*	0.73*	0.99*	0.94*	0.98*	0.73*
Standing idle	0.89*	0.87*	0.83*	0.75*	0.96*	0.92*	0.92*	0.76*
Tongue rolling/playing	0.75*	0.83*	0.94*	0.75*	0.68*	0.66*	0.89*	0.71*
Manipulating fence	0.64*	0.72*	0.75*	0.73*	0.64*	0.70*	0.74*	0.718
Manipulating feeder	0.90*	0.89*	0.95*	0.76*	0.88*	0.86*	0.94*	0.74*
Manipulating pen mate	0.75*	0.65*	0.36*	0.71*	0.67*	0.67*	0.57*	0.71*
Manipulating floor	0.68*	0.58*	0.88*	0.63*	0.54*	0.73*	0.13	0.59*
Play behaviour	0.50*	0.76*	0.75*	0.66*	0.53*	0.46*	0.49*	0.56*
(Sham) ruminating	0.79*	0.53*	0.86*	0.71*	0.54*	0.25*	0.57*	0.66*
Comfort behaviour	0.47*	0.63*	0.53*	0.70*	0.75*	0.71*	0.71*	0.73*
Self licking[1]	0.63*	0.69*	0.64*	0.72*	0.73*	0.68*	0.61*	0.72*
Walking	0.40*	0.48*	0.59*	0.68*	0.38*	0.43*	0.65*	0.67*

[1] Self licking is also included in comfort behaviour.
* $P < 0.05$.

The results of the behavioural observations around feeding are a little more ambiguous. Most inter-observer and test-retest reliabilities were significant and correlation coefficients were higher at farm level than at pen level. Again, this is relevant for the development of an animal welfare monitoring system aimed at estimating animal-based parameters at farm level. Due to the tight time schedule of an observation day, a habituation period of five minutes prior to starting an observation was achievable, maximally. Indeed, calves were not adapted completely to humans standing in front of their pen by then, but all observations were conducted according to the same procedure and therefore assumed to be comparable.

Most inter-observer reliabilities for behaviour around feeding were significant. The procedure for these observations was designed in such a way that two observers could make behavioural recordings without speaking to each other. There was 30 s per pen for recording in small groups and 120 s for large. In that time, observers had to record for each individual calf, the behaviour observed. Recordings, therefore, were not conducted exactly simultaneously.

https://doi.org/10.1017/S0962728600000786 Published online by Cambridge University Press

Hence, differences between observers could occur, especially for behavioural elements of short duration. Nevertheless, the procedure for these observations was robust enough to be reliable. The present study was conducted with the aim of examining the reliability of observation methods, and not to identify factors that may affect behaviour. A larger scale study would be needed in order to address this latter aim.

Conclusion and animal welfare implications

Inter-observer reliability was high for behavioural tests and observations around feeding. Test-retest reliability was high for all behavioural tests with the exception of the RRT. The RRT also had a low distinctive power between farms indicating that the value of this measurement is low to detect differences between farms. The RRT is an unreliable test and, therefore, unsuitable as a monitor of animal welfare. For the other behavioural tests, variation between farms was significant. Although inter-observer and test-retest reliability of the CET_F was high, this test was not feasible in practice due to an over-reliance on the farmer. The CET_U was feasible and gave similar results as the CET_F. Therefore, we would advocate against the use of the CET_F in an animal welfare monitoring system. A constraint for the CET_U is that it requires to be performed at an age when interaction with calves in a relatively small enclosure is safe. All the other behavioural tests and the observations around feeding were feasible to perform on-farm, although certain minor practical constraints did exist. It can be concluded that the UET, NOT, HAT, CET_U and the observation method for spontaneous behaviour are all reliable tools to be utilised in an animal welfare monitoring system. Assessing animal welfare with reliable tools is a must as important decisions may ultimately be made on the strength of these; this has implications for both farmers and animals.

Acknowledgements

The present study is part of the Welfare Quality® research project that has been co-financed by the European Commission, within the 6th Framework Programme, contract No FOOD-CT-2004-506508. The text represents the authors' views and does not necessarily reflect the position of the Commission who will not be liable for the use made of such information. This research was financed by the Dutch Ministry of Agriculture, Nature and Food Quality, the Product Board Animal Feed and the Product Boards for Livestock, Meat and Eggs. We gratefully acknowledge our collaboration with Van Drie Group, Alpuro Group and Denkavit Group. We would like to thank the farmers for their hospitality and kindness in receiving us at their farm, and Bas Engel and Willem Buist for their help with the data analysis.

References

Andersen IL, Bøe KE, Foerevik G, Janczak AM and Bakken M 2000 Behavioural evaluation of methods for assessing fear responses in weaned pigs. *Applied Animal Behaviour Science 69*: 227-240

Anonymous 2001 Scientists' assessment of the impact of housing and management on animal welfare. *Journal of Applied Animal Welfare Science 4*: 3-52

Arave CW, Albright JL, Armstrong DV, Foster WW and Larson LL 1992 Effects of isolation of calves on growth, behavior, and first lactation milk yield of Holstein cows. *Journal of Dairy Science 75*: 3408-3415

Boissy A and Bouissou MF 1995 Assessment of individual differences in behavioural reactions of heifers exposed to various fear-eliciting situations. *Applied Animal Behaviour Science 46*: 17-31

Bokkers EAM and Koene P 2001 Activity, oral behaviour and slaughter data as welfare indicators in veal calves: a comparison of three housing systems. *Applied Animal Behaviour Science 75*: 1-15

de Passillé AM, Rushen J, Ladewig J and Petherick C 1996 Dairy calves' discrimination of people based on previous handling. *Journal of Animal Science 74*: 969-974

de Wilt JG 1985 *Behaviour and welfare of veal calves in relation to husbandry systems*. PhD Thesis, Landbouwhogeschool Wageningen, The Netherlands

EFSA 2006 The risks of poor welfare in intensive calf farming systems. An update of the Scientific Veterinary Committee Report on the Welfare of Calves. *European Food Safety Authority Annex to EFSA Journal 366*: 1-36, 144

Fagen RM 1981 *Animal Play Behaviour.* Oxford University Press: New York, USA

Forkman B, Boissy A, Meunier-Salaun MC, Canali E and Jones RB 2007 A critical review of fear tests used on cattle, pigs, sheep, poultry and horses. *Physiology & Behavior 92*: 340-374

Fraser AF and Broom DM 1997 *Farm Animal Behaviour and Welfare.* CAB International: Wallingford, UK

Genstat 2005 *VSN International Ltd, Version 8.1.0.152.* Venstat: UK

Graml C, Waiblinger S and Niebuhr K 2008 Validation of tests for on-farm assessment of the hen-human relationship in non-cage systems. *Applied Animal Behaviour Science 111*: 301-310

Grignard L, Boivin X, Boissy A and Le Neindre P 2001 Do beef cattle react consistently to different handling situations? *Applied Animal Behaviour Science 71*: 263-276

Heeres JJ, Ruis-Heutinck LFM and van Reenen CG 2000 The effect of roughage supply on zootechnical parameters, meat quality, behaviour and pathological status of the digestive organs in veal calves. In: Van Arendonk JAM, Hofer A, van der Honing Y, Madec F, Sejrsen K, Pullar D, Bodin L, Fernández JA and Bruns EW (eds) *51st Annual Meeting of EAAP* pp 199. 21-24 August 2000, The Hague, The Netherlands

Hemsworth PH, Price EO and Borgwardt R 1996 Behavioural responses of domestic pigs and cattle to humans and novel stimuli. *Applied Animal Behaviour Science 50*: 43-56

Jago JG, Krohn CC and Matthews LR 1999 The influence of feeding and handling on the development of the human-animal interactions in young cattle. *Applied Animal Behaviour Science 62*: 137-151

Jensen MB and Kyhn R 2000 Play behaviour in group-housed dairy calves, the effect of space allowance. *Applied Animal Behaviour Science 67*: 35

Jensen MB, Vestergaard KS and Krohn CC 1998 Play behaviour in dairy calves kept in pens: the effect of social contact and space allowance. *Applied Animal Behaviour Science 56*: 97

Lensink BJ, Boivin X, Pradel P, Le Neindre P and Veissier I 2000 Reducing veal calves' reactivity to people by providing additional human contact. *Journal of Animal Science 78*: 1213-1218

Lensink BJ, Raussi S, Boivin X, Pyykkonen M and Veissier I 2001 Reactions of calves to handling depend on housing condition and previous experience with humans. *Applied Animal Behaviour Science 70*: 187-199

Lensink BJ, van Reenen CG, Engel B, Rodenburg TB and Veissier I 2003 Repeatability and reliability of an approach test to determine calves' responsiveness to humans: 'a brief report'. *Applied Animal Behaviour Science 83*: 325-330

https://doi.org/10.1017/S0962728600000786 Published online by Cambridge University Press

Lidfors LM 1993 Cross-sucking in group-housed dairy calves before and after weaning off milk. *Applied Animal Behaviour Science 38*: 15-24

Martin P and Bateson P 1993 *Measuring Behaviour, an Introductory Guide*. Cambridge University Press: Cambridge, UK

Newberry RC, Wood-Gush DGM and Hall JW 1988 Playful behaviour of piglets. *Behavioural Processes 17*: 205-216

Rushen J, Taylor AA and de Passillé AM 1999 Domestic animals' fear of humans and its effect on their welfare. *Applied Animal Behaviour Science 65*: 285-303

Smulders D, Verbeke G, Mormède P and Geers R 2006 Validation of a behavioral observation tool to assess pig welfare. *Physiology & Behavior 89*: 438-447

Terosky TL, Wilson LL, Stull CL and Stricklin WR 1997 Effects of individual housing design and size on special-fed Holstein veal calf growth performance, hematology, and carcass characteristics. *Journal of Animal Science 75*: 1697-1703

van Reenen CG, Engel B, Ruis-Heutinck LFM, van der Werf JTN, Buist WG, Jones RB and Blokhuis HJ 2004 Behavioural reactivity of heifer calves in potentially alarming test situations: a multivariate and correlational analysis. *Applied Animal Behaviour Science 85*: 11-30

van Reenen CG, O'Connell NE, van der Werf JTN, Korte SM, Hopster H, Jones RB and Blokhuis HJ 2005 Responses of calves to acute stress: Individual consistency and relations between behavioral and physiological measures. *Physiology & Behavior 85*: 557-570

van Vuuren AM, Stockhofe N, Heeres-van der Tol JJ, Heutinck LFM and van Reenen CG 2004 Effect of roughages added to the milk replacer diet of veal calves on behavior and gastric development. In: Joshi NP and Herdt TH (eds) *12th International Conference on Production Diseases in Farm Animals*.19-22 July 2004, East Lansing, Michigan, USA

Waiblinger S, Menke C and Fölsch DW 2003 Influences on the avoidance and approach behaviour of dairy cows towards humans on 35 farms. *Applied Animal Behaviour Science 84*: 23-39

https://doi.org/10.1017/S0962728600000786 Published online by Cambridge University Press