

How to measure dairy cows' responsiveness towards humans in breeding and welfare assessment? A comparison of selected behavioural measures and existing breeding traits

Asja Ebinghaus*, Silvia Ivemeyer*, Vanessa Lauks*, Laura Santos^a, Kerstin Brügemann^a,
Sven König^b, Ute Knierim*

*Section of Farm Animal Behaviour and Husbandry, Faculty of Organic Agricultural Sciences, University of Kassel, Witzenhausen, Germany, Tel. +49 (0)561 804-1748, Fax +49 (0)561 804-1646, email: ebinghaus@uni-kassel.de

^a Section of Animal Breeding, Faculty of Organic Agricultural Sciences, University of Kassel, Witzenhausen, Germany

^b Institute of Animal Breeding and Domestic Animal Genetics, Faculty of Agricultural Sciences, Nutritional Sciences, and Environmental Management, Justus Liebig University, Gießen, Germany

Abstract

For dairy cattle breeding as well as animal welfare science reliability and validity issues regarding measures of cows' responsiveness towards humans are relevant. The objective of this study was to investigate reliability and validity aspects of four selected behavioural measures at the feeding gate (BM) and to compare them with existing breeding traits and with observations of cows' behaviour during milking. This included the analysis of criterion validity at animal and at herd level and of repeatability of the BM over time. Data on avoidance distances (AD), tolerance to tactile interaction (TTI), release behaviour (RB), and qualitative behaviour assessment (QBA) were collected on a total of 33 dairy farms. The breeding traits average milk flow (AMF), milking speed (MS), and milking temperament (MT) had been recorded per test-day or during linear description by milk testing or breeding associations. Inter- and intra-test associations were investigated by Spearman rank or Pearson correlation analysis. At animal level, significant correlations of different strengths were found within

BM ranging from $r_s = 0.28$ ($p < 0.01$, $n = 1890$) between AD and TTI to $r_s = 0.74$ ($p < 0.01$, $n = 582$) between RB and QBA, and between MS and MT ($r_s = 0.62$, $p < 0.01$, $n = 269$). No significant correlations were found between BM and breeding traits. For analyses at herd level, medians and percentages of certain categories were calculated and the average number of stepping (STEP) and kicking (KICK) per cow during milking was recorded. Between all BM closer correlations were found than at animal level. STEP correlated with KICK ($r_s = 0.80$, $p < 0.01$, $n = 24$), while no further significant correlations were found between BM, STEP or KICK and breeding traits. For analysis of repeatability BM were applied repeatedly on three farms following a three-week interval. Correlations ranged from $r_s = 0.33$ ($p < 0.01$, $n = 67$) for TTI to $r_s = 0.73$ ($p < 0.01$, $n = 23$) for QBA. Apparently, behavioural measures at the feeding gate reflect partly similar and partly different aspects of cows' responsiveness towards humans. The present study underlines methodological problems that exist with MT and MS, and questions the relationship between milkability, milking behaviour and responsiveness in the barn. The analyses show that AD, TTI, and RB appear to be vulnerable to short- or mid-term influences. Overall, QBA turned out to be a promising measure for breeding and welfare assessment.

Keywords: dairy cattle, human-animal relationship, reliability, validity, breeding traits

1. Introduction

The level of fearfulness of dairy cows towards humans has an impact on their welfare including health as well as on dairy production (reviewed by Hemsworth, 2003; Rushen and de Passillé, 2010; Waiblinger et al., 2006). Moreover, fearful or highly responsive cows may reduce work safety and quality for the stockperson, e.g. during milking routine (Breuer et al., 2000) or when moving the cows to claw trimming (Lindahl et al., 2016). The level of fearfulness is determined by experiences the animal has made, in interaction with their individual genetic disposition. The animals' responses towards humans are indicators of the human-animal relationship (HAR), receiving increasing attention in animal welfare science for several years. Also in dairy cattle breeding, functional traits reflecting

the animals' responsiveness towards humans are gaining interest. In behavioural traits, such as milking temperament (MT), estimated heritabilities are ranging from $h^2 = 0.07$ to $h^2 = 0.47$ (reviewed by Haskell et al., 2014), making them amenable for breeding selection.

To assess responsiveness, fear and confidence towards humans, a broad spectrum of behavioural measures has been developed and applied in various animal welfare studies (reviewed by Forkman et al., 2007; Waiblinger et al., 2006). A widely used HAR-measure in dairy cows is the avoidance distance (AD) towards an unfamiliar person approaching the cow at the feeding place or inside the barn. Aspects of its reliability and criterion validity in individual animals and at herd level have repeatedly been investigated (e.g. Ebinghaus et al., 2016; Rousing and Waiblinger, 2004; Windschnurer et al., 2008). In addition, stepping and kicking behaviour during milking is suggested to indicate the quality of HAR at herd level (e.g. Dodzi and Muchenje, 2011; Rousing et al., 2004). In research focussing on the genetics of animals' behaviour towards humans, different handling tests are routinely used in order to assess the animals' temperament. They are applied during confinement in a chute as well as after release from restraint. While the animal is restrained in a head gate, the behavioural reaction is assessed on a scale from quiet to extremely excited (chute score / crush score). Various studies have combined this test with the assessment of exit behaviour – either on a multistage scale or by measuring the exit speed. These tests have been developed and applied to assess beef cattle (e.g. Burrow et al., 1988; Burrow and Dillon, 1997; Cafe et al., 2011; Grandin, 1993; Hoppe et al., 2010; Lanier and Grandin, 2002) predominantly, but have also been conducted to assess dairy cows more recently (e.g. Dodzi and Muchenje, 2011; Sutherland and Huddart, 2012; Sutherland et al., 2012). Aspects of reliability and criterion validity have likewise been investigated in beef cattle (e.g. Curley et al., 2006; Grignard et al., 2001) and in dairy cows (Gibbons et al., 2011).

In recent years, qualitative behaviour assessment (QBA) developed by Wemelsfelder et al. (2000, 2001) has been adopted to assess the animals' behavioural reactions and body language in various handling situations (dairy cows: Ebinghaus et al., 2016; calves: Ellingsen et al., 2014; beef cattle: Sant'Anna and da Costa, 2013; Stockman et al., 2011, 2012).

To genetically improve dairy cows' manageability, breeding associations routinely record the milking temperament (MT), which is related to the cows' behaviour during milking routine. This is subjectively assessed by the animal owner using a multistage scale from very nervous to very quiet (Adamczyk et al., 2013). Along with MT, breeding associations in Germany and other countries also record the cows' milkability by means of subjective classification of milking speed (MS). Alternatively, milking speed is measured as average milk flow per minute (AMF) (e.g. Interbull, 2009; VIT, 2016), showing a moderate genetic background (Santos et al., 2015). AMF is recorded monthly via Lactocorder, or MS classified by the animal owner within linear assessment of exterior traits once during the first lactation. Which traits are recorded on-farm, differs regionally in Germany. If recorded in combination, AMF and MS are summarised with a weighting of 50 : 50 (VIT, 2016). Milkability traits and MT are not included in the total merit index within dairy cattle breeding programmes as yet, but reported as relative breeding values (VIT, 2016).

However, information on reliability and validity aspects of these breeding traits is insufficient (reviewed by Adamczyk et al., 2013; Haskell et al., 2014). On a phenotypic level, AMF is influenced by anatomical and physiological factors and by milking management, such as pre-milking operations (e.g. Bruckmaier and Blum, 1996; Sandrucci et al., 2007). To lesser degrees it is also influenced by the cows' responsiveness towards humans. Furthermore, MT and MS scores based on animal owner interviews may not be sufficiently reliable, particularly on farms with larger herd sizes and with specialised work organisation (VIT, 2016). Thus, there is an urgent need for more reliability and validity information and for improved breeding traits referring to dairy cows' behaviour towards humans (Adamczyk et al., 2013). At the same time, also studies investigating the HAR in the context of animal welfare science have to deal with partly low levels of reliability of current measures and uncertainties regarding their validity (de Passillé and Rushen, 2005).

In a preceding pilot study, four selected or modified behavioural measures of dairy cows' responsiveness towards humans at the feeding gate (BM) have been identified as practically applicable within a linear assessment for breeding evaluation on three different farms, and as sufficiently repeatable in terms of inter- and intra-observer reliability (Ebinghaus et al., 2016). These measures

were the avoidance distance (AD), the tolerance to tactile interaction (TTI), the behaviour during and after release from restraint (RB), and the general cow's responsiveness during a simulated linear assessment expressed as a QBA-score.

In order to recommend HAR-measures for breeding and welfare assessment, the objective of the present study was to extend reliability and validity testing of these four BM under varying on-farm conditions and to compare them with the existing breeding traits AMF, MS and MT as well as observations of the cows' behaviour during milking. This included the analysis of criterion validity of all measures at individual animal and at herd level, as well as of repeatability of the BM over time.

2. Farms, animals, material and methods

2.1 Farms and animals

Data were collected in Middle and Northern Germany during the winter period 2014/2015 and 2015/2016 on 33 dairy farms (24 organic and eight conventional farms) equipped with loose housing. A total of 24 farms provided summer pasture for all cows, five farms for dry cows only, and four farms were managed with zero grazing. Nine farms used automatic milking systems (AMS), the others milked in fishbone or tandem milking parlours. While 11 organic farms kept horned cows, the other 14 organic and eight conventional farms kept dehorned or genetically hornless cows. All herds consisted of > 50 % Holstein Friesian cows. Herd sizes ranged from 29 to 530 cows (mean = 102, sd \pm 105, median = 81).

Since the cows had to be fixed in the feeding gate for the application of the behavioural measures TTI, and RB, all farms were at least partly equipped with self-locking feeding gates.

Behavioural assessments were conducted once on each farm at one or a maximum of two consecutive days, depending on the herd size. To test intra-test consistency, application of behavioural measures was repeated on three farms after a period of about three weeks (21, 24, and 25 days, respectively).

2.2 Behavioural measures at individual animal level

The following measures were applied at individual animal level: 1) avoidance distances towards an unfamiliar person at the feeding place (AD), 2) tolerance to tactile interaction (TTI), 3) behaviour during release from restraint in the feeding gate (RB), and 4) cows' behaviour and body language in a simulated linear assessment by means of Qualitative Behaviour Assessment (QBA).

Lactating cows of all parities (26.5% primiparous), and dry cows only when kept in the same group with lactating cows were tested by using the measures AD, TTI, and RB. Pregnant heifers were not tested. Since QBA requires a high level of concentration from the observer, this method was carried out on a sample of 12 to 30 cows with an equal distribution of different parities per farm.

Data collection was conducted by seven trained observers altogether (six female, one male; all wearing green overalls and gumboots, differently experienced in working with cows and in behavioural observation). Inter-observer reliabilities for all observers were tested prior to data collection and acceptable agreements ($r_s = 0.71 - 0.94$) were reached. For analysis of intra-test consistency, recordings of TTI, RB, and QBA were done by the same observer, whereas for organisational reasons AD was applied by different trained observers at first and repeated measurement.

The assessments started in the morning after feeding, when the cows were fixed in the feeding gate. The measures were always applied in the same order: at the beginning, one observer located on the feeding bunk measured the AD at the feeding place according to the Welfare Quality[®] protocol for cattle (Welfare Quality[®], 2009). Subsequently, another observer (blind to the assessment of AD) applied the TTI and RB tests and carried out the QBA according to Ebinghaus et al. (2016). For the TTI and RB tests the cows remained fixed in the feeding gate. The observer simulated a linear assessment, looking at the test cow from behind, left and right hand side for about 30 seconds, and then approached the cow from one side and stroked three times along the back and down the flank. The cows' behavioural reaction was rated on a 5-point scale, with 1: cow stays calm, 2: cow lowers hindquarters at least over the duration of two strokes, or steps maximum twice, 3: cow steps maximum five times, 4: cow steps more than five times or kicks at least once, 5: cow reacts violently, touching

barely possible. RB was then assessed during and after opening the feeding gate by the observer, and rated on a similar 5-point scale, with 1: cow stays calm, leaves the feeding place hesitantly, 2: cow leaves walking in intermediate speed, 3: cow leaves walking fast, 4: cow leaves running or jumping, 5: cow reacts violently, opening the feeding gate barely possible. AD, TTI and RB were recorded using digital record sheets in Microsoft Excel on a Sony Xperia Z2 tablet computer.

Afterwards, for the QBA the observer rated the cows' body language with respect to the human-animal interactions during the entire TTI and RB tests. For the assessment a fixed list of 20 descriptors was used, which had specifically been developed for this purpose by Ebinghaus et al. (2016) based on Wemelsfelder et al. (2009). The degree of the expression of each descriptor was marked on a visual analogue scale ranging from 0 mm (minimum) to 125 mm (maximum), using the software QBA App 1.0.7 (www.egenes.co.uk/qba) on a Sony Xperia Z2 tablet computer.

The multivariate QBA data of all assessed cows were reduced to dimensions by a principal component analysis (PCA, eigenvalue > 1, maximum two components, correlation matrix, without rotation, SPSS 24). For further analyses, the cows' individual scores of the PC1 were used.

2.3 Breeding traits related to dairy cows' responsiveness towards humans

Measures of average milk flow (AMF), scores for milking speed (MS) and milking temperament (MT), measured per test-day during milk recordings or recorded during linear description for conformation traits during the first lactation of individual cows by the breeding association between 2007 and 2015, were requested at VIT IT-solutions for animal production, Verden, Germany. Data on AMF were available for 67 cows of 11 different farms; data on MS and MT were available for 269 and 270 cows of 19 farms. AMF had been recorded in milk kg/min. MS and MT had been assessed by subjective evaluation by the animal owner on a 5-point scale from 1 (very slow milking / very nervous) to 5 (very fast milking / very quiet) (W. Ruten, VIT, personal communication, 12.09.2016).

2.4 Behavioural measures and breeding traits at herd level

In order to analyse inter-test correlations at herd level, medians were calculated for AD (ADmedian), QBA (QBAmmedian), and AMF (AMFmedian). With regard to AD, also the percentages of cows that allowed touching (AD% touch) and of those that avoided the approaching human at ≥ 100 cm (AD% ≥ 100) were considered. The percentages of stronger reacting cows with scores > 2 in the TTI (TTI% >2) and RB test (RB% >2), and the percentages of slower milking and more nervous cows with scores < 4 for the breeding traits MS (MS% <4) and MT (MT% <4) were calculated. Only those farm level data were used for which at least five observations per farm were available (AD, TTI, RB, QBA: n = 33; AMF: n = 6; MS, MT: n = 15).

The frequency of stepping (STEP) and kicking (KICK) behaviour per cow during milking was recorded in the milking parlour. Since on most farms identification of individual cows in the milking parlour was impossible or excessively time-consuming, analysis was only done at herd level. Observations and analyses were done according to Waiblinger et al. (2002) and Ivemeyer et al. (2011). STEP was defined as weight shifting with an elevation of the hind foot of ≤ 15 cm; KICK was defined as a distinct raising of the hind foot of > 15 cm. Each STEP and KICK was counted when the milker was working with a cow or standing within a radius of 0.5 m. If a farm had more than one regular milker (i.e. milking at least four times a week), a maximum of two milking routines with different regular milkers was observed, and the data were summarised at herd level. Weighted averages were used: non-qualified and non-permanent staff was weighted with the factor 0.7; all others were weighted with the factor 1.0. On farms using automatic milking systems (AMS), STEP and KICK were not recorded. For data collection during milking Noldus Pocket Observer 3.2 software was used on Sony Xperia Z2 tablet computers.

2.5 Statistical data analysis

Data were analysed using IBM SPSS statistics 24. For all measures and breeding traits normal distribution was ascertained visually via QQ plots. Inter-test correlations at individual animal and herd

level were analysed using the Pearson correlation for normally distributed data and the Spearman rank correlation for non-normally distributed data.

Intra-test consistencies of AD, TTI, RB and QBA were analysed at individual animal level by calculating Spearman rank correlation coefficients.

Correlation coefficients are referred to as follows: < 0.2 as very weak and negligible, $0.2 - 0.4$ as low, $0.4 - 0.7$ as moderate, $0.7 - 0.9$ as high, and > 0.9 as very high (Martin and Bateson, 1986: 134; Rowntree, 1991).

3. Results

Principal component analysis (PCA) of the QBA data explained 67.5 % of variance on the first principal component (PC1) and 7.4 % on the second principal component (PC2). PC1 was characterised by descriptors relating to relaxation/attraction/trust on the negative and descriptors relating to fear/distress/aversion on the positive end (Fig. 1).

Figure 1 near here

Descriptive statistics for all behavioural measures and breeding traits at individual animal level and at herd level are shown in Table 1 and 2.

Table 1 near here

Table 2 near here

3.1 Inter-test correlations at individual animal level

Significant correlations of different strengths were found among the behavioural measures AD, TTI, RB and QBA, as well as between MT and MS. No significant correlations were found between the behavioural measures and breeding traits (Tab. 3).

Table 3 near here

3.2 Inter-test correlations at herd level

At herd level, significant correlations of different strengths were found between all behavioural measures, except for STEP and KICK. STEP correlated only significantly with KICK. A marginally significant low correlation was found between STEP and QBAmmedian. No significant correlations were found between behavioural measures and breeding traits and within breeding traits, except for one marginally significant negative correlation between ADmedian and AMFmedian (Tab. 4).

Table 4 near here

3.3 Intra-test consistencies at individual animal level

Repeated assessments after a time interval of three weeks showed significant correlations of different strengths for all behavioural measures (Tab. 5).

Table 5 near here

4. Discussion

In order to identify appropriate HAR-measures for on-farm welfare assessment and as breeding traits, the present study investigated aspects of validity in terms of phenotypic correlations between and within existing breeding traits and behavioural measures as well as repeatability of behavioural measures over time.

4.1 Criterion validity

The most salient results were the largely absent associations among the existing breeding traits and with the behavioural HAR-measures. The only significant (moderate) correlation at individual animal level was found between the scores for milking speed (MS) and milking temperament (MT), which was stronger than in previous studies in Australian dairy cows ($r = 0.18$; Beard, 1993) and in New Zealand dairy cows ($r = 0.20 - 0.44$; Cue et al., 1996; Sosa-Ferreira et al., 1990). Descriptive statistics show that the rating scales for both traits were applied only to a limited extent (Table 2): more than 50 % of cows received intermediate scores and extreme scores were barely used, which was similarly

found in larger data sets of Canadian dairy cows (Sewalem et al., 2011) and Swiss dairy cows (Ilahi and Kadarmideen, 2004). It is possible that this tendency towards intermediate scores contributed to an apparently closer correlation. Additionally, infrequent contact to individual cows due to herd sizes and milking systems has probably played a part in improper allocation of MS and MT scores: six of the 19 farms, from which data on cows' MS and MT was available, worked with herd sizes > 100, and further six farms worked with automatic milking systems (AMS). In addition, animal owners might be less experienced in scoring on a standardised scale than independent experts. Animal owners may also be biased by their generalised assessment of a cow's quality and performance. This would explain the non-significant weak correlation between MS and measured average milk flow (AMF), although both traits refer to milk flow linearly from slow to fast milking. However, it should be taken into account that the sample size (n = 38 cows) was lower in relation to high sample sizes in behavioural measures at animal level (n = 582-1,890 cows).

Also between behavioural measures and breeding traits no significant correlations were found, neither at individual animal level nor at herd level. It must be noticed, however, that low sample sizes in the case of AMF at herd level (n = 6 farms) likely reduce the conclusiveness of results. Strong effects of single farms within this low sample size cannot be ruled out. The single marginally significant correlation coefficient of 0.771 (ADmedian – AMFmedian, p = 0.072) reflects a tendency that herds with cows showing shorter avoidance distances at the feeding place also have a faster average milk flow. Low to moderate correlations between HAR-measures and milk yield (Breuer et al., 2000; Waiblinger et al., 2002) as well as significant effects of aversive handling on milk yield and amount of residual milk (Rushen et al., 1999) point into the same direction. Nevertheless, milk flow is also affected by other factors, such as physiology and milking management. Sandrucci et al. (2007) found influences by lactation number, days in milk (DIM) and pre-milking operations, such as teat stimulation and attachment delay. In general, it has to be considered that data regarding AMF, MS, and MT had been recorded in the first lactation and thus partly some years before the behavioural measures have been applied for this study. Individual responsiveness, management or husbandry conditions might have changed and affected these results.

Surprisingly, also STEP and KICK behaviour in the milking parlour (which could only be recorded at herd level) showed neither associations with MT and milk flow measures, nor with any other behavioural HAR-measure, except for one marginally significant low correlation between STEP and QBAmmedian. Further associations with the avoidance distance (AD) would have been expected according to Waiblinger et al. (2002), who found low to moderate associations between stockpersons' behaviour and both cows' behavioural responses in the barn and in the milking parlour. However, in the investigated herds the cows' behaviour during milking was apparently influenced by other factors than the behaviour towards humans in the barn. Frequencies of stepping and kicking might have been affected by functionality of the milking machine or quality of teat cups. Similarly, Ivemeyer (2010) found only weak associations between the cows' behaviour during milking (KICK) and AD ($r_s = 0.29$). At individual animal level, Rousing et al. (2004) found a negative association between KICK and the percentage of cows that allowed touching in the AD test (odds ratio 3.5). However, associations between STEP and the percentage of cows that avoided an approaching human at > 2 m were only found in cows with major integument lesions (odds ratio 3.2) (Rousing et al., 2004). As a consequence, the results presented in this study do not support the general notion that milking ease is affected by HAR, at least not in the range of cows' fearfulness or responsiveness found in this study.

Results of the majority of HAR-tests in the barn at individual animal level showed moderate to strong correlations, implying that they partly reflect similar and partly different aspects of the HAR as previously discussed by Ebinghaus et al. (2016). Even though behavioural measures were all conducted at the feeding place involving a person unfamiliar to the cow, the measures varied regarding the location of the human and regarding the intensity of interaction. AD recorded the cows' behavioural reaction towards a human approaching from the feeding bunk outside the barn, while records of the cows' tolerance to tactile interaction (TTI), release behaviour (RB), and QBA were taken by a human approaching the cow inside the barn. AD involved mere approach, whereas the other three measures involved forced handling – in the case of TTI with physical human-animal contact. In fact, between AD and TTI, the measures that differed the most regarding the test situation, only a low correlation was found. However, since previous studies found strong correlations between AD at the

feeding place and AD in the barn, both at herd level when considering medians and percentages of cows that could be touched (Windschnurer et al., 2008; $r_s = 0.79$ and $r = 0.82$, respectively) and at individual animal level (Ebinghaus et al., 2016; $r_s = 0.77$), not only the location of the human, but the whole test setting appears to make a difference. This suggests that the cow perceives a human approaching in the barn in a different way while restrained in the feeding gate or when a close human-animal interaction is forced.

Since TTI, RB, and QBA were assessed by the same person, a bias in scoring RB and QBA cannot be ruled out. However, the lower correlations between TTI and RB compared to correlations between AD and QBA assessed independently from each other, do not indicate such a bias.

Although conducted by the same observer, at the same location, and involving forced human-animal interaction, TTI and RB correlated only moderately. Deviations might be related to the level of control of the situation, which is suggested to be an important aspect for the animal when being handled (Boivin et al., 2003; Waiblinger et al., 2004). The cow stayed restrained in the feeding gate during recording TTI, but was released and free to decide the distance kept to the human during RB recording.

The cows' responses also depend on previous experiences related to the specific setting. For instance, for AD, apart from the general HAR, it might be relevant whether the cows experienced frequent manual feeding as a positive human-animal interaction or whether feeding is routinely done by machine; thus a frontally approaching human might be unexpected and frightening to the cow. An approach at the barn-side feeding alley might be associated for the cow with potentially stressful handling, such as artificial insemination or veterinary treatment (Waiblinger et al., 2004). On the other hand, if stockpersons regularly conduct inspection rounds through the herd, clean bedding and walkways manually, the animals might be familiar with the presence of humans. These interrelationships should be kept in mind and possibly be investigated further as potential sources of influence, limiting validity of the HAR-measures.

QBA in different handling situations might be a good way to avoid such bias, and in fact, this measure showed the highest correlations with all other HAR-measures. Even though closer correlations of

QBA with TTI and RB partly may have been due to assessment by the same observer, QBA likely captures joint aspects, too. In fact, also AD, which was assessed in a blinded manner by another observer, showed strongest correlations with QBA.

Additionally, deviations between the different measures may be due to different assessment methods. AD uses quantitative distance measurement, TTI and RB the observation of defined behaviours, QBA the qualitative assessment of body language. Hence, the different measures record different components of the behavioural reaction or exclude others. Again, with QBA it appears possible to capture most aspects of response.

Additionally, it was conspicuous that herd medians or percentages of cows in specific categories of AD, TTI, RB, and QBA correlated stronger with each other than the data at individual animal level. These results suggest that at farm level and in a summarized form all the different aspects of the cows' responses measured by the different tests allow rather general conclusions on the quality of the human-animal relationship (HAR).

4.2 Repeatability of behavioural measures over time

Except for QBA, repeatability of the measures on animal level over time was not satisfactory when a limit of $r_s = 0.7$ is set (Knierim and Winckler, 2009). However, for the repeated AD measurements ($r_s = 0.60$) the change between observers for organisational reasons must be taken into account, despite good agreements between the observers ($r_s = 0.86 - 0.93$). Therefore, these results cannot be directly compared with the other measures recorded by the same person. However, De Rosa et al. (2003) found a similar repeatability of AD measurements on three occasions with time intervals of 15 – 20 days of $W = 0.43 - 0.64$, although they interpreted it as satisfactory. Repeated measurements on two occasions with a time interval of 4 – 5 days in a study by Rousing and Waiblinger (2004) showed a Kappa coefficient_{weighted} = 0.503, which is just a moderate agreement according to the general interpretation of Kappa (e.g. Fleiss et al. 2003). On the other hand, at herd level, Winckler et al. (2007) found strong correlations between five consecutive measurements of AD over one year ($r_s = 0.79 - 0.91$).

The present results of the TTI and RB tests support results of a study on the repeatability of the chute and flight speed test in dairy cows: Gibbons et al. (2011) found a low to moderate repeatability over four-week intervals of the flight speed test recorded by speed measurement ($r_s = 0.31 - 0.47$), but no significant correlation of repeated chute tests.

A previous study investigated the intra-day variation of QBA with regard to the emotional state at herd level in dairy cows and found consistent QBA-scores in six of ten assessed farms (Gutmann et al., 2015). The reasonable intra-day repeatability found by Gutmann et al. (2015) as well as the good mid-term repeatability found in the present study, could be explained by the methodically broader perspective on the animals' responses. Since the method of QBA does not measure or count specific behaviours, but uses various descriptors for behavioural expressions, spontaneous random reactions might have a smaller effect.

Moreover, effects of less regular handling routines, such as claw trimming, artificial insemination or veterinary treatment, might have an additional influence on the cows' behavioural reaction towards humans in the short- or mid-term, leading to limited repeatability. Further influencing factors could be physiological aspects, such as days in milk (DIM) or oestrus stage. Recording rules such as exclusion of cows in oestrus can likely help to increase consistency of information on the cows' responsiveness or the HAR; however, systematic investigations are not available.

Conclusions

The lack of phenotypic correlations among existing breeding traits and with behavioural HAR-measures poses questions concerning their validity. From the four investigated alternative measures – AD, TTI, RB, and QBA – the latter turned out to be the most promising with regard to repeatability over time and criterion validity. Apparently, QBA captures different aspects of the HAR most comprehensively and is less vulnerable to effects due to a certain test setting. However, in order to be recommended as behavioural measure for dairy cattle breeding programmes, further validity aspects, such as effects of routine procedures or lactation status, heritability and genetic correlations with other traits have to be investigated. Closer correlations between AD, TTI, RB, and QBA at herd rather than

on individual level suggest that in this summarized form they all allow general conclusions on the quality of the HAR on the farm. Within the range of responsiveness towards humans found on the investigated farms, the cows' behaviour during milking was apparently stronger influenced by other factors than by the HAR.

Acknowledgements

We are very grateful to all participating farmers for their openness and cooperativeness, and to Françoise Wemelsfelder and Tomasz Krzyzalewski (SRUC, Scotland) for support in digital recording of QBA data.

The present study is part of the LOEWE research cluster “Animals – Humans – Society: Interdisciplinary Animal Studies” at the University of Kassel, Germany and has been funded by the Hessen State Ministry of Higher Education, Research and the Arts.

References

- Adamczyk, K., Pokorska, J., Makulska, J., Earley, B., Mazurek, M., 2013. Genetic analysis and evaluation of behavioural traits in cattle. *Livestock Science* 154, 1-12.
- Beard, K., 1993. Genetic evaluation for milking speed, temperament, likability and survival in Australia. *Interbull Bulletin* 8, 1-4.
- Boivin, X., Lensink, J., Tallet, C., Veissier, I., 2003. Stockmanship and farm animal welfare. *Animal Welfare* 12, 479-492.
- Breuer, K., Hemsforth, P.H., Barnett, J.L., Matthews, L.R., Coleman, G.J., 2000. Behavioural response to humans and the productivity of commercial dairy cows. *Applied Animal Behaviour Science* 66, 273-288.
- Bruckmaier, R.M., Blum, J.W., 1996. Simultaneous recording of oxytocin release, milk ejection and milk flow during milking of dairy cows with and without prestimulation. *Journal of Dairy Research* 63, 201-208.
- Burrow, H. M., Dillon, R.D., 1997. Relationships between temperament and growth in a feedlot and commercial carcass traits of *Bos indicus* crossbreds. *Animal Production Science* 37, 407-411.
- Burrow, H.M., Seifert, G.W., Corbet, N.J., 1988. A new technique for measuring temperament in cattle. *Proc. Australian Society of Animal Production* 17, 154-157.
- Cafe, L.M., Robinson, D.L., Ferguson, D.M., McIntyre, B.L., Geesink, G.H., Greenwood, P.L., 2011. Cattle temperament: persistence of assessments and associations with productivity, efficiency, carcass and meat quality traits. *Journal of Animal Science* 89, 1452-1465.
- Cue, R.I., Harris, B. L., Rendel, J. M., 1996. Genetic parameters for traits other than production in purebred and crossbred New Zealand dairy cattle. *Livestock Production Science* 45, 123-135.
- Curley, K.O., Paschal, J.C., Welsh, T.H., Randel, R.D., 2006. Technical note: Exit velocity as a measure of cattle temperament is repeatable and associated with serum concentration of cortisol in Brahman bulls. *Journal of Animal Science* 84, 3100-3103.

- de Passillé, A.M., Rushen, J., 2005. Can we measure human-animal interactions in on-farm animal welfare assessment? Some unresolved issues. *Applied Animal Behaviour Science* 92, 193–209.
- De Rosa, G., Tripaldi, C., Napolitano, F., Saltalamacchia, F., Grasso, F., Bisegna, V., Bordi, A., 2003. Repeatability of some animal-related variables in dairy cows and buffaloes. *Animal Welfare* 12, 625–629.
- Dodzi, M.S., Muchenje, V., 2011. Avoidance-related behavioural variables and their relationship to milk yield in pasture-based dairy cows. *Applied Animal Behaviour Science* 133, 11–17.
- Ebinghaus, A., Ivemeyer, S., Rupp, J., Knierim, U., 2016. Identification and development of measures suitable as potential breeding traits regarding dairy cows' reactivity towards humans. *Applied Animal Behaviour Science* 185, 30-38.
- Ellingsen, K., Coleman, G.J., Lund, V., Mejdell, C.M., 2014. Using qualitative behaviour assessment to explore the link between stockperson behaviour and dairy calf behaviour. *Applied Animal Behaviour Science* 153, 10-17.
- Fleiss, J.L., Levin, B., Paik, M.C., 2003. *Statistical Methods for Rates and Proportions*. John Wiles & Sons, Hoboken, pp. 598-626.
- Forkman, B., Boissy, A., Meunier-Salaün, M.C., Canali, E., Jones, R.B., 2007. A critical review of fear tests used on cattle, pigs, sheep, poultry and horses. *Physiology & Behavior* 92, 340-374.
- Gibbons, J.M., Lawrence, A.B., Haskell, M.J., 2011. Consistency of flight speed and response to restraint in a crush in dairy cattle. *Applied Animal Behaviour Science* 131, 15-20.
- Grandin, T., 1993. Behavioral agitation during handling of cattle is persistent over time. *Applied Animal Behaviour Science* 36, 1-9.
- Grignard, L., Boivin, X., Boissy, A., Le Neindre, P., 2001. Do beef cattle react consistently to different handling situations? *Applied Animal Behaviour Science* 71, 263-276.
- Gutmann, A. K., Schwed, B., Tremetsberger, L., Winckler, C., 2015. Intra-day variation of Qualitative Behaviour Assessment outcomes in dairy cattle. *Animal Welfare* 24, 319-326.

- Haskell, M.J., Simm, G., Turner, S.P., 2014. Genetic selection for temperament traits in dairy and beef cattle. *Frontiers in genetics* 5, 368, <https://doi.org/10.3389/fgene.2014.00368>.
- Hemsworth, P.H., 2003. Human-animal interactions in livestock production. *Applied Animal Behaviour Science* 81, 185-198.
- Hoppe, S., Brandt, H.R., König, S., Erhardt, G., Gauly, M., 2010. Temperament traits of beef calves measured under field conditions and their relationships to performance. *Journal of Animal Science* 88, 1982-1989.
- Ilahi, H., Kadarmideen, H. N., 2004. Bayesian segregation analysis of milk flow in Swiss dairy cattle using Gibbs sampling. *Genetics Selection Evolution* 36, 1.
- Interbull, 2009. Genetic Evaluation for workability traits, August 2009.
http://www.interbull.org/web/static/mace_evaluations_archive/Workability/wo-aug09.html
(accessed 15.02.2017).
- Ivemeyer, S., 2010. Einfluss der Mensch-Tier-Beziehung auf die Eutergesundheit von Milchkühen [Effect of human-animal relationship on udder health of dairy cows]. PhD thesis, Faculty of Organic Agricultural Sciences, University of Kassel, Germany.
- Ivemeyer, S., Waiblinger, S., Knierim, U., 2011. Effect of human-animal relationship and management on udder health in Swiss dairy herds. *Journal of Dairy Science* 94, 5890-5902.
- Knierim, U., Winckler, C., 2009. On-farm welfare assessment in cattle – validity, reliability and feasibility issues and future perspectives with special regard to the Welfare Quality® approach. *Animal Welfare* 18, 451-458.
- Lanier, J.L., Grandin, T., 2002. The relationship between *Bos taurus* feedlot cattle temperament and cannon bone measurements. In: *American Society of Animal Science Proceedings, Western Section*, 53, 97-98.

- Lindahl, C., Pinzke, S., Herlin, A., Keeling, L.J., 2016. Human-animal interactions and safety during dairy cattle handling – Comparing moving cows to milking and hoof trimming. *Journal of Dairy Science* 99, 2131-2141.
- Martin, P., Bateson, P., 1986. *Measuring behaviour*. Cambridge University Press, Cambridge.
- Rousing, T., Bonde, M., Badsberg, J.H., Sørensen, J.T., 2004. Stepping and kicking behaviour during milking in relation to response in human–animal interaction test and clinical health in loose housed dairy cows. *Livestock Production Science* 88, 1-8.
- Rousing, T., Waiblinger, S., 2004. Evaluation of on-farm methods for testing the human-animal relationship in dairy herds with cubicle loose housing systems – test-retest and inter-observer reliability and consistency to familiarity of test person. *Applied Animal Behaviour Science* 85, 215-231.
- Rowntree, D., 1991. *Statistics without tears: a primer for non-mathematicians*. Penguin Books, London.
- Rushen, J., de Passillé, A.M.B., Munksgaard, L., 1999. Fear of people by cows and effects on milk yield, behavior and heart rate at milking. *Journal of Dairy Science* 82, 720-727.
- Rushen, J., de Passillé, A.M., 2010. The importance of good stockmanship and its benefits for the animals. In: Grandin, T. (Ed.), *Improving animal welfare: a practical approach*. CAB International, Wallingford, 50-63.
- Sandrucci, A., Tamburini, A., Bava, L., Zucali, M., 2007. Factors affecting milk flow traits in dairy cows: results of a field study. *Journal of Dairy Science* 90, 1159-1167.
- Sant’Anna, A.C., da Costa, M.J.P., 2013. Validity and feasibility of qualitative behavior assessment for the evaluation of Nellore cattle temperament. *Livestock Science* 157, 254-262.
- Santos, L., Brügemann, K., Simianer, H., König, S., 2015. Alternative strategies for genetic analyses of milk flow in dairy cattle. *Journal of Dairy Science* 98, 8209-8222.

- Sewalem, A., Miglior, F., Kistemaker, G.J., 2011. Genetic parameters of milking temperament and milking speed in Canadian Holsteins. *Journal of Dairy Science* 94, 512-516.
- Sosa-Ferreya, C., Ahlborn-Breier, G., Blair, H., 1990. Multivariate Estimates Of Genetic Parameters For Traits Other Than Production In New Zealand Dairy Cows. In: AAABG Conference Proceedings 8, 335-337, <http://www.aaabg.org/livestocklibrary/1990/ab90065.pdf> (accessed 15.02.2017).
- Stockman, C.A., Collins, T., Barnes, A.L., Miller, D., Wickham, S.L., Beatty, D.T., Blache, D., Wemelsfelder, F., Fleming, P.A., 2011. Qualitative behavioural assessment and quantitative physiological measurement of cattle naïve and habituated to road transport. *Animal Production Science* 51, 240-249.
- Stockman, C.A., McGilchrist, P., Collins, T., Barnes, A.L., Miller, D., Wickham, S.L., Greenwood, P.L., Cafe, L.M., Blache, D., Wemelsfelder, F., Fleming, P.A., 2012. Qualitative Behavioural Assessment of Angus steers during pre-slaughter handling and relationship with temperament and physiological responses. *Applied Animal Behaviour Science* 142, 125-133.
- Sutherland, M.A., Huddart, F.J., 2012. The effect of training first-lactation heifers to the milking parlor on the behavioral reactivity to humans and the physiological and behavioral responses to milking and productivity. *Journal of Dairy Science* 95, 6983-6993.
- Sutherland, M.A., Rogers, A.R., Verkerk, G.A., 2012. The effect of temperament and responsiveness towards humans on the behavior, physiology and milk production of multi-parous dairy cows in a familiar and novel milking environment. *Physiology & Behavior* 107, 329-337.
- Waiblinger, S., Boivin, X., Pedersen, V., Tosi, M.V., Janczak, A.M., Visser, E.K., Jones, R.B., 2006. Assessing the human-animal relationship in farmed species: A critical review. *Applied Animal Behaviour Science* 101, 185–242.
- Waiblinger, S., Menke, C., Coleman, G., 2002. The relationship between attitudes, personal characteristics and behaviour of stockpeople and subsequent behaviour and production of dairy cows. *Applied Animal Behaviour Science* 79, 195–219.

- Waiblinger, S., Menke, C., Korff, J., Bucher, A., 2004. Previous handling and gentle interactions affect behaviour and heart rate of dairy cows during a veterinary procedure. *Applied Animal Behaviour Science* 85, 31-42.
- Welfare Quality®, 2009. Welfare Quality® assessment protocol for cattle (online). Welfare Quality® Consortium, Lelystad, <http://www.welfarequality.net/network/45848/7/0/40> (accessed 15.02.2017).
- Wemelsfelder, F., Hunter, E.A., Mendl, M.T., Lawrence, A.B., 2000. The spontaneous qualitative assessment of behavioural expressions in pigs: first explorations of a novel methodology for integrative animal welfare measurement. *Applied Animal Behaviour Science* 67, 193-215.
- Wemelsfelder, F., Hunter, T.E. Mendl, M.T., Lawrence, A.B., 2001. Assessing the ‘whole animal’: a free choice profiling approach. *Animal Behaviour* 62, 209-220.
- Wemelsfelder, F., Millard, F., De Rosa, G., Napolitano, F., 2009. Qualitative Behaviour Assessment. In: Forkman, B., Keeling, J. (Eds.), *Assessment of Animal Welfare Measures for Dairy Cattle, Beef Bulls and Veal Calves*, Welfare Quality® Reports No. 11, 215-224.
- Windschnurer, I., Schmied, C., Boivin, X., Waiblinger, S., 2008. Reliability and inter-test relationship of tests for on-farm assessment of dairy cows’ relationship to humans. *Applied Animal Behaviour Science* 114, 37-53.
- Winckler, C., Brinkmann, J., Glatz, J., 2007. Long-term consistency of selected animal-related welfare parameters in dairy farms. *Animal Welfare* 16, 197-199.
- VIT (2016): Estimation of Breeding Values for Milk Production Traits, Somatic Cell Score, Conformation, Productive Life and Reproduction Traits in German Dairy Cattle. Vereinigte Informationssysteme Tierhaltung, Verden, http://www.vit.de/fileadmin/user_upload/vit-fuers-rind/zuchtwertschaetzung/milchrinder-zws-online/Zws_Bes_eng.pdf (accessed 15.02.2017).

Table 1: Descriptive statistics of behavioural measures and breeding traits at individual animal level

individual level	n	mean	median	min – max	25 – 75%	sd
AD (cm)	2,017	36	20	0 - 200	5 - 50	43
QBA (PC1-score)	588	-0.006	-0.017	-1.804 - 2.393	-0.909 - 0.765	1.012
AMF (kg/min)	67	2.47	2.50	0.71 - 5.00	1.73 - 3.09	0.94
		score 1	score 2	score 3	score 4	score 5
TTI (% of cows)	1,897	36.9	31.5	20.1	9.1	2.4
RB (% of cows)	1,892	29.0	42.7	23.2	4.1	1.0
MS (% of cows)	269	1.5	7.1	57.6	31.6	2.2
MT (% of cows)	270	1.9	5.6	54.4	33.3	4.8

AD = avoidance distance, QBA = qualitative behaviour assessment (PC1-score = score regarding first principal component), AMF = average milk flow, TTI = tolerance to tactile interaction (score), RB = release behaviour (score), MS = milking speed (score), MT = milking temperament (score); definition of scores: see text

Table 2: Descriptive statistics of behavioural measures and breeding traits at herd level

herd level	n	mean	median	min - max	25 - 75%	sd
ADmedian	33	19	15	0 - 70	5 - 30	17
AD%touch	33	27	23	0 - 71	11 - 40	18
AD%\geq100	33	9	5	0 - 40	3 - 14	9
QBAmmedian	33	-0.026	0.159	-1.254 - 1.089	-0.507 - 0.443	0.640
TTI%$>$2	33	31	31	0 - 54	25 - 39	12
RB%$>$2	33	27	27	3 - 54	15 - 36	13
STEP	24	1.75	1.64	0.64 - 3.63	1.32 - 2.33	0.75
KICK	24	0.80	0.70	0.03 - 2.18	0.27 - 1.31	0.64
AMFmedian	6	2.61	2.65	1.66 - 3.47	2.03 - 3.16	0.64
MS%$<$4	15	69	78	18 - 88	66 - 84	20
MT%$<$4	15	61	67	18 - 82	55 - 72	19

ADmedian = median of avoidance distance, AD%touch = percentage of cows that could be touched in the AD test; AD% \geq 100 = percentage of cows that avoided the approaching human at \geq 100 cm,

QBAmmedian = median of qualitative behaviour assessment PC1-score, TTI%>2 = percentage of cows with tolerance to tactile interaction score > 2, RB%>2 = percentage of cows with release behaviour score > 2, AMFmedian = median of average milk flow, MS%<4 = percentage of cows with a milking speed score < 4, MT%<4 = percentage of cows with a milking temperament score < 4

Table 3: Inter-test correlations at individual animal level between behavioural measures (AD, TTI, RB, QBA) and breeding traits (MT, MS, AMF), analysed by calculation of Spearman rank correlation coefficients (r_s)

		TTI	RB	QBA	MT	MS	AMF
AD	r_s	0.281**	0.463**	0.539**	-0.046	-0.017	0.028
	p	< 0.001	< 0.001	< 0.001	0.455	0.783	0.828
	n	1,890	1,885	586	264	263	65
TTI	r_s		0.470**	0.683**	-0.068	0.039	-0.070
	p		< 0.001	< 0.001	0.279	0.535	0.599
	n		1,890	583	254	253	59
RB	r_s			0.738**	-0.095	0.006	-0.005
	p			< 0.001	0.133	0.920	0.971
	n			582	254	253	58
QBA	r_s				-0.063	0.029	-0.213
	p				0.567	0.791	0.411
	n				86	86	17
MT	r_s					0.617**	0.007
	p					< 0.001	0.968
	n					269	38
MS	r_s						0.186
	p						0.263
	n						38

** significant at the 0.01 level (two-tailed)

AD = avoidance distance (cm), TTI = tolerance to tactile interaction (score), RB = release behaviour (score), QBA = qualitative behaviour assessment (PC1-score), MT = milking temperament (score), MS = milking speed (score), AMF = average milk flow (kg/min)

Table 4: Inter-test correlations at herd level between behavioural measures and breeding traits, analysed by calculation of Pearson correlation or Spearman rank correlation (s) coefficients

measure		TTI%>2	RB%>2	QBAmmedian	STEP	KICK	AMFmedian	MS%<4	MT%<4
ADmedian	r/r _s	0.655^{**}	0.751^{s**}	0.654^{s**}	0.263 ^s	0.206 ^s	-0.771 ^{s1}	-0.124 ^s	0.183 ^s
	p	< 0.001	< 0.001	< 0.001	0.215	0.334	0.072	0.659	0.513
	n	33	33	33	24	24	6	15	15
AD%touch	r/r _s	-0.575^{s**}	-0.678^{s**}	-0.638^{s**}	-0.298 ^s	-0.272 ^s	0.657 ^s	-0.220 ^s	-0.304 ^s
	p	< 0.001	< 0.001	< 0.001	0.158	0.198	0.156	0.431	0.270
	n	33	33	33	24	24	6	15	15
AD%≥100	r/r _s	0.569^{s**}	0.667^{s**}	0.527^{s**}	-0.012 ^s	0.010 ^s	-0.543 ^s	-0.238 ^s	0.176 ^s
	p	0.001	< 0.001	0.002	0.954	0.961	0.266	0.393	0.530
	n	33	33	33	24	24	6	15	15
TTI%>2	r/r _s		0.737^{**}	0.667^{s**}	0.041 ^s	0.022 ^s	-0.314 ^s	0.020 ^s	0.206 ^s
	p		< 0.001	< 0.001	0.848	0.920	0.544	0.945	0.462
	n		33	33	24	24	6	15	15
RB%>2	r/r _s			0.711^{s**}	-0.042 ^s	-0.034 ^s	-0.600 ^s	-0.054 ^s	0.311 ^s
	p			< 0.001	0.845	0.875	0.208	0.849	0.259
	n			33	24	24	6	15	15
QBAmmedian	r/r _s				0.352 ^{s1}	0.258 ^s	-0.257 ^s	0.064 ^s	-0.045 ^s
	p				0.091	0.223	0.623	0.820	0.874
	n				24	24	6	15	15
STEP	r/r _s					0.799^{s**}	-1.000 ^s	0.067 ^s	-0.256 ^s
	p					< 0.001	na	0,864	0.506
	n					24	2	9	9

KICK	r/r _s			-1.000 ^s	0.250 ^s	-0.192 ^s
	p			na	0.516	0.620
	n			2	9	9
AMFmedian	r/r _s				0.000 ^s	-0.100 ^s
	p				1.000	0.873
	n				5	5
MS%<4	r/r _s					0.351 ^s
	p					0.199
	n					15

** significant at the level of 0.01 (two-tailed), 1 marginally significant on the level 0.10 (two-tailed),
na = not available due to low sample size

ADmedian = median of avoidance distances, AD%touch = percentage of cows that could be touched in the AD test; AD% \geq 100 = percentage of cows that avoided the approaching human at \geq 100 cm, TTI% $>$ 2 = percentage of cows with tolerance to tactile interaction score $>$ 2, RB% $>$ 2 = percentage of cows with release behaviour score $>$ 2, QBAmedian = median of PC1-scores of qualitative behaviour assessments, AMFmedian = median of average milk flow per minute, MS% $<$ 4 = percentage of cows with milking speed score $<$ 4, MT% $<$ 4 = percentage of cows with milking temperament score $<$ 4

Table 5: Spearman rank correlations (r_s) between repeated assessments for four behavioural measures

measure	n	r _s	p
AD (avoidance distance)	74	0.603	$<$ 0.001
TTI (tolerance to tactile interaction)	67	0.326	0.007
RB (release behaviour)	67	0.556	$<$ 0.001
QBA (qualitative behaviour assessment)	23	0.728	$<$ 0.001

Figure 1: Component plot of the PCA loadings of 20 QBA descriptors to assess dairy cows' responsiveness towards humans, based on correlation matrix, without rotation (n = 622)