A Proposal for Three Modifications for the Welfare Quality© Protocol for Dairy Cattle.

Frank J.C.M. van Eerdenburg*, Jan Hulsen †, Bert Snel ‡, Jan van den Broek* and Arjan Stegeman*

*Dept Farm animal health, Faculty of Veterinary Medicine, Utrecht University, Yalelaan 7, 3584 CL Utrecht, the Netherlands.
†Vetvice/CowSignals, Moerstraatsebaan 115, 4614 PC Bergen op Zoom, the Netherlands
‡DLV Advies, Postbus 546, 7400 AM Deventer, the Netherlands
Email: F.J.C.M.vanEerdenburg@UU.nl

ABSTRACT

To evaluate the Welfare Quality® assessment protocol (**WQ**), it was applied on 60 dairy farms in the Netherlands, with good, moderate and poor welfare in equal numbers, according to the veterinary opinion. Three (5%) farms scored Not Classified, 52 (86.7%) Acceptable and 5 (8.3%) Enhanced, no farm received a score Excellent. Because similar results were reported previously, it was concluded that the discriminative capacity of the WQ protocol was low under Dutch circumstances. It became clear that, as was also reported by others, only a few measures and criteria had a major impact on the end result. The farmers stated that assessment of the welfare level was very informative and valuable. However, they were not convinced that the result of the Qualitative Behavior Assessment (**QBA**) is something to be taken seriously and they did not consider this test a proper component in the determination of the level of welfare of their animals. We modified the WQ protocol in 3 ways: the calculation of the cleanliness of water points and integument alterations was changed in a more quantitative way and the QBA was omitted. After these modifications to the WQ protocol, the discriminative capacity appeared substantially higher. We, therefore, propose to modify the WQ protocol accordingly.

Keywords: Welfare assessment, water supply, integument alteration, Qualitative Behavior Assessment (QBA), acceptance by farmers.

INTRODUCTION

Ethical considerations and care for the animal, and an increasing concern of the general public, make animal welfare an issue on a dairy farm, [1]. Furthermore, an incentive for farmers to improve the conditions for the cows on their farms can be that an increase in the level of animal welfare is correlated with a higher milk yield [2,3]. How to measure animal welfare in an overall, objective, way, however, has been the subject of discussion for a long period of time and still is. For a long time, the focus has been on the measuring of biological functioning, which is necessary, but not sufficient. Welfare issues also relate to affective states, such as pleasure or suffering from pain [4]. Furthermore, for dairy cattle, the expression of natural behavior and access to pasture are also very important [5,6]. Although a 'gold standard' is still lacking, several protocols have been developed that measure welfare at a dairy farm. The Welfare Quality assessment protocol®[7] (**WQ**) is one of the most extensive ones and uses mainly animal based measures. These are parameters that are measured directly on/from the animals, like skin lesions or behavior, and not in the environment (resource based parameters). In total, 33 measures are taken on a farm that are integrated in 12 criteria. These 12 criteria are then further grouped into 4 principles: Good feeding, Good Housing, Good Health and Appropriate Behavior. Finally, an end qualification is calculated and this can be Excellent, Enhanced , Acceptable or Not Classified (Welfare Quality assessment protocol® , 2009). However, the WQ protocol is extensive and time consuming (almost a full day is needed for its execution), which has hampered its implementation as a routine, on farm, welfare check [8,9]. Furthermore, the relative contribution of certain measures/criteria is disputed [9-12]. In the study of Heath et al. (2014) it appeared that they could classify the final outcome of the WQ protocol correctly with the result for "Absence of prolonged thirst" in 88% of the farms. This problem is not only present in the dairy cattle protocol, but also in the chicken protocol, where Buijs et al. (2016) reported that the overall classification for 95% of the flocks could be explained by two measures only ('drinker space' and 'stocking density'). In the present

study, the WQ protocol for dairy cattle (2009) has been executed on 60 Dutch dairy farms, of various levels of welfare, to evaluate practicality and acceptance by the farmers. During the study, it became evident that the WQ protocol had a low discriminative capacity on the 60 farms as most of them were classified as acceptable. This corroborates the results of De Vries et al.[11], Heath et al. (2014), De Graaf et al. (ILVO, Gent, Belgium, personal communication) in Belgium, who classified 94 out of 111 farms as acceptable (none not classified or excellent) and Toma et al. [13] who had 9 farms with good, 25 acceptable and 1 not classified in a study in Scotland. We propose 3 modifications to the original WQ protocol [14], to increase the discriminative power and acceptance by the farmers.

MATERIALS & METHODS

The WQ protocol (2009) was applied on 60 dairy farms in the Netherlands. Four large veterinary practices, spread over the Netherlands in order to avoid possible regional effects, were asked to make a list of their dairy farmer clients. Each farm was classified as good, average or bad, based on the availability of good quality food & water, quality of housing, health and behavior. This mark was based on the impression of all the dairy cattle veterinarians of each practice, no specific assessment was done at this time (because there is no 'gold standard'). These were large practices with each more than 5 dairy cattle veterinarians, so it was not a subjective, individual opinion. Furthermore, because there were 4 practices involved, the influence of an individual opinion was minimized. Out of the lists, randomly, 60 farms were selected in such a way that in each of the 4 practices there were 5 good-, 5 average- and 5 bad farms. This was not used as a 'gold standard' or a representative sample of the Dutch dairy farms, but just to get a diverse quality of farms, in order to evaluate the assessment protocol over the full range of animal welfare status. The selected farmers were asked if they would be willing to participate in the project and, if not, the next farmer on the list was addressed. This occurred twice.

Of each practice at least one veterinarian was trained to execute the WQ protocol during a three day course, provided by the Welfare Quality consortium. The observers did not assess farms that they regularly visit and advise. One of the other observers assessed those farms.

The WQ protocol (2009) consists of several steps. It starts with measuring 33 parameters (indicators) that are converted into 12 criteria. These 12 criteria form the basis for the calculation of a score for 4 main principles: Feeding, Health, Housing and Behavior. Finally, an end qualification is computed: Not classified, Acceptable, Enhanced or Excellent. In this study, initially, the original protocol and calculations were used (Welfare Quality protocol for dairy cattle, 2009). After the measurements and calculations with the original protocol, 3 modifications were applied that are described below. These modifications did not require a new assessment, because they are alternative calculations of the original measures.

1) Absence of Thirst

The first modification of the WQ protocol, was the introduction of a weighted score for cleanliness of the drinkers. A clean drinker scored 1, a partially dirty 2, and a dirty one 3 points. After giving the score for the rest of the drinking related parameters measured, the total is divided by the average score for the cleanliness (see fig. 1). This number was then used in the calculations according to the original WQ protocol [14].

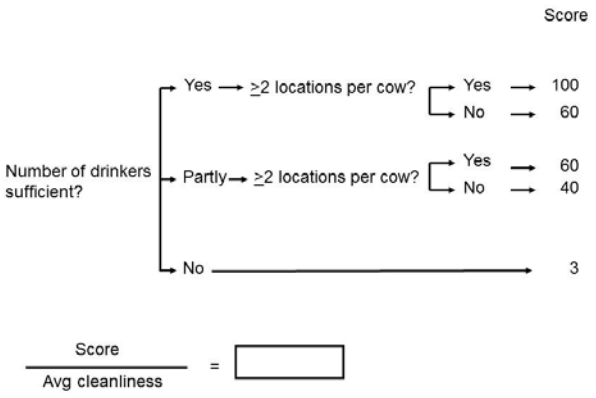

Figure 1: The scoring for water in the modifed WQ protocol for dairy cattle. For the determination of the number being sufficient, the requirements of the WQ protocol (2009) were used. Then it was checked if there were at least 2 drinking locations available per cow (WQ protocol, 2009). The cleanliness was scored in points per drinker: clean = 1; partly dirty = 2 and dirty = 3. The average of all drinkers is computed and used in the calculation.

2) Integument Alterations (hairless patches (HP) and lesions/swellings)

In the modified protocol, the HPs and lesions/swellings were assessed and counted as in the WQ protocol (2009). However, the average number of HPs, lesions and swellings per cow in the group was used in the calculations. Because a lesion or swelling is a more severe impairment for the welfare of the cow, it receives more weight in the calculations, similar to the WQ protocol (2009).

The Index for integument alterations was calculated as:

$$I = 100 - \frac{(2HP+5(lesions+swellings))x10}{5}$$

If $I < 65$ the score becomes: $(0,43 \times I) + (0,0065 \times I^2) + (0,00013 \times I^3)$

If $I > 65$ the score becomes: $29,9 – (0,94 \times I) + (0,015 \times I^2) + (0,00002 \times I^3)$

Where HP is the average number of HP's per cow and lesions + swellings are also the average number of lesions and swellings per cow. This index was then used instead of the one for integument alterations from the original WQ protocol (2009) in the calculations.

3) Qualitative Behavior Assessment (QBA)

The QBA was omitted in the modified protocol. Since this was the component with the largest weight for the principle of Appropriate Behavior score, the values for μ for the other criteria in the Choquet integral were doubled.

Statistical analysis

To model the relation between the four principles and the welfare scores, a multinomial regression model (base-line category logit) was used [15]. Residuals were checked. Two logits were modeled: the log-odds for acceptable versus not classified (**AvsNC**) and the log-odds for enhanced versus not-classified (**EvsNC**). Akaike's Information Criterion (**AIC**) was used for model reduction. For the important effects, Wald confidence intervals were calculated. Note that the Wald intervals need not be in accordance with the AIC.

RESULTS & DISCUSSION

The results for the original WQ protocol (2009) were: 3 (5%) farms with score Not Classified, 52 (86.7%) with a score Acceptable and 5 (8.3%) Enhanced, no farm received a score Excellent (Table 1). Since the farms were selected as having bad, average or good welfare, in equal numbers, this was not expected. Because there is not a 'gold standard' for animal welfare assessment, some degree of subjectivity is inevitable when weighing different measures [16]. So it could be that the farms were not selected in an appropriate way. However, analysis of the measurements of the farms showed that there were indeed substantial differences between parameters on farms (Table 1; figures 2-4). This way of selecting of the farms was comparable with Botreau et al [17], who used the 'general impression' of the observers of the farms in their study to compare 5 procedures that could form the basis of the calculations of aggregation of the measures in the WQ protocol. In the end, the way of computing that matched the 'general impression' of the observers best, was implemented in the WQ protocol as final step to categorize the farms [17]. In the present study, not just the general impression of one person was used, but several persons based their opinion on the availability of good quality food & water, quality of housing, health and behavior.

Due to the way of calculating and aggregation of the measures in the WQ protocol most farms received an 'Acceptable' as the end result. This implied that the WQ protocol (2009) did have little discriminative capacity under Dutch circumstances. As can be seen in figure 2-4, several farms had a substantial amount of problems, e.g. 18 farms had 10% or more severely lame cows, a disorder with a substantial impact on animal welfare. This appeared to be acceptable for the WQ protocol. Similar findings have been reported previously. In a study in England and Wales by Heath et al. (2014), all the 92 farms they assessed had a result as acceptable (35 farms) or enhanced (57 farms). Data from de Graaf et al. in Belgium (ILVO, Gent, Belgium, personal communication) confirmed this. Out of 111 farms they assessed 94 as acceptable vs 17 as enhanced (none not classified or excellent). Furthermore, Toma et al. (2017) had 9 farms categorized as enhanced, 25 acceptable and 1 not classified in a study in Scotland. Due to (almost) singular Hessian matrices the housing effects could not be estimated for the WQ and the WQ-modified scores. The principles Good Feeding and Appropriate Behavior were the variables related to the WQ-scores according to AIC. The Odds ratio's and their confidence intervals are presented in table 2.

Table 1: Overview of the principle- and end score of WQ and the modified version

	WQ - Original					WQ - Modified				
Farm	Feed	Hous	Health	Behav	Score	Feed	Hous	Health	Behav	Score
1	85	53	16	37	A	81	53	26	35	A
2	30	52	16	38	A	25	52	27	32	A
3	40	53	21	32	A	56	53	33	29	A
4	12	52	19	47	NC	12	52	34	45	A
5	85	65	21	39	E	81	65	36	33	E
6	40	53	18	45	A	15	53	26	38	A
7	63	53	23	43	A	63	53	41	34	A
8	27	53	16	50	A	36	53	30	44	A
9	15	64	44	52	A	15	64	56	46	A
10	15	53	17	24	NC	15	53	26	23	NC
11	62	53	28	42	A	53	53	39	35	A
12	32	64	23	38	A	38	64	36	35	A
13	24	52	16	51	A	24	52	25	49	A
14	15	65	23	39	A	15	65	30	28	A
15	25	53	19	37	A	31	53	33	33	A
16	23	69	41	18	A	37	69	57	7	NC
17	25	64	29	22	A	20	64	44	10	NC
18	62	59	35	36	E	58	59	55	33	E
19	37	59	24	47	A	48	59	32	37	A
20	65	72	25	57	E	60	72	41	55	E
21	22	52	32	17	A	22	52	52	9	NC
22	9	69	37	25	NC	9	69	60	9	NC
23	47	53	19	39	A	43	53	34	42	A
24	36	52	41	20	A	34	52	60	8	NC
25	35	64	22	27	A	31	64	38	8	NC
26	82	52	33	20	A	82	52	45	10	NC
27	44	61	24	20	A	37	61	43	7	NC
28	41	69	54	18	A	37	69	66	9	NC
29	10	52	25	26	A	10	52	40	7	NC
30	24	69	28	15	A	9	69	38	8	NC
31	40	52	27	42	A	40	52	33	45	A
32	23	53	22	40	A	46	53	42	33	A
33	26	52	26	29	A	42	52	36	20	A
34	23	53	29	18	A	25	53	31	8	NC
35	35	53	21	18	A	35	53	37	4	NC
36	33	56	31	22	A	31	56	53	7	NC
37	34	53	19	29	A	50	53	35	34	A
38	40	68	37	25	A	27	68	42	9	NC
39	40	67	40	20	A	40	67	47	5	NC
40	38	52	26	47	A	38	52	44	32	A
41	35	52	27	41	A	60	52	49	26	A
42	43	53	45	50	A	35	53	44	40	A
43	30	62	34	55	E	32	62	58	51	E
44	25	56	27	31	A	25	56	36	26	A
45	38	57	34	26	A	23	57	53	6	NC
46	35	52	36	22	A	51	52	35	4	NC
47	33	58	30	24	A	33	58	52	7	NC
48	34	57	46	45	A	49	57	54	34	A
49	32	53	18	29	A	29	53	33	41	A
50	27	52	27	44	A	32	52	48	26	A
51	28	52	16	33	A	28	52	32	33	A
52	28	52	37	43	A	18	52	43	38	A
53	33	56	48	36	A	33	56	60	30	E
54	31	58	54	47	A	26	58	57	36	E
55	43	52	41	27	A	43	52	56	7	NC
56	20	52	17	35	A	20	52	29	56	A
57	32	52	37	34	A	44	52	57	9	NC
58	59	52	37	35	A	56	52	51	37	A
59	60	54	24	68	E	60	54	52	62	E
60	65	54	28	23	A	65	54	59	9	NC

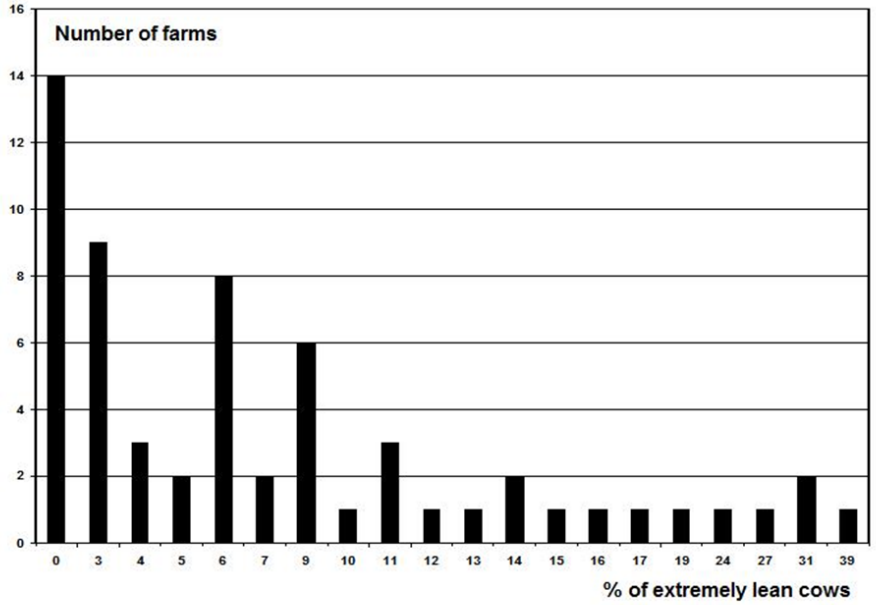

Figure 2: Frequency distribution of the percentage of extremely lean cows on the 60 farms.

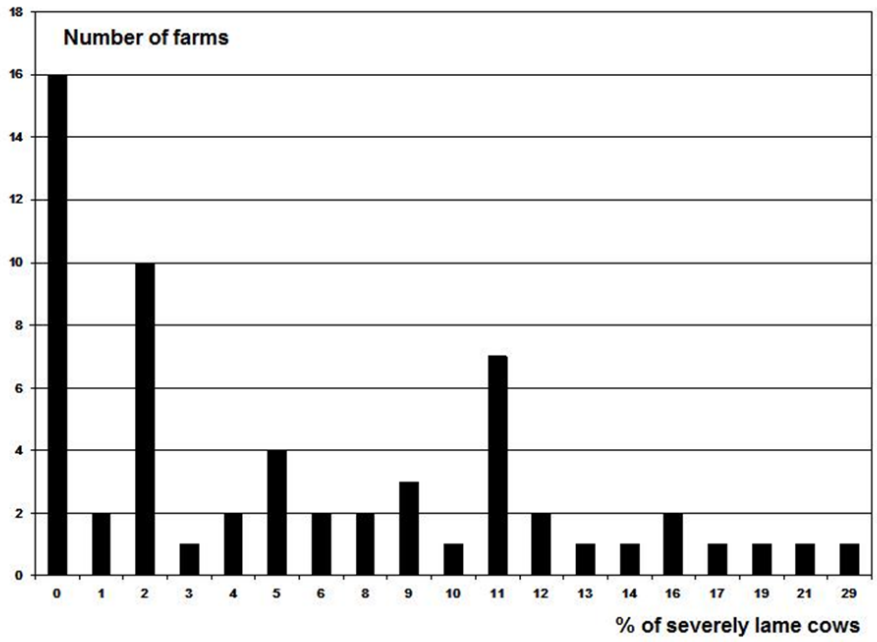

Figure 3: Frequency distribution of the percentage of severely lame cows on the 60 farms.

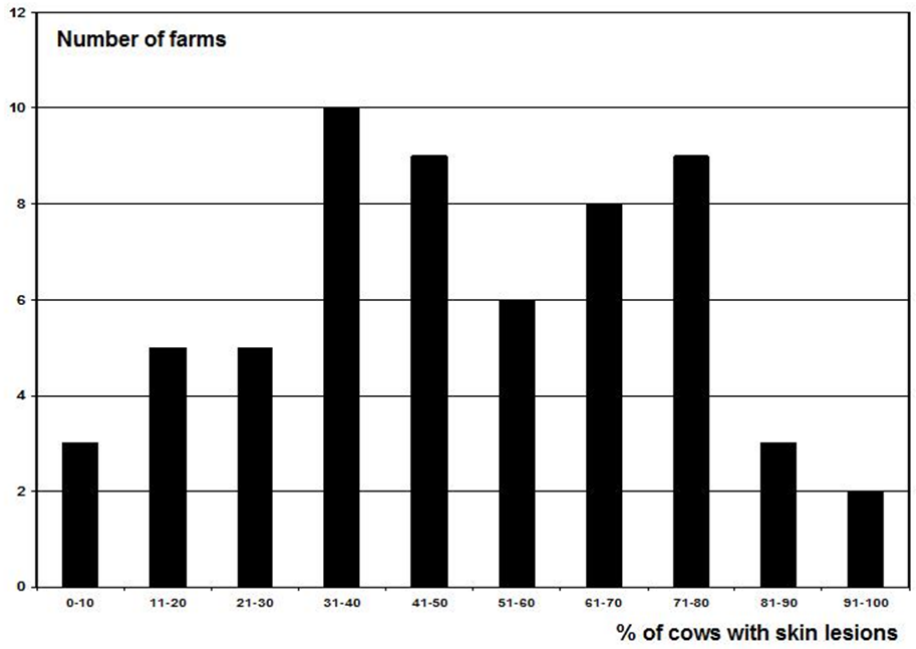

Figure 4: Frequency distribution of the percentage of cows with at least 1 skin lesion on the 60 farms.

Table 2: Odds ratio's and their confidence intervals of WQ

	AvsNC		EvsNC	
	OR	confidence interval	OR	confidence interval
Feeding	1.51	.99-2.31	1.68	1.10-2.59
Behavior	1.02	.87-1.18	1.34	.93-1.92

AvsNC = Acceptable vs Not Classified
EvsNC = Enhanced vs Not Classified

Apparently, mainly the principles Good Feeding and Appropriate Behavior were determining the WQ end score. The other two appeared of minor importance. De Vries et al. (2013) also reported that a limited number of welfare measures had a strong influence on the WQ classification of dairy herds. This was confirmed in the study of Heath et al. (2014), where 88% of the farms could be classified correctly with "absence of prolonged thirst", a component of the first principle, only. De Graaf et al. (2016) reported also that absence of prolonged thirst and the QBA were the most influential measures. Heath et al. (2014) suggested that the protocol could be shortened to just 15 minutes with the same outcome. Furthermore, during the execution of the WQ protocol, several questions were raised by the observers and farmers about the parameters measured, tests and calculations used. Therefore, the original WQ protocol was modified in 3 ways:

1) Cleanliness of the Drinkers

If on a farm there is 1 dirty (or partially dirty) drinker, not all drinkers are clean. The question in the calculation of the WQ protocol is: "Are the drinkers clean?" (WQ 2009 pp 95). This than has to be answered as 'No'. Resulting in a maximum score for absence of prolonged thirst of 32 out of 100 points. This implies that on a farm with 100 cows with 12 water bowls with sufficient flow and of adequate size, on at least 2 different locations, the score for absence

of prolonged thirst will be 32 points if 1 of the drinkers is (partially) dirty and 11 are clean. These are more clean drinkers than required by the WQ protocol. On another farm with 100 cows with 7 water bowls with sufficient flow and of adequate size, on at least 2 different locations, the score for absence of prolonged thirst will be 60 points if all 7 drinkers are clean. This implies that the WQ protocol (2009) considers the water supply almost twice as good when there are 4 clean drinkers less available for the animals. This is, in our opinion, not correct. And since, in practice, a farmer cannot clean each drinker several times a day, often one of the drinkers will be (partially) dirty when the assessor is at the farm. This implies that, in practice, the maximum score for absence of prolonged thirst will be 32 points. Even when the score for absence of prolonged hunger is maximal (100 points) the score for the first principle will be 40.16. This is not even considered 'enhanced' by the WQ protocol. In our modified WQ protocol, therefore, the weighted score for cleanliness of the drinkers (see M&M section) was introduced. In this way the cleanliness of a single drinker cannot determine the score for absence of prolonged thirst and thus the score for the first principle.

2) Integument Alterations (hairless patches and lesions/swellings)

For this criterion the WQ protocol takes into consideration if a cow has one, or more, hairless patches (HP), swellings or lesions. The classification in the WQ protocol (2009) is as follows: "Percentage of animals with no integument alteration (no HP, no lesion/swelling). Percentage of animals with mild integument alterations (at least one HP, no lesion/swelling). Percentage of animals with severe integument alterations (at least one lesion/swelling)". However, the number of these alterations nor the severity is taken into account. A cow with 20 HP is the same in the calculations as one with just 1. And a lesion of 20 cm^2 is the same as one of 3 cm^2. This seems not right, because it will make a difference for the level of pain experienced by the cow if a lesion is 3 or 20 cm^2. So in the modified WQ protocol the average number of HP/lesions/swellings per cow is used in the calculations.

3) Qualitative Behavior Assessment (QBA)

QBA is a method based upon the integration by observers of perceived expression of animal behavior, using descriptors such as 'calm', 'aggressive', 'sociable' or 'indifferent' [18]. A description or definition of these animal behavioral expressions is lacking, so it is left to the observer what the perception of these descriptors is. This is remarkable, because for the rest of the parameters measured in the WQ protocols everything is well defined and photos are presented to illustrate the definitions. However, with the QBA this is completely different. For the QBA there is no description at all of what to look for and how to assess. Terms like 'indifferent' and 'content' are very subjective and, without a proper definition or guidance, the score will vary substantially among assessors [19,20]. In our experience it appeared to be highly subjective and variable. During the training sessions there was almost no agreement between the assessors. The usefulness of the QBA has been seriously disputed and considered "insufficiently reliable as a tool for welfare assessment in dairy cattle" by Bokkers et al. [21]. There are reports that QBA can be a reliable method for the assessment of the emotional state of the cows [18], but the QBA scores did not have a meaningful pattern of relationship with other WQ measures in a Danish study [22] and in the study of Hubbard and Scott (2011). This is not surprising because, as Tuyttens et al. [20] have pointed out, there can be substantial bias in the observer reliability when executing the QBA.

Another important aspect, in this respect, is that the aim of WQ [14] is improvement of animal welfare on dairy farms. In order to achieve this, one has to motivate the farmer to improve the situation on his farm [23]. If they think the protocol makes sense, they will take the outcome seriously. If not, nothing will be done. All the farmers in our study did not consider the QBA a proper test for determining the level of welfare of their animals, and were not convinced that the result is something to be taken serious. And because Hubbard and Scott (2011) reported that farmers and scientists use the same measures for the determination of animal welfare, the opinion of the farmers should not be neglected in this respect. Whay et al. [24] reported that it was possible to make steps forward in a project to reduce lameness on dairy farms, but that it was very difficult to convince the farmers. Therefore, we avoid any disputed measurement in the modified protocol. Because of the 3 reasons mentioned above, the QBA was omitted in the modified protocol. We think that measuring social interactions and activities would be a better way to assess the emotional state of the cows. Those can be defined and objectively assessed. Since the QBA is the component with the largest weight for the principle of Appropriate Behavior, the values for μ for the other criteria in the Choquet integral were doubled, else the

maximum score for this principle would not reach 100 points.

All other measures and calculations remained the same as in the original WQ protocol (2009). Since the cleanliness of the drinkers was already scored per drinker and the number of integument alterations was already quantified according to the instructions of the WQ protocol (2009), no new measurements needed to be performed. A new score was calculated after these modifications for the 60 farms and resulted in 23 farms with score Not Classified, 30 Acceptable and 7 Enhanced, no farms were scored Excellent. In Table 3 is shown that the principles Feeding, Health and Behavior were the variables related to the WQ-modified scores according to AIC. Because the WQ protocol (2009) is the fruit from a large panel of experts, it can be seen as a kind of 'standard'. However, as is mentioned in the publication of the protocol, it is a living document and the protocol might be changed according to current insights.

Table 3: Odds ratio's and their confidence intervals of WQ-modified

	AvsNC		EvsNC	
	OR	confidence interval	OR	confidence interval
Feeding	1.24	.68-2.23	1.36	.75-2.49
Health	1.10	.42-2.96	1.41	.34-3.67
Behavior	3.54	.43-4.64	4.07	.49-33.94

AvsNC = Acceptable vs Not Classified
EvsNC = Enhanced vs Not Classified

CONCLUSIONS

In order to improve the discriminative capacity, to level the importance of each of the 4 principles, and to increase the acceptance of the WQ protocol for dairy cows, we propose the 3 modifications described above: Change the calculation of the cleanliness of water points and integument alterations in a more quantitative way and omit the QBA.

ACKNOWLEDGEMENTS

The study was financed by the Dutch Ministry of Economic affairs, Agriculture and Innovation.

REFERENCES

1. Blokhuis, H.J.; Jones, R.B.; Geers, R.; Miele, M.; Veissier, I. Measuring and monitoring animal welfare: Transparency in the food product quality chain. *Animal Welfare* 2003, *12*, 445-455.
2. Christensen, T.; Stott, A.; Lawrence, A.; Sandøe, P. In *What can economists do for animal welfare?*, UFAW International Animal Welfare Symposium, Portsmouth, UK 28-29th June, 2011; Portsmouth, UK p2.
3. Van Eerdenburg, F.J.C.M.; Vázquez-Flores, S.; Saltijeral-Oaxaca, J.; Sossidou, E.N. A cow comfort monitoring scheme to increase the milk yield of a dairy farm. In *Livestock housing*, Aland, A.; Banhazi, T., Eds. Wageningen Academic Publishers: Wageningen, The Netherlands, 2013; pp 55-74.
4. Broom, D.M. In *Sentience and pain in relation to animal welfare.*, XVII International Congress on Animal Hygiene, Košice, Slovakia, 2015; Venglovsky, J., Ed. Košice, Slovakia.
5. Von Keyserlingk, M.A.G.; Rushen, J.; de Passillé, A.M.; Weary, D.M. Invited review: The welfare of dairy cattle—key concepts and the role of science. *Journal of Dairy Science* 2009, *92*, 4101-4111.
6. Von Keyserlingk, M.A.G.; Cestari, A.A.; Franks, B.; Fregonesi, J.A.; Weary, D. Dairy cows value access to pasture as highly as fresh feed. *Scientific Reports* 2017, *7*, Article number: 44953
7. Welfare Quality Consortsium, L., the Netherlands. *Welfare quality assessment protocol for cattle*; 2009.
8. Andreasen, S.N.; Sandøe, P.; Forkman, B. Can animal-based welfare assessment be simplified? A comparison of the welfare quality® protocol for dairy cattle and the simpler and less time consuming

protocol developed by the danish cattle federation. *Animal Welfare* 2014, *23*, 81-94.
9. Heath, C.A.; Browne, W.J.; Mullan, S.; Main, D.C. Navigating the iceberg: Reducing the number of parameters within the welfare quality(®) assessment protocol for dairy cows. *Animal* 2014, *8*, 1978-1986.
10. De Graaf, S.; Ampe, B.; Buijs, S.; Andreasen, S.N.; De boyer Des Roches, A.; van Eerdenburg, F.J.C.M.; Haskell, M.J.; Kirchner, M.K.; Mounier, L.; Radeski, M.*, et al.* In *Sensitivity of the integrated welfare quality scores of the dairy cattle protocol to changes in individual measures.*, Benelux ISAE conference, Berlicum, the Netherlands, 2016; Reimert, I.; Wijnen, J.; Buijs, S.; Bolhuis, E., Eds. Berlicum, the Netherlands, p 13.
11. de Vries, M.; Bokkers, E.A.; van Schaik, G.; Botreau, R.; Engel, B.; Dijkstra, T.; de Boer, I. Evaluating results of the welfare quality multi-criteria evaluation model for classification of dairy cattle welfare at the herd level. *J Dairy Sci* 2013, *96*, 6264-6273.
12. Buijs, S.; Ampe, B.; Tuyttens, F.A.M. Sensitivity of the welfare quality® broiler chicken protocol to differences between intensively reared indoor flocks: Which factors explain overall classification? *Animal* 2017, *11*, 244–253.
13. Toma, L.; Haskell, M.J.; Barnes, A.P.; Stott, A.W. In *Relationship between animal welfare, production and environmental performance of dairy farms.*, 7th International Conference on the Assessment of Animal Welfare at Farm and Group Level, Ede, the Netherlands, September 5-8, 2017; De Jong, I.C.; Koene, P., Eds. Wageningen Academic Publishers: Ede, the Netherlands, 2017; p 39.
14. Welfare-Quality-Consortium. *Welfare quality® assessment protocol for cattle.* . Welfare Quality® Consortium Lelystad, the Netherlands., 2009.
15. Bilder, C.R.; Loughin, T.M. *Analysis of categorical data with r*. CRC Press: Boca Raton, FL, USA, 2015.
16. Spoolder, H.; De Rosa, G.; Hörning, B.; Waiblinger, S.; Wemelsfelder, F. Integrating parameters to assess on-farm welfare. *Animal Welfare* 2003, *12*, 529-534.
17. Botreau, R.; Veissier, I.; Perny, P. Overall assessment of animal welfare: Strategy adopted in welfare quality®. *Animal Welfare* 2009, *18*, 363-370.
18. Rousing, T.; Wemelsfelder, F. Qualitative assessment of social behaviour of dairy cows housed in loose housing systems. *Appl Anim Beh Sci* 2006, *101*, 40-53.
19. Hubbard, C.; Scott, K. Do farmers and scientists differ in their understanding and assessment of farm animal welfare? *Animal Welfare* 2011, *20*, 79-87.
20. Tuyttens, F.A.M.; de Graaf, S.; Heerkens, J.L.T.; Jacobs, L.; Nalon, E.; Ott, S.; Stadig, L.; Van Laer, E.; Ampea, B. Observer bias in animal behaviour research: Can we believe what we score, if we score what we believe? *Animal Behaviour* 2014, *90* 273-280.
21. Bokkers, E.A.; De Vries, M.; Antonissen, I.C.M.A.; De Boer, I.M. Inter- and intra-observer reliability of experienced and inexperienced observers for the qualitative behaviour assessment in dairy cattle. *Animal Welfare* 2012, *21*, 307-318.
22. Andreasen, S.N.; Wemelsfelder, F.; Sandøe, P.; Forkman, B. The correlation of qualitative behavior assessments with welfare quality® protocol outcomes in on-farm welfare assessment of dairy cattle. *Applied Animal Behaviour Science* 2013, *143*, 9-17.
23. Kielland, C.; Skjerve, E.; Osterås, O.; Zanella, A.J. Dairy farmer attitudes and empathy toward animals are associated with animal welfare indicators. *J Dairy Sci.* 2010, *93*, 2998-3006.
24. Whay, H.R.; Barker, Z.E.; Leach, K.A.; Main, D.C.J. Promoting farmer engagement and activity in the control of dairy cattle lameness. *The Vet Journal* 2012, *193*, 617-621.