JZAR www.jzar.org JOURNAL OF ZOO AND AQUARIUM RESEARCH

JZAR Research article

OPEN ACCESS

Research article

# Findings and refinements from a decade of applying the Cetacean Welfare Assessment ("C-Well") to multiple species on public display

**Isabella L.K. Clegg and Rhiannon L. Schultz**

[1]*Animal Welfare Expertise, London, UK*

*Correspondence: Isabella Clegg, email; hello@animalwelfareexpertise.com*

**Keywords:** animal-based indicators, animal welfare, C-Well Assessment, cetaceans, welfare assessment, Zoo-Well Assessment

**Article history:**
Received: 21 Feb 2025
Accepted: 17 Jul 2025
Published online: 31 Jul 2025

**Abstract**

Zoo animal welfare assessments can be applied to multi-species risk evaluations to species-specific, animal-based protocols. The latter are more difficult to develop and implement due to time-intensive data collection and difficulties with standardisation. The Cetacean Welfare Assessment (C-Well) was published a decade ago and has been applied globally more than 30 times across five cetacean species, with many facilities conducting repeated assessments, and has undergone multiple updates to align with the latest research. This paper reviews the C-Well's application over the past decade, highlights key insights from the collected data, and details the measure composition in the C-Well version 2.0. The effectiveness, limitations and future directions of the project are explored in detail, including how its scope can be broadened. The C-Well's evolution has recently been leveraged to develop the Zoo-Well approach which uses a core framework designed to allow species-specific welfare assessments to be readily developed, and which has been conducted successfully on a dozen species to date.

## Introduction

Despite significant progress in the last decades regarding consensus on defining animal welfare, researchers are still wrestling with how to transcribe something subjective, i.e. animal emotions, into an objective and standardised format (Hampton et al. 2023). Many different approaches to solving this problem have been trialled, with varying degrees of success. The aim to use our knowledge of welfare states to improve animals' lives inevitably means any welfare measurement techniques should be applicable on the ground by organisations that care for animals. The field has largely agreed that welfare concerns the balance of affective or emotional states, that the absence of negative states does not result in the presence of positive ones (Boissy et al. 2007), and that multi-disciplinary resource and animal-based measures should be included in assessments (Botreau et al. 2007b; Veissier et al. 2008). The Five Domains Model (Mellor 2017) has stood out as the most valuable in helping to design welfare assessments (Hampton et al. 2023).

### *Zoo animal welfare assessments*

Zoological organisations that house terrestrial or marine wild animals for public display (hereafter zoos) cover a wide range of environments from small city zoo collections to expansive drive-through 'safari parks.' While zoos have increased efforts in assessing the welfare of their residents, their key challenge is the diversity of species present in a single collection. Species-specific welfare assessments only exist for a handful of species, where the next-best approach might be to apply a 'welfare risk assessment' i.e. using expert and/or stakeholder opinions on welfare measures to generate consensus on scores for each measure. This is how the Five Domains Model was designed to function, and many argue it is not appropriate for direct application as a welfare assessment despite zoo industry bodies encouraging members to do so (Hampton et al. 2023).

Following the publication of suggested protocol for welfare risk assessment in zoos (Sherwen et al. 2018) involving a majority of resource-based measures and a few animal-based measures, several facilities have implemented welfare risk assessments (I. Clegg, personal observation) or similar impact assessment frameworks (Kagan et al. 2015; Jones et al. 2022) into internal welfare monitoring programs.

The few species-specific assessments that do exist for captive wild animals have generally been adapted from farm animal assessment protocols, and most of them follow the well-established WelfareQuality® Project blueprint, e.g. Dorcas gazelles *Gazella dorcas* (Salas et al. 2018), farmed foxes and mink (Mononen et al. 2012), pygmy blue-tongue skink *Tiliqua adelaidensis* (Benn et al. 2019), and bottlenose dolphins *Tursiops truncatus* (Clegg et al. 2015). Although more time-intensive, the advantages of species-specific as opposed to generic risk assessments are that the former are more comprehensive, involve more measures based on objective and quantifiable data, cover all aspects of animal welfare, and are tailored to the specific needs of each species (Salas et al. 2024). Nevertheless, their impact on individual animals' welfare is dependent on how they are applied in practice and how those responsible for the animals use the evidence to effect sustainable change.

***The Cetacean Welfare ("C-Well") Assessment approach***

Concurrent to, and probably as a result of, the social licence questions surrounding cetacean facilities, efforts have increased into developing assessments of cetacean welfare (Clegg et al. 2015; 2017; 2018; Serres et al. 2020a; Delfour et al. 2021; Lauderdale et al. 2021;). The C-Well Assessment was published in 2015 as the first multi-disciplinary, comprehensive assessment available for bottlenose dolphins *Tursiops truncatus*, and included both animal-based and resource-based measures (Clegg et al. 2015). A few years later an Animal Welfare Decision Tree was developed and applied to bottlenose dolphins (von Fersen et al. 2018), and most recently the Dolphin Welfare Evaluation Tool was established by the European Association for Aquatic Mammals (EAAM), designed for use by internal staff members of cetacean facilities (Baumgartner et al. 2024). In addition, two assessment tools for wild cetaceans have been developed in the last 5 years (Nicol et al. 2020; Serres et al. 2020) as well as some unpublished frameworks from different zoo and industry body organisations.

The original C-Well Assessment project aims were to firstly develop assessment indices and protocols to aid in the objective measurement of welfare, and secondly to serve as a model to support future "benchmarking, industry best practices, and certification" efforts (Clegg et al. 2015). In the first version of the assessment, there were 36 welfare measures (58% animal-based) specific to bottlenose dolphins, covering the categories of Good Feeding, Good Housing, Good Health and Appropriate Behaviour, following the WelfareQuality® structure (Blokhuis 2008). Given that at the time of publication there were no other proposed protocols for assessing cetacean welfare, the collection of standardised data was expected to facilitate a positive feedback loop and lead to the evolution of the C-Well approach over time.

***Evolution of the C-Well Assessment***

Since its publication a decade ago, the C-Well has been applied 35 times to five different cetacean species, including those facilities that have undergone repeated assessments. It has been applied in zoos, marine parks, Swim-with-the-dolphin facilities, and 'sanctuary' environments (defined here as those not focused on commercial activities, who do not attempt to breed the animals, and who aim to improve welfare from the animals' original

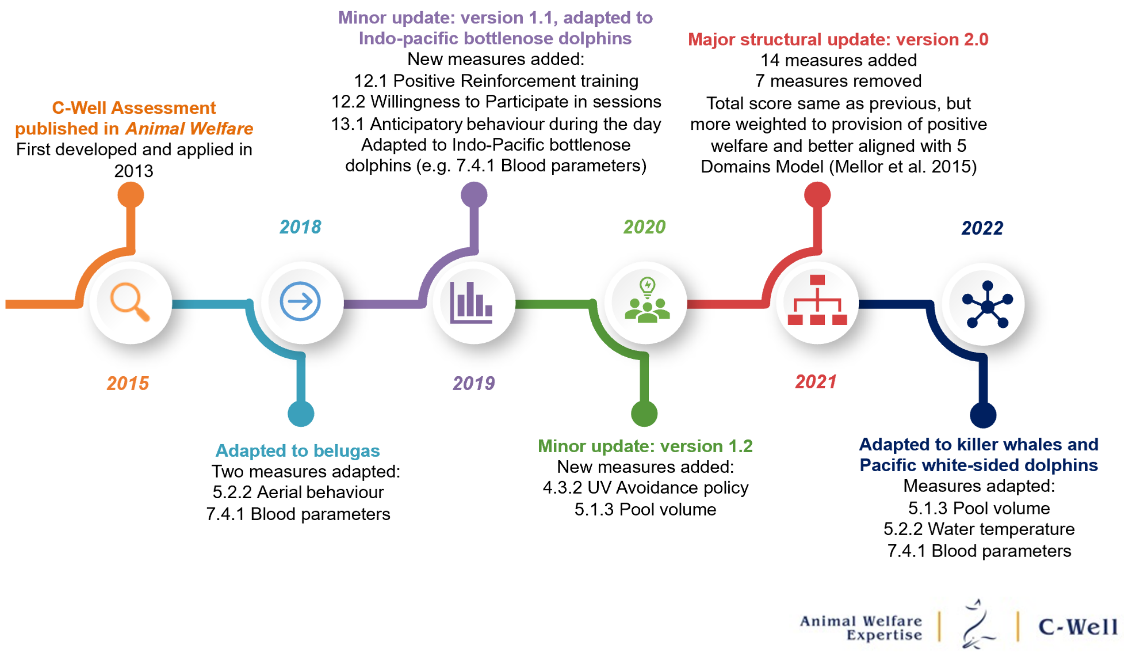


**Figure 1.** Evolution of the Cetacean Welfare ("C-Well") assessment since its inception (graphic credit to the authors)

location (Bruck 2024)). 15 organisations in 9 countries have used the assessment to estimate the individual welfare status of 159 cetaceans. Six assessors have been applying the assessment, all trained by the organisation Animal Welfare Expertise.

Over the years, the C-Well measures have been periodically reviewed and updated to ensure they remain in line with the latest research and best practice standards (Figure 1). The assessment has undergone several minor updates, which involve changes to only a few measures, and one major structural update in 2021 (version 2.0) to better align it with the Five Domains Model and ensure an appropriate focus on opportunities for positive welfare ( Mellor and Beausoleil 2015; Mellor 2016; Littlewood et al. 2023). The C-Well measures have also been adapted to Indo-Pacific bottlenose dolphins *Tursiops aduncus*, belugas *Delphinapteras leucas*, Pacific white-sided dolphins *Lagenorhynchus obliquidens* and killer whales *Orcinus orca*.

***Objectives of the paper***

Given that the C-Well assessment has evolved since its development a decade ago, and has been applied repeatedly across a range of organisations and species, the key objective of this paper was to review the findings and discuss insights into cetacean welfare. Since several of the organisations where the assessment was applied chose to remain anonymous, analyses using identifying characteristics of the data are not possible at this time and we focussed instead on highlighting the potential impact this type of assessment data could have on the theory and practice of cetacean welfare. A second objective was to present the updated measures included in the C-Well version 2.0 framework.

## Objective 1: Findings from C-Well assessment applications

***Sample population***

While the C-Well assessment has been applied at 15 organisations to date, the data from 11 of those facilities are pooled and presented here (as per consent requests) and cover four cetacean species. The relatively small number of organisations holding captive cetaceans worldwide means they are easy to distinguish using basic information, and as such all identifying characteristics of the data have been removed in this paper including names, location, and species of cetacean. The age (at the time of their first C-Well assessment) and sex of the 111 animals included in this paper's dataset are shown below (Figure 2), comprising 47 males and 64 females. The C-Well was not conducted on any animals younger than two years old to safeguard the mother-calf relationship, and due to limitations in collecting accurate welfare data. If considering the total number of applications of the assessment, which includes repeated assessments, the current dataset includes 246 individual applications, and thus total welfare scores.

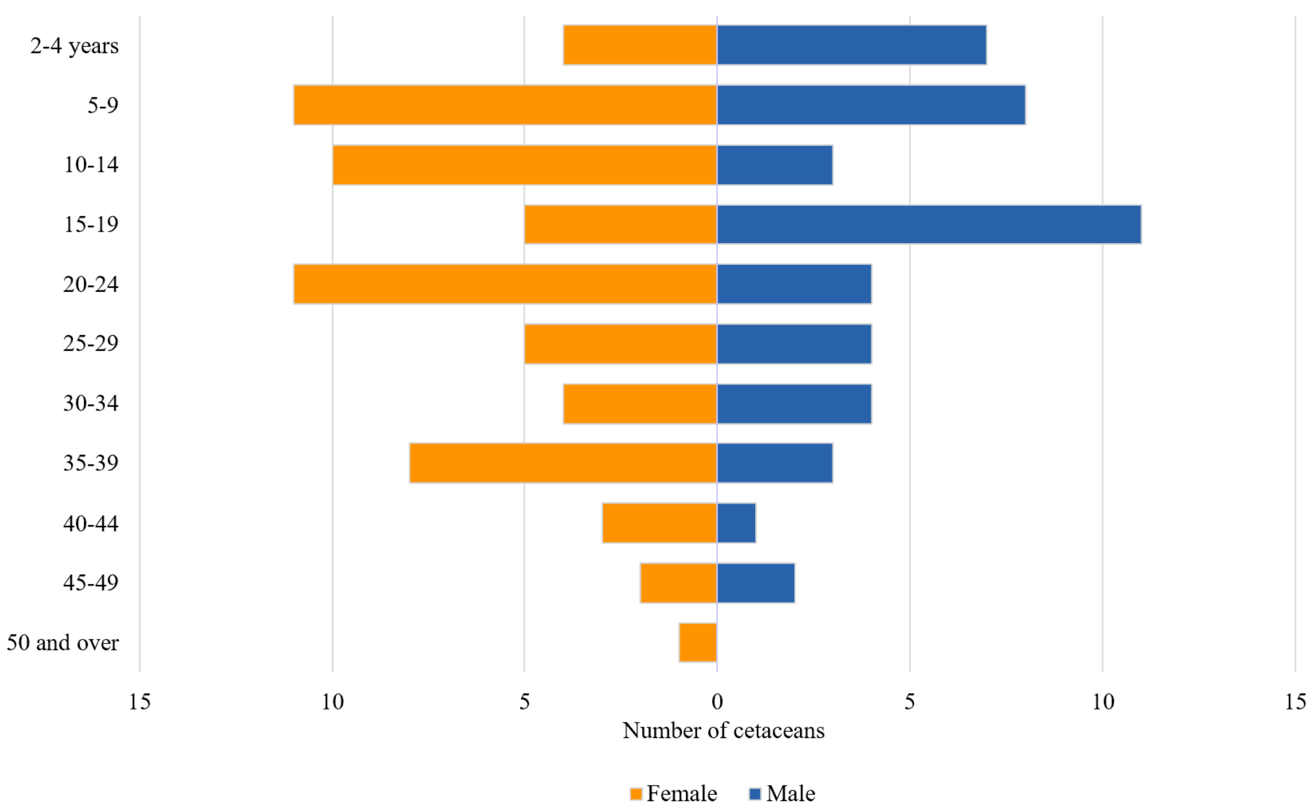


**Figure 2.** Population pyramid for cetaceans to whom the C-Well assessment has been applied up to 2024, and whose facilities consented to share their data (n=111; 47 females and 64 males).

***Application of the C-Well assessment***

As described in the conflict of interest and ethical declarations, data for this paper was collected when the welfare consultancy Animal Welfare Expertise (AWE), was contracted by a host facility to apply the C-Well assessment. The agreement was that the AWE welfare assessors, or those who were collecting data on their behalf, had full access to observe the animals, interview staff and review records, and that a welfare report and recommendations for improvement would be produced using the results. Assessors were under no obligation to produce a certain result for certification or legal compliance, and the host facilities retained full control of the results once generated. Further detail is provided in Supplementary Information 1 about the welfare assessors' roles and training. Facilities chose to apply the C-Well assessment as a management tool and were therefore incentivised to supply the most accurate welfare data to the C-Well assessors, so that meaningful and valuable recommendations could be made. In terms of repeated assessments, five facilities decided to apply the assessment at least twice, three facilities have applied it four times, and one facility has applied it seven times.

The C-Well assessment generally takes five days to apply to 10 individual dolphins, with the most time-intensive element being the behavioural observations used to establish activity budgets. To make even conservative welfare interpretations using the behavioural data, the minimum amount of observational data collected from each animal during periods without training sessions is 20 five-minute observations (100 minutes). The vast majority of the welfare data collected as part of the C-Well assessments is non-invasive (predominantly observational and records-based). The protocol for the few husbandry measures dictates that assessors only use results (e.g. blood samples) taken as part of the facility veterinarian's routine management. Where these routine procedures did not overlap with the C-Well assessment period, the relevant measures were excluded from the assessment and overall scores.

***Aggregation and interpretation of welfare scores***

A current limitation of any welfare assessment that functions by generating a total numerical score is that there are no agreed-upon methods of score aggregation. In order to interpret and compare results across multiple sites, including for certification and regulatory purposes, we need some way of moving beyond descriptive evaluations of welfare towards prescribed norms (Botreau et al. 2007b; Hampton et al. 2023). Score aggregation methods attempted in existing schemes range from informal estimations by experts, to minimum requirements and sums of ranked scores, and finally to the use of models and calculations to attribute different weightings, or priorities, to certain measures. Linked to aggregation, methods for determining the weighting loads are also not yet standardised or well-established (Botreau et al. 2007b; Browning 2022).

The score aggregation process associated with the original C-Well assessment is based on summing scores assigned using a three-level scale (similar to the 'minimum requirement' approach described above). Each C-Well measure is assigned one score from a choice of three, criteria for which were established during the initial development phase and updated subsequently, and which represent good welfare (score 0), adequate or sub-optimal welfare (score 1), or poor welfare (score 2) (Clegg et al. 2015). The total score for the assessment is then a simple sum of these measures' scores, and when presented to stakeholders was most often converted to a percentage for ease of interpretation.

The C-Well score aggregation process itself has not changed since the original publication, but a notable refinement has impacted how the total score provides an estimation of an animal's welfare state. In the updated version 2.0, the assessment structure was revised so that of the 44 measures in total, an equal number (22) are assigned to the first three functional domains of welfare (nutrition, environment and health; generally associated with meeting basic needs) as well as to the latter domains of behaviour, enrichment and training/human-animal interactions, whose scope covers aspects relating to opportunities for positive welfare. This structure is directly in line with the latest updates to the Five Domains model (Mellor et al. 2020), and is a proxy weighting adjustment in that the choice and numbers of measures themselves, which feed into the final aggregate score, dictate the welfare 'priorities' of an assessment (Browning 2022; Fischer et al. 2024). This restructuring was accompanied by revisions to the scoring criteria of each measure to ensure they exceeded simply assessing the absence of negative states, and only awarded the "good welfare" score 0 when the presence of positive states was evidenced.

***Results from the C-Well dataset***

The anonymity restrictions for this publication limits the power of analysis, but given there are 111 individual cetaceans' results from 246 applications of the assessment, within which dozens of multi-disciplinary welfare measures have been applied, we believe there is still value in presenting some selected results to showcase the types of questions these data could answer in the future. Nevertheless, limitations such as the unweighted measures and the changing composition of the C-Well assessment over the years mean any interpretations must be extremely conservative. Analyses were conducted using the programme R, version 4.4.2 (R Development Core Team 2025).

Research questions in the following sections were chosen to highlight the potential of the dataset, with an exploration into the variance of scores across the six welfare Domains in Supplementary Information 2.

*C-Well scores by age class and sex*

Mean total C-Well scores were split by sex and calculated for each age category of cetacean, then plotted alongside standard deviations to descriptively visualise the variance across the sample population (Figure 3). There was very little difference in mean C-Well scores between males and females overall (80.3%±7.83 SD for males, versus 80.6%±6.58 SD for females), and mean scores for age categories were also similar with the exception of the geriatric (35 years +) animals, where females appeared to have higher scores in general.

*Changes in total scores following repeated C-Well assessments*

To understand whether the C-Well approach is a useful tool for improving cetacean welfare, repeated application(s) of the assessment were investigated to determine if they result to any changes in the scores over time.It was hypothesised that the total scores would improve overall with repeated assessments, based on the fact that detailed recommendations for improving specific welfare parameters are given alongside C-Well results to the facilities.

To answer this question accurately, only data from version 2.0 of the assessment were included to improve validity of comparisons by using the same set of measures. In addition, C-Well scores were only included where the assessment had been repeated between 2 and 4 times. There were only a few animals that had received repeated assessments beyond 4 times, which would have led to biased and non-normal averages being used. These filters resulted in 95 repeated assessment scores from 47 cetaceans (3 species), where 21 animals had the assessment applied twice, 38 animals had it applied three times, and 34 animals four times.

We used the *lmer* function of the *lme4* package in R (Bates et al. 2017) to fit a linear mixed-effects model (LMM) to examine

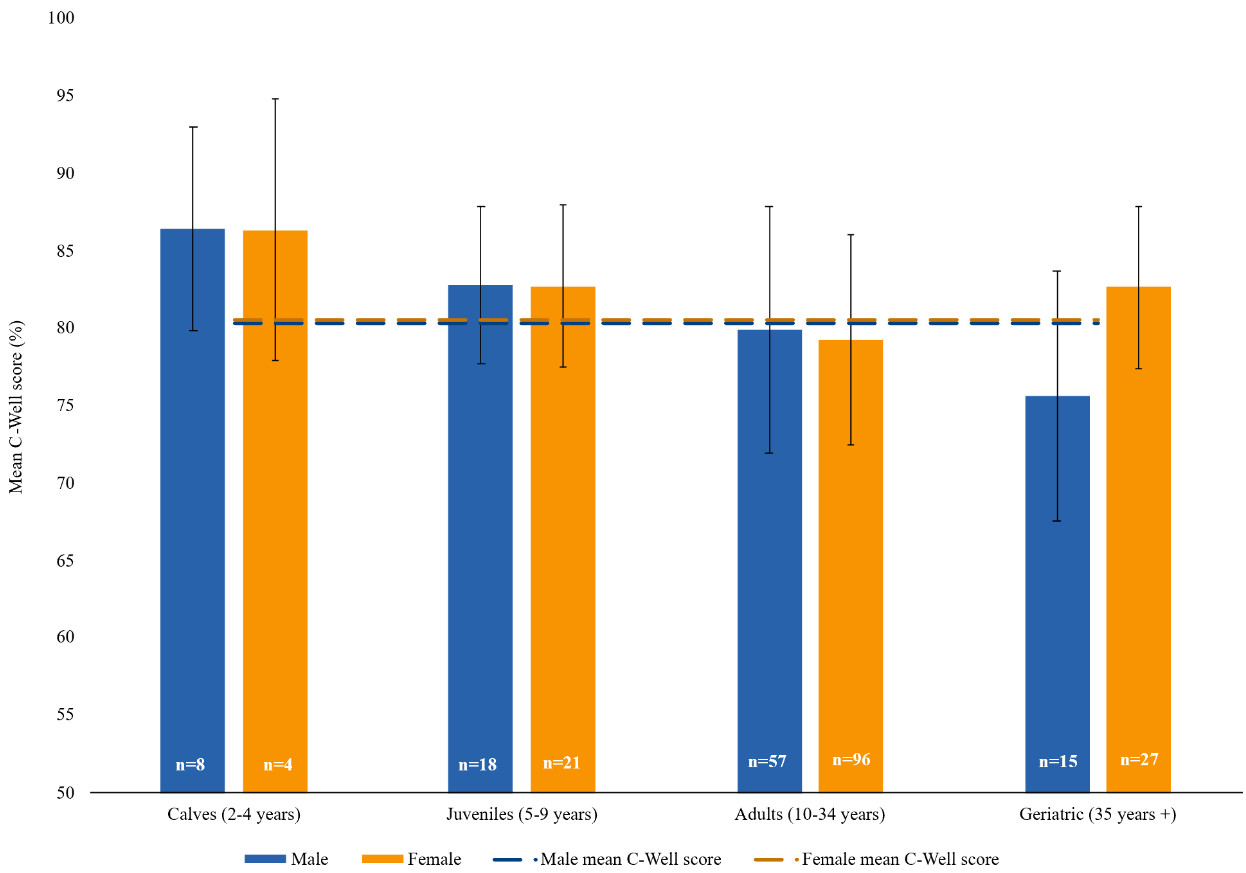


**Figure 3.** Mean C-Well Assessment scores (as percentages) categorised by sex and age class, generated from 246 applications of the assessment on 111 individuals, with black lines showing standard deviation. Numbers in white at bottom of bars show number of individuals in each class. Dotted lines show mean C-Well scores for males and females across all ages.

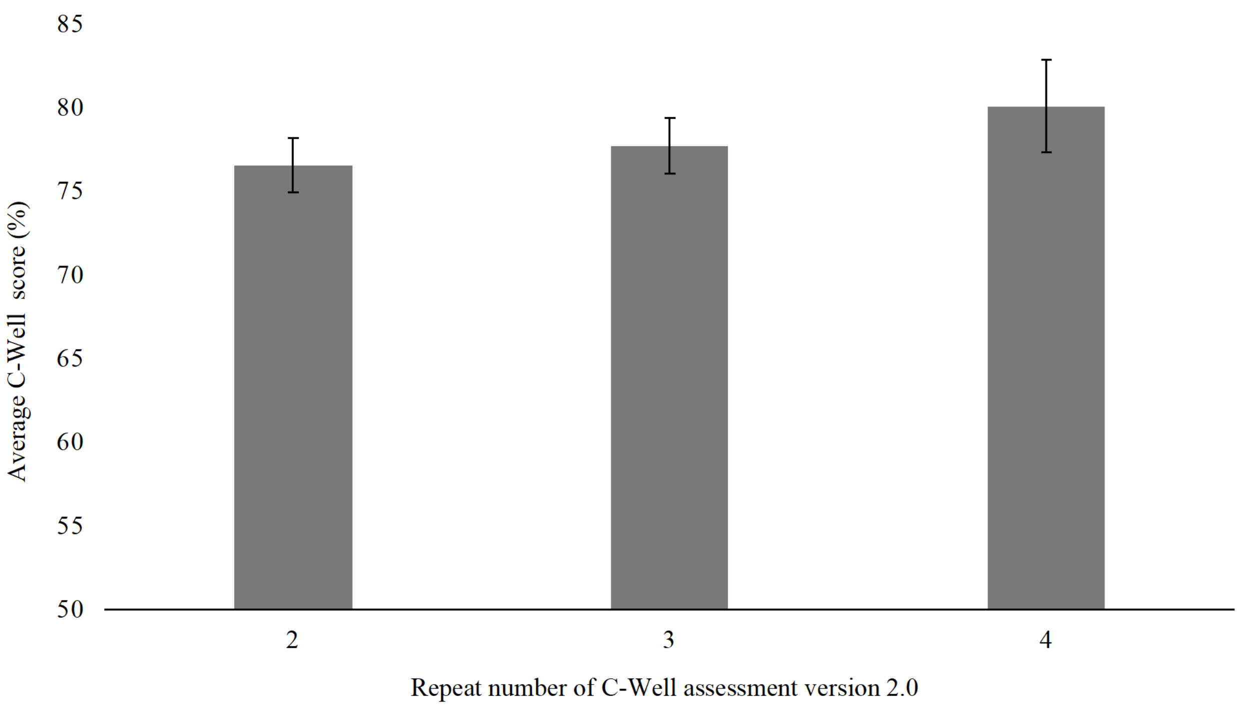


**Figure 4.** Average C-Well Assessment version 2.0 scores (as percentages) for the second, third and fourth repeats of the assessment, generated from 95 assessment applications on 47 individuals. 21 individuals had the assessment applied twice, 38 had it applied three times, and 34 had it applied four times. Black lines show 95% confidence intervals. Linear mixed-effect modelling indicated a statistically significant positive relationship between the C-Well score and repeat number (β=2.1459±0.5892 SE, P<0.001).

**Table 1.** Four C-Well measures, all applied to individual animals, that were included as predictors in the Linear Mixed-Effect Model (LMM) for the dependent variable of total C-Well score.

| C-Well measure | Units | Methods of application during C-Well Assessment | Evidence of link to welfare |
|---|---|---|---|
| Pool volume | $m^3$ per animal | Pool volume available to the individual animal based on the pool configuration that they spent the majority of their time in (over the full 24 hours of the day) during the C-Well assessment | (Miller et al. 2021; Ugaz et al. 2013) |
| Group size | Number | The number of animals in the group (including the focal) based on the group configuration that they spent the majority of their time in (over the full 24 hours of the day) during the C-Well assessment | (Connor 2007; Stanton and Mann 2012; Serres et al. 2020b) |
| Anticipatory behaviour | Percentage of observed time | The percentage of visible time spent showing anticipatory behaviour during five minute continuous focal observations conducted during the C-Well. Each animal was observed for at least 100 minutes per assessment, using the ZooMonitor platform to collect data (Wark et al. 2019), where observations were randomly spread across daylight hours (7am to 5pm). Anticipatory behaviour was defined as "an animal directs its eyes and sometimes orients its body towards an area, usually where caretakers appear, or where a gate usually opens, and/or towards familiar staff walking around the pools. Does not include any anticipatory pattern swimming, as this was defined within route-tracing behaviour". | (Watters, 2014; Clegg and Delfour 2018; Clegg et al. 2023) |
| Affiliative behaviour | Percentage of observed time | The percentage of visible time spent showing affiliative behaviour during five minute continuous focal observations conducted during the C-Well. Each animal was observed for at least 100 minutes per assessment, using the ZooMonitor platform to collect data (Wark et al. 2019), where observations were randomly spread across daylight hours (7am to 5pm). Affiliative behaviour was defined as "positive social behaviours between two or more animals less than 2 body lengths apart including synchronous or group swimming, play, rubbing and gentle tactile contact (but not focussed on genitals)". | (Clegg et al. 2017; Dudzinski and Ribic 2017; Holmes et al. 2024) |

the relationship between 'number of repeat assessments' (independent variable) and 'total C-Well score' (dependent variable), accounting for repeated measures by including a random factor for each individual ('CetaceanID') and facility ('Facility'). We verified homogeneity of variances by plotting fitted values versus residuals (Faraway 2006) and confirmed normality of residuals and linearity of the data through visualising Q-Q and scatter plots.

The results indicate a statistically significant positive relationship between number of repeat assessments and total C-Well scores, where higher numbers of repeated assessments were associated with higher C-Well scores, with an average increase of 2 percentage points with every assessment ($\beta$=2.1459±0.5892 SE, $P<0.001$, Figure 4). The random factor CetaceanID effectively accounted for individual-level variability, which was substantial (variance=21.817, SD=4.671), with a subsequent calculation of the intraclass correlation coefficient (ICC) being 0.65, indicating that 65% of the variance in C-Well scores was associated with individual differences. The random factor Facility only accounted for 4.6% of variability (variance=1.661, SD=1.289).

*Correlation of total C-Well scores to selected welfare parameters*

Those practically applying cetacean welfare assessments are interested in finding specific measures that significantly contribute to overall welfare, where such analyses can ultimately lead to weightings being able to be assigned across a set of indicators (Fischer et al. 2024). Between 36 and 44 resource- and animal-based welfare measures were collected during applications of different versions of the C-Well, and the size of the current dataset is not sufficient for investigating all of them in terms of validity and impact on total scores. We therefore selected two resource-based and two animal-based measures that we hypothesised might have a significant effect on total C-Well scores. The 'predictors' were pool volume per animal, group size, anticipatory behaviour and affiliative behaviour, and data for these individually-applied measures were available for all 246 applications of the assessment on 111 cetaceans. Descriptions are given below of how these measures were defined and applied and the evidence linking them to cetacean welfare (Table 1). Given that the total C-Well score is derived from 44 measures, which includes each of the selected predictors for the model, minor collinearity and circularity are unavoidably introduced. However, circularity effects would be minimal given each predictor contributes approximately 2% to the dependent variable value, and Variance Inflation Factors (VIF) were calculated where all predictors had VIF values below 2, indicating low multicollinearity.

To test whether any of these four measures explained variance in the total C-Well scores, we used an LMM approach. First, to test whether we should include all measures in our model, we constructed separate LMMs with total C-Well scores as the dependent variable for each of the four measures in Table 1 as independent variables, with 'Facility' and 'CetaceanID' as two random factors. These were compared to an LMM with pool volume, group size, anticipatory behaviour and affiliative behaviour included together as independent variables. We used the *AIC()* function to facilitate stepwise deletion and comparisons between models, with lower AICs indicating more explanation of

**Table 2.** Results of Akaike Information Criterion (AIC) comparison that included pool volume, group size, anticipatory behaviour and affiliative behaviour as single independent variables or all together in one model.

| Model | Degrees of Freedom | AIC |
|---|---|---|
| Pool volume only | 5 | 1627.6723 |
| Group size only | 5 | 1592.5920 |
| Anticipatory behaviour only | 227 | 598.8744 |
| Affiliative behaviour only | 222 | 633.8024 |
| All four measures included | 248 | 515.9463 |

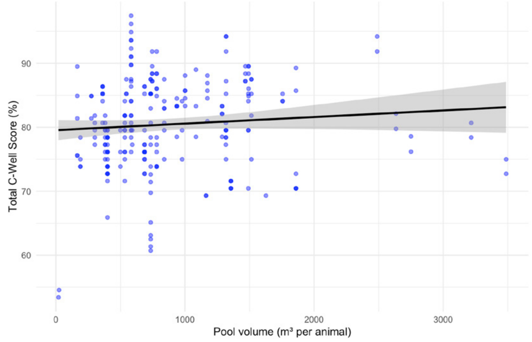


**Figure 5.** Regression plot of individual total C-Well Assessment scores (as percentages, from all C-Well versions) and pool volume available to the animal for the majority of time during the assessment. Data points represent 246 assessment applications on 111 individuals. LMM analysis indicated a statistically significant positive relationship between the C-Well score and pool volume, as indicated by the black regression line (β=1.9028±0.4939 SE, P<0.001) with the grey area representing 95% confidence intervals.

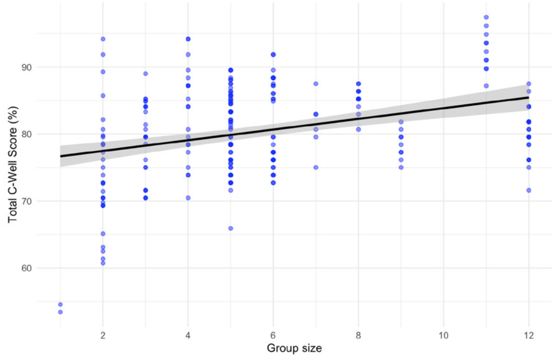


**Figure 6.** Regression plot of individual total C-Well Assessment scores (as percentages, from all C-Well versions) and group size (including focal) available to the animal for the majority of time during the assessment. Data points represent 246 assessment applications on 111 individuals. LMM analysis indicated a statistically significant positive relationship between the C-Well score and group size, as indicated by the black regression line (β=3.8823±0.6768 SE, P<0.001) with the grey area representing 95% confidence intervals.

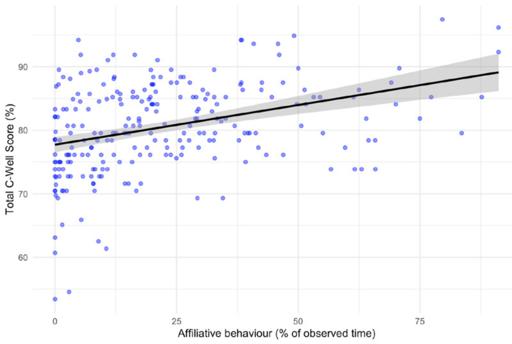


**Figure 7.** Regression plot of individual total C-Well Assessment scores (as percentages, from all C-Well versions) and affiliative behaviour (measured as percentage of observed and visible time) shown during the assessment (ethogram definitions in Table 1). Data points represent 246 assessment applications on 111 individuals. LMM analysis indicated a statistically significant positive relationship between the C-Well score and affiliative behaviour, as indicated by the black regression line (β=1.9560±0.4371 SE, P<0.001) with the grey area representing 95% confidence intervals.

variance in total C-Well scores. The results show that the full model with all four measures included had the lowest AIC (515.95), thus providing the best overall fit compared to models with single independent variables, or predictors (Table 2). Comparing the single predictor models suggests that anticipatory and affiliative behaviours were the two strongest predictors, with pool volume and group size explaining less of the variance.

We then ran the full model with the four predictors to examine the variance explained by the random effects, and tested model assumptions. CetaceanID explained very little variance, so we removed it and retested AIC which was lower, confirming this action. The effect of anticipatory behaviour on total C-Well scores was non-significant, so we attempted removing it from models, but after AIC increased in this reduced model we chose to leave it in the final composition. Residual plots confirmed homoscedasticity, linearity, and normality of residuals.

The final LMM results indicate that pool volume, group size, and affiliative behaviour are significant predictors of total C-Well scores, whereas anticipatory behaviour did not have a significant relationship (Table 3). That is to say that larger pool volumes (Figure 5), larger group sizes (Figure 6) and more time spent

**Table 3.** LMM results for effect of pool volume, affiliative behaviour, group size, and anticipatory behaviour on overall C-Well scores (dependent variable). 'Facility' (i.e. name of cetacean facility) was included as a random intercept (with CetaceanID removed due to negligible variance).

| Predictor | Estimate | Std. Error | t-value | P value |
|---|---|---|---|---|
| Intercept | 79.99 | 1.2613 | 63.414 | *<0.001 |
| Pool Volume | 1.9028 | 0.4939 | 3.852 | *<0.001 |
| Group Size | 3.8823 | 0.6768 | 5.736 | *<0.001 |
| Affiliative Behaviour | 1.956 | 0.4371 | 4.475 | *<0.001 |
| Anticipatory Behaviour | 2.7003 | 5.7318 | 0.471 | 0.638 |

**Table 4.** The domains, criteria and measures included in the C-Well Assessment version 2.0, which has currently been applied to common bottlenose dolphins, Indo-pacific bottlenose dolphins, belugas, killer whales, and Pacific white-sided dolphins. Scoring thresholds for selected measures differ between species. Asterisks denote which measures are animal-based by design, with two asterisks indicating those measures scored through the collection of continuous behavioural data in the animal's 'free-time.'

| Domains | Welfare Criteria | | C-Well Assessment measures version 2.0 |
|---|---|---|---|
| Nutrition | 1 | Food intake | 1.1 Body Condition Score* |
| | | | 1.2 Frequency of weight measurements |
| | | | 1.3 Diet |
| | 2 | Hydration | 2.1 Hydration management |
| Environment | 3 | Resting comfort | 3.1 Resting behaviour** |
| | 4 | Thermal comfort | 4.1 Water temperature and diet |
| | | | 4.2 Protection from UV exposure |
| | | | 4.2.1 Shade presence |
| | | | 4.2.2 UV avoidance policy |
| | 5 | Appropriate environment | 5.1 Topography |
| | | | 5.1.1 Complexity of enclosure |
| | | | 5.1.2 Pool volume |
| | | | 5.1.3 Interconnecting pools |
| | | | 5.2 Water quality |
| | | | 5.2.1 Water quality parameters |
| | | | 5.2.2 Frequency of water quality testing |
| Physical health | 6 | Physical injuries | 6.1 Percentage cover of rakes and wounds* |
| | | | 6.2 Wounds from enclosure* |
| | 7 | Disease | 7.1 Respiratory system |
| | | | 7.1.1 Frequency of coughing* |
| | | | 7.2 Eye diseases |
| | | | 7.2.1 Discoloration* |
| | | | 7.2.2 Squinting* |
| | | | 7.3 Skin diseases |
| | | | 7.3.1 Skin abnormalities* |
| | | | 7.3.2 Mouth abnormalities* |
| | | | 7.4 Blood parameters* |
| | 8 | Pain or stress induced by management procedures | 8.1 Husbandry behaviour training* |
| Behaviour | 9 | Expression of social behaviours | 9.1 Affiliative behaviour** |
| | | | 9.2 Sexual behaviour** |
| | | | 9.3 Agonistic behaviour** |
| | | | 9.4 Social group size |
| | | | 9.5 Social group management |
| | 10 | Abnormal Repetitive Behaviours | 10.1 Abnormal Repetitive Behaviours** |
| | | | 10.2 Pattern swimming** |
| | 11 | Positive human-animal relationship | 11.1 Response to trainer while not under stimulus control* |
| | | | 11.2 Off-session interaction policy |
| | 12 | Anticipatory behaviour | 12.1 Average anticipatory behaviour during day** |
| | 13 | Staff behavioural observations | 13.1 Facility behavioural observation policy |
| Training | 14 | Training management | 14.1 Presence of Positive Reinforcement training |
| | | | 14.2 Daily duration of training sessions |
| | | | 14.3 Variability of training sessions |
| | | | 14.4 Variability of guest-facing sessions |
| | 15 | Animal response to training | 15.1 Willingness to Participate (WtP) in training sessions* |

**Table 4.** Continued.

| Domains | Welfare Criteria | | C-Well Assessment measures version 2.0 |
|---|---|---|---|
| Enrichment | 16 | Application of enrichment | 16.1 Enrichment variability |
| | | | 16.2 Enrichment frequency |
| | | | 16.3 Enrichment novelty |
| | | | 16.4 Enrichment engagement records |
| | | | 16.5 Enrichment safety protocol |
| | | | 16.6 Average enrichment engagement time** |

showing affiliative behaviours (Figure 7) were all correlated to higher overall C-Well scores for individuals.

## Objective 2: Presentation of version 2.0 of the C-Well framework

***Approach to updating the C-Well assessment***

While the C-Well has been updated a number of times since its inception (Figure 1), efforts were made to not update the structure too frequently to protect the standardised nature of the assessment, considered as one of its strengths. Decisions about how to update the measures and their methods were based on discussions between C-Well assessors, where external expert feedback was sought on a case-by-case basis for some measures. A full log of the changes made to the C-Well assessment measures from 2015 to 2024, with rationale and references, can be found in Supplementary Information 3.

***Table of measures for C-Well assessment version 2.0***

As in the original publication of the C-Well assessment version 1.0 (Clegg et al. 2015), the full set of measures in version 2.0 of the C-Well is included here (Table 4). While measure composition varied for different species in past versions (see Supplementary Information 2), in version 2.0 there are no differences to the measure names, but there are species differences in the scoring criteria of several measures (e.g., 5.1.2 Pool volume, 7.4 Blood parameters).

The proportion of animal-based measures (denoted by asterisks in Table 4) has decreased from 21 out of 36 (58%) in the C-Well version 1.0 to 19 out of 44 (43%) in version 2.0. However, the number of measures based on quantitative behavioural data taken during observations in the animal's 'free-time' (outside of training and other sessions) has increased from one measure in C-Well 1.0 to eight measures in C-Well 2.0 (denoted by two asterisks in Table 4), reflecting efforts to increase objectivity.

## Discussion

The C-Well process, methods and results presented above can be viewed as an example of a species-specific, animal-based approach to monitoring zoo animal welfare, ultimately aiming to lead to welfare improvements where possible. The following discussion aims to critically evaluate the findings, explore the challenges and limitations, and outline potential future directions so that the approach could be adapted to other zoo species if desired.

***Effectiveness of the C-Well assessment approach***

To evaluate whether the C-Well approach is effective, we will consider whether it represents an objective assessment of cetacean welfare and whether the results are leading to measurable improvements. In addressing the former question, it is prudent to question the C-Well's validity, accuracy and reliability (Browning 2022). Validity concerns whether the assessment measures a target state or not – in this case, individual cetacean welfare. Following suggestions in the literature (Browning 2022; Fischer et al. 2024), we have investigated this through testing correlations between selected measures and changes in the target state, i.e., overall C-Well score, our measure of welfare. We found that pool volume per animal, group size, and affiliative behaviour correlated significantly with C-Well scores, whilst anticipatory behaviour did not. Our experimental design was not a causal one, so we cannot conclude that the three former parameters were the agents that led to the change in C-Well scores. We can, however, interpret that pool volume, group size and affiliative behaviour are valid predictors of overall welfare as measured by the C-Well, which fits with current understanding on cetacean welfare in that these gregarious animals would likely experience more positive welfare states from large enclosures and group sizes as this would offer them increased choice of activity and social partner, leading to higher levels of affiliative behaviour (Clegg et al. 2017; Rose et al. 2017).

Affiliative behaviour in particular has strong evidence to support its construct validity as a welfare measure, given that many wild cetacean studies have found higher affiliation rates to correlate to survival and increased fitness (Stanton and Mann 2012; Ellis et al. 2017; Holmes et al. 2024), and captive research showing that affiliative behaviours correlated with optimistic cognitive bias (Clegg et al. 2017), which is a well-established test of overall welfare. The fact that anticipatory behaviour was not found to be a significant predictor of overall welfare score will be discussed in the limitations section below.

The accuracy of the C-Well approach, i.e. how close the recorded welfare state is to the true state, could be considered using the standard errors for the selected welfare measure predictors (Table 3), which are relatively small and suggest they have a precise effect on the overall C-Well score. Accuracy is otherwise difficult to evaluate in welfare assessments or indices because a) welfare is a subjective and dynamic state whose true value is only known to the animal itself, and b) no one measure of welfare can be used to test the accuracy of assessments.

The reliability of the C-Well assessment, i.e. how consistent and repeatable it is, can be considered in terms of how consistent the data collection is within and between different assessors. Repeatability over time would not apply to the dynamic state of welfare, and some studies have suggested that the results of welfare measures such as affiliative behaviour frequency can significantly change over less than six months, with no other substantial changes to the environment (Clegg et al. 2017). This is partially supported by the fact that our individual animal random factor explained very little variance in our model with the welfare measure predictors, although it should be noted that individual variation was found in the model testing the effect of repeated assessments on overall C-Well scores. Inter-Observer Reliability for the complete C-Well was not tested between the C-Well assessors, although components of the assessment were applied by two or more assessors simultaneously on the same individual animals and results compared informally. Regarding the behavioural observation component, all assessors conducted practice and video observations, where IOR was verified using the platform ZooMonitor, before conducting the C-Well assessment for the first time.

To answer whether the C-Well is effective as a catalyst for welfare improvements, we turn to our analysis of the sub-set of data that included between two and four repeated assessments. These results showed that as more repeat assessment were applied, overall C-Well scores significantly increased. As part of the service provided with application of the C-Well, a detailed report including recommendations for how to improve scores was given to facilities after every assessment. Given the results we assumed that these recommendations were being implemented to a certain extent. Again, however, this analysis could not indicate causality, since scores could also be influenced by many other factors such as social group, management or other environmental changes. Nevertheless, this result tentatively suggests that the C-Well process itself may be effective in bringing about welfare improvements.

Lastly, it is worth noting that although validity, reliability, and accuracy are important to optimise in any framework, some research suggests that evaluations of welfare assessment approaches should be specific to their varied purposes, whether those are to describe a pre-existing situation (descriptive models), assess against a set of standards (normative models), or to help people make better decisions (prescriptive models) (Botreau et al. 2007a). It may be that the C-Well approach is more similar to a 'prescriptive model' since there are no pre-defined norms or stable states to measure: these models should not be evaluated by their empirical validity but instead by their pragmatic value and role in decision-making. This should be considered in further detail as the C-Well project progresses and the potential increases for meaningful evaluations.

Given the points of this discussion, those responsible for cetaceans on public display might consider whether they should apply management changes based on the initial findings presented here: for example, using resources to prioritise larger pool volumes per animal and group sizes, while using affiliative behaviour frequencies to monitor individual welfare. We would suggest that because all these indicators have strong construct validity i.e. they can be easily aligned with established theoretical constructs of cetacean welfare, these findings could start to be applied in practice. However, caution and managerial experience must be exercised when making changes since parameters like group size should not be used without considering related factors like compatibility, control of breeding, and pool volume. The suggested differences in welfare scores between Domain and age/sex categories could also be used conservatively to prioritise certain resource areas or demographics, again while overlaying context and anecdotal evidence related to the specific cetacean individuals.

### *Challenges and limitations of the C-Well approach*

Although the methods of applying the C-Well measures have benefitted from a decade of application, the global nature of cetacean facilities, spanning cultures, languages and regulations, means that standardisation of the qualitative measures can be challenging. In these cases, like with many welfare assessments, the final decision on how to award a welfare score rests with the assessor, but introduces the risk of bias (Wemelsfelder and Mullan 2014; Salas et al. 2024). To mitigate this risk, future C-Well efforts should include more IOR testing opportunities, and continue to transition the more subjective measures towards protocols that rely on quantitative data. However, there will likely always be a place for qualitative measures in the assessment, as it has been suggested that these parameters are uniquely able to capture some types of holistic, contextual data that are key to welfare insights (Wemelsfelder et al. 2000; Stockman et al. 2011; Phillips et al. 2017;).

Another methodological challenge encountered in the C-Well is the individual variation, which may be inherent to some of the animal-based measures included, that may limit welfare interpretations. An example might be anticipatory behaviour (definition in Table 1), which is considered a key cross-species, non-linear welfare indicator where moderate levels indicate positive states but excessive levels suggest negative states (Watters 2014; Makowska and Weary 2016; Anderson et al. 2020), and initially validated with dolphins (Clegg and Delfour 2018). However, anticipatory behaviour is also known to include a broad range of behaviours and emotional states, sometimes crossing over into Abnormal Repetitive Behaviour (ARB), and is impacted by a wide range of internal and external stimuli (Mason and Rushen 2008; Krebs et al. 2022). Our analyses found that anticipatory behaviour was not a significant predictor of overall welfare scores, and the wide standard error values suggest high variability. This may be because behavioural data was only collected over a week or two during assessments, and may not be representative of true frequencies. Until there is more research on anticipatory behaviour and its potential link to cetacean welfare, a limitation of the C-Well assessment is that the scoring thresholds have to remain extremely conservative.

Some practical challenges faced during the C-Well applications included needing to rapidly learn individual identification of animals, sometimes with limited visual fields over the pools. To mitigate this risk, C-Well assessors trained with facility staff and had to achieve 100% accuracy before proceeding to behavioural data collection. Conducting remote assessments was possible in a few cases but required a close level of collaboration with the facility, and the acceptance that bias would be introduced where facility staff were responsible for collecting data. Remote assessments were most successful where live video systems were able to be used by assessors to collect data themselves.

Limitations also exist concerning interpretation of the C-Well results, including difficulties in determining causality in dynamic and complex systems, making conclusions where measures are not weighted, and bench-marking results without a fully representative sample of cetacean facilities. Our results here found that higher pool volumes were correlated with higher welfare scores, but we cannot conclude whether it was the volume itself that may have improved other welfare parameters, or if it was more likely to be the increased depth and access to multiple pools that generally accompanies increased volumes (Miller et al. 2021). As discussed previously, aggregation of total scores is a central limitation to all welfare assessments (Botreau et al. 2007a; Hampton et al. 2023), where all solutions involve a set of ethical decisions that could be debated (Sandøe et al. 2019). The C-Well's current system of unweighted totalling of scores for each measure risks over-estimations of some measures, which may matter less to the animal, and under-estimations of the importance of other results. The C-Well results are further limited by the fact that since the assessment is voluntary, and must be funded by the host facility or another stakeholder, there has likely been an over-representation in the dataset of better-resourced and higher welfare facilities around the world. Only a few assessments have been conducted on behalf of governments or NGOs on facilities where welfare is suspected to be reduced; this means that formal benchmarking using C-Well results has not yet been possible.

A final challenge of the C-Well that spans methodology, practicality, and interpretation of data is striking a balance between collecting enough data to make meaningful conclusions whilst ensuring that the process remains practical, easy to implement, and affordable enough to encourage implementation. This issue has also been identified in farm and other zoo welfare assessments (Veissier et al. 2008; Botreau et al. 2009; Jones et al. 2022;), where it has been recognised there is no 'optimal' duration or size of dataset, but moreover different approaches exist that will be suited to the range of purposes of assessments. In the case of the C-Well, a minimum amount of behavioural data collected on individual animals was set to ensure that somewhat representative activity budgets can be established and used (Tallo-Parra et al. 2023). This behavioural data approach, however, means that currently, only specialised assessors with years of experience with cetaceans can be used to apply the assessment, which may be limiting its adoption (Jones et al. 2022).

***Advancing the C-Well framework***

*Future updates and applications*

An iterative process involving periodic review and updates is recommended for developing welfare assessments (Browning 2022; Jones et al. 2022), which the C-Well has followed thus far. Future refinement efforts will continue incorporating more animal-based measures as the evidence base grows. For instance, Qualitative Behavioural Assessments (QBA) were originally used in the WelfareQuality® framework, but were not included in the 2015 C-Well or subsequent versions due to a lack of validated QBAs for cetaceans. Multiple projects are now working on this (e.g. Warner et al. 2022) and therefore QBAs may be able to be included in the next assessment update, particularly since they can provide richer, more nuanced insights into zoo animal welfare (Rose and Riley 2019; Skovlund et al. 2023).

As has been attempted through the publication of this paper, the C-Well project will continue to uphold transparency as a core principle, particularly because it encourages standardised application and pooling of data. As was recently acknowledged by a group of welfare researchers: "…there will never be a perfect solution to the problem of converting and summating numerical scores from subjective assessments. Each solution will be based on ethical and methodological assumptions that can be debated. However, what can be achieved is a solution where these assumptions are made transparent" (Hampton et al. 2023). The C-Well will continue to be applied at those facilities that request it, and has started to be used in other sectors such as the tourism industry and as part of Environmental, Social and Governance efforts in corporate transactions. This may lead to a more representative set of results with which to consider benchmarks and best practice standards. Once the evidence-base is more robust, a longer term aim is to expand the scoring system to include best practice and aspirational scoring thresholds to better facilitate sustainable, animal-centric changes.

*Weighting and further validation*

The gold standard of the C-Well framework would be to only include scientifically validated welfare indicators that have undergone a weighting process. For this, the standardised protocols need to be applied across many more animals. Hampton et al. (2019) suggest more than 300) across time, and more comprehensive analyses must take place to analyse the indicators' relationships with overall welfare scores (Hampton et al. 2023; Fischer et al. 2024). This study begins to demonstrate the feasibility of such an approach, but the dataset remains small and constrained by anonymity. Several other validation methods would be to use the Delphi method to gather expert opinion (Veasey 2020; Whittaker et al. 2021), and correlating C-Well findings to cognitive bias results (Clegg 2018; Browning 2022).

*The Zoo-Well framework*

The C-Well assessment is inherently limited in that it is only applicable to a small group of animals in relation to all species housed in zoos. As was the case with adapting the WelfareQuality approach to establish the original C-Well, the latter has recently been further adapted to apply to any zoo animal in what has been termed the "Zoo-Well framework" (owned by Animal Welfare Expertise). Similar to how the C-Well includes some adaptable measure thresholds to apply to multiple cetacean species, the Zoo-Well framework includes a core set of measures which can easily undergo species- or taxa-specific adaptation. Additional measures can be added based on the literature and expert opinion, which we have sought in our development of species Zoo-Well assessments thus far. Zoo-Well assessments have been applied to polar bears *Ursus maritimus*, koala *Phascolarctos cinereus*, dhole *Cuon alpinus*, chimpanzees *Pan troglodytes*, gorillas *Gorilla gorilla*, and seven pinniped species. While species-specific, animal-based protocols have been upheld as optimal models to assess zoo animal welfare but considered too resource-intensive to develop (Sherwen et al. 2018; Tallo-Parra et al. 2023; Salas et al. 2024), the Zoo-Well leverages the findings and advances of the C-Well project to facilitate the streamlined development of such protocols.

## Conclusion

The Cetacean Welfare Assessment was designed to bridge the gap between research and the practice of managing cetaceans for public display, aiming to provide objective information on individual welfare states. Following a decade of applications and version updates, initial findings suggest that repeated assessments correlate to improved overall scores, and that pool volume, group size, and affiliative behaviour are significant predictors of total scores with no significant effect for anticipatory behaviour. The measures included in C-Well version 2.0 were presented to indicate how protocols must keep changing in line with the latest research. Whilst the C-Well could be evaluated as an effective prescriptive model to aid decision-making on cetacean welfare, its limitations must be recognised. Finally, the C-Well has recently facilitated the development of the Zoo-Well framework, where the evolution of the former assessment has been leveraged to develop an adaptable, practical protocol that is being readily applied on a species-specific basis.

## Acknowledgements

We would like to sincerely thank Drs Jill Richardson and Holli Eskelinen for their guidance, input and collaboration on the first version of the C-Well Assessment published in 2015. We are very much indebted to all the organisations who were involved with the development of, or chose to apply, the C-Well Assessment on their animals, and salute their forward-thinking attitude in terms of monitoring and enhancing welfare. Some organisations preferred not to be named, but those that were happy to receive our grateful acknowledgement for their collaboration were: Kolmårdens Djurpark, Coffs Coast Wildlife Sanctuary (CCWS), South African Association for Marine Biological Research (SAAMBR), John G. Shedd Aquarium, and Resorts World Sentosa. We would also like to thank all the C-Well assessors for their commitment and motivation to evaluate cetacean welfare.

## Ethics statement

The data collected during cetacean welfare assessments and presented in this article were gathered as part of a contracted consultancy with the client, not for research purposes. All data included in the welfare assessments was either non-invasive and adhered to the Guidelines for the Treatment of Animals in Behavioural Research and Teaching published by the Association for the Study of Animal Behaviour (http://asab.nottingham.ac.uk/downloads/guidelines2006.pdf), or was included after the client organisations' veterinary teams had collected it for their own health management purposes.

## Conflict of interest

The authors received no funding to complete this publication. To reinforce our approach of presenting the data with full transparency and clarity, the co-authors' organisation Animal Welfare Expertise is an animal welfare consultancy that was contracted by organisations to conduct the C-Well Assessment for management optimisation purposes, i.e. not primarily for research objectives. The organisations' agreement to publish the data was not contingent on Animal Welfare Expertise applying the assessment, and conversely Animal Welfare Expertise was not required by the organisations to publish the results. Written consent was obtained for every organisation concerning the use of their fully anonymised data for this publication.

## References


Anderson C., von Keyserlingk M.A.G., Lidfors L.M., Weary D.M. (2020) Anticipatory behaviour in animals: A critical review. *Animal Welfare* 29(3): 231–238. https://doi.org/10.7120/09627286.29.3.231

Bates D., Maechler M., Bolker B., Walker S., Christensen R.H.B., Singmann H., Dai B., Eigen C. (2017) lme4: Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1): 1–113. https://doi.org/10.18637/jss.v067.i01

Baumgartner K., Hüttner T., Clegg I.L.K., Hartmann M.G., Garcia-Párraga D., Manteca X., Mercera B., Monreal-Pawlowsky T., Pilenga C., Ternes K., Tallo-Parra O., Vaicekauskaite R., von Fersen L., Yon L., Delfour F. (2024) Dolphin-WET—Development of a welfare evaluation tool for bottlenose dolphins (*Tursiops truncatus*) under Human Care. *Animals* 14(5): 701. https://doi.org/10.3390/ani14050701

Benn A.L., McLelland D.J., Whittaker A.L. (2019) A review of welfare assessment methods in reptiles, and preliminary application of the welfare quality® protocol to the pygmy blue-tongue skink, *Tiliqua adelaidensis*, using animal-based measures. *Animals* 9(1): 27.. https://doi.org/10.3390/ani9010027

Blokhuis H.J. (2008) International cooperation in animal welfare: the Welfare Quality® project. *Acta Veterinaria Scandinavica* 50(1): S10. https://doi.org/10.1186/1751-0147-50-S1-S10

Boissy A., Manteuffel G., Jensen M.B., Moe R.O., Spruijt B., Keeling L.J., Winckler C., Forkman B., Dimitrov I., Langbein J., Bakken M., Veissier I., Aubert A. (2007) Assessment of positive emotions in animals to improve their welfare. *Physiology and Behavior* 92(3): 375–397. https://doi.org/10.1016/j.physbeh.2007.02.003

Botreau R., Bonde M., Butterworth A., Perny P., Bracke M.B.M., Capdeville J., Veissier I. (2007a) Aggregation of measures to produce an overall assessment of animal welfare. Part 1: a review of existing methods. *Animal* 1(08): 1179–1187. https://doi.org/10.1017/S1751731107000535

Botreau R., Veissier I., Butterworth A., Bracke M.B.M., Keeling L.J. (2007b) Definition of criteria for overall assessment of animal welfare. *Animal Welfare* 16(2): 225–228.

Botreau R., Veissier I., Pern P. (2009) Overall assessment of animal welfare: Strategy adopted in Welfare Quality? *Animal Welfare* 18(4): 363–370.

Browning H. (2022) Assessing measures of animal welfare. *Biology and Philosophy* 37(36). https://doi.org/10.1007/s10539-022-09862-1

Bruck J.N. (2024) The Cetacean Sanctuary: A sea of unknowns. *Animals* 14(2): 335.https://doi.org/10.3390/ani14020335

Clegg I.L.K. (2018) Cognitive bias in zoo animals: An optimistic outlook for welfare assessment. *Animals* 8(7): 104. https://doi.org/10.3390/ani8070104

Clegg I.L.K., Borger-Turner J.L., Eskelinen H.C. (2015) C-Well: The development of a welfare assessment index for captive bottlenose dolphins (*Tursiops truncatus*). *Animal Welfare* 24(3): 267–282. https://doi.org/10.7120/09627286.24.3.267

Clegg I.L.K., Delfour F. (2018) Cognitive judgement bias is associated with frequency of anticipatory behavior in bottlenose dolphins. *Zoo Biology* 37(2): 67–73. https://doi.org/10.1002/zoo.21400

Clegg I.L.K., Domingues M., Ström E., Berggren L. (2023) Cognitive foraging enrichment (but not non-cognitive enrichment) improved several longer-term welfare indicators in bottlenose dolphins. *Animals* 13(2): 238. https://doi.org/10.3390/ani13020238

Clegg I.L.K., Rödel H.G., Boivin X., Delfour F. (2018) Looking forward to interacting with their caretakers: Dolphins' anticipatory behaviour indicates motivation to participate in specific events. *Applied Animal Behaviour Science* 202: 85–93. https://doi.org/10.1016/j.applanim.2018.01.015

Clegg I.L.K., Rödel H.G., Delfour F. (2017) Bottlenose dolphins engaging in more social affiliative behaviour judge ambiguous cues more optimistically. *Behavioural Brain Research* 322(A): 115–122. https://doi.org/10.1016/j.bbr.2017.01.026

Connor R.C. (2007) Dolphin social intelligence: complex alliance relationships in bottlenose dolphins and a consideration of selective environments for extreme brain size evolution in mammals. *Philosophical Transactions of the Royal Society B* 362: 587–602. https://doi.org/10.1098/rstb.2006.1997

Delfour F., Vaicekauskaite R., García-Párraga D., Pilenga C., Serres A., Brasseur I., Pascaud A., Perlado-Campos E., Sánchez-Contreras G. J., Baumgartner K., Monreal-Pawlowsky T. (2021) Behavioural diversity study in bottlenose dolphin (*Tursiops truncatus*) groups and its implications for welfare assessments. *Animals* 11(6): 1715. https://doi.org/10.3390/ani11061715

Dudzinski K.M., Ribic C.A. (2017) Pectoral fin contact as a mechanism for social bonding among dolphins. *Animal Behavior and Cognition* 4(1): 30–48. https://doi.org/10.12966/abc.03.02.2017

Ellis S., Franks D.W., Nattrass S., Cant M.A., Weiss M.N., Giles D., Balcomb K.C., Croft D.P. (2017) Mortality risk and social network position in resident killer whales: Sex differences and the importance of resource abundance. *Proceedings of the Royal Society B,* 284(2017): 1313.

Faraway J.J. (2006) *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. Boca Raton, FL: CRC Press.

Fischer B., Pempek J.A., Flint J., Wittum T., Flint M. (2024) Application of a One Welfare-Based Ecosystem Model to assess wild collection for public aquariums. *Aquatic Conservation: Marine and Freshwater Ecosystems* 34(2): 1–17. https://doi.org/10.1002/aqc.4098

Hampton J.O., Hemsworth L.M., Hemsworth P.H., Hyndman T.H., Sandøe P. (2023) Rethinking the utility of the Five Domains model. *Animal Welfare* 32(62): 1–10. https://doi.org/10.1017/awf.2023.84

Hampton J.O., Mackenzie D.I., Forsyth D.M. (2019) How many to sample ? Statistical guidelines for monitoring animal welfare outcomes. *PLoS ONE* 14(1): 1–18. https://doi.org/10.1371/journal.pone.0211417

Holmes K.G., Krützen M., Ridley A.R., Allen S.J., Connor R.C., Gerber L., Stamm C.F., King, S.L. (2024) Juvenile social play predicts adult reproductive success in male bottlenose dolphins. *PNAS* 121(25): 1–9.

Jones N., Sherwen S.L., Robbins R., McLelland D.J., Whittaker A.L. (2022) Welfare assessment tools in zoos: From theory to practice. *Veterinary Sciences* 9(4): 170. MDPI. https://doi.org/10.3390/vetsci9040170

Kagan R., Carter S., Allard S. (2015) A universal animal welfare framework for zoos. *Journal of Applied Animal Welfare Science* 18(Sup1): S1–S10.

Krebs B.L., Chudeau K.R., Eschmann C.L., Tu C.W., Pacheco E., Watters J.V. (2022) Space, time, and context drive anticipatory behavior: Considerations for understanding the behavior of animals in human care. *Frontiers in Veterinary Science* 9: 1–12. Lauderdale L.K., Walsh M.T., Mellen J.D., Granger D.A., Miller L.J. (2021) Environmental enrichment, training, and habitat characteristics of common bottlenose dolphins (*Tursiops truncatus*) and Indo-Pacific bottlenose dolphins (*Tursiops aduncus*). *PLoS ONE* 16(8).

Littlewood K.E., Heslop M.V., Cobb M.L. (2023) The agency domain and behavioral interactions: assessing positive animal welfare using the Five Domains Model. *Frontiers in Veterinary Science* 10: 1284869. https://doi.org/10.3389/fvets.2023.1284869

Makowska I.J., Weary D.M. (2016) Differences in anticipatory behaviour between rats (*Rattus norvegicus*) housed in standard versus semi-naturalistic laboratory environments. *PLoS ONE* 11(1). https://doi.org/10.1371/journal.pone.0147595

Mason G., Rushen J. (2008) *Stereotypic animal behaviour: fundamentals and applications to welfare and beyond* (G. J. Mason & J. Rushen, Eds.). CABI Publishing. https://doi.org/10.1079/9780851990040.0227

Mellor D.J. (2016) Updating animal welfare thinking: Moving beyond the " Five Freedoms " towards " A Life Worth Living ". *Animals* 6(21): 1–20. https://doi.org/10.3390/ani6030021

Mellor D.J. (2017) Operational details of the five domains model and its key applications to the assessment and management of animal welfare. *Animals* 7(8): 60. https://doi.org/10.3390/ani7080060



http://doi/org/10.19227/jzar.v13i3.929

Mellor D.J., Beausoleil N. (2015) Extending the 'Five Domains' model for animal welfare assessment to incorporate positive welfare states. *Animal Welfare* 24(3): 241–253. https://doi.org/10.7120/09627286.24.3.241

Mellor D.J., Beausoleil N.J., Littlewood K.E., McLean A.N., McGreevy P.D., Jones B., Wilkins C. (2020) The 2020 five domains model: Including human–animal interactions in assessments of animal welfare. *Animals* 10(10): 1–24. https://doi.org/10.3390/ani10101870

Miller L.J., Lauderdale L.K., Mellen J.D., Walsh M.T., Granger D.A. (2021) Relationships between animal management and habitat characteristics with two potential indicators of welfare for bottlenose dolphins under professional care. *PLoS ONE* 16(8). https://doi.org/10.1371/journal.pone.0252861

Mononen J., Møller S.H., Hansen S.W., Hovland, A.L., Koistinen T., Lidfors L., Malmkvist J., Vinke C.M., Ahola L. (2012) The development of on-farm welfare assessment protocols for foxes and mink: The WelFur project. *Animal Welfare* 21(3): 363–371. https://doi.org/10.7120/09627286.21.3.363

Nicol C., Bejder L., Green L., Johnson C., Keeling L., Noren D., Van der Hoop J., Simmonds M. (2020) Anthropogenic threats to wild cetacean welfare and a tool to inform policy in this area. *Frontiers in Veterinary Science* 7: 1–12. https://doi.org/10.3389/fvets.2020.00057

Phillips C.J.C., Tribe A., Lisle A., Galloway T.K., Hansen K. (2017) Keepers' rating of emotions in captive big cats, and their use in determining responses to different types of enrichment. *Journal of Veterinary Behavior* 20: 22–30. https://doi.org/10.1016/j.jveb.2017.03.006

R Development Core Team. (2025) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/"

Rose N.A., Hancock Snusz G., Brown D.M., Parsons E.C.M. (2017) Improving captive marine mammal welfare in the United States: Science-based recommendations for improved regulatory requirements for captive marine mammal care. *Journal of International Wildlife Law and Policy* 20(1): 1–35. https://doi.org/10.1080/13880292.2017.1309858

Rose P., Riley L. (2019) The use of qualitative behavioural assessment in zoo welfare measurement and animal husbandry change. *Journal of Zoo and Aquarium Research* 7(4): 150–161.

Salas M., Manteca X., Abáigar T., Delclaux M., Enseñat C., Martínez-Nevado E., Quevedo M. Á., Fernández-Bellon H. (2018) Using farm animal welfare protocols as a base to assess the welfare of wild animals in captivity—Case study: Dorcas gazelles (*Gazella dorcas*). *Animals* 8(7): 111. https://doi.org/10.3390/ani8070111

Salas M., Tallo-Parra O., Manteca X. (2024) Evidence-based zoo animal welfare assessment: Putting science into practice. *Animal Welfare* 32(e62): 1–10. https://doi.org/10.19227/jzar.v12i4.810

Sandøe P., Corr S., Lund T., Forkman B. (2019) Aggregating animal welfare indicators: Can it be done in a transparent and ethically robust way? *Animal Welfare* 28(1): 67–76. Serres A., Boys R.M., Beausoleil N.J., Platto S., Delfour F., Li S. (2024) The first standardized scoring system to assess the welfare of free-ranging Indo-Pacific humpback dolphins (*Sousa chinensis*). *Aquatic Conservation: Marine and Freshwater Ecosystems* 34(11): e70004. https://doi.org/10.1002/aqc.70004

Serres A., Hao Y., Wang D. (2020a) Swimming features in captive odontocetes: Indicative of animals' emotional state? *Behavioural Processes* 170: 103998. https://doi.org/10.1016/j.beproc.2019.103998

Serres A., Robeck T., Deng X., Steinman K., Hao Y., Wang D. (2020b) Social, reproductive and contextual influences on fecal glucocorticoid metabolites in captive Yangtze finless porpoises (*Neophocaena asiaeorientalis asiaeorientalis*) and bottlenose dolphins (*Tursiops truncatus*). *Journal of Zoological and Botanical Gardens* 1(1): 24–41. https://doi.org/10.3390/jzbg1010003

Sherwen S.L., Hemsworth L.M., Beausoleil N.J., Embury A., Mellor D.J. (2018) An animal welfare risk assessment process for zoos. *Animals* 8(130): 1–16. https://doi.org/10.3390/ani8080130

Skovlund C.R., Kirchner M.K., Contiero B., Ellegaard S., Manteca X., Stelvig M., Tallo-Parra O., Forkman B. (2023) Qualitative Behaviour Assessment for zoo-housed polar bears (*Ursus maritimus*): Intra- and inter-day consistency and association to other indicators of welfare. *Applied Animal Behaviour Science* 263: 2–10. https://doi.org/10.1016/j.applanim.2023.105942

Stanton M.A., Mann J. (2012) Early social networks predict survival in wild bottlenose dolphins. *PLoS ONE* 7(10): 1–6. https://doi.org/10.1371/journal.pone.0047508

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 naive and habituated to road transport. *Animal Production Science* 51(3): 240–249. https://doi.org/10.1071/AN10122

Tallo-Parra O., Salas M., Manteca X. (2023) Zoo animal welfare assessment: Where do we stand? *Animals* 13(12): 1966.. https://doi.org/10.3390/ani13121966

Ugaz C., Valdez R.A., Romano M.C., Galindo F. (2013) Behavior and salivary cortisol of captive dolphins (*Tursiops truncatus*) kept in open and closed facilities. *Journal of Veterinary Behavior* 8(4): 285–290. https://doi.org/10.1016/j.jveb.2012.10.006

Veasey J.S. (2020) Assessing the psychological priorities for optimising captive Asian elephant (*Elephas maximus*) welfare. *Animals* 10(1): 1–13. https://doi.org/10.3390/ani10010039

Veissier I., Butterworth A., Bock B., Roe E. (2008) European approaches to ensure good animal welfare. *Applied Animal Behaviour Science* 113(4): 279–297. https://doi.org/10.1016/j.applanim.2008.01.008

von Fersen L., Encke D., Huttner T., Baumgartner K. (2018) Establishment and implementation of an animal welfare decision tree to evaluate the welfare of zoo animals. *Aquatic Mammals* 44(2): 211–220. https://doi.org/10.1578/AM.44.2.2018.211

Wark J.D., Cronin K.A., Niemann T., Shender M.A., Horrigan A., Kao A., Ross M.R. (2019) Monitoring the behavior and habitat use of animals to enhance welfare using the ZooMonitor App. *Animal Behavior and Cognition* 6(3): 158–167. https://doi.org/10.26451/abc.06.03.01.2019

Warner E., Brando S., Wemelsfelder F. (2022) Recognising emotional expressions in captive bottlenose dolphins: Can lay observers agree using qualitative behavioural assessment? *Journal of Zoo and Aquarium Research* 10(3): 139–148. https://doi.org/10.19227/jzar.v10i3.601

Watters J.V. (2014) Searching for behavioral indicators of welfare in zoos: Uncovering anticipatory behavior. *Zoo Biology* 33(4): 251–256. https://doi.org/10.1002/zoo.21144

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(3): 193–215. https://doi.org/10.1016/S0168-1591(99)00093-3

Wemelsfelder F., Mullan S. (2014) Applying ethological and health indicators to practical animal welfare assessment. *Scientific & Technical Review* 33(1): 111–120.

Whittaker A.L., Golder-Dewar B., Triggs J.L., Sherwen S.L., McLelland D.J. (2021) Identification of animal-based welfare indicators in captive reptiles: A delphi consultation survey. *Animals* 11(7): 2010. https://doi.org/10.3390/ani11072010