

Table 3. Evaluation factor scoring for each HAR method evaluated

<b>Evaluation factor</b>	<b>Reliability</b>	<b>Robustness</b>	<b>Practical application</b>	<b>Validity and accuracy</b>	<b>Feasibility (safety, financial, long term study)</b>
<b>Method 1: Latency and distance parameters</b>	<p>Good - observers using distance meters could potentially miss subtle cues from animals.</p> <p>Could be difficult to know specifically whether the animal is responding to the human cue or coincidentally performing the required behaviour by chance or for another reason.</p>	<p>Fair – could be difficult for multi-zoo comparison if different behaviour requests are used from zoo keepers at different HAI events, which may not be directly comparable.</p>	<p>Good – requires little formal training, can last a short time and be recorded from a distance. Physical existing landmarks can be used to measure distance. .</p>	<p>Good – influence of animal sociality must be considered, as differences in behavioural responses to keeper commands have been found between socially and solitary housed species (Ward and Melfi, 2013).</p> <p>Estimations of distance parameters are likely to introduce errors and inconsistencies between or within studies.</p>	<p>Good – performed during routine HAI events (e.g. daily training, movement between enclosures, feeding events) therefore safe for human and animal, suitable for long term study.</p> <p>Distance meters could be used as an alternative to estimations and physical marks in enclosure, at little costs.</p>
<b>Method 2: Avoidance test</b>	<p>Good – inter-observer reliability has been found to be high irrespective of whether person is experimenter or observer. Individual animal reactions to avoidance distance test were shown as moderately repeatable when tested by different experimenters (Windschnurer et al 2009).</p>	<p>Fair – currently no zoo studies using this method. Possibly due to ethical implications of creating a situation to potentially elicit a fear response.</p> <p>Would not be suitable for all zoo species, especially species with protected contact therefore not sufficient to fully test an animal's avoidance response.</p>	<p>Poor – Time consuming on a large fam scale, requires specific training by observer to properly move into area, recognise first avoidance reaction (Battini et al 2016).</p>	<p>Good – has been validated for several agricultural species (Battini et al 2016).</p> <p>Validated by showing its sensitivity to gentle HAIs (Windschnurer et al 2009).</p> <p>Interpretation of animal response can also be difficult if animal did not move and neither approached nor avoided the human (Rousins and Wablinger 2004)</p>	<p>Poor – safety concerns for both human and animal with some zoo species.</p>
<b>Method 3: Voluntary animal approach</b>	<p>Good - easily performed, however, curiosity of a novel event such as a human's presence may increase the motivation to approach. This must be considered for use with animals in a zoo setting which rarely have human contact.</p>	<p>Fair – may be unsuitable for animals which rarely have human contact as it is measuring animal curiosity rather than fear.</p> <p>The safety risks for participants will also prohibit the use of this test for some zoo species.</p>	<p>Fair – may require time to carry out test and train observers to be able to identify approaches, whilst ensuring safety.</p> <p>Test is dependent on animal management and accessibility, i.e. animal species and temperament will dictate whether this test can be performed from inside the animal's enclosure or not.</p>	<p>Good - curiosity of a novel event such as a human's presence may increase the motivation to approach.</p>	<p>Good – minimal financial cost, potential safety implications to both animal and human however this test could be adapted to been used with the presence of a physical barrier.</p>
<b>Method 4: Reaction to handling</b>	<p>Poor – several potential confounding variables which could reduce reliability and repeatability; handlers reacting differently based on animal response, physical and social environment. More testing is required.</p>	<p>Poor – test relies on animal being suitable for handling by humans in a safe manner. Due to the variation in handling of animals in zoos this test would not be suitable for all zoo species.</p>	<p>Poor – this test could be considered invasive and therefore has ethical implications.</p> <p>This test is species dependant meaning t cannot be performed for all species within a zoo.</p>	<p>Poor – has been used in conjunction with physiological measures, such as faecal glucocorticoid metabolites (Baird et al 2016), however these can be influenced by other variables such as feeding habits, diurnal variations and life history.</p> <p>Requires some standardisation in how the animal is handled as the variation between handling styles and skills of handlers could influence reliability.</p>	<p>Poor – this test could be considered invasive and therefore has ethical implications.</p> <p>Safety concerns for both animal and human.</p> <p>Animals which are not handled as part of daily routine would require additional time from keepers, and cause unnecessary stress, therefore not suitable for long-term monitoring.</p>

<b>Method 5: Qualitative behaviour assessment</b>	Excellent - Inter-observer reliability has been tested using a variety of statistical methods to rate zookeeper QBA scoring of animals; overall high levels of agreement have been found for zoo keepers assessing cheetahs (Wielebnowski 1999) and black rhinos (Carlstead et al 1999a).	Excellent – videos of animals can be obtained for all zoo species, dependant on practical application, enclosure design, video recording equipment.	Poor - due to the requirement of multiple observers to analyse clips, the practical application of QBA can be challenging and time consuming.	<p>Good – has been validated in previous studies and used as a cost-effective approach in monitoring farm animal welfare.</p> <p>Free-choice profiling allows observers to integrate subtle movements, posture and aspects of the context in which the behaviour occurs into an animal’s overall style of behaviour, evaluating the “animal-as-a-whole” (e.g. bold, shy, hostile) (Wemelsfelder et al 2000, 2001).</p> <p>High levels of observer agreement in scoring found when testing inter-observer reliability.</p>	<p>Poor – requires organising and conducting two phases of focus/observer groups to analyse clips which is time consuming and challenging.</p> <p>Cost of obtaining a video recording device.</p>
---	---	---	---	---	---