NATIONAL MARINE BIOLOGICAL ANALYTICAL QUALITY CONTROL SCHEME

DEVELOPMENT OF THE NMBAQC VIDEO RING TEST

Envision Mapping Ltd

Newcastle upon Tyne
UK

March 2010

Prepared for	**The NMBAQC Committee (Dr Matt Service, AFBINI)**
Title	Development of the NMBAQC Video Ring Test
Abstract	
Contract reference:	2007-1017-NMBAQC-Ring Test
Report identification:	2007-1017- ISS - Development of the NMBAQC Video Ring Test

Document control

Rev.	Originator	Date	Status	Checked & Approved
0	Judy Foster-Smith	08/09/09	Draft	J.E.
1	Judy Foster-Smith	22/09/09	Draft	I.S.S
2	Judy Foster-Smith	06/10/09	Final Draft	
3	Ian Sotheran	26/02/10	Submitted	

File index: \\Exchange\Projects\2007-1017-NMBAQC-Ring Test\REPORT\Final Report\Submitted Versions\Envision Report - Development Of The NMBAQC Video Ring Test (22nd February 2010).Docx

Report:

Prepared by Judy Foster-Smith & Ian Sotheran

Envision Mapping Ltd.
Stephenson House
Horsley Business Centre
Horsley , Northumberland
NE15 0NY
UK

T:+44 (0)1661 854 250
F:+44 (0)1661 854 361

NOTES:

Contents

1 INTRODUCTION 5

2 TEST 1 5

2.1 Aims 5

2.2 Methods 5

2.3 Assessment 5

2.4 Results 6

2.5 Feedback 7

2.6 Lessons learnt: 7

3 TEST 2 7

3.1 Aims 7

3.2 Methods 7

3.3 Assessment 10

3.4 Results 10

3.4.1 Analysis of Substrates 10

3.4.2 Analysis of Biota 11

3.5 Feedback 12

3.6 Discussion and lessons learnt 12

3.7 The Use of Still Images 13

4 TEST 3 13

4.1 Aims 13

4.2 Methods 14

4.3 Assessment 14

4.4 Results 14

4.4.1 Analysis of Substrates 14

4.4.2 Analysis of Biota 15

5 ISSUES OUTSTANDING (AT THE TIME OF THE WORKSHOP) 15

5.1.1 Workshop Summary 16

5.1.2 Workshop Action Points 16

6 RECOMMENDATIONS FOR THE FUTURE DIRECTION OF THE NMBAQC VIDEO RING TEST 16

6.1 Issues identified by the development of the test and workshop 16

6.2 Preparation 17

6.3 Training 17

6.4 The Test 17

6.5 Testing and Assessment 18

6.6 'Standard Operating Procedures' for Video Analysis 18

7 REFERENCES 19

8 APPENDICES 20

1 Introduction

This document outlines the design and development of the proposed NMBAQC 'Video Ring Test' carried out by Envision Mapping Ltd. It briefly describes the trial stages of the test, highlighting in particular:

- The content of each trial test;
- What was learnt from each of the trial tests;
- How the test was refined at each stage and
- What the outstanding issues are.

Full details of the process is given in the interim reports on Test 1 and Test 2 (Foster-Smith *et al.* 2008 and Sotheran *et al.*, 2009 respectively).

2 Test 1

Test 1 was circulated to participants during May 2008 for completion by end June 2008. The participants' results were submitted on time. Timescales for the development of the first test were short and following discussion with NMBAQC, in-house video footage collected by Envision Mapping or SeaMap research group were used for the test.

2.1 Aims

The aims of Test 1 were to:

- Establish the general abilities of the participants;
- Produce information that would help to refine the content of the test.

2.2 Methods

Hard copies of the 'Guidelines' for carrying out the test, 'Data Entry' forms and 'Feedback' questionnaires were sent to each of the contacts representing the participating organisations. In addition, DVDs containing the following were circulated:

- 10 x 1-minute video clips (all provided by Envision Mapping Ltd. - i.e. they were standardised as far as possible in that they had been produced using the same equipment by the same team);
- A set of 'Guidelines' explaining how to carry out the test;
- Electronic (.pdf) 'Data Entry' forms (see Appendices 2 (i) and 3(i));
- Electronic (.pdf) 'Feedback' questionnaires (see Appendix 4 (i));
- A set of 'Analysis Tools' to aid video analysis (see e.g. Appendix 5 (i));
- A reference list containing references to publications for aiding the identification of marine organisms, and for describing work relating to video survey and analysis.

The data entry forms were based on the current JNCC benthic still and video data entry spreadsheet. Participants were expected to analyse all aspects of all 10 video clips and record all organisms observed.

2.3 Assessment

Test 1 took a disproportionately long time to analyse because the majority of submissions were returned as hard copy, and the data had to be transcribed and re-organised before the analysis could take place. Statistical analysis was carried out using PRIMER to produce Bray-Curtis similarity values.

Test 1was not marked in the sense of providing participants with, for example, a percentage mark to indicate their absolute performance. Instead, the submissions were judged relative to each other, based on the degree of similarity between them. This method of judging performance was incorporated into the development of the test, as initially it was unclear what the responses would

be. Additionally, to provide an absolute measure of performance would have entailed making a subjective decision as to what was a correct or incorrect answer, when it was felt a test should be objective and independent from any subjective bias.

2.4 Results

21 participants from 11 organisations took part in this trial test.

The most notable aspect of the results was that there was a lot of discrepancy in the responses, as follows:

- The time taken to analyse the video clips varied hugely: individual participants took a mean time of between 15 and 60 minutes to analyse a 1-minute clip.
- There was relatively little similarity between participants in their assessment of the abundance of the different substrates. The mean similarity ranged between 25.8 (Sullom Voe 2) and 60.5 (Menai Strait 1).
- There was relatively little similarity in the sets of abundance assessments allocated by the different participants for the different organisms identified; they ranged between 7.8% (Poole Bay) and 66.0% (Isle of Wight).
- There was also difference of opinion regarding the 'biological zones'. 2 of the video sites were allocated all 4 of the different zones; 6 of the sites were thought to be from 3 different zones and the remaining 2 sites were allocated 2 different zones. In other words there was no total agreement as to which biological zones any of the videos represented.
- The same was true of the 'life form' allocations. For only one of the 10 video sites was there total agreement about which life form it represented (Sullom Voe 1).
- Similarly, the allocation of 'biotopes' varied enormously, with (as an example) as many as 9 different biotopes having been allocated for one of the sites (Poole Bay).

The likely reasons for these discrepancies are that:

- The participants' degree of experience of video survey and video analysis varied widely: between 'none' and '12 years'!
- There was a variety of different in-house or 'standard' analysis techniques used, many of which are not documented in SOPs.
- There was a large range of types of equipment used. Each organisation used different hardware and, in all, seven different versions of software were used for observing the videos. As a result, there would have been inconsistencies in resolution and therefore participants would effectively have been analysing different visual outputs.
- There was a lack of clarity of some of the instructions given.
- Very little use was made of the resources available. (For instance, in all, participants referred to only 8 books and 3 websites in total for identification of organisms).
- Participants used a variety of protocols and techniques to analyses the video.

Other interesting features that came out of the test include:

- From the results of all the participants, a total of 124 organisms were 'identified' from the ten minutes of video (between 12 and 38 per 1-minute video clip).
- The 10 most-recognised species were all fauna while the 10 least-recognised species were all flora.
- It appeared that several of the participants tried to allocate the organisms that they observed in the video to the lowest taxonomic level, i.e. they tried to give species names to organisms seen in the video when even expert taxonomists would not have done so. Consequently, their responses were often incorrect. They would have been more successful had they kept to a taxonomic level that they were 100% confident in.

2.5 Feedback

Collecting and collation of feedback was vital to the development of the test. Most of the participants found doing the test rather onerous. In particular, the feedback pointed to errors and omissions in the data entry forms and the lack of clarity in the guidance provided. In addition, there were several comments about the 'poor' quality of the video used and the lack of scale given.

2.6 Lessons learnt:

The main lessons learnt from Test 1 were that:

- The test needed to be simplified;
- Video clips being analysed needed to be longer than one minute in length, to be of better quality and to have a minimum set of metadata (including site, depth, date, ownership);
- Test submissions should have been in electronic form to help both the data entry and the analysis. Some people had not realised that electronic (.pdf) forms had been available on the CD provided, and they had not been able to get all of their information into the appropriate boxes on the hard copy version. Data entry in electronic format can allow sufficient space for responses. In addition, it can be controlled, for instance, by programming the data entry process to prevent blanks cells being returned by the candidates, or to prevent percentage totals (i.e. for substrate and species cover) being submitted unless they amount to 100.
- Refinement of the 'Guidance' document was also required to clarify instructions about how to complete the test.

3 Test 2

Test 2 was circulated to participating organisations on 14th November 2008 for completion by 19th December 2008. Footage for this test was donated by a variety of sources; Countryside Council for Wales, Agri-Food and Biosciences Institute and Scottish Natural Heritage.

3.1 Aims

The aims of Test 2 were three-fold:

1. To continue refining the basic elements of the Video Ring Test, i.e. to test candidates':
    - Substrate recognition skills;
    - Substrate abundance assessment skills;
    - Species ID skills;
    - Species abundance assessment skills;

2. To test the effectiveness of using still images in helping to analyse video;

3. To introduce and trial a marking scheme.

3.2 Methods

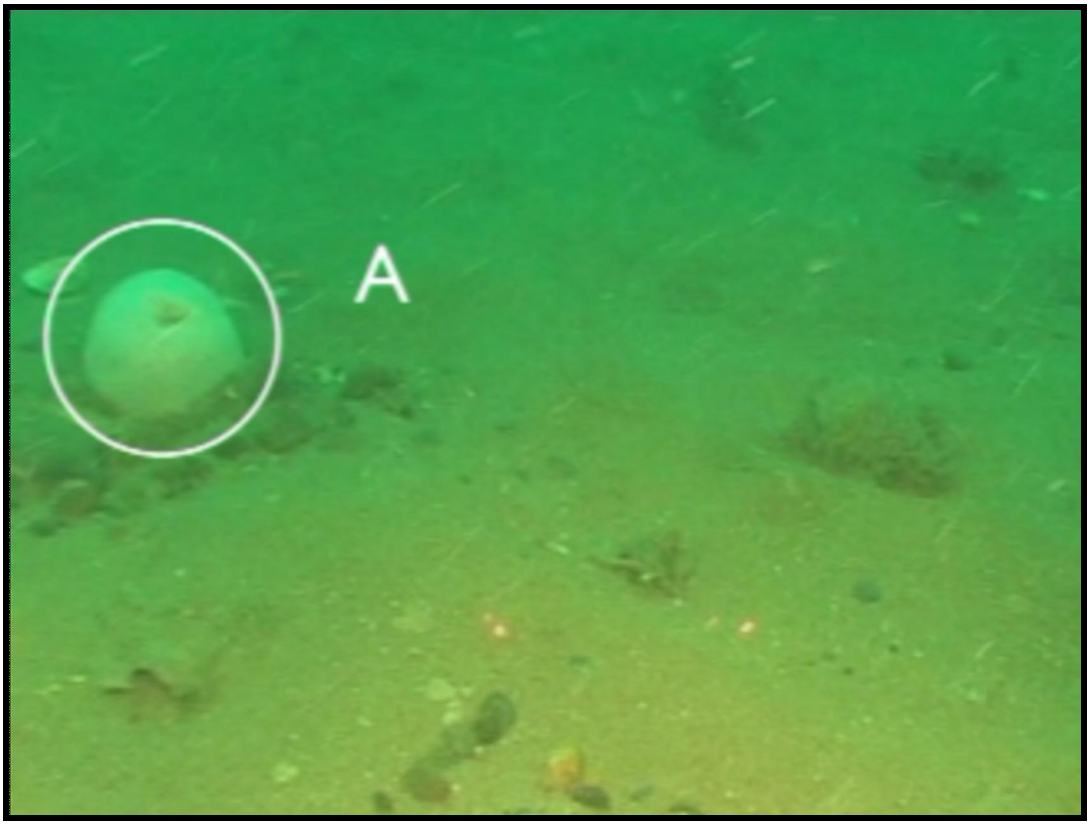
The methods for Test 2 were a revised version of those used in Test 1. The main improvement was the creation of a purpose-built on-line website for the test (see Figure 1). All of the guidance, data entry forms, questionnaires, analysis tools and reference list required for completion of the test were made available to download from this and the forms could be completed and submitted online.

Other refinements included:

- Clarifying that there should be only one submission from each participating organisation (instead of several as had been the case for Test 1);
- providing 3-minute, rather than 1-minute, long video clips;
- supplying more metadata (see Appendix 1) (only 'site' had been supplied with the video clips for Test 1);
- greatly simplifying the Data Entry forms (see Appendices 2[i versus ii] and 3[i versus ii]);
- deleting the 'Life Form' and 'Biotope' allocation exercise from the test, arguing that these are not strictly part of the process of analysing video but are more an interpretation of the data obtained as a result of video analysis. (We recommend that these should be part of another QA exercise that would need to involve a considerable training element).
- providing a series of labelled (target) organisms for identification and abundance assessment (see, *e.g.* Figure 2) instead of requiring respondents to recognise and make abundance assessment for all organisms present in the video clips, as was the case for Test 1;
- revising and simplifying the Feedback questionnaire (Appendices 4(i versus ii));
- revising some of the Analysis Tools i.e. the SACFOR scale (Appendices 5(i versus ii);
- introducing a 'Rugosity Index' as an indicator of habitat complexity to replace the substrate 'features' sections that had been omitted from the Test 2 Substrate data entry form (Appendix 6);
- introducing two new questionnaires: 'Video quality (for grading the quality of video) and 'Training requirements';
- up-dating the Reference list to include those extra references (i.e. ones that had not been provided as part of the test) that participants had used.

Figure 1. Front page of the Video Ring Test website

Figure 2. An example of how the target organisms were labelled.



The amount of work required by participants was greatly reduced in Test 2. Instead of being required to view all 10 of the video clips for analysing substrates and organisms, they were required only to view the first 4 clips (1 – 4) for substrate analysis and only the second 4 clips (5 – 8) for biological analysis. Clips 9 and 10 were used for assessing the effectiveness of using still images in the video analysis process.

3.3 Assessment

The analysis process was much more refined and straightforward than it had been for Test 1. This was partly due to the use of on-line submissions, but was also because aspects of the Test had been designed with a particular answer in mind.

This time, instead of assessing the participants' performance relative to one another, the responses given by each of the different organisations were compared with the modal response returned for each of the various tasks within the test.

3.4 Results

Although the submission date was in mid-December, several participants had not made their submissions until early January 2009. 18 Participant organisations were sent Test 2 but only 9 fully completed and 1 partially completed submissions were received.

3.4.1 Analysis of Substrates

The participants all achieved a ≥69% for substrate analysis. This included substrate recognition and abundance assessment, presence/absence of certain substrate features, and rugosity assessment. The main difficulty people had was distinguishing between some of the substrates e.g. gravel/course/fine sand.

3.4.2 Analysis of Biota

A sliding scale needed to be used for marking the taxonomic identifications. Since the 'Test' (target) organisms were 'known' by the assessors, the answers could be marked 'right' (5 marks) or 'wrong' (0 marks) or given a mark in between, depending on how close the answer was to the appropriate Taxonomic group. (See Table 1 for an illustration of the mark scheme used). However, allocation of marks to the differing responses in this section of the test needs more thought and discussion. For instance, should marks be allocated for assignment of any taxonomic level within the appropriate Phylum (e.g. Class Crinoidea in the case of the Test organisms being *Antedon bifida*) or should only the species name be accepted as being the correct answer?

Table 1. Illustration of mark scheme used for assessment of taxonomic identification section of Test 2.

		Taxonomic Assignment			
Test Organism		**Species Name**	**Genus Name**	**Higher Taxonomic Level**	**General Description**
Antedon bifida		*Antedon bifida*	*Antedon* sp.	Family: Antedonidae Class: Crinoidea	Dark pink with feathery branches
Mark allocated		**5**	**4**	**3-2**	**1**
Branching erect sponge*		*Haliclona oculata*	*Stelligera* sp.	Class: Demospongiae	Branching erect sponge
Mark allocated		**0**	**0**	**3**	**5**

* It is sometimes impossible to distinguish between species, (e.g. *Stelligera stuposa*, *Haliclona oculata* and *Raspailia hispida* on video footage) and so a broad taxonomic category or general description may sometimes need to be used in this aspect of the Ring Test.

Overall, 65% of the identifications allocated by participants were correct.

'Abundance of organism' estimates submitted by participants were assessed on the basis of comparison with the mode. As expected, the more abundant the organisms were the more variable these estimates were between respondents. For example, where there was just one *Asterias rubens* in the video clip, estimates ranged between 1 and 2 (just 1unit different) but where there were lots of *Aquipecten opercularis,* estimates ranged between 21 and 150 (129 units different). It is suggested, therefore, that there should be a sliding scale for marking this aspect of the Ring Test that would depend on the size and abundance of species present. This could take the form that large, infrequently seen but distinctive organisms have a narrow margin of error or range in which answers are acceptable and those small, cryptic or inconspicuous organisms have a wide margin of error or range in which answers are acceptable. In addition, there needs to be some means of being able to assess how people cope with organisms that definitely cannot be identified to species level on video. Marking of organism abundance estimates was complicated by the fact that some people used % cover and others used the SACFOR abundance scale for the same target organism. In order to provide consistent results the values given were translated to percentage cover based upon the SACFOR scale.

The marks for the different components of the test were given equivalent weightings (see e.g. Table 2 for the biological components of the test). However, this may need more thought and discussion, as differing weightings can alter the results significantly. (See Test 2 Report for full details)

Table 2. Summary of marks allocated for different components of the BIOTA ANALYSIS part of Test 2.

Organisation Code	Total % Mark for ID	Total % Mark or SACFOR Assessment	Total % Mark for Count & %Cover	Total % Mark for Biota Section of Test (based on mean weighting of preceding columns)
QBA	88	78	42	**69**
QBZ	83	72	33	**63**
NTOA	87	73	46	**69**
NTOB	72	88	63	**74**
NTOO	85	77	46	**69**
QAA	77	83	38	**66**
QAC	65	75	46	**62**
QAD2	52	27	13	**31**
QAE	85	47	50	**61**
QAI	72	67	58	**66**
QAL	80	67	83	**77**
QAM	93	77	46	**72**
QAN	63	70	25	**53**
Weighting	**33.3**	**33.3**	**33.3**	

3.5 Feedback

Information received from the participants about the different aspects of the test through the feedback questionnaire was extremely useful, not only for refining the test itself, but for providing details of other aspects associated with the test. For instance, they gave some indication of the types of training that was felt should be incorporated into the QA procedure and of the resources used, the metadata required and the factors required for grading video footage.

3.6 Discussion and lessons learnt

A number of lessons were learnt from Test 2 in relation to refining the test, as follows:

- In Test 2 it was possible for candidates to submit a partially completed test response and this is clearly not satisfactory. The Video Ring Test needs to have a fool-proof way of getting candidates to complete the entire test.
- There was a need to address the marking scheme. Using the mode (the 'majority' response - the value that occurs most frequently within the sample) as the yardstick is not the ideal method of assessment because it will vary according to the capabilities of the particular individuals participating in the Test. Also, it makes the precarious assumption that 'the majority is correct'. While these weaknesses had been recognised during the development of Test 2, it was decided to go ahead and use this method of analysis because it is actually very difficult to agree an absolute 'right' answer for many aspects of video footage. Video analysis carried out by several experienced members of Envision staff confirmed this to be the case. However, they agreed that the Video Ring Test would need to have a more suitable yardstick associated with it and that this would need to be discussed at the Workshop. In addition, it was recognised that other aspects of the marking scheme (e.g. weighting options for different parts of the Test, ranges of correct answers (e.g. for taxonomic identifications, pass marks) needed to be discussed further at the Workshop.
- There was a need to address how to deal with the difficulty of recognising different sediments (in particular, gravel/course/fine sand) on video.

- The 'Substrate data entry' component needed modification (*i.e.* taking out 'Other substrates' and adding 'Shell' as a substrate type).
- Methods of assessing the abundance of organisms needed to be addressed. The instruction to use either '% cover' (*e.g.* for encrusting/certain attached species) or 'counts' (*e.g.* for mobile species) had proved confusing; some participants used one method for all of the target organisms while others used both methods, thus making the analysis of the data unnecessarily complex.
- In Test 2 the organisms that were targeted for identification tended to be relatively common and distinct and the participants allocated the correct identification in the majority (65%) of cases. It may be that this aspect of Test 2 had been too easy; in the Video Ring Test it may be necessary to involve a range of increasingly more difficult specimens for candidates to identify so that their ID skills are properly tested.
- The length of time taken to complete the Test needs to be considered. The average time taken to complete Test 2 was around 9 hours.
- Training (e.g. in species and substrate recognition) is regarded as an important aspect of the whole NMBAQC Video Ring Test QA process.

3.7 The Use of Still Images

One of the aims of Test 2 was to test the effectiveness of using still images in helping to analyse video. This was carried out using 2 video clips taken from the same site. During the Test, the first of these clips was analysed twice by participants. The second clip was also analysed twice but, this time, still images (taken at the same time as the video clip) were viewed between each analysis.

The results of this exercise showed some evidence:

1. of a small increase in the number of species that were recognised by participants and
2. that participants adjusted their abundance assessments of some of the different substrate types (pebble, gravel and sand). However, the changes were not all in the same direction.

Presumably these results were because (i) the participants noticed organisms on the still images that were impossible to recognise on the video and (ii) they felt more sure about the nature of the substrate once they had viewed the still images (i.e. because these were of better resolution than the video footage).

The differences in the analyses of the video clip before and after viewing the still images were not as significant as was expected.

4 Test 3

This Test was circulated at the end of March 2009 for completion by end April 2009.

The third test considered the recommendations from Tests 1 and 2 and the majority of the test was online/electronic which enabled all results to be returned electronically.

Footage for this test was donated by a variety of sources, Countryside Council for Wales, Agri-Food and Biosciences Institute and Scottish Natural Heritage.

The test was simplified from test 2 and only incorporated substrate and biota ID with organisms identified on video footage with overlaid graphics.

4.1 Aims

The aims of Test 3 were to:

1. Improve the on-line version of the Ring Test by:
    a. refining data entry forms (see e.g. Appendices 7 and 8);
    b. providing a means of making the data entry compulsory;

c. providing a 'Resolution Test Card' to enable participants to ensure that they have an adequate resolution on the hardware/software system that they use for carrying out the Test.

2. Test the candidates':
    a. Substrate recognition skills;
    b. Substrate abundance assessment skills;
    c. Species ID skills;
    d. Species abundance assessment skills.

3. Compare methods of assessing the abundance of organisms.

4. Refine the 'Analysis Tools' (*i.e.* the Rugosity Index and the SACFOR scale) and up-date the Reference list (38 entries for Test 3 compared with 34 for Test 2 and 15 for Test 1).

5. Refine the marking/assessment scheme.

4.2 Methods

Details of the Guidance, Resources and Data Entry forms for completion of Test 3 can be viewed at:

http://www.envisionmapping.com/nmbaqc/test3.asp

4.3 Assessment

One of the main topics of discussion was the marking scheme used and the incorporation of local expertise and knowledge in that the results of users with local knowledge or data ownership should be given high value and that these answers should be considered correct. For Test 3 the owner or provider of the video footage was considered correct and participants' results were judged against these.

The scheme devised was based on an expert result and the value entered by the candidate was checked against this. Full details of the marking scheme and assement are provide in Appendix 12.

4.4 Results

4.4.1 Analysis of Substrates

Candidates showed some agreement with the expert result when examining substrate types and their coverages but this is often lower than 60% agreement and no candidate scored over 70% agreement with the 'expert' score.

Recognising modifying features of habitats appears to be consistently accurate for candidates, with the lowest score being 21 out of 30 (70%) agreement with the 'expert' result and the remaining values showed over 80% agreement.

It appears the familiarity and clarification of the Rugosity score value enables the large majority of candidates to be in agreement and scores were consistently accurate.

4.4.1.1 Issues

Several issues were encountered with both the data collected and the 'marking' of the results:

- Percentage cover is an arbitrary scale;
- Expert response is considered correct;
- Confusion of substrates i.e. coarse sand and gravels;

- Each element is equally weighted for the overall result but a breakdown does show specific results for each element of the test.

4.4.2 Analysis of Biota

Candidates appear to be able to identify organisms to a suitable taxonomic level from video. Percentage cover and counts of organisms showed some agreement. Where organisms had high numbers the count was highly variable, and where it was more appropriate to estimate percentage cover rather than counts, counts were variable and often inaccurate.

The overall % marks (using equally weighted components) ranged between 56% and 81%. On the basis of a 50% pass mark, no candidate would have 'failed' the Test.

Candidates are able to get an idea of their performance relative to others from the results; they will also be able to discover where their relative strengths and weaknesses are in relation to the different aspects of the analysis of benthic video biota. It is important to supply marks for the different components of the Test so that candidates can identify any parts of the video analysis process which might require remedial training.

4.4.2.1 Issues

- The issues raised in previous tests regarding estimates of abundance either in terms of percentage covers or count was attempted by candidates submitting values for both, and the results show that depending upon on the form of the organism being identified both measures of abundance are appropriate for different organisms. Future tests should attempt to clarify which estimate of abundance is most appropriate for each organism.
- 'Expert' type answers were used for this assessment and this does seem to give sensible results but the validity of the 'expert' should be investigated further and candidates' feedback should be sought on how they performed against the 'expert' results.
- The marking scheme was discussed during the final workshop but there is still ambiguity around the weighting scheme, and whether the marking schemes employed are suitable and acceptable. This should be considered and agreed by the NMBAQC committee prior to any further tests.

5 Issues outstanding (at the time of the Workshop)

The main issues which remained to be resolved at this stage in the development process were categorised under the following four headings and these were presented as Workshop Discussion Topics at the NMBAQC Video Ring Test Workshop which took place in May 2009:

- How much detail needs to be analysed from video?
- What quality of video (and associated metadata) is required?
- How should the Ring Test be marked?
- What training is required before the Test can be attempted?

The workshop also aimed to address four aspects of the Video Ring Test scheme in the context of two themes:

THEME 1: Fisheries and Monitoring Issues and Objectives
THEME 2: Mapping and Biodiversity Monitoring Issues and Objectives

- What Tools and Resources are required?
- What is Best Practice for video analysis?
- How should the Ring Test be assessed?
- What are the Training and Testing requirements?

Breakout groups discussed and debated these issues and reported back to the workshop with a view to reaching a consensus opinion.

A final session reviewed all the issues discussed and summarised the workshop findings and extracts below:

5.1.1 Workshop Summary

There should be 'stages of development' of the whole process, for example:

i. Review existing video analysis procedures;
ii. Produce a 'Guide to Video Analysis' manual;
iii. Train potential video analysis contractors;
iv. Create and carry out a simple Ring Test (which would be only part of the whole QA process);
v. Hold a workshop to review manual/training/testing procedures;
vi. Hold regular workshops to review.

It was felt that it is important to keep the momentum of the project going and that the cohesion of the group attending the workshop should be kept by involving everyone at the next stage.

5.1.2 Workshop Action Points

It was agreed that the participant organisations present would:

Provide a 3-minute video clip appropriate for video analysis together with:

i. A task for analysis (e.g. counting of *Nephrops* burrows; biotope allocation);
ii. A method for analysis;
iii. A set of 'correct' answers.

The outcome of the Workshop discussions is provided in the NMBAQC Video Ring Test Workshop Proceedings (Sotheran, 2009). Conclusions and outcomes from the workshop are incorporated into the recommendations below.

6 Recommendations for the Future Direction of the NMBAQC Video Ring Test

As a result of the trial Video Ring Test procedure, recommendations for the future development of the test are as follows:

6.1 Issues identified by the development of the test and workshop

The development of the test identified several issues, some of which were addressed and incorporated as the test developed and some are carried forward, the points raised and addressed by each test are summarised in the sections above and the main points carried forwards are detailed below:

Video is used for a wide range of purposes and functions, thirty-three were identified in the workshop and these involve a wide range of analytical procedures and techniques for the review and extraction of data from video footage. It may be possible to produce a best practice document for video review in which all video analysis disciplines must be considered and incorporated.

Marking the scheme proved to have its problems and whilst marking schemes were trialled and tested, one of the outcomes of the workshop was that the correct answers should be expert led.

The identification of substrate and estimation of percentage covers was found to be difficult and noted by some candidates and it was felt more resources to enable consistency would be useful.

There is a wide range of knowledge and experience of video analysis spread across a range of disciplines and feedback from the test and workshop identified a requirement for the development of training materials and resources.

Technology is advancing quickly in the field of video and its application to underwater video, and it is important that any quality assurance methodologies or procedures have suitable review periods and are phase incorporated to enable any advances to be assessed.

6.2 Preparation

- There should be a review of existing video analysis procedures;
- A 'Guide to Video Analysis' Best Practice Manual should be produced if no existing document or SOP is found to be suitable;
- There should be regular workshops to review the manual and training and testing procedures;
- The Video Ring Test process should be supported by centralised management structure for dealing with materials and information relating to video analysis techniques;

6.3 Training

- Training materials should be made available to potential Video analysis contractors to enable them to work to the Best Practice guidelines. The guidelines would address a number of problem areas identified from the ring tests including;
    - A complete list of available resources should be produced and made available via the NMBAQC website.
    - 'In situ' species identification;
    - Substrate/Habitat recognition;
    - Enumeration techniques (i.e. counting, assessing SACFOR and % cover) - may need to reinvent abundance assessment techniques specifically for video analysis;
    - New technologies (to keep abreast of developments that improve quality of benthic video and its analysis).
    - If 'Biotope', 'Life Form' or similar habitat recognition from video is to form part of a ring test then, specific documentation and training materials should be developed and made available.

    Where persistent problem areas are identified through further ring tests or other means, targeted workshops should be held to assist video analysis contractors and future ring tests.

6.4 The Test

- There should be two types of Video Ring Test:
    - A general (standard) test
    - A specialist (purpose-driven test);
- The test should be carried out on-line;
- 'Life Form' and 'Biotope' allocation should not form part the test. (We argue that these are not strictly part of the process of analysing video but are more an interpretation of the data

obtained as a result of video analysis.) We recommend that these should be part of a separate QA exercise that would need to involve a considerable training element.

- For taxonomic identification, annotated video footage should be incorporated to highlight the target species to ensure consistent identification.
- Video clips for analysis should be no less than 3 minutes long;
- Video clips for analysis should be accompanied by appropriate metadata; the appropriateness should be judge against the purpose of the test, if the metadata provided would give an advantage or elucidate to the correct answer then this would be inappropriate.
- Video footage should be of suitable or typical quality and bias should not be towards 'high' quality footage as this would skew the test towards high quality footage and not footage of a quality likely to be encountered on a contract work basis.
- Where feasible a scale bar or indication of scale should be provide on the video footage to assist the candidates.

6.5 Testing and Assessment

- Testing should be carried out twice yearly: early spring and late autumn on fixed dates to accommodate workloads around survey periods and to ensure personnel changes with an organisation are accounted for and are current. Tests that are more frequent were considered a over-burden on workloads and less frequent tests would not accommodate personnel changes.
- Feedback should be given as soon as possible after the test, on fixed dates;
- Feedback should include details of candidates' strengths and weaknesses so that remedial action can be taken where necessary;
- The test should be marked by NMBAQC appointed assessors;
- The 'yardsticks' (for each component of the test) against which to assess the performance of candidates need to be set for each video clip used for the test BEFORE the test is carried out;
- Advice from local (i.e. local to the site from which the video was taken) experts should be sought in setting yardsticks;
- Consideration is required to determine the level at which Pass/Fail marks are given. From the development of the test substrate marks over 70% was achieved by more than 90% of participants in test 2 and 80% in test 3, for biota in test 2 85% of candidates scored over 60% and in test 3 70% of candidates scored over 60%. These levels could be used as a starting point for future tests and reviewed at regular intervals.

6.6 'Standard Operating Procedures' for Video Analysis

- In order for there to be consistency between results from various organisations the type of hardware to be used for video analysis should follow standard recommendations and should be consistent as possible amongst participating organisations;
- Only those people who have been on appropriate suitable training sessions or have other appropriate training should be eligible as test candidates;
- Video clips to be analysed should be graded for quality;
- Video clips should have a minimum set of metadata (MESH metadata standards);
- Videos should include an indication of scale;
- The specific purpose of the analysis should be clearly defined;

- Associated still images should be used where practicable;
- Analysis procedures (e.g. reviewing of video clips, data entry, enumeration) should follow the proposed 'Guidance to Video Analysis' best practice manual (see 'Preparation' section above);
- Standard data entry forms (as on on-line test) should be used;
- There should be 'in-house' quality checks as a minimum, such as review by second analyst, but preferably with independent reviews.

7 References

Foster-Smith, J., Benson, A. and Sotheran, I.S.. (2008) *Development of an NMBAQC Scheme Video and Photographic Ring Test: Interim Report - Results of Test 1*. An **Envision Mapping Ltd.** report for the NMBAQC Committee, c/o The Agri-Food and Biosciences Institute (AFBI), Belfast.

Sotheran, I.S. (2009) *NMBAQC Video Ring Test: Workshop Proceedings*. An **Envision Mapping Ltd.** report for the NMBAQC Committee, c/o The Agri-Food and Biosciences Institute (AFBI), Belfast.

Sotheran, I.S. (2009.) *Development of an NMBAQC Scheme Video and Photographic Ring Test: Interim Report - Results of Test 3*. An **Envision Mapping Ltd.** report for the NMBAQC Committee, c/o The Agri-Food and Biosciences Institute (AFBI), Belfast. (Attached as Appendix 9).

Sotheran, I.S., Foster-Smith, J. and Benson, A. (2009) *Development of an NMBAQC Scheme Video and Photographic Ring Test: Interim Report - Results of Test 2*. An **Envision Mapping Ltd.** report for the NMBAQC Committee, c/o The Agri-Food and Biosciences Institute (AFBI), Belfast.

8 Appendices

Appendix 1. Metadata supplied with the Test 2 video clips.

Video Clip Number	1	2	3	4	5	6	7	8	9	10
Test 2 Exercise	Substrate Analysis	Substrate Analysis	Substrate Analysis	Substrate Analysis	Biological Analysis	Biological Analysis	Biological Analysis	Biological Analysis	Use of Stills	Use of Stills
Location	Eastern English Channel	Eastern English Channel	Eastern English Channel	Stanton Banks – South Outer Hebrides	Eastern English Channel	East Antrim - North Channel	Eastern English Channel	Off Weymouth - Western English Channel	Eastern English Channel	Eastern English Channel
Owner	CEFAS	CEFAS	CEFAS	AFBI	CEFAS	AFBI	CEFAS	ENVISION	CEFAS	CEFAS
Year of Survey	2006	2006	2006	2006	2006	2006	2006	2007	2006	2006
Depth	53m	45m	44m	160m	52m	22m	52m	23m	61m	61m

Appendix 2. Simplification of the Data Entry forms

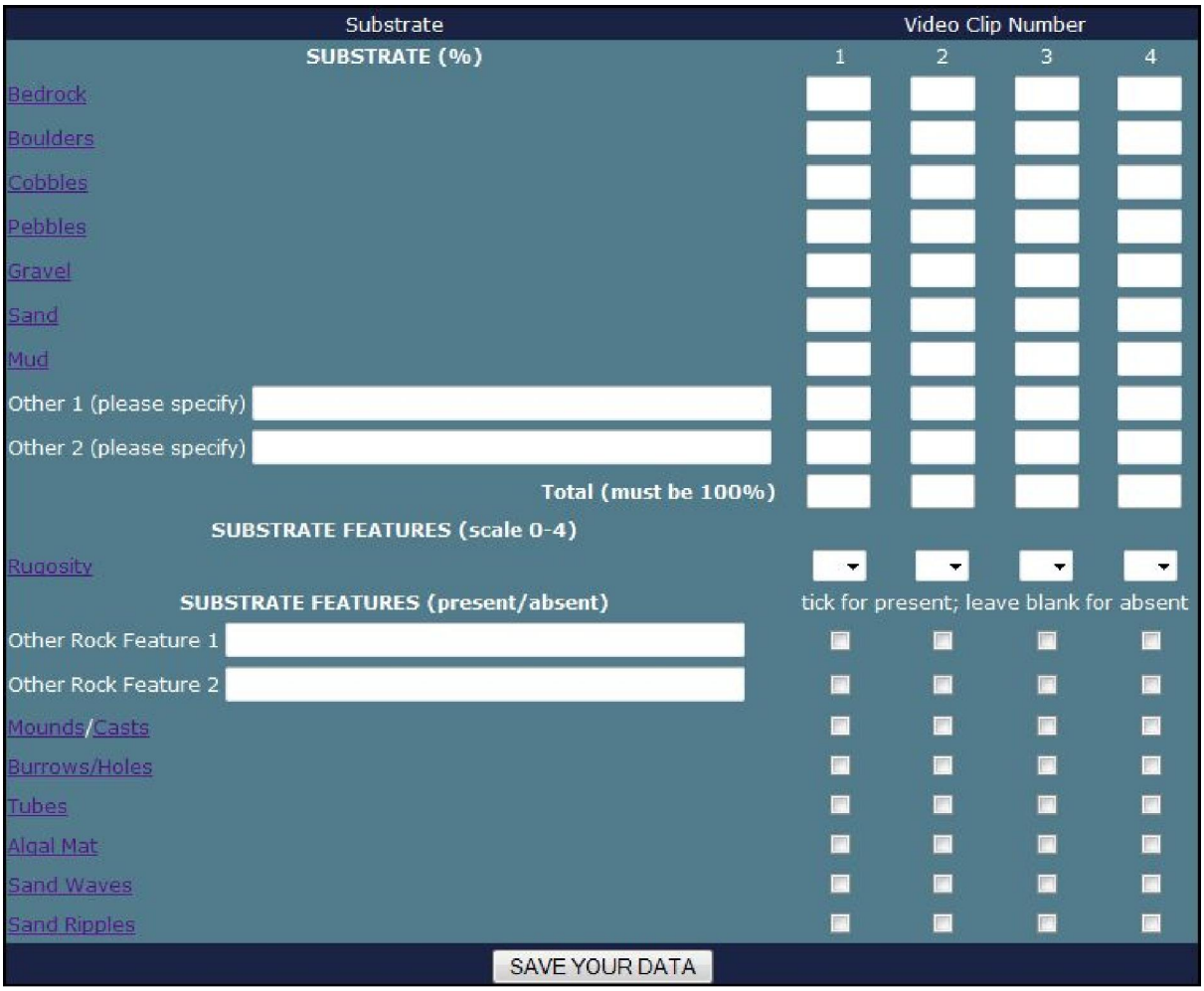
Appendix 2i) Test 1 Substrate Data Entry form

NAME OF ORGANISATION										
NAME OF PARTICIPANT										
DATE TEST CARRIED OUT										
VIDEO CLIP	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***
1. SURVEY DETAILS										
AREA OF SURVEY	***Treshnish Isles***	***Isle of Wight***	***Menai Strait***	***Menai Strait***	***Poole Bay***	***Sullom Voe***	***Sullom Voe***	***North Norfolk Coast***	***North Norfolk Coast***	***North Norfolk Coast***
YEAR OF SURVEY	***2003***	***1994***	***1994***	***1994***	***2007***	***2004***	***2004***	***2007***	***2007***	***2007***
PURPOSE OF SURVEY	***Habitat Mapping***	***Habitat Mapping***	***Habitat Mapping***	***Habitat Mapping***	***Habitat Mapping***	***Habitat Mapping***	***Habitat Mapping***	***Habitat Mapping***	***Habitat Mapping***	***Habitat Mapping***
2. HABITAT DETAILS										
Biological Zone (see MarLIN Website Glossary for definitions)										
Upper Infralittoral										
Lower Infralittoral										
Upper Circalittoral										
Lower Circalittoral										

Substrate (%) (Total must add up to 100%)										
Bedrock - Horizontal (0-40°)										
Bedrock - Steep face (40-80°)										
Bedrock - Vertical face (80-100°)										
Bedrock - Underhangs										
Boulders (over 1024mm)										
Boulders (512 to 1024mm)										
Boulders (256 to 512mm)										
Cobbles (64 to 256mm)										
Pebbles (16 to 64mm)										
Shells – empty										
Shells - live e.g. Mytilus										
Stony gravel (4 to 16mm)										
Shelly gravel (4 to 16mm)										
Dead Maerl										
Live Maerl										
Coarse sand (1 to 4mm)										
Medium sand (0.25 to 1mm										
Fine sand (0.063 to 0.25mm)										
Mud (less than 0.063mm)										
Biogenic Reef										
Metal										
Concrete										
Wood										
Wreckage										
Trees_branches										
Algae										
Peat										
Rock Features (scale of 1-5)										
Surface Relief (Even - Rugged)										
Texture (Smooth - Pitted)										
Boulder/Cobble/Pebble Shape (Rounded to Angular)										
Rock Features (P/A)										
Fissures										
Gully										
Cave										
Tunnel										
Boulder/Cobble on Rock										
Boulder/Cobble on Sediment										

Boulder Holes										
Scour										
Sediment on Rock										
Sediment Features (scale of 1-5)										
Surface Relief (Even - Uneven)										
Sediment Features (P/A)										
Mounds/Casts										
Burrows/Holes										
Tubes										
Algal Mat										
Waves/Dunes										
Ripples										
3. SPECIES LIST AND ABUNDANCE										
See 'Species Data' Sheet										
4. LIFE FORM										
Life Form Name (See 'Life Form List' Sheet)										
Life Form Abundance										
5. BIOTOPE CLASSIFICATION (JNCC [Connor et al.] V. 04.05)										
Biotope Name 1										
Biotope Code 1										
Confidence Level in allocation of Biotope 1 (scale 1-5: low to high)										
Biotope Name 2										
Biotope Code 2										
Confidence Level in allocation of Biotope 2 (scale 1-5: low to high)										
6. QUALITY OF VIDEO										
Overall visual quality of video (poor / moderate / good)										
Speed of Footage (poor / moderate / good)										
Movement of Camera (poor / moderate / good)										
Lighting (poor / moderate / good)										
Visibility (poor / moderate / good)										
Resolution/Definition (poor / moderate / good)										
Completeness of video sample (complete/adequate/incomplete)										

Appendix 2ii) Test 2 Substrate Data Entry form



Appendix 3. Revision of the Biological Data Entry form (Test 1 versus Test 2)

Appendix 3i) Test 1 Biological Data Entry form

NB - Complete at Species level as far as possible.
If not possible then use a Generic Taxonomic description.

Video Clip Code	Species Name	Species Code	Species Abundance (SACFOR)	Characteristic Species (Y/N)	Confidence Level in identification (scale 1-5: low-high)	If Confidence less than 3, please state reason	Generic Taxonomic Description (e.g. 'Sponge'/ 'Bryozoan')	Generic Taxonomic Abundance (SACFOR)

Appendix 3ii) Test 2 Biological Data Entry form

Biological Data

VIDEO	SPECIMEN	Taxonomic Assignment: Species Name (e.g. *Balanus crenatus*)	Genus Name (e.g. *Balanus* sp.)	Higher Taxonomic Level	General Description	Abundance Assessment: Estimated Abundance (SACFOR)	Count	Estimated % Cover
5	A							
5	B							
5	C							
6	D							
6	E							
6	F							
7	G							
7	H							
7	I							
8	J							
8	K							
8	L							

SAVE YOUR DATA

Appendix 4. Revision of the Feedback questionnaire (Test 1 versus Test 2)

Appendix 4i) Test 1 Feedback questionnaire

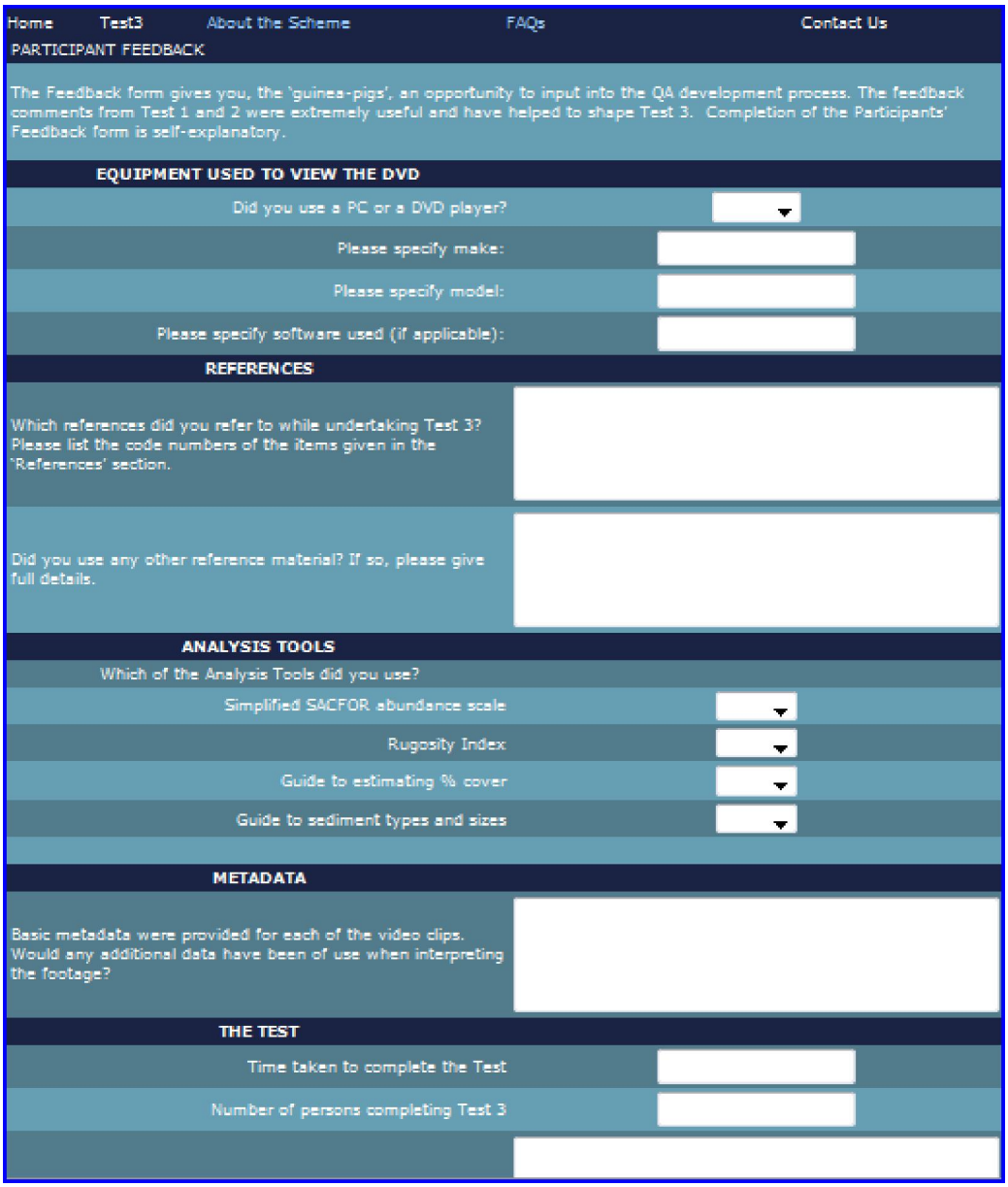
TIME TAKEN TO COMPLETE THE TEST	*Do not account for time spent completing this form, or User Registration form*
Clip 1	minutes
Clip 2	minutes
Clip 3	minutes
Clip 4	minutes
Clip 5	minutes
Clip 6	minutes
Clip 7	minutes
Clip 8	minutes
Clip 9	minutes
Clip 10	minutes
Total	minutes

EQUIPMENT USED TO VIEW THE DVD	
DVD PLAYER *(indicate whether PC or DVD player was used)*	DVD PLAYER / COMPUTER
Please specify make:	
Please specify model:	
Please software used if applicable:	
SCREEN	
Please specify make:	
Please specify model:	
Please specify screen resolution:	

TOOLS/RESOURCES USED					
Did you use species identification guides or online resources?	YES/NO				
Please specify which:					
Did you use sediment identification guides?	YES/NO				
Please specify which:					
	1 -5 (1-Not At All 5-Very)				
Would an independent stills image be useful?	1	2	3	4	5
Would grab sample associate with the clip be useful?	1	2	3	4	5
Background references used, if any?					

TEST FEEDBACK					
	1 -5 (1-Not At All 5-Very)				
How onerous did you find the test?	1	2	3	4	5
How useful was the resource CD?	1	2	3	4	5
Would you fill the forms in online?	YES/NO				
If no please give a reason:					
Do you use or are required to use Marine Recorder?	YES/NO				
Do you feel anything was omitted from the test form?	YES/NO				
If yes, what?					
Do you feel anything was unnecessary on the test form?	YES/NO				
If yes, what?					

Appendix 4ii) Test 2 Feedback questionnaire



Home Test3 About the Scheme FAQs Contact Us

PARTICIPANT FEEDBACK

The Feedback form gives you, the 'guinea-pigs', an opportunity to input into the QA development process. The feedback comments from Test 1 and 2 were extremely useful and have helped to shape Test 3. Completion of the Participants' Feedback form is self-explanatory.

EQUIPMENT USED TO VIEW THE DVD

Did you use a PC or a DVD player?	
Please specify make:	
Please specify model:	
Please specify software used (if applicable):	

REFERENCES

Which references did you refer to while undertaking Test 3? Please list the code numbers of the items given in the 'References' section.	
Did you use any other reference material? If so, please give full details.	

ANALYSIS TOOLS

Which of the Analysis Tools did you use?

Simplified SACFOR abundance scale	
Rugosity Index	
Guide to estimating % cover	
Guide to sediment types and sizes	

METADATA

Basic metadata were provided for each of the video clips. Would any additional data have been of use when interpreting the footage?	

THE TEST

Time taken to complete the Test	
Number of persons completing Test 3	

Appendix 5. Revision of the SACFOR Scale (Test 1 versus Test 2)

Appendix 5i) Test 1 'JNCC' SACFOR Scale

SACFOR abundance scale used for both littoral and sublittoral taxa from 1990 onwards

NB. Read notes below prior to use of scale

Growth form	Size of individuals/colonies							
% cover	Crust/meadow	Massive/Turf	<1cm	1-3 cm	3-15 cm	>15 cm	Density	
>80%	S		S				>1/0.001 m2 (1x1 cm)	>10,000 / m2
40-79%	A	S	A	S			1-9/0.001 m2	1000-9999 / m2
20-39%	C	A	C	A	S		1-9 / 0.01 m2 (10 x 10 cm)	100-999 / m2
10-19%	F	C	F	C	A	S	1-9 / 0.1 m2	10-99 / m2
5-9%	O	F	O	F	C	A	1-9 / m2	
1-5% or density	R	O	R	O	F	C	1-9 / 10m2 (3.16 x 3.16 m)	
<1% or density		R		R	O	F	1-9 / 100 m2 (10 x 10 m)	
					R	O	1-9 / 1000 m2 (31.6 x 31.6 m)	
						R	<1/1000 m2	

Use of the MNCR SACFOR abundance scales

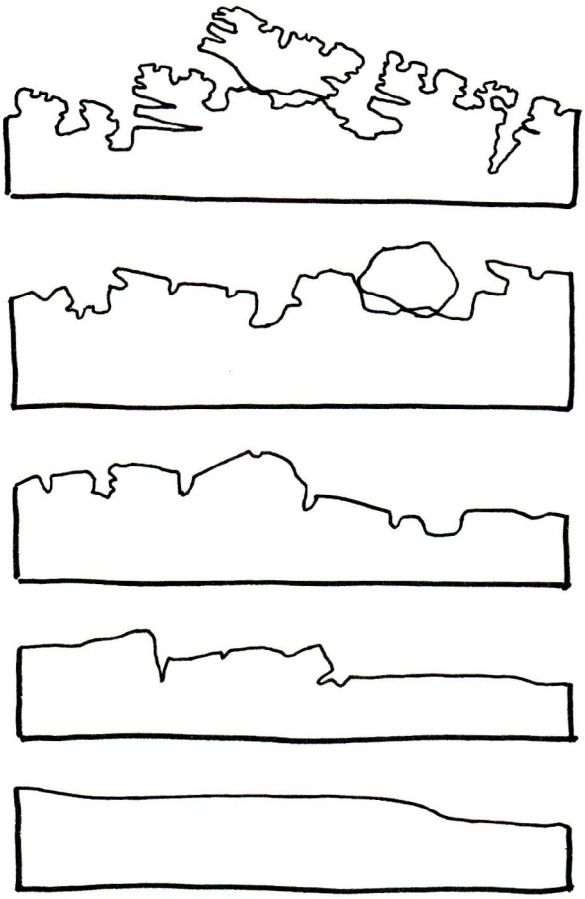
The MNCR cover/density scales adopted from 1990 provide a unified system for recording the abundance of marine benthic flora and fauna in biological surveys. The following notes should be read before their use:

1. Whenever an attached species covers the substratum and percentage cover can be estimated, that scale should be used in preference to the density scale.
2. Use the massive/turf percentage cover scale for all species, excepting those given under crust/meadow.
3. Where two or more layers exist, for instance foliose algae overgrowing crustose algae, total percentage cover can be over 100% and abundance grade will reflect this.
4. Percentage cover of littoral species, particularly the fucoid algae, must be estimated when the tide is out.
5. Use quadrats as reference frames for counting, particularly when density is borderline between two of the scale.
6. Some extrapolation of the scales may be necessary to estimate abundance for restricted habitats such as rockpools.
7. The species (as listed above) take precedence over their actual size in deciding which scale to use.
8. When species (such as those associated with algae, hydroid and bryozoan turf or on rocks and shells) are incidentally collected (i.e. collected with other species that were superficially collected for identification) and no meaningful abundance can be assigned to them, they should be noted as present (P).

Appendix 5ii) Test 2 'Seasearch' SACFOR Scale

Abundance	***Encrusting and turf species*** ***e.g. encrusting algae/sponge, jewel anemones, hydroids, barnacles, mussels, seaweeds***	***Small plants and animals (1-5cm)*** ***e.g. worms, small sponges, anemones, cup-corals, shells, solitary sea squirts***	***Large plants and animals (> 5cm)*** ***e.g large sponges, sea fans and pens, large anemones, crabs and lobsters, starfish, fish,***
Superabundant	*80-100% cover*	*10,000 per m²*	*100 per m²*
Abundant	*40-80% cover*	*1000 per m²*	*10 per m²*
Common	*20-40% cover*	*100 per m²*	*1 per m²*
Frequent	*10-20% cover*	*10 per m²*	*1 per 10m²*
Occasional	*5-10% cover*	*1 per m²*	*1 per 100m²*
Rare	*< 5% cover*	*< 1 per m²*	*<1per 1000m²*

Appendix 6. Diagram illustrating the 'Rugosity Index' introduced in Test 2



RUGOSITY INDEX

Level 4 (Extreme rugosity)

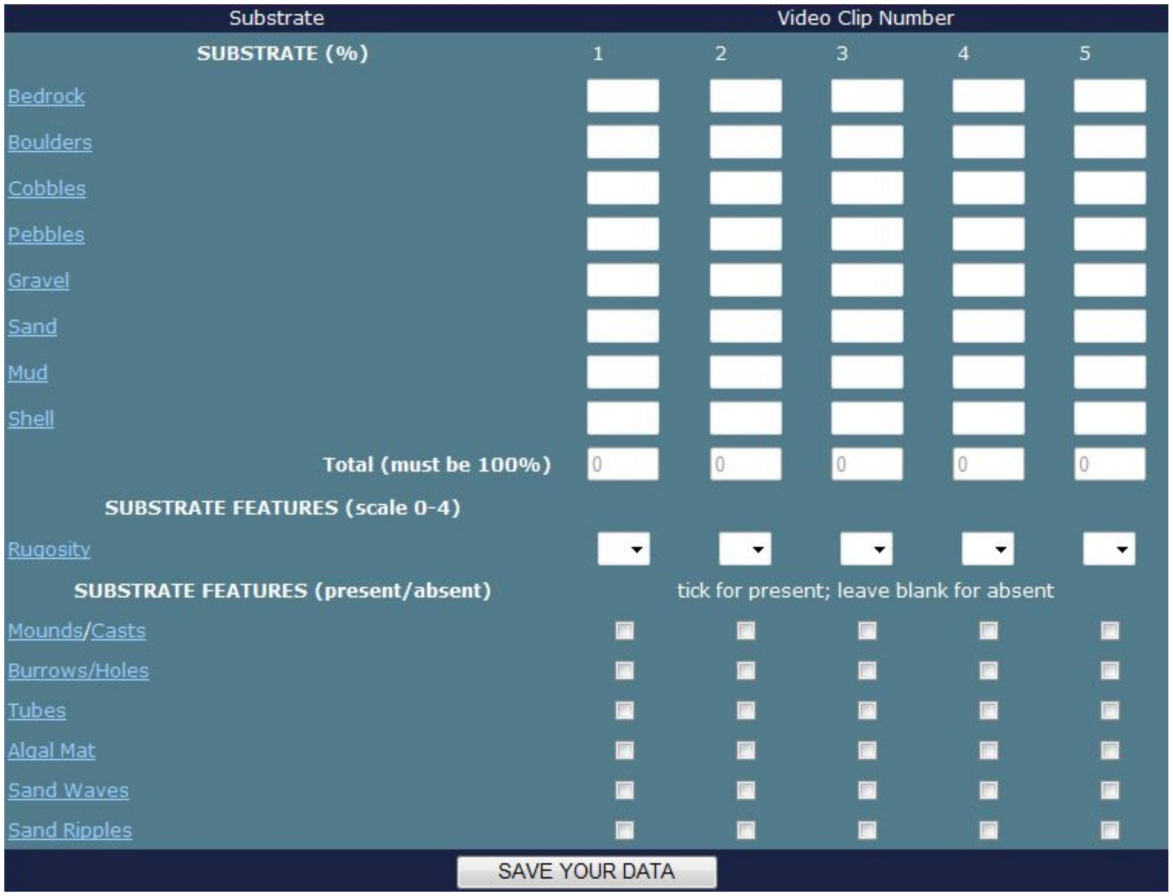
Level 3 (High rugosity)

Level 2 (Moderate rugosity)

Level 1 (Low rugosity)

Level 0 (No rugosity)

Appendix 7. Test 3 Substrate Data Entry form (cf. Appendix 2ii)



Appendix 8. Test 3 Biological Data Entry form (cf. Appendix 3ii)

Biological Data

VIDEO	SPECIMEN	Taxonomic Assignment: Level of Identification	Taxonomic Assignment: Organism Name	Abundance Assessment (Complete all three if possible): Estimated Abundance (SACFOR)	Abundance Assessment: Count	Abundance Assessment: Estimated % Cover
6	A					
7	B					
7	C					
7	D					
8	E					
8	F					
9	G					
10	H					
10	I					
11	J					
11	K					
12	L					
12	M					
13	N					
13	O					

SAVE YOUR DATA

Appendix 9.

Sotheran, I.S. (2009.) *Development of an NMBAQC Scheme Video and Photographic Ring Test: Interim Report - Results of Test 3*. An **Envision Mapping Ltd.** report for the NMBAQC Committee, c/o The Agri-Food and Biosciences Institute (AFBI), Belfast.

Appendix 10. Report and results of test 1

Appendix 11. Report and results of test 2

Appendix 12. Report and results of test 3

Appendix 13. Workshop Proceedings & Presentations