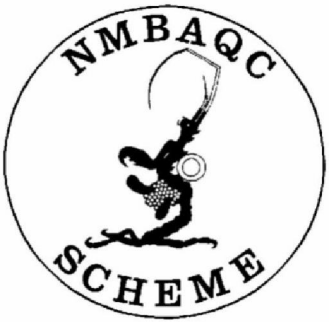
ENVISION

envision.uk.com

NATIONAL MARINE BIOLOGICAL ANALYTICAL QUALITY CONTROL SCHEME



NMBAQC VIDEO RING TEST INTERIM REPORT: RESULTS OF TEST 2

Envision Mapping Ltd.

Newcastle upon Tyne

UK

March 2009

Prepared for	**AFBI-NMBAQC**
Title	NMBAQC Video Ring Test Interim Report: Results of Test 2
Abstract	Interim Report for development of NMBQAC video ring test 2
Contract reference:	2007-1017-AFBI-Develoment of a Video and Photographic Ring Test
Report identification:	2007-1017-ISS-NMBAQC-Ring Test 2

Document control

Rev.	Originator	Date	Status	Checked & Approved
0	Ian Sotheran	26/02/09	Internal	
1	Ian Sotheran	09/03/09	First Draft	JE
2	Ian Sotheran	11/03/09	Final Draft	J F-S

File index: \\Exchange\Projects\2007-1017-NMBAQC-Ring Test\TEST 2\Report\Final Test 2 Report\NMBAQC Video Ring Test 2 Report V4 (31st March 2009) With Recommendations.Docx

Report:

Prepared by Ian Sotheran, Alison Benson and Judy Foster-Smith

Envision Mapping Ltd.

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

NOTES:

Contents

1 Introduction

The second test for the development of a ring test for NMBAQC Video analysis has taken place and the results have been collected, collated and marked.

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

2 Aims of Test 2

In addition to the general aims of developing the NMBAQC video ring test several issues were aimed to be addressed with this test.

- Use of still images alongside video analysis
- A consistent marking scheme
- Test candidates' species identification skills/ability
- Test candidates' physical habitat identification skills/ability
- Test candidates' ability to estimate percentage cover/abundance

3 Completion of Test

The second test was designed to be completed and submitted online via a purpose-built website which contained online and downloadable resources. The DVD containing video footage was circulated by post.

18 participants/organisations were sent Test 2 but only 9 fully completed responses and 1 partially completed test were received. Several of these tests requested extensions to the deadlines which were approved.

Five Envision staff completed parts of the Test for comparison and to provide additional data for marking and analysis.

This interim report will be edited and circulated to participants along with their participant ID so that they can assess their own performance in the Test. No participant IDs other than that which belongs to the candidate will be circulated to them.

4 Results

4.1 Video Analysis of Substrates

4.1.1 Marking Scheme

The scheme devised was based upon the value entered by the candidate which was checked against a range of values calculated from all the results returned.

For each substrate type the mode of the returned results was calculated along with the percentile values at 30% and 70% to give a 'correct' range of values. If a candidate's value was within the acceptable range a score of 1 was given and a 0 score was given if the value was outside the acceptable range. This was carried out for each of the 7 substrate types and for each of the video clips used for this part of the Test, which gave a maximum possible score of 28. The percentage of correct answers was then calculated to give each candidate a score (Table 1).

A similar approach was taken for the marking of substrate features. A decision as to whether a feature was present or not was based on the majority response submitted by the candidates. Each candidate's response was then compared to this majority decision. If it matched a score of 1 was given and, where no match occurred, a 0 mark was given. Again, the percentage of correct answers for this part of the Test was calculated (Table 1).

An overall percentage mark was then calculated for the candidates' assessment of substrate abundance and the presence of substrate features combined (Table 1).

The 'Rugosity index' results were also marked on the basis of a comparison with a majority response: if a candidate's response was ±1 index point different from the mode value then a score of 1 was given and, if the score was outside these values, a 0 mark was given (Table 1).

4.1.2 Substrate Test Results

Table 1: Marks allocated for the identification of the physical substrates

ID	SUBSTRATE SCORE	SUBSTRATE PERCENTAGE	FEATURES SCORE	FEATURES PERCENTAGE	**OVERALL PERCENTAGE**	RUGOSITY SCORE
QBA	19/28	68%	22/24	91.67%	**78.85%**	4/4
QBZ	15/28	54%	21/24	87.50%	**69.23%**	0/4
NTOA	24/28	86%	23/24	95.83%	**90.38%**	4/4
NTOB	19/28	68%	22/24	91.67%	**78.85%**	4/4
NTOO	22/28	69%	23/24	95.83%	**86.54%**	4/4
QAA	19/28	68%	24/24	100.00%	**82.69%**	4/4
QAC	23/28	82.14%	23/24	95.83%	**88.46%**	4/4
QAD2	19/28	67.86%	22/24	91.67%	**78.85%**	0/4
QAE	18/28	64.29%	20/24	83.33%	**73.08%**	4/4
QAI	19/28	67.86%	19/24	79.17%	**73.08%**	4/4
QAK	19/28	67.86%	22/24	91.67%	**78.85%**	0/4
QAL	23/28	82.14%	24/24	100.00%	**90.38%**	4/4
QAM	22/28	78.57%	24/24	100.00%	**88.46%**	4/4
QAN	20/28	71.43%	21/24	87.50%	**78.85%**	4/4

4.1.3 Findings

The results returned by candidates for percentage cover of substrate types varied widely with confusion between pebbles, cobbles and gravel substrates. Despite this, the majority of candidates fall within a ±20% bracket around the mode of the results submitted.

Recognising substrate features proved to be more consistent between candidates, with a high percentage match amongst results. The lowest was 79% (19/24 matches) and three candidates matched 100% with the majority responses.

Rugosity levels were again estimated consistently between candidates with some candidates scoring zero due to incomplete submissions rather than incorrect answers.

4.1.4 Issues

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

- Incomplete forms & data quality assurance;
- Percentage range value is arbitrary;
- Pass mark is required;
- Difficult to assign correct or incorrect answer;
- Confusion of substrates i.e. coarse sand and gravels;
- 'Other features' are ambiguous, inconsistent and does not allow for consistent marking.

4.1.5 Recommendations

- Changes to the website for 'other' classes to be altered per take and possibly a drop down box to ensure consistency;
- Agree percentage range;
- Agree pass mark;
- Discuss variation in results and difficulty in correctly assigning substrate estimates.

4.2 Video Analysis of Biota

4.2.1 Marking Scheme

The marking scheme for the analysis of Biota part of the Benthic Video Ring Test is still in its infancy and the marks allocated for Test 2 are provisional. The marks are not to be taken as a true reflection of the performance of the candidate. Instead, they should be regarded as a product of a trial process that is likely to require substantial modification before it provides a reliable indication of a candidates ability to analyse benthic video.

Several aspects of the marking scheme used here require considerable discussion and it is hoped that these can be addressed at the NMBAQC video Ring Test Workshop in May.

4.2.1.1 Marking scheme - Identification

This part of the scheme was to test candidates' abilities to accurately assign the organisms that they observe in benthic video to a particular taxon. The organisms labelled in the video clips for Test 2 (i.e. the 'Test organisms') were of known, pre-determined taxa. It was, therefore relatively straight forward to recognise the 'spot on' correct answers. However, there was a variety of other types of answer given, which, it was felt, justified a mark. For instance, the first labelled organism (in Video Clip 5) was *Asterias rubens* and one of the respondents assigned this to 'Class Asteroidea'. Clearly this is not an incorrect answer and it was felt that marks for this sort of response should be built into the marking scheme. Similarly, if it is clearly impossible to identify an organism which appears in the video to the species level (either because of the video resolution or because it is only possible to identify certain organisms, for example, under the microscope (e.g. some of the hydroids or encrusting bryozoans) then the correct answer would be in a higher taxonomic category. In other

words, in certain cases, being less definite about an identification is, in fact, more correct. This also needs to be built into the marking scheme. Marks were therefore allocated as follows:

- 5 marks for the correct taxon allocation
- 3 – 4 marks for a close taxon allocation, the mark depending on how close the taxon assigned by the candidate (e.g. the right genus or family) was to the correct one.
- 1 – 2 marks if a species identification was incorrect but if the species given was in a taxonomic group close to the correct one (e.g. the right Class or nearer), the mark depending on how close.

The marks for this part of the Test are presented in Table 2

4.2.1.2 Marking Scheme – Estimated SACFOR Abundances

The estimated abundances of the Test organisms could be assigned to one of six abundance categories: S, A, C, F, O, or R. The abundance assessments were marked on the basis of the mode which, in turn, was derived from the whole Test 2 response sample. For example, the last test organism (No. 12) was thought to be 'Abundant' (A) by 2 candidates, 'Common' (C) by 4 candidates, 'Frequent' (F) by 3 candidates and 'Occasional' by 2 candidates. Thus, in this case, 'Common' was taken as the modal response. Therefore, each of the candidates who assigned the abundance of this organism as 'Common' were given the top mark (5); those who assigned it to one category either side of this (i.e. Abundant or Frequent) were given 3 marks. Otherwise a '0' mark was allocated (see Table 3).

4.2.1.3 Marking Scheme – Counts and Percentage Cover

It has proved difficult to come up with an entirely satisfactory marking scheme for the counts and percentage cover for the different video specimens. The counts allocated by the candidates varied considerably, particularly, as one would expect, when the organisms were more abundant. For example, in the case of Specimen 1, *Asterias rubens*, where there were very few individuals present in the video clip, the estimated counts ranged between 1and 2 (1difference), whereas in the case of the more abundant Specimen 3, *Aequipecten opercularis*, count scores ranged between 21and 150 (129 difference) (Appendix 3).

The range of counts given for all the different Test Organisms are presented (see Appendix 3) to give an indication of the variability of the responses given by candidates, but probably the most useful indication of the 'right' count is the mode, and marks are given on the basis of how close a candidate's count is to this. However, this does involve making the precarious assumption that 'the majority is right'.

2 marks were allocated to those respondents who gave a score equivalent to the mode (the 50 percentile) or within the 10 percentile each side of it (i.e. between the 40 and 60 percentiles), and 1mark was allocated to those who gave a score between the 10 and 20 percentiles each side of the mode (i.e. between the 30 and 40 percentiles on the one side and between the 60 and 70 percentiles on the other). So for instance, for Test Organism Number 6, the range of 'count' scores given by respondents was 14 (7 – 21). The response mode was 7, the 40 percentile was 7, the 60 percentile was 9, the 30 percentile was 6, and the 70 percentile was 14 (all rounded to nearest whole numbers). Therefore those candidates that had given a count between '7' and '9' were given 2 marks and those that gave a count of '6' or of between '10' and '14' (inclusive) were given 1 mark. All others were given '0'.

Percentage cover estimates were marked in the same way.

Because some candidates assessed certain Test organisms on the basis of counts while others assessed the same ones on the basis of percentage cover (see Appendices 3 and 4 - Organism Numbers 6,7,8, and 9), the marks allocated for their abundance estimates have been combined into a single table (Table 4). However, having the two options for completing this part of the Test has implications on the reliability of the mode as an indicator of the 'correct' answer because, in some

cases, the mode is calculated from the full number of responses (13), whilst in others it is calculated from as few as 3 responses.

4.2.2 Biota Test Results

The responses given for the identification of the labelled Test organisms are given in Appendix 1. The associated marks allocated are presented below (Table 2).

Table 2: Marks allocated for the identification of the labelled Test organisms

	Test Organism Number													
Organisation Code	1	2	3	4	5	6	7	8	9	10	11	12	Total Mark	% Mark
QBA	5	5	5	5	5	5	5	5	0	5	5	3	**53**	**88**
QBZ	5	5	5	5	1	5	5	3	3	5	5	3	**50**	**83**
NTOA	5	3	5	5	1	5	5	5	5	5	5	3	**52**	**87**
NTOB	5	5	5	5	1	4	5	0	0	5	5	3	**43**	**72**
NTOO	5	5	5	5	5	3	5	5	5	5	0	3	**51**	**85**
QAA	5	5	5	5	5	5	5	0	3	5	0	3	**46**	**77**
QAC	5	3	5	3	0	5	5	0	5	5	0	3	**39**	**65**
QAD2	1	3	1	5	0	5	5	3	3	5	0	0	**31**	**52**
QAE	5	3	5	5	5	5	5	5	5	5	0	3	**51**	**85**
QAI	2	3	5	5	3	5	5	5	5	5	0	0	**43**	**72**
QAL	5	3	5	5	3	4	5	5	5	5	0	3	**48**	**80**
QAM	5	5	5	5	3	5	5	5	5	5	5	3	**56**	**93**
QAN	5	5	5	5	0	5	5	0	0	5	0	3	**38**	**63**

The SACFOR Abundance estimates made by the candidates are presented in Appendix 2. The associated marks allocated are presented below (Table 3**Error! Reference source not found.**).

Table 3: Marks allocated for SACFOR Abundance assessment

	Test Organism Number													
Organisation Code	1	2	3	4	5	6	7	8	9	10	11	12	Total Mark	% Mark
QBA	5	5	5	5	3	3	5	3	0	5	3	5	**47**	**78**
QBZ	3	5	5	5	0	5	5	0	5	5	5	0	**43**	**72**
NTOA	5	5	5	3	0	5	5	0	3	5	3	5	**44**	**73**
NTOB	5	5	5	5	5	5	5	0	5	5	5	3	**53**	**88**
NTOO	3	5	5	5	5	3	3	3	3	3	5	3	**46**	**77**
QAA	5	5	5	3	5	3	5	3	3	5	3	5	**50**	**83**
QAC	3	5	5	3	3	3	5	5	3	5	0	5	**45**	**75**
QAD2	0	0	0	0	0	0	5	0	3	5	3	0	**16**	**27**
QAE	0	3	0	3	3	3	3	0	5	5	0	3	**28**	**47**
QAI	0	5	3	3	5	3	5	5	3	5	0	3	**40**	**67**
QAL	3	3	3	5	5	3	3	5	0	5	5	0	**40**	**67**
QAM	5	5	5	5	5	5	3	5	0	5	3	0	**46**	**77**
QAN	5	5	3	3	0	5	5	0	5	5	3	3	**42**	**70**

The Abundance estimates given by candidates on the basis of 'Counts' and 'Percentage cover' are given in Appendices 3 and 4 respectively. The associated marks allocated are presented below (Table 4).

Table 4: Marks allocated for Counts and Percentage Cover assessment

Organisation Code	Test Organism Number																									
	1		2		3		4		5		6		7		8		9		10		11		12			
	Count	%Cover	Count	%Cover	Count	%Cover	Count	%Cover	Count	%Cover	Count	%Cover	Count	%Cover	Count	%Cover	Count	%Cover	Count	%Cover	Count	%Cover	Count	%Cover	TOTAL MARK (Max =24)	Total % Mark
QBA	2		2		0		0			0	0		1			2		0		2	1			0	10	42
QBZ	0		0		0		2		0		1		0		0		0			2	1			2	8	33
NTOA	2		0		0		0			2	0					2		0		2	1			2	11	46
NTOB	2		1		2		2			0	0			0		2		2		2	0			2	15	63
NTOO	2		2		1		0			2	0		0		2		0			0	0			2	11	46
QAA	2		0		1		0			2	0		2		0		0			0	2			0	9	38
QAC	2		1		0		2				2		0		0		2			2	0			0	11	46
QAD2	0		1		1		0				1												0		3	13
QAE	2		0		0		2			0		0		2		0		0		2	2			2	12	50
QAI	2		2		0		2			2		2	1		1		0			2	0			0	14	58
QAL	2		0		2		2			2		2		2		0		2		2	2			2	20	83
QAM	2		0		0		2		0	2	0		0			0		2		2	1			0	11	46
QAN	2		2		0		0				2										0				6	25

Table 5: Summary of marks allocated for the difference components of the BIOTA ANALYSIS part of Test2

Organisation Code	Total % Mark for ID	Total % Mark or SACFOR Assessment	Total % Mark for Count & %Cover	Total % Mark for Biota Section of Test
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	

4.2.3 Findings

4.2.3.1 Identification

The candidates' ability to identify the organisms that had been labelled in the video clips was generally high, with all of them gaining marks of between 52% and 93% for this exercise (Table 2). As would have been expected most of the candidates were able to identify the large solitary organisms (e.g. *Aequipecten opercularis*) to species level, and there was clearly less confidence about assigning the smaller, less distinct species (e.g. the fine hydroids and encrusting sponges) to the lower taxonomic levels (Appendix 1). Overall, 65% of the identification responses were correct. Thus, from the point of view of testing identification skills, most of the video organisms selected for the Test were probably too easily recognisable (8 of the 12 Test organisms were large and distinctive and commonly found in British waters). It is suggested that organisms requiring a greater range of identification skills are selected for Test 3.

4.2.3.2 SACFOR Abundance assessment

The SACFOR Abundance estimates given by the candidates were quite variable, with between 2 and 4 SACFOR categories given for each of the Test organisms (Appendix 2). This is reflected in the marks allocated. On the basis of the marking scheme used here (including the associated assumption) (Section 4.2.1.2), candidates were awarded between 27% and 88% for their SACFOR Abundance assessment skills. The fact that there was so much variability in the responses may point to the need for some kind of training in SACFOR Abundance assessment. This is also mentioned by the candidates themselves in the 'Participant Feedback' information gathered (See Section 5).

4.2.3.3 Counts and Percentage Cover

The 'Counts' and 'Percentage cover' estimates are presented in Appendices 3 and 4, respectively. Two of the candidates (QAD2 and QAN) did not attempt to assign a percentage cover estimate to any of the Test organisms. The two types of abundance estimates should have been mutually exclusive since candidates were asked EITHER to assess counts of individual organisms OR to estimate their percentage cover, as appropriate, depending on the kind of organism concerned. The responses showed that this was, in fact the case, apart from one candidate (QAM) who gave both a count and a percentage cover assessment for Test Organism Number 5. However, the abundance of certain organisms (Test Organism Numbers 6, 7, 8, and 9) was estimated using counts by some candidates and by using percentage cover by others. Consequently the marks for these two

elements of the Test have been combined into one Table (Table 4). Final marks for these abundance assessments ranged between 13% and 83%. However, the candidates' decisions to assess abundance on the basis of 'count' or 'percentage cover' have clearly complicated the marking process and the reliability of the ultimate marks, and this needs to be addressed in Test 3.

4.2.3.4 Final Marks for the BIOTA ANALYSIS part of Test 2

The summary of marks given for the BIOTA ANALYSIS part of TEST 2 is given in Table 5. It is to be noted that the weightings for the different components of the Test are equal. However, it may be more appropriate to give different weightings to the different components. The weightings to be used need to be agreed at the NMBAQC Video Ring Test Workshop.

The overall % marks (using equally weighted components) ranged between 31% and 77%. On the basis of a 50% pass mark, just one candidate (of 13) would have 'failed' the Test.

Candidates will be able to get an idea of their performance relative to others from Table 5; 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 the parts of the video analysis process for which they might require remedial training.

4.2.4 Issues

- Agreement is required on the weighting of marks. Weightings for top/average/low marks for identification of organisms is at present entirely subjective. Marks between 1 and 5 have been allocated here (see section 4.2.1.1) but these could equally have ranged between 1and 3, or 1 and 10, or even 1 and 50. The marks given need to allow sufficient distinction to differences in performance between responses.
- A major issue in marking abundance estimates is knowing what the correct answer is. Because of the nature of video material it is often impossible to decide on what the abundance of a particular organism is – it can be one person's word against another. The approach used here for marking estimated semi-quantitative (SACFOR) abundance, counts and percentage cover relies on using the 'majority' response as the 'correct' answer; it does not involve an 'absolute' correct answer. This is clearly unsatisfactory as (a) it makes the assumption that the category most frequently assigned to (the mode) is the correct one (which is not necessarily the case - the majority may, in fact, be incorrect!) and (b) the 'majority' response will be variable, depending on the group of candidates concerned. In future Tests there needs to be, as far as possible, a predetermined 'correct' answer. This issue will need to be addressed during the NMBAQC Video Ring Test Workshop.
- Agreement is required on whether counts or % cover ought to be used in assessing quantitative abundances and whether marks should be allocated when one (e.g. count) has been used when it would have been more appropriate to use the other (% cover). A statistical complication arises in the process of allocating marks to answers that relate to modes that have been derived using different sample sizes (see section 4.2.1.3).
- Agreement is required on the appropriate weightings to be used for the different components of the Test (i.e. in addition to those for the allocation of marks mentioned in the first 'issue' listed above). Using different 'sets' of weightings for the different components of the Test can have a huge impact on the final marks.
- Agreement as to what the 'Pass Mark' should be is required. This will be critical to the NMBAQC Testing process. It could be an absolute figure (e.g. 50%), or it could vary from one Testing session to another if the notion of having a fixed a proportion of success:failure rate is introduced.

4.2.4.1 Recommendations for Test 3

- Include species which require increasing levels of identification skills for the 'organism Identification' part of the Test.
- Video clips with known organisms and 'known' (i.e. predetermined) SACFOR abundances, counts and % cover, should be used in order to facilitate the marking process.
- The implications of asking candidates to assess abundance in terms of 'count' OR 'percentage cover' for the marking process need to be addressed.
- Recommend that candidates thoroughly check their responses before submitting results to avoid losing marks. For instance, one candidate had clearly mistakenly exchanged data for two of the labelled video specimens, entering details for *Eucratea loricata* instead of for *Nucella lapillus* and vice versa.

4.2.4.2 Suggested Workshop Topics

- These largely relate to the issues identified in section 4.2.4 above.
- Deciding on numbers (i.e. weightings) of marks to allocate;
- How to arrive at the 'correct' answer for abundance assessments;
- Assessment of abundances using 'Counts' and 'Percentage cover' – should they be mutually exclusive?
- Weightings of the different elements of the Test, e.g. what weighting should be given to the following: Taxon ID, SACFOR assessment, Count/%Cover assessment??
- Deciding on a pass mark;
- Abundance assessment Training requirements.

5 Feedback Results

5.1 Feedback response

Nine organisations responded to the second NMBAQC Video Analysis Ring Test, and all respondents gave feedback to some degree. The three areas of feedback given were on the video quality for each video clip, the training and experience needed to complete the test, and finally general feedback on equipment and resources used, and all other aspects of undertaking the test and entering the results online. The comments that were received are summarised in the Tables and discussion points below.

5.2 Results

5.2.1 Video Quality

The participants were asked to answer 10 questions on the quality of each video clip, giving either a positive answer by checking a box, or a negative answer by leaving it blank. These are summarised in Table 6, with a potential mark out of 9 for each question per video clip (9 respondents), and finally presented as a total mark (out of a potential 90 positive answers) and the corresponding percentage. Particularly low scores indicating an unsatisfactory feature of video quality (≤5/9 or ≤ 50%) have been highlighted in red.

Table 6: Summary of Feedback for Video Clip Quality

Take No.	Location	Owner	Consistent camera height	Can determine substrate?	Can distinguish individual orgs?	Smooth camera	lighting bright enough?	lighting consistent?	water clarity good?	suitable speed for analysis?	speed consistent?	angle of camera consistent?	Total (out of 90)	Percentage positive
1	English Channel	CEFAS	9	9	6	6	8	9	9	7	7	9	79	88
2	English Channel	CEFAS	9	9	5	9	9	8	8	5	8	9	79	88
3	English Channel	CEFAS	8	8	8	8	9	8	9	7	6	9	80	89
4	Outer Hebrides	AFBI	0	8	7	2	7	5	6	5	2	1	43	48
5	English Channel	CEFAS	9	9	8	5	8	7	9	8	5	9	77	86
6	East Antrim	AFBI	0	2	4	0	4	0	8	3	0	2	23	26
7	English Channel	CEFAS	8	7	7	2	8	7	9	7	5	9	69	77
8	Wey-mouth	Envision	0	9	8	0	7	6	8	2	1	0	41	46
9	English Channel	CEFAS	9	7	7	8	8	9	9	5	8	9	79	88
10	English Channel	CEFAS	9	7	6	9	8	8	9	5	8	9	78	87

5.2.2 Training Needs

Participants were asked to list any training that they thought might be necessary to complete the video analysis in this test, and whether they, themselves, had experienced this kind of training. The types of training were collated into common themes and types, and the number of respondents who mentioned each type were counted and the tally noted. This was then represented as a percentage of the total number of possible respondents.

Table 7: Summary of Recommended Training and Experience of Participants

Type of Training	No. of respondents who suggested training	No. of respondents who have experienced training	Percent of respondents who mentioned training
Species identification (from video/stills)	**4**	**2**	**44%**
Species identification (in situ)	1	1	11%
algae ID	1	0	11%
sponge ID	1	0	11%
Seasearch surveyor course - species ID	1	1	11%
phylum specific ID courses	1	1	11%
Hydroids and Bryozoans	1	0	11%
Enumeration techniques (percentage cover)	**3**	**0**	**33%**
Abundance scales	1	1	11%
substrate scales	1	0	11%
Supervised analysis/with expert mentoring	**3**	**2**	**33%**
Diving experience of area	**3**	**2**	**33%**
NMBAQC ringtests (undertaking tests and feedback with answers e.g. annotated species/substrates	**3**	**3**	**33%**
Substrate types	**3**	**0**	**33%**
finer fractions (mud/silt & surface sediment)	1	0	11%
Seasearch surveyor course - general habitat ID	1	1	11%
Biotope allocation (Eunis/MNCR) training	**2**	**0**	**22%**
Field experience of video analysis and ground truthing/ROV	**2**	**2**	**22%**
Rugosity	**1**	**0**	**11%**
Trawl samples	**1**	**1**	**11%**
Grab samples (sediment)	**1**	**1**	**11%**
Practical workshops	**1**	**1**	**11%**

5.2.3 General Feedback

Comments on various aspects of the test are summarised as follows:

Equipment and Materials Used

- 67% of respondents used PCs to view video clips, 33% used DVD players.
- A total of 14 different reference materials were used to aid video analysis of this test, with the most popular being the MARLIN and Marine Life Encyclopaedia websites, the Hayward and Ryland 'Handbook for Marine Fauna' and Naylor's 'Great British Marine Animals' (all used by more than one respondent).
- 33% of respondents used no references at all.

Metadata

- 44% of respondents thought that the metadata provided was sufficient (location, owner, year of survey and depth). However, 33% suggested providing the length of tow (to calculate SACFOR) and other metadata mentioned was an indication of scale, time of year (seasonality), location (open water, estuarine, coastal), and equipment used.

Undertaking the Test

- Time taken to complete the test ranged from 3.5 hours to 15 hours, with an average time of 8.65 hours.
- The number of people completing the test from each organisation ranged from 1 to 6, with an average of 2 per organisation. 44% of responses were completed by 1 person only.

Stills and Stills Analysis

- Stills were generally considered to be a useful tool for any video analysis. However, one respondent questioned their use when the details could not be recorded separately from the video analysis, and others questioned the 'control' aspect of clip 9.

Video Footage Quality

- 44% of respondents felt that the majority of video clips were of good quality. However, some points mentioned were that:
    - grading of video footage is necessary but difficult, as it is still subjective, and depends on the purpose of analysis;
    - some of the clips were quite blurry and needed better resolution, and this should have been better addressed in grading the quality of the footage (respondents suggested resolution tests, test cards for testing the equipment handling of resolution);
    - other video footage formats were requested such as .AVI files, or tapes;
    - footage from a wider variety of habitats was requested, including more hard/algal-covered substrate.

Guidance

- 78% of respondents thought that the guidance provided was either 'good' or 'adequate'. However, some points mentioned were that:
    - the location of the stills for analysis of clip 10 was not clear, as was the information given on scale (not on the website);
    - the completion of the 'Biota' section and in particular the use of the 'higher taxonomic level' field caused confusion and needs better explanation.

Online Submission of Test

- The time-out function was too short and interrupted the entering of data (55%);
- There were some complications with the submission of feedback entries not being accepted (22%);
- Text entry couldn't include inverted commas, and the auto-fill was a problem on some browsers.

Analysis Tools

- 100% of respondents used all the analysis tools, and most (66%) considered them useful;
- Two respondents mentioned the issue of vertical scale for rugosity, and whether it should be assumed independent of scale i.e. a boulder field is no more/less rugose than cobble and gravel.

Abundance Assessments and Scale Issues

- Most difficulties encountered were with the SACFOR scale (considered too blunt a tool), which cannot be addressed by this test alone, but could perhaps be better explained in the guidance:
    - Needs a range of count per area (1-9 per m^2, rather than 1 per m^2);
    - Needs a 'P' or 'Present' category;
    - It was not always obvious which category to use (encrusting/small/large) for SACFOR or counts.

Substrate and Biota Data Entry

- 55% of respondents felt the data entry for the substrate was good or adequate (only issue was the lack of a 'shell' category, but this could have been included in 'other substrate type');
- Issues with the recording of analysis of the biota video clips lay mainly with the 'higher taxonomic level':
    - where there was confusion there was no space to indicate what was done;
    - main confusion concerned the completion of the 'higher taxonomic level' field;
    - if identifying species just to hydroid level, do you then count all hydroids or just the species indicated;
    - there should be lookup tables for life-form categories;
    - difficult to differentiate between sponges and to know how to deal with that;
    - when it is possible to identify one creature to species level do you then assume all of that genus are the same species even if not sure because of video quality?

5.3 Findings

5.3.1 Video Quality

The results for the video quality show that most clips received positive answers on around 80% or higher of the questions, and therefore seem to be of adequate quality for the level of analysis requested in this test.

Three video clips have overall percentages of positive answers of less than 50% (Clip 4, 6 and 8) which would appear to make them of less than satisfactory quality. However, looking at the questions asked, the two most crucial questions (highlighted in pink) are whether the substrate and individual organisms indicated could be identified (the main purpose of this test), and Clip 4 and 8 in fact have sufficiently high scores for these two questions, and lose marks only for variables such as the consistency of camera angle, movement and speed, which may make viewing less easy but does not necessarily hinder analysis of the substrate and biota.

The only clip that scores a low percentage overall, and low marks for the two more important questions concerning substrate and biota determination is Clip 6, of East Antrim, where respondents felt that the footage was too bright and blurry to distinguish the sediments and individual organisms that appeared.

5.3.2 Training Needs

The most commonly mentioned training needs were species identification (through video and stills, or in situ, and especially for algae, sponges, hydroids and bryozoans), substrate identification, enumeration techniques and the use of scales, supervised analysis with experts, diving experience, and the use of 'practise' NMBAQC tests with answers and feedback.

5.3.3 General Feedback

In general, it seems that the second NMBAQC ring test for video analysis and its online submission was well received and considered to be an improvement on the first test. Of course, certain issues still need to be clarified and/or solved for the third and final test and for workshop discussion, which are summarised in the following sections.

5.4 Issues

5.4.1 Video Quality

5.4.1.1 Feedback

Although it is possible to ask a whole range of questions to determine video quality and grading on various aspects of the footage, the feedback received highlighted that this is still a highly subjective process, and depends on the purpose of the video analysis. From the questions asked in Test 2, it was concluded that the only really significant questions to ask, for the purposes of the video analysis concerned, are: 1) is it possible to determine the substrate, and 2) is it possible to distinguish individual organisms?

5.4.1.2 Resolution

One further aspect that was mentioned by several participants was improving and assessing the resolution of the video footage in some way. Within the scope of these tests, we can only use the video footage available to us, and we will strive to find footage of the best quality possible. However, footage from underwater towed video surveys often has less than optimal resolution is , and so this reflects 'real life' situations. Nonetheless, as a means of assessing resolution for rating video analysis, methods for testing the quality of resolution have been suggested (i.e. asking participants to state the frame in which objects of differing sizes can be seen, and/or test cards for testing the quality of the equipment being used for its resolution capability).

5.4.1.3 Type of Footage

Respondents also requested that footage be made available in different formats (tapes or .avi files). However, this would further complicate comparison of analysis between participants, and footage from a wider variety of habitats and locations (again, this will depend upon availability).

5.4.2 Training

The training techniques mentioned by participants were obvious recommendations, such as species and substrate ID, enumeration and use of scale techniques, *in situ* experience, and guidance from experts. However, it is worth nothing that respondents felt that assessment tests such as this NMBAQC process could, with feedback and answers provided, also be of benefit for training.

5.4.3 Other Feedback

5.4.3.1 Metadata

The main comments were several requests for length of tow to be provided, to aid use of the SACFOR scale. Certainly, the usual recommendation for video footage to be recorded in line with the MESH metadata standards (minimum requirements) would be sufficient, according to this test.

5.4.3.2 Use of Stills

Using stills to aid video analysis was welcomed, as they undoubtedly give more information on the substrate and any organisms found there. However, when this cannot be recorded separately, it is difficult to relate this information back to that recorded for the video analysis – although you can identify more information within the instant that the still image was taken, can you then assume that the information is the same for the rest of the video clip that does not have associated still images?

5.4.3.3 Online Submission of the Test

Respondents encountered difficulties with submitting feedback and the time-out function with the online system for Test 2.

5.4.3.4 Analysis Tools

Problems were encountered regarding an indication of scale on the vertical axis of the rugosity scale, or whether this measurement should be independent of scale, and also with using the SACFOR scale, (although this was a function of this measurement tool rather than the video analysis test), and guidance could possibly be made clearer).

5.4.3.5 Substrate and Biota Data Entry

The recording of substrate type and quantity during Test 2 was generally considered satisfactory; however, problems arose during the identification of organisms which could not be identified to species or genus level. In particular, the use of the 'higher taxonomic level' field was an area of confusion for participants, who were unsure of:

- Whether to fill in the lowest possible taxonomic term (phylum, class, order etc.), or the common term, or the most descriptive life-form category (yellow encrusting sponge, hydroid/bryozoans turf etc.);
- When identifying an organism only to 'sponge' or 'hydroid' level, do you then count all the organisms that could be included under this category, or just the species indicated (especially where it is difficult to distinguish between different organisms because of video quality)?
- When it is possible to identify one organism to species level, do you then assume all similar organisms are of the same species, even if the video quality makes the identification unsure?

5.5 Recommendations

5.5.1 Recommendations for Test 3

- Include the extra references used by respondents in Test 2 in the reference list for Test 3;
- Include length of tow, where possible, in the metadata;
- Warn participants that the average time to complete the test was 8.65 hours and that this should be factored in to any schedule for the 5-6 weeks given for Test 3 to be returned;
- If stills are to be used, information should be recordable (and marked?) within the Test;
- Include questions on the resolution of video clips in the feedback forms, as well as a resolution test and/or test card for resolution capability of user equipment in Test 3;
- To assess quality of video footage, only ask whether it was possible to distinguish the substrate and organisms that are to be identified;
- Try to obtain video footage from a wider variety of habitats and locations;
- Give a list of potential training sources for video analysis;
- Remove time-out function on website;
- Correct feedback submission problems on website;
- Check the auto-fill feature of fields on the online test (is this purely dictated by browser?);
- Clarify explanation of the rugosity tool in terms of vertical scale;
- Give clearer guidance on interpretation of the SACFOR scale (for organisms that are seen only once ('R' or 'P'?), provide a range of count per area (1-9 per m2, or all counts per m2) and which category to use for various organisms (encrusting/small/large);
- Clarify guidance and online fields on the use of 'higher taxonomic level' (or some similar term) for species identification, where the organism cannot be identified to species or genus level. Perhaps the use of lookup tables might aid the completion of this field, but this could be leading if the range of species is restricted, or technically difficult to account for a wide range of species names, common names etc.

5.5.2 Recommendations for Workshop Discussion

- Quality of Video Footage – how to assess if the footage is fit for purpose, and to maximise and assess resolution for viewing;

- Training – is there a need for compulsory training for video analysis, or just recommend types of existing training opportunities;
- Stills – is it possible to translate back information from stills to video analysis without recording separately for still images (i.e. the information is only definitely correct for that instant when the still image was taken);
- Metadata – obviously as much metadata as possible is preferable, but is use of the MESH metadata standard minimum requirements for data collection sufficient?
- Use of the SACFOR scale – ranges of count per area, the inclusion of a 'Present' or 'P' category, and the use of the encrusting/small/large category which is difficult to decide between (e.g. do you use largest extent of species, or those being recorded from video?);
- Species Identification – where an organism cannot be identified to species or genus level:
    - Should the organism be identified either to higher taxonomic level, or common term, or life form (often more informative that taxonomic term)?
    - Should all the organisms of this 'higher taxonomic level' then be counted, especially where it is difficult to distinguish between individuals because of video quality?
    - Should participants be penalised for guessing to species level when this level of identification is not possible?
    - When it is possible to identify one creature to species level, do you then assume all similar organisms are the same species, even if video quality makes the identification unsure?

6 Appendices

Appendix 1: The identification given to the 12 labelled Test organisms by the candidates

Organisation Code	Test Specimen Number											
	1	2	3	4	5	6	7	8	9	10	11	12
QBA	Asterias rubens	Pagurus sp.	Aequipecten opercularis	Echinus esculentus	Class: Hydrozoan	Alcyonidium diaphanum	Pentapora foliacea	yellow encrusting sponge	green sponge	Mytilus edulis	Nucella lapillus	Sertularia sp.
QBZ	Asterias rubens	Pagurus sp.	Aequipecten opercularis	Echinus esculentus	Sertularia argentea	Alcyonidium diaphanum	Pentapora foliacea	Halichondria sp.		Mytilus edulis	Nucella lapillus	Hydrozoan or fine algae?
NTOA	Asterias rubens	Family Paguridae	Aequepecten opercularis	Echinus esculentus	Eucratea loricata	Alcyonidium diaphanum	Pentapora foliacea	yellow sponge	yellow sponge	Mytilus edulis	Nucella lapillus	Sertularia cupressina
NTOB	Asterias rubens	Pagurus sp.	Aequipecten opercularis	Echinus esculentus	Vesicularia spinosa	Alcyonidium sp.	Pentapora fascialis	yellow encrusting sponge		Mytilus edulis	Nucella lapillus	Sertularia sp.
NTOO	Asterias rubens	Pagurus sp	Aequipecten opercularis	Echinus esculentus	Sertularia sp	Acyonidium gelatinosum	Pentapora foliacea	Indet. sponge 1. Possibly Suberites sp.	Indet. Sponge 2. Possibly Raspalia sp.	Mytilus edulis	Buccinum undatum	Haleciidae/Hydroid 1. Possible *Halecium* sp.
QAA	Asterias rubens	Pagurus sp	Aequipecten opercularis	Echinus esculentus	Class Hydrozoa/ Hydroid Species A	Alcyonidium diaphanum	Pentapora foliacea	Phylum Porifera	Family Axinellidae/Branching Porifera	Mytilus edulis	Buccinium undatum	Class Hydrozoa
QAC	Asterias rubens	Pagurus bernhardus	Aequipecten opercularis	Echinus sp.	Bryozoan	Alcyonidium diaphanum	Pentapora foliacea	Unknown - Res	Phylum Porifera: Arborescent	Mytilus edulis	Buccinum undatum	Phylum; CHROMOPHYCOTA
QAD2	Henricia oculata	Pagurus bernhardus	Pecten maximus	Echinus esculentus	Cellaria	Alcyonidium diaphanum	Pentapora fascialis	Order: Halichondrida	Order: Halichondrida	Mytilus edulis	Eucratea loricata	Buccinum undatum
QAE	Asterias rubens	Family; Paguridae	Aequipecten opercularis	Echinus esculentus	Subclass; Leptothecatae	Alcyonidium diaphahum	Pentapora fascialis	Phylum; Porifera yellow	Phylum; Porifera yellow	Mytilus edulis	Buccinum undatum	Class; Chlorophyceae
QAI	Class: Asteroidea	Family: Paguridae	Aequipecten opercularis	Echinus esculentus	Bryozoan/hydroid tuft	Alcyonidium diaphanum	Pentapora fascialis	Phylum Porifera/ yellow mound sponge	erect yellow fingery sponge	Mytilus edulis	Buccinum undatum	Sargassum muticum
QAL	Asterias rubens	Pagurus bernhardus	Aequipecten opercularis	Echinus esculentus	erect hydroid / bryozoan	Alcyonidium sp./ possibly A. Diaphanum	Pentapora foliacea	small orange-yellow sponge-like entity	small yellow arborescent sponge	Mytilus edulis	Buccinum undatum	Phylum: Rhodophycota
QAM	Asterias rubens	Pagurus sp.	Aequipecten opercularis	Echinus esculentus	Bushy branched bryozoan, possibly *Bugula* sp? Out of focus and unable to get to species level with resolution of video.	Alcyonidium diaphanum	Pentapora foliacea	Porifera/Yellow encrusting sponge	P = Porifera; C = Demospongiae; (O = Halichondrida?; F = Halichondridae?) Branching, tassled beige sponge, possibly Halichondria bowerbanki	Mytilus edulis	Nucella lapillus	P = Cnidaria; C = Hydrozoa; (O = Conica?; F = Sertullariidae?)/ Branching hydroid, possibly *Halecium?*
QAN	Asterias rubens	Pagurus sp.	Aequipecten opercularis	Echinus esculentus	Bugula plumosa	Alcyonidium diaphanum	Pentapora foliacea	Suberites sp.		Mytilus edulis	Buccinum undatum	Sertularia sp.

Appendix 2: SACFOR estimates given by candidates for each of the 12 Test Organisms

Organisation Code	Test Organism Number											
	1	2	3	4	5	6	7	8	9	10	11	12
QBA	R	O	C	O	F	F	C	O	R	S	R	C
QBZ	O	F	C	O	C	O	C	F	F	S	O	
NTOA	R	O	C	R	C	O	C	C	C	S	F	C
NTOB	R	O	C	O	O	O	C	C	F	S	O	A
NTOO	O	F	C	O	O	R	F	O	O	A	O	F
QAA	R	F	C	F	O	F	C	O	O	S	F	C
QAC	O	F	C	F	F	F	C	R	C	S	C	C
QAD2					A		C	C	C	S	F	
QAE	F	R	O	F	F	R	F	F	F	S	C	F
QAI	F	F	A	F	O	R	C	R	O	S	C	F
QAL	O	R	F	O	O	R	F	R	R	S	O	O
QAM	R	O	C	O	O	O	F	R	R	S	R	O
QAN	R	O	F	R	A	O	C	F	F	S	R	A
Number of Categories	3	3	4	3	4	3	2	4	4	2	4	4

Appendix 3: The 'Counts' allocated by the candidates for each of the 12 Test Organisms

Organisation Code	Test Organism Number											
	1	2	3	4	5	6	7	8	9	10	11	12
QBA	1	5	77	5		15	40				16	
QBZ	2	6	24	3	46	12	39	55	21		16	
NTOA	1	3	25	1		5					10	
NTOB	1	4	40	3		14					6	
NTOO	1	5	30	1		3	20	5	5		2	
QAA	1	10	37	4		3	48	9	54		12	
QAC	1	4	50	2		7	100	1	33		50	
QAD2	2	4	41	1		6						23
QAE	1	3	21	2							11	
QAI	1	5	150	2			50	3	40		30	
QAL	1	6	38	3							14	
QAM	1	6	28	3	20	21	65				16	
QAN	1	5	43	1		7					1	
MODE	1	5	38	2	33	7	48	5	33		13	23
40 PERCENTILE	1	5	36	2	30	7	43	4	28		11	23
60 PERCENTILE	1	5	40	3	36	9	49	7	36		15	23
30 PERCENTILE	1	4	29	2	28	6	40	3	23		10	23
70 PERCENTILE	1	5	42	3	38	13	53	8	39		16	23
Max	2	10	150	5	46	21	100	55	54		50	23
Min	1	3	21	1	20	3	20	1	5		1	23
Range	1	7	129	4	26	18	80	54	49		49	0

Appendix 4: Percentage Cover scores allocated by candidates for each of the 12 Test organisms

Organisation Code	Test Organism Number											
	1	2	3	4	5	6	7	8	9	10	11	12
QBA					15			5	1	80		25
QBZ										90		20
NTOA					5			5	5	90		10
NTOB					3		6	4	2	90		10
NTOO					5					60		10
QAA					5					95		25
QAC										80		40
QAD2												
QAE					20	3	15	15	15	80		20
QAI					5	1				90		15
QAL					5	1	15	1	2	90		10
QAM					5			1	2	80		7
QAN												
MODE					5	1	15	5	2	90		15
40 PERCENTILE					5	1	13	4	2	80		10
60 PERCENTILE					5	1	15	5	2	90		20
30 PERCENTILE					5	1	11	3	2	80		10
70 PERCENTILE					5	2	15	5	4	90		20
MAX					20	3	15	15	15	95	0	40
MIN					3	1	6	1	1	60	0	7
RANGE					17	2	9	14	14	35	0	33