United Nations Educational, Scientific and Cultural Organization

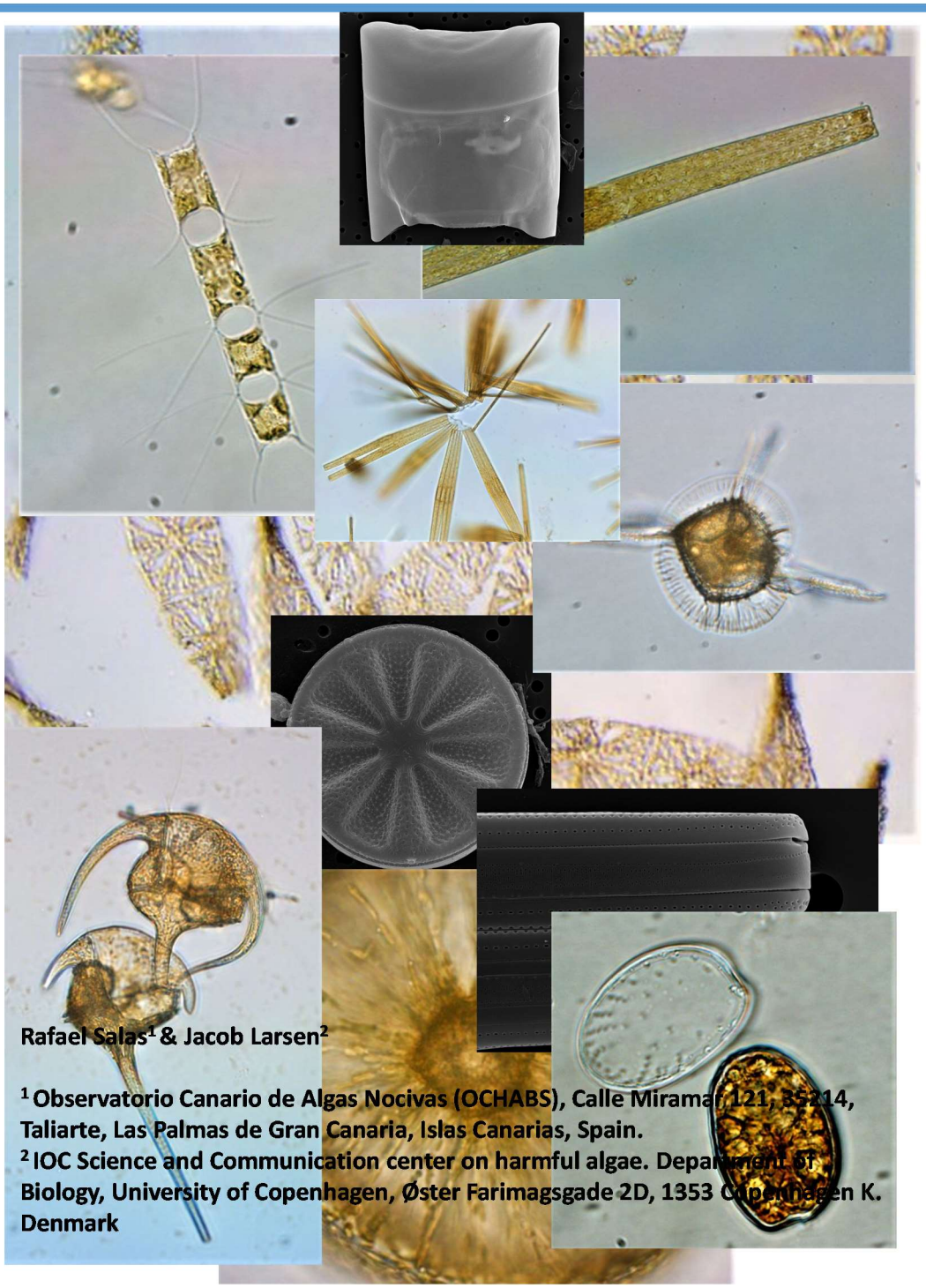
Intergovernmental Oceanographic Commission

IPI
INTERNATIONAL PHYTOPLANKTON INTERCOMPARISON

OCHABs
Observatorio Canario de algas nocivas

ULPGC
Universidad de Las Palmas de Gran Canaria

INTERNATIONAL PHYTOPLANKTON INTERCOMPARISON (IPI)
Proficiency testing in the abundance and composition of marine microalgae 2022 report



Rafael Salas[1] & Jacob Larsen[2]

[1] Observatorio Canario de Algas Nocivas (OCHABS), Calle Miramar, 21, 35214, Taliarte, Las Palmas de Gran Canaria, Islas Canarias, Spain.
[2] IOC Science and Communication center on harmful algae. Department of Biology, University of Copenhagen, Øster Farimagsgade 2D, 1353 Copenhagen K. Denmark

Table of Contents:

1. Summary of results

- In 2022, 84 analysts across 45 laboratories around the world participated in the IPI 2022 exercise. European countries accounted for 66% of the total participation, 5% came from South America, 10% from African countries, 6% from Oceania and 13% from Asia.

- 8 species were used in total in the samples. There were three dinoflagellates and five diatoms.

- The dinoflagellates were ***Alexandrium pacificum*** R.W.Litaker, 2014**,** ***Prorocentrum rhathymum*** Loeblich III, Sherley & Schmidt, 1979, ***Coolia sp.*** Meunier, 1919.

- The Diatom species were ***Guinardia striata*** (Stolterfoth) Hasle, 1996, ***Chaetoceros peruvianus*** Brightwell, 1856, ***Actinoptychus splendens*** (Shadbolt) Ralfs ex Pritchard, 1861, ***Synedropsis sp.*** G.R.Hasle, L.K.Medlin & E.E.Syvertsen, 1994 and ***Lampriscus sp.*** A. Schmidt in A. Schmidt et al., 1882.

- The robust average and standard deviation for each measurand was calculated using the Q/Hampel method in ProLab Plus statistical software. The expanded standard deviation was input manually into the program to take into consideration the heterogeneity of the samples. This expanded standard deviation was calculated using the consensus value through the iterative process and the between sample standard deviation from the homogeneity and stability test.

- All measurands passed the expanded criterion for homogeneity and stability according to ISO13528:2015 except for *A.splendens*, which did not pass the adequate homogeneity or the significant heterogeneity criterion.

- There were a very small number of warning and action signals across measurands. 3 Red flags (0.5%), 21 (3.4%) yellow flags and 8 (1.3%) non-detection flags (Grey triangles) from 616 results is evidence of good performance overall.

- 4 analysts weren't successful at the overall test from 77 returned results. Analysts 43, 47, 62 and 90 failed the quantitation of at least 3 or 4 items which requires training and improvement for the next round. 58 analysts had all the measurands (8) within the tolerance limits, 11 analysts had one failed measurand and 4 analysts two.

- The hardest identification was Lampriscus sp with only 42 analysts identifying correctly to genus level. 28 other analysts used *Odontella* as the answer and this was given as correct because of the lack of literature in the genus and also because sometimes it is considered in the small amount of literature available for the genus as *Odontella* or *Biddulphia.*

- The most non-detected species in the samples were *Coolia* and *Alexandrium.* 4 and 3 analysts each did not detect one or the other and just one non-detection for *prorocentrum.* Generally, dinoflagellates were harder to identify than diatoms. 7 non-detections on 3 dinoflagellates and zero non-detections on 5 diatoms.

- In 2022, all analysts passed the qualitative test. 51 analysts identified correctly all measurands. 12 analysts identified incorrectly 1 measurand, 4 analysts 2 measurands, 6 analysts had a non-detection and one incorrect identification, 3 analysts had one non-detection each and one analyst had two non-detections.

- Overall, from 616 possible correct identifications, there were a total of 586 correct answers at genus level (95%) and for the 4 species that could have been easily identified to species level (*Actinoptychus, Guinardia, Prorocentrum and* Chaetoceros) 250 correct answers from a possible 308 correspond to 81% correct. There were also 8 non-detections (1.3%) and 22 (3.6%) incorrect identifications.

- There were 83 attempts at the OTGA assessment, the median overall grade was 92.6%. 73.5% of analysts performed above the proficiency threshold of 90% and 18.1% of all analysts between 80-90%. 7.2% above 70% and another 1.2% below 70% requiring improvement.

- The OTGA facility index shows that the worst answered question in the test was Q11 (66.7%) a numerical question and the best Q15 (100%) another numerical question.

2. Introduction

The IPI Proficiency testing scheme is designed to test the ability of analysts to correctly identify and enumerate marine phytoplankton species in lugol's preserved water samples using the Utermöhl method. As in previous years, samples have been produced using laboratory cultures.

8 species were used in the IPI2022 samples. The dinoflagellates were *Alexandrium pacificum* R.W.Litaker, 2014, *Prorocentrum rhathymum* Loeblich III, Sherley & Schmidt, 1979, *Coolia sp.* Meunier, 1919. The Diatom species were *Guinardia striata* (Stolterfoth) Hasle, 1996, *Chaetoceros peruvianus* Brightwell, 1856, *Actinoptychus splendens* (Shadbolt) Ralfs ex Pritchard, 1861, *Synedropsis sp.* G.R.Hasle, L.K.Medlin & E.E.Syvertsen, 1994 and *Lampriscus sp.* A. Schmidt in A. Schmidt et al., 1882.

From 2021 to 2025, the IPI program is hosted by the Canary Islands HAB Observatory (OCHABS) in Las Palmas, Gran Canaria, Spain with the continued collaboration of the IOC Science and Communication Centre on Harmful Algae and in association with NMBAQC in the UK. The collaboration with the IOC UNESCO Centre for Science and Communication of Harmful algae in Denmark date back to 2011. This collaboration involves the use of algal cultures from the Scandinavian Culture Collection of Algae and Protozoa in Copenhagen, the elaboration of an online marine phytoplankton taxonomy assessment and the organization of an annual training workshop to discuss the results of the intercomparison exercise and to provide guidance on phytoplankton taxonomy.

The taxonomic assessment is set up in the online platform 'Ocean Teacher Global academy' hosted by the IODE (International Oceanographic Data and information Exchange) office based in Oostende, Belgium, a project office of the IOC.

In 2022, 84 analysts in 45 laboratories from across the world participated in the IPI exercise. European countries accounted for 66% of the total participation, 5% from South America, 10 % from African countries, 6% from Oceania and 13% from Asia (Figure 1). 23 countries are represented in this intercomparison exercise. The list of participating laboratories can be found in Annex IV of the annex report and a breakdown of participation from each country in figure 2.

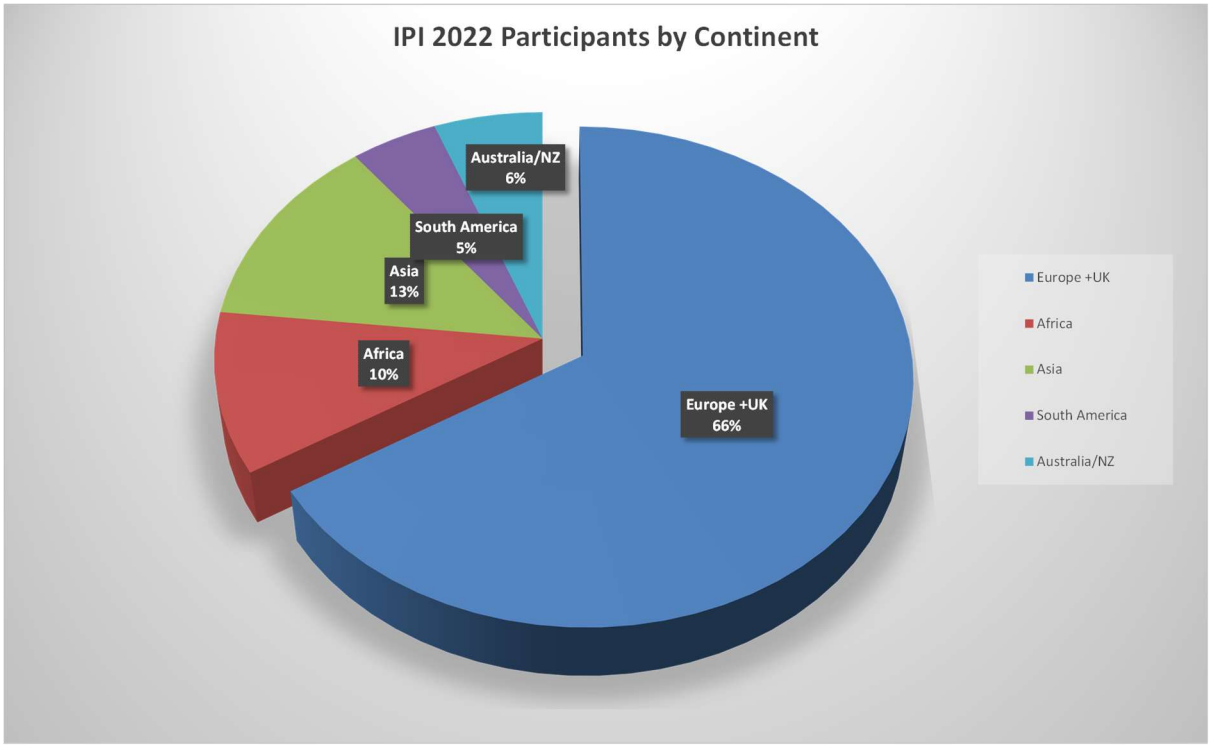


Figure 1: Participants by continent IPI2022

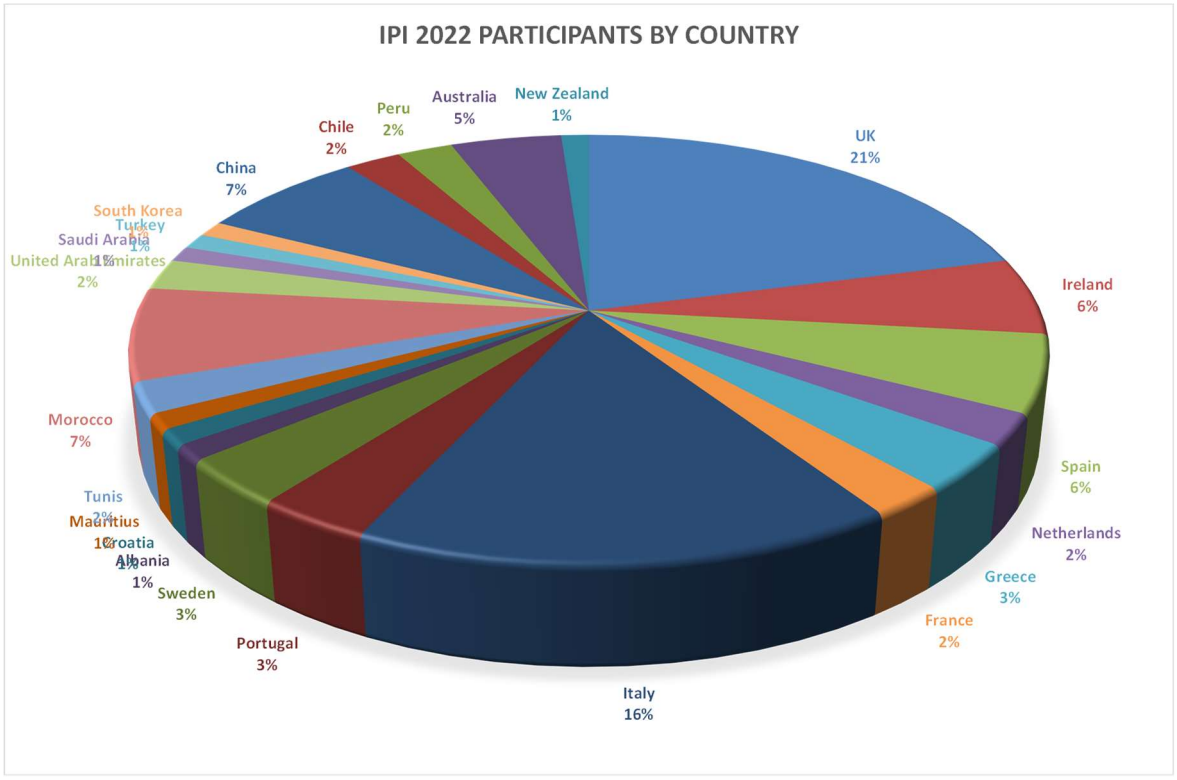


Figure 2: Participants by country IPI 2022

This intercomparison exercise has been coded in accordance with defined protocols for the purposes of quality traceability and auditing. The code assigned to the current study is OCHABS-

IPI-2022. The number of IPI participants has increased significantly since 2011 and the influence of the test has also been widened to many regions across the globe (figure 2).

Many laboratories participate on a regular basis and several analysts have more than 15 result contributions since 2005. In 2022, it is the first time we have a laboratory from South Korea, Turkey and Saudi Arabia participating in the scheme (Figure 2).

Pre-registration to the IPI intercomparison is through our dedicated website www.iphy.org to provide a structured and user-friendly single point source of information relating to the IPI. Here, laboratories can find information about the IPI scheme and the schedule for the year.

3. Materials and Methods

3.1 Sample preparation, homogenization and inoculation

The seawater used in this study was collected at Ballyvaughan pier, Galway Bay, Ireland, filtered through 47mm GF/C Whatmann filters (Whatmann™, Kent, UK), autoclaved (Systec V100, Wettenberg, Germany) and preserved using neutral Lugol's iodine solution (Clin-tech, Dublin, Ireland).

The materials were produced from a number of isolated strains. A stock solution for each of the species was prepared using 50ml glass screw top bottles (Duran®, Mainz, Germany). Then, a working stock to the required cell concentration was prepared using a measured aliquot from each stock solution into a 2l Schott glass bottle. The stock solution containing all the species for each specific batch, were homogenized using the 2L Inversina (Bioengineering AG, Wald, Switzerland), which uses the Paul-Schatz rotation method and sub-divided into four replicate working stocks containing 400 ml each. These working stocks were homogenized again before inoculation for 3 minutes at speed setting number 4 or roughly 73 rpm.

5 ml amber glass ampoules (Wheaton, New Jersey, USA) were used to store the inoculum. 3ml aliquots of the homogenized materials were inoculated into each ampoule containing 100µl of neutral lugol's iodine. This was carried out using an automatic eppendorf multipipette Xstream (0-50ml) (Eppendorf, Hamburg, Germany), set to dispense accurately 3 ml per sample. Once all the samples were inoculated, ampoules were purged with nitrogen gas to stop oxidation and

sealed using a flame torch. The ampoules were submerged into a water bath to test that they were sealed properly.

Each ampoule was labeled with a sequential number and each box of ampoules was also labeled to differentiate sample sets produced from different working stocks (IPI2022 batches #1, #2, #3 & #4) and store in the fridge (2-5 °C) in the dark until further transport to the participating laboratories.

Participants must carry out preparatory steps before the samples can be analysed. Analysts had to accurately pipette or dispense 47 ml of seawater including lugol's iodine into the sterilin tubes, open the ampoule by the break-line carefully and pipette out its contents including a rinsing step into the sterilin tube. Once the sterilin tube is inoculated with the 3ml ampoule, the tube is ready for homogenization and analysis.

3.2 Culture material, treatments and replicates.

All the cultures used in this study have been collected in the Canary Islands except one. Most species were identified through light microscopy techniques using an inverted microscope Olympus BX-53 (Olympus, Southend-on-Sea, UK) and a bench-top SEM Hitachi FlexSEM 1000 (Hitachi, Maidenhead, UK).

The cultures are checked by light microscopy in relation to their condition, shape, size and quality of their fixation using lugol's. Chain formers are also examined for their ability to stay in chains after preservation. At this point some other preliminary cultures may be discarded if they don't achieve the desired standard for the test. Images under the LM and SEM are taken of all the potential candidate species at high magnification as a record for the species in the test.

A total of 576 ampoules were produced for this study. Each participant was sent a set of four replicates. 84 analysts in 45 laboratories were sent a total of 336 ampoules. Each sample set consisted of a padded brown envelope containing 4 ampoules, 4 x 50 ml skirted centrifuge tubes and 4 plastic droppers.

3.3 Cell concentrations

Preliminary cell counts from individual stock solutions were carried out using a 1 ml glass Sedgewick-Rafter cell counting chamber (Pyser-SGI, Kent, UK) to establish the approximate cell concentration for each species.

These approximate cell concentrations were used to decide the volume of the aliquot for each species and the final concentration required for the working stock. Microscopic analysis of an aliquot of all the working stocks together, allow us to preview how the final samples will appear before a final decision is made on cell concentrations and number of species to be inoculated.

3.4 Sample randomization

All samples were allocated randomly to the participants using Minitab® Statistical Software Vr16.0 randomization tool.

3.5 Forms and instructions

The instructions and forms required for this test are available at www.iphyi.org for download in the menu item IPI documents and are also sent via e-mail to all registered participants including their unique identifiable laboratory and analyst code. Here you can find a counting guide in pdf format to advise in the identification and counting of the species. Also, a short video is uploaded onto our website in the IPI documents under sample preparation, showing how to prepare the samples prior to analysis.

Form 1 (Annex I) is required to confirm the receipt of materials, the number and condition of samples and the correct sample code. Form 2 (Annex II) in Excel format is required to record the species composition in the samples and to calculate their abundance. All participants are asked to read and follow the instructions for the test (Annex III in separate annex report) before commencing.

At the end of the exercise and with the publication of this report, analysts will be issued with a statement of performance certificate (Annex V in separate annex report) which is tailored

specifically for each test. This is an important document for auditing purposes and ongoing competency.

3.6 Statistical analysis

Statistical analysis was carried out using PROlab Plus version 2022.7.25.0 dedicated software for the statistical analysis of intercalibration and proficiency testing exercises from Quodata, and Microsoft office Excel 2016.

We follow the standard ISO normative 13528:2015, which describes the statistical methods to be used in proficiency testing by inter-laboratory comparisons. Here, we use this standard to determine and assess the homogeneity and stability of the samples, how to treat outliers, determining assigned values and calculating their standard uncertainty. Comparing these values with their standard uncertainty and calculating the performance statistics for the test through graphical representation and the combination of performance scores.

The statistical analysis of the data and final scores generated from this exercise has been carried out using the consensus values from the participants. The main transformation is the use of iteration to arrive at robust averages and standard deviations for each test item. This process allows for outliers and missing values to be dealt with, and it also allows for the heterogeneity of the samples to be taken into consideration when calculating these values.

3.7 IPI Ocean teacher online taxonomic assessment

The online taxonomic assessment or HAB quiz was organized and set up by Jacob Larsen (IOC UNESCO, Centre for Science and Communication on Harmful Algae, Denmark) and Rafael Salas (OCHABS, Canary Islands, Spain). The exercise was prepared in the web platform 'Ocean teacher'. The Ocean teacher training facility is run by the IODE (International Oceanographic Data and information Exchange) office based in Oostende, Belgium. The IODE and IOC organize some collaborative activities among them, the IOC training courses on toxic algae and the IPI online HAB quiz. The online quiz uses the open-source software Moodle Vr2.0 (https://moodle.org).

This year, participants were sent information from ioc.training@unesco.org to register to the OTGA website. The preparatory phase consisted of an online quiz made available on the IOC/OceanTeacher e-Learning Platform.

In order, to access the quiz, participants had to create an account on OceanTeacher (www.oceanteacher.org). Once they received confirmation of their account, each participant then was able to enroll to the course. Participants that already have an account on OT were able, instead, to enroll directly using the link and enrolment key to the course/quiz. Note that OceanTeacher send automatic messages once enrolled to the course and these may be considered SPAM, so please make sure to regularly check your SPAM box.

Additionally, the participant's name was added to the official participants list of this year's HAB-IPI Exercise on the UNESCO/IOC's event calendar on https://oceanexpert.org. Participants were invited to create or update their profile on the Ocean Expert Directory. This is used for UNESCO-IOC statistics on Capacity Development only. Please note that the OceanTeacher e-Learning Platform and the Ocean Expert Directory are two different and independent websites.

In case of any issues using the OceanTeacher e-Learning Platform participants could contact us on ioc.training@unesco.org ; and in case of any questions regarding content, they could contact IPI on rsalas@observatoriocanariohabs.com

The test itself consisted of 18 questions (see Annex XVI). Question types used in the quiz were 'matching type' (Q 1-3-8-17-18) which have dropdown menus including a selection of answers that analysts must choose from, 'multiple choice' (Q 4-5-9-10) where the participant must fill in the right option from those given, and it penalizes wrong choices. The amount of this deduction depends on the number of possible answers and ranges from 5% to 25% per wrong answer. There were also 'numerical' questions (Q 12-13-15-16) where analysts had to count the cells in the images provided and 'drag and drop' types (Q 2-6-7-13-14) where objects must be dropped onto place holders. All questions had equal value and the quiz had a maximum grade of 100% for a perfect score. The online quiz can only be submitted once. After submission, no changes can be made. However, analysts can login and out as many times as they wish throughout the allocated time periods and make changes. The changes are saved and can be accessed at a later stage, as long as participants don't press submit.

4. Results

4.1 Homogeneity and stability study

The homogeneity and stability test in 2022 included 8 measurands (Table 1) and most of them except for *A.splendens* satisfied at least the ISO13528:2015 requirements for significant heterogeneity which allows the standard deviation to be greater than 30%. Also, all materials passed the stability assessment according to the expanded criterion. This means, as in previous years that the materials are not adequately homogeneous but not significantly heterogeneous, except for the one measurand above.

The procedure for a homogeneity and stability test is recorded in annex b of ISO13528:2015. The assessment criteria for suitability, is also explained there. See Annex VI in the annex report to see all the results from the homogeneity and stability test for each measurand.

The calculations have been carried out using ProLab Plus version 2022.7.25.0 and the reports for homogeneity and stability are given separately for each measurand. The top of the report gives you information on the measurand, mean and analytical standard deviation for the homogeneity analysis and the homogeneity and stability mean comparison in the stability analysis. The reports, also show the target standard deviation for each measurand, which in this case was calculated manually using the consensus results of the participants and taking into consideration the heterogeneity of the samples, as will be explained later.

The middle part of the report gives you the results of the different tests. ProLab Plus calculates whether the data has passed the criteria for the F-test and ISO13528:2015 test for homogeneity and significant heterogeneity. The bottom part of the report is the actual graphical representation of the sample results as box plots. The homogeneity test shows the 10 samples that were analyzed and calculates the heterogeneity standard deviation (SD between samples) and the analytical standard deviation (SD within samples). The stability test graph shows the 10 homogeneity sample results and the 3 stability test sample results, thirteen in total and compare their mean values (Annex VI of annex report).

According to ISO 13528:2015, the heterogeneity standard deviation (s(sample)) between the proficiency test items should not exceed 30 % of the standard deviation for the proficiency

assessment. If the homogeneity test fails, the heterogeneity standard deviation is then, taken into consideration, when calculating the standard deviation for the measurand. The consensus values new heterogeneity standard deviation (STD) was used for all measurands as most items failed the adequate homogeneity criterion except for *Alexandrium pacificum, Coolia monotis* and *Chaetoceros peruvianus* (table 1). However, no significant heterogeneity was found according to the expanded criterion except for *Actinoptychus splendens*.

The within sample difference of the homogeneity cell counts for *Actinoptychus splendens* suggests that homogenization was not achieved here, with large variance in cell counts between replicates. This is also refuted in the results of the participants showing the largest variance between replicates for these species (see Annex XIII: Graphical summary of results in the annex report for *Actinoptychus splendens*. Hence, the proficiency test items cannot be considered fully homogeneous but not significantly heterogeneous (Table 1) except for *Actinoptychus splendens*.

Measurands	Cochran outliers	F-test	ISO 13528:2015 test for adequate homogeneity	ISO 13528:2015 - test for significant heterogeneity	Stability test ISO 13528:2015	Stability test - expanded criterion
Actinoptychus splendens	no outliers found	Not OK	Not OK	Not OK	Not OK	Ok
Alexandrium pacificum	no outliers found	Ok	Ok	Ok	Not OK	Ok
Coolia monotis	no outliers found	Ok	Ok	Ok	Not OK	Ok
Chaetoceros peruvianus	no outliers found	Ok	Ok	Ok	Ok	Ok
Guinardia striata	no outliers found	Ok	Not OK	Ok	Not OK	Ok
Lampriscus sp.	no outliers found	Ok	Not OK	Ok	Ok	Ok
Prorocentrum rathymum	no outliers found	Ok	Not OK	Ok	Not OK	Ok
Synedropsis sp.	no outliers found	Ok	Not OK	Ok	Not OK	Ok

Table 1: IPI2021 Homogeneity and stability results according to ISO13528:2015

As most analysts achieved good Z-scores for *Actinoptychus splendens*, there was no need to disregard this result except for one analyst (10) with a questionable score. Since we cannot be sure of the reliability of the homogenate, this result was waived for this analyst as n/a (not applicable) as there is reasonable doubt about the significant heterogeneity for this measurand. In relation to the stability test, all items were considered stabled according to the expanded criterion (table 1).

4.2 Outliers and missing values

Outliers in the data have been addressed by using the robust analysis as set out in Annex C algorithm A + S of ISO 13528:2015 and through the Q/Hampel algorithm is ProLab Plus which

truncates outlier values to +3 or -3 values. The robust estimates for this exercise have been derived by iterative calculation, that is, by convergence of the modified data (Annex VIII: Robust mean + SD iteration ISO13528 in the separate annex report) for each measurand.

In relation to missing values, the standard proposes that participants must report 0.59 n replicate measurements, so in the case of three replicates, at least two replicate results from each measurand must be obtained from each participant for the data to be included in the statistical calculations. If this rule is not fulfilled results from these participants won't be included in the calculation of statistics that affect other laboratories, but they may be used for the calculation of their own, for example z-scores.

Analysts that did not detect a particular species in the samples was given a 'non-detected' flag in their identification score and a +3 Z-score in their certificate. These Z-scores were signaled as 'Grey triangles' in the summary of Z-scores (Annex IX: Summary of Z-scores for all measurands in the annex report).

4.3 Analysts' Data

The full table of participants' results can be found in Annex VII in the annex report. The average count for each measurand was used to calculate the robust averages and standard deviations by iteration (Annex VIII in annex report). These values were then used to calculate the confidence limits for the Z-scores (See Annex IX).

For the purpose of this exercise we have used the consensus standard deviation from the participants and we have calculated the new standard deviation for each test item by adding the between samples standard deviation from the homogeneity test according to the formula below (A) from ISO13528:2015. The calculations are generated by iteration and can be found for each measurand in the annex report in annex VIII.

$$\sigma_{r1} = \sqrt{\sigma_r^2 + s_s^2}$$

(A)

Where;

σ_{r1} =the new SD for the homogeneity test

σ_r =between samples Standard deviation and

S_s= the robust standard deviation for the test

4.4 Assigned value and its standard uncertainty.

The assigned values (robust mean and standard deviation) for a test material are calculated as explained before from the consensus values of the participants (Annex VIII in annex report). The standard uncertainty of the assigned value can then be calculated using the equation (B) below.

B) $$u_X = 1{,}25 \times s^* / \sqrt{p}$$

Where;

u_x= Standard uncertainty of the assigned value,

s^*= robust standard deviation for the test

p= number of analysts

Species	A.pacificum	P.rhathymum	A.splendens	Coolia sp.	C.peruvianus	G.striata	Lampriscus sp.	Synedropsis sp.
Robust mean x*	3011	5361	1376	4797	15703	6339	1485	3501
Robust Stdev s*	692	1463	478	1058	4994	1320	292	604
Standard Ux	101	210	68	155	711	188	42	86
n=	74	76	77	73	77	77	77	77
if Ux < 0.3xSTdev	208	439	143	317	1498	396	88	181
then Ux is negligible	neg	neg	neg	neg	neg	neg	neg	neg
The equation is satisfied in all cases								
Cumulative distribution function cut off points for normal distribution								
*x *-1.5s**	1973	3167	659	3210	8212	4359	1047	2595
*x *+1.5s**	4049	7556	2093	6384	23194	8319	1923	4407
Homogeneity test	A.pacificum	P.rhathymum	A.splendens	Coolia sp.	C.peruvianus	G.striata	Lampriscus sp.	Synedropsis sp.
Reference value mean	4022	7570	1427	6308	15920	7962	1358	3062
Reference value stdev	379	727	934	600	1874	1660	285	243
	Comparison with assigned value							
	A.pacificum	P.rhathymum	A.splendens	Coolia sp.	C.peruvianus	G.striata	Lampriscus sp.	Synedropsis sp.
*x *-X*	1011	2209	51	1511	217	1623	127	439
Uncertainty of diff.	142	297	96	219	1006	266	59	122
2* Uncertainty of diff.	284	593	193	438	2012	532	118	243
If diff. Is more than twice its Uncertainty then rule is not satisfied								

Table 2: Assigned values and standard uncertainties for the test.

If *Ux* is less than 0.3 times the standard deviation for the test, then this uncertainty is negligible for the test material. In our case, all our test materials satisfy the equation (Table 2).

4.4 Calculation of performance statistics

We are following the statistical methods laid out in ISO13528:2015 to calculate the performance statistics for the test. The results of the exercise have been processed using the consensus values of all the analysts to form the basis of their final Z-scores. Since 2014, we are using the statistical software program ProLab Plus to calculate the descriptive statistics for the test and the performance characteristics including the graphical representation of all the results.

The performance statistics for the exercise have been calculated using ProLab Plus Version 2022.7.25.0. The summary table of all the Z-scores can be found in Annex IX of the annex report. The performance statistics (Annex XII) show the results by measurand and analyst of all the results for the test including the Z-scores and outliers, the statistical method used for the data (Q/Hampel), means and standard deviations, measures of repeatability and reproducibility for each measurand, number of participants and other relevant information on the test. The graphical summary for each measurand by analyst can be found in Annex XIII of the annex report.

For 2022, we used the Q/Hampel algorithm to calculate the Z-scores and Standard deviation for the test. This year we have used the SDPA calculated by the program to generate our Z-scores and Standard deviations for each measurand (Annex XII).

4.4.1 Z-scores

The quantitative Z-scores derived using the robust averages and standard deviations can be found in Annex IX. Any results in blue are within the specification of the test (+/-2SD). The yellow triangles indicate warning signals (outside +/-2SDs but inside +/-3SDs), red triangles indicate action signals (outside +/-3SDs). If the analyst failed to identify one or various species in the samples, these appear as 'Grey triangles' and a +3SD score. All qualitative scores are included for the final evaluation of analysts.

There were a very small number of warning and action signals across measurands. 3 Red flags (0.5%), 21 (3.4%) yellow flags and 8 (1.3%) non-detection flags (Grey triangles) from 616 results is evidence of good performance overall.

4 analysts weren't successful at the overall test from 77 returned results. Analysts 43, 47, 62 and 90 failed the quantitation of at least 3 or 4 items which requires training and improvement for the next round. 58 analysts had all the measurands (8) within the tolerance limits, 11 analysts had one failed measurand and 4 analysts two.

Quantitatively, The measurand *A.pacificum* appears to be not only the most difficult organism to count in the samples, with 6 yellow and 2 red flags, but also one of the most difficult to detect in the samples (3 non-detections) compared with *Coolia monotis* 3 yellow flags and 4 non-detections. Probably the second most difficult. There is evidence to suggests that some analysts that failed to identify one or the other, also obtained a yellow or red flag in the other cell count. The reason is that those that identified *A.pacificum* and failed to identify *Coolia* in the samples, probably counted all *Coolia* cells as part of their *Alexnadrium* cell count or viceversa.

4.5. Relative Laboratory Performance (RLP) and Rescaled Sum of Z-scores (RSZ) and Lischer plots

The chart of RLP against RSZ (Annex XIV) expresses some combination statistics from the test. This shows the sum of all the Z-scores for the test as a dot in a graph. Each dot represents one analyst and all their pooled results. RSZ is based on the standardized sum of all the z-scores for each analyst and it can be interpreted as a single Z-score: that is an evaluation across all samples and measurands. The position of the dot indicates whether the analyst is committing systematic laboratory bias. This is independent of a pass or fail for the test and only indicates whether the analyst results vary from the others significantly. The x axis gives a measure of the overall mean of all the results and the y axis measures the deviation of these results. The green area represents where analysts should be if there was no bias. A large bias to the right or left indicates that your mean Z-scores may be overestimated or underestimated according to the SDPA.

Laboratories dotted within the green colored area are within the values required to pass the test, but they still may show some bias. Those outside these areas are showing a systematic bias in their counting. Laboratories to the right of zero have an overall tendency to overestimate values and to the left to underestimate them which suggests some kind of methodology bias which should be explored, investigated and corrected by the laboratory themselves.

The RLP is the mean length of all the Z-scores for each analyst and is derived from the sum of the squared mean length of all the Z-scores. The height indicates whether your results reproducibility is good or not. Large standard deviations indicate greater variability in your counts.

The plots of repeatability standard deviations or Lischer plots (Annex XV in the annex report) are somewhat similar but measurands are plotted individually, instead of all combined. Here, you may be able to glean other problems more specific to the identification and counting of certain species. Perhaps, a tendency to underestimate particular species or group of species or have a particular difficulty with dinoflagellates or diatoms for example. These graphs show how you did compared to everyone else in a very interactive way.

It works in a similar way to the RLP plot but uses the 95% Confidence limit and 99% and 99.9% limits to indicate whether your score is within which level. This will give you an idea of your mean and repeatability standard deviation compared to the rest. Lischer plots, assume that the data is normally distributed and the null hypothesis is that there are no differences between the analyst means and standard deviations compared to the consensus at the 95% level of confidence (Green area). If there are differences, then your results will be outside of this green area. The spread of the data will show you how the distribution of the data looks for all the analysts. Results high into the y axis show poor repeatability among replicates and the x axis shows your mean compared to the robust means and that of the other analysts, that is how close your results are to the consensus mean.

4.6 Qualitative sample data

At least 75% of identification results must be correct to pass this particular test in conjunction with 75% of your quantitative results. Generally we use a 80% pass rate but this year there were only 8 measurands in the samples instead of the usual 10 or more. The identification of measurands in the samples are given a 'correct', 'incorrect' and 'non-detected' flag to the analysts. This parameter is an important component of this test and analysts must be able to recognize the species at least to genus level for all species.

Analyst performance on the correct composition of species in the samples was generally quite good (Table 3). To pass the qualitative test, analysts had to identify correctly at least 75% of the

measurands, that is at least 6 of the species in the samples. In 2022, all analysts passed the qualitative test. 51 analysts identified correctly all measurands. 12 analysts identified incorrectly 1 measurand, 4 analysts 2 measurands, 6 analysts had a non-detection and one incorrect identification, 3 analysts had one non-detection each and one analyst had two non-detections.

Analyst code	Alexandrium pacificum	Prorocentrum rhathymum	Actinoptychus splendens	Coolia sp.	Chaetoceros peruvianus	Guinardia striata	Lampriscus sp.	Synedropsis sp.
1	Correct	Correct	Correct	Correct	Correct	Incorrect	Correct	Correct
2	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
3	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
4	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
5	Correct	Correct	Correct	Incorrect	Correct	Correct	Correct	Correct
6	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
7	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Incorrect
8	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
9	Correct	Correct	Correct	Correct	Correct	Correct	Incorrect	Correct
10	Correct	Correct	incorrect	Correct	Correct	Correct	Correct	Correct
11	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
13	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Incorrect
14	Correct	Correct	incorrect	not detected	Correct	Correct	Correct	Correct
15	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
16	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Incorrect
17	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
18	Correct	Correct	Correct	Correct	Correct	Correct	Incorrect	Incorrect
19	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
20	Correct	Correct	Correct	Incorrect	Correct	Correct	Incorrect	Correct
21	Correct	Correct	Correct	Incorrect	Correct	Correct	Incorrect	Correct
22	Correct	Correct	Correct	Incorrect	Correct	Correct	Incorrect	Correct
23	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
24	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
25	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
26	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
27	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
28	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
29	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
30	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
31	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
32	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
33	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
34	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
35	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
36	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
37	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
43	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
44	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
45	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
46	Correct	Correct	Correct	Incorrect	Correct	Correct	Correct	Correct
47	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
50	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
51	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
52	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
53	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
54	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
55	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
56	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
57	Correct	Correct	Correct	Correct	Correct	Correct	Incorrect	Correct
58	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
59	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
60	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
61	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Incorrect
62	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
63	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
64	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
65	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
67	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
68	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
70	Correct	Correct	Correct	Correct	Correct	Correct	Incorrect	Correct
71	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
72	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
73	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
74	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
75	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
76	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
77	Correct	Correct	Correct	not detected	Correct	Correct	Correct	Correct
78	Correct	Correct	Correct	Correct	Correct	Correct	Incorrect	Correct
79	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
80	not detected	Correct	Correct	Correct	Correct	Correct	Correct	Incorrect
81	not detected	Correct	Correct	Correct	Correct	Correct	Correct	Incorrect
82	Correct	Correct	Correct	not detected	Correct	Correct	Correct	Correct
83	not detected	Correct	Correct	Correct	Correct	Correct	Correct	Correct
84	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
85	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
86	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
90	Correct	no detected	Correct	not detected	Correct	Correct	Correct	Correct

Table 3: Qualitative results IPI2022 by Analyst and Measurand. Not- detected (ND)

Actinoptychus splendens, Guinardia striata, Prorocentrum rhathymun and *Chaetoceros peruvianus* were identified easily by most analysts. There was only one incorrect identification for *G.striata* and two incorrect identifications for *A.splendens.* Generally most analysts were also able to identify these organisms to species level correctly for *G.striata* (71), *P.rhathymun* (72), if we consider *P.mexicanum* as synonym and also correct. For *A.splendens* 45 analysts identified correctly to species. For *C.peruvianus* 62 analysts were correct to species level and all analysts to genus level.

Dinoflagellates were generally more difficult to identify than diatoms in this test. The results show this trend with approximately 14 incorrect flags (yellow or red) for 3 dinoflagellate species compared to approximately 12 incorrect flags for all the diatoms (5) plus 8 non-detections for dinoflagellates and none for diatoms.

The hardest identification was Lampriscus sp with only 42 analysts identifying correctly to genus level. 28 other analysts used *Odontella* as the answer and this was given as correct because of the lack of literature in the genus and also because sometimes it is considered in the small amount of literature available for the genus as *Odontella* or *Biddulphia.* No special difficulties were found with the identification of all the other diatoms in the sample. All analysts detected all diatoms in the samples but not all identified correctly. There were 7 incorrect identifications for *Synedropsis* and *Lamprisucs* probably the two most challenging identifications. *Coolia* was the most difficult of the dinos with 5 incorrect identifications. Generally, *Gambierdiscus* was the most likely incorrect substitution, another benthic dinoflagellate.

There were no difficulties identifying *Prorocentrum* to genus level, but there were differences at species level. 20 analysts identified as *P.mexicanum* instead of *P.rhathymum.* Both are considered as synonyms in the current literature search, although there is evidence that there are most likely different species and should be considered separately.

Other species like *Alexandrium pacificum* was easily identifiable to genus level but not to species level, this is normal using light microscopy as the separating taxonomic characters must be viewed under Scanning Electron Microscopy. 74 analysts identified correctly to genus level but there were 3 analysts were not able to detect the species in the samples. *Coolia monotis* was also easily identified to genus level by most analysts (68), but there were 5 incorrect identifications and 4 non-detections.

Species	Identification
Guinardia striata	71
Guinardia sp.	5
NR	7
Not detected	0
Incorrect ID	1
Total analysts	84

Species	Identification
A.pacificum	**8**
A.catanella	24
A. tamarense	31
A. minutum	8
A. tamutum	1
A. mediterraneum	1
A. ostenfeldii	1
NR	7
Not detected	3
Incorrect ID	0
Total analysts	**84**

Species	Identification
Synedropsis sp.	**37**
Synedropsis hyperborea	27
Synedropsis laevis	6
Incorrect IDs	**7**
Fragillaria	4
Bacillaria	1
Thalassionema	1
Asterionellopsis	1
NR	7
Not detected	0
Incorrect ID	7
Total analysts	84

Species	Identification
Chaetoceros peruvianus	**62**
Chaetoceros (Phaeoceros) sp.	12
Chaetoceros (Hyalochates) sp.	1
Chaetoceros concavicornis	1
Chaetoceros danicus	1
NR	7
Not detected	0
Incorrect ID	0
Total analysts	84

Species	Identification
Actinoptychus splendens	**45**
Actinoptychus sp.	23
Actinoptychus senarius	4
Actinoptychus boliviensis	1
Actinoptychus glabratus	1
Actinoptychus helycopelta	1
NR	7
Not detected	0
Incorrect ID	2
Total analysts	84

Species	Identification
Prorocentrum rhathymum	**52**
Procentrum mexicanum	20
Prorocentrum compressum	1
Prorocentrum elegans	1
Prorocentrum lima	1
Prorocentrum sp.	1
NR	7
Not detected	1
Incorrect ID	0
Total analysts	**84**

Species	Identification
Coolia sp.	28
Coolia monotis	33
Coolia canariensis	1
Coolia malayensis	1
Coolia palmyrensis	1
Coolia tropicalis	2
NR	7
Not detected	4
Incorrect ID	5
Total analysts	**84**

Species	Identification
Lampriscus sp.	**27**
Lampriscus orbiculatum	8
Lampriscus shadboldtianum	6
Lampriscus fasciculatus	1
Odontella	28
Incorrect IDs	**7**
Guinardia flaccida	3
Bellerochea	2
Coscinodiscus	1
Mediopyxis helysia	1
Not detected	0
Total analysts	84

Table 4: Qualitative data by measurand.

Overall, from 616 possible correct identifications, there were a total of 586 correct answers at genus level (95%) and for the 4 species that could have been easily identified to species level (Actinoptychus*, Guinardia, Prorocentrum and* Chaetoceros) 250 correct answers from a possible 308 correspond to 81% correct. There were also 8 non-detections (1.3%) and 22 (3.6%) incorrect identifications.

4.7 Ocean Teacher 2022 online taxonomic assessment

The test itself consisted of 18 questions (see Annex XVI in the annex report) and annex XVII show the overall results and grades of the participants. There were 83 attempts at the OTGA assessment, the median overall grade was 92.6%. 73.5% of analysts performed above the proficiency threshold of 90% and 18.1% of all analysts between 80-90%. 7.2% above 70% and another 1.2% below 70% requiring improvement (Table 5).

Ocean Teacher IPI 2022 Exercise Results					
A. Code	Total (%)	A. Code	Total (%)	A. Code	Total (%)
25	100.0	33	96.8	21	91.6
67	100.0	40	96.8	18	91.2
70	100.0	29	96.6	14	90.5
84	100.0	11	96.5	59	90.4
10	99.5	13	96.0	19	90.2
16	99.5	41	96.0	73	89.9
20	99.5	42	96.0	26	89.8
35	99.5	3	95.9	46	89.8
75	99.4	64	95.9	9	89.5
86	99.4	65	95.8	43	88.5
15	99.2	34	95.5	23	87.3
27	99.2	45	95.1	60	85.8
30	99.2	6	94.8	66	85.1
53	99.2	63	94.7	76	84.8
37	98.9	50	94.4	57	84.6
4	98.7	51	94.4	78	84.4
17	98.4	52	94.4	22	83.5
36	98.4	74	93.7	7	83.2
58	98.2	38	93.6	80	81.3
28	98.1	55	93.2	44	81.2
31	98.1	72	93.2	5	79.7
8	97.9	71	93.1	79	77.7
85	97.9	54	92.9	90	77.4
39	97.6	47	92.7	62	75.9
32	97.3	83	92.5	56	75.8
2	97.2	24	92.0	82	75.7
1	97.0	68	92.0	81	68.7
77	96.9	61	91.7	Average	92.6

Table 5: Ocean Teacher scores by analyst code

Q#	Question type	Question name	Attempts	Facility index	Standard deviation	Intended weight	Effective weight
1	Matching	IPI2022 Alexandrium terminology	83	94.35%	12.49%	5.56%	5.65%
2	Drag and drop onto image	IPI 2022 Alexandrium chain vs non-chain formers	83	92.86%	7.21%	5.56%	4.00%
3	Matching	IPI 2022 Gymnodiniales acrobase	83	87.62%	16.40%	5.56%	5.75%
4	Multiple choice	General dinoflagellate life cycle 1 2013	83	89.29%	31.12%	5.56%	8.19%
5	Multiple choice	General dinoflagellate life cycle 2 2013	83	97.62%	15.34%	5.56%	4.03%
6	Drag and drop onto image	IPI 2022 Gambierdiscus Taxonomy	83	98.07%	6.72%	5.56%	3.02%
7	Drag and drop onto image	IPI 2022 Ostreopsis Tabulation	83	85.29%	21.56%	5.56%	7.23%
8	Matching	Chaetoceros terminology 1 IPI2017	83	97.62%	7.86%	5.56%	3.32%
9	Multiple choice	Thalassiosira characters IPI2018	83	89.37%	22.41%	5.56%	8.38%
10	Multiple choice	Thalassiosiraceae family IPI2018	83	92.92%	19.88%	5.56%	8.31%
11	Numerical	Enumeration 2 IPI 2017	83	66.67%	47.42%	5.56%	13.42%
12	Numerical	Enumeration 3 IPI 2017	83	98.81%	10.91%	5.56%	4.72%
13	Drag and drop onto image	IPI 2022 Dinoflagellates and their toxins	83	98.69%	3.73%	5.56%	2.60%
14	Drag and drop onto image	IPI 2022 Toxic Peridiniales or Gonyaulacales	83	96.67%	9.10%	5.56%	4.47%
15	Numerical	Enumeration 5 BEQ15	83	100.00%	0.00%	5.56%	0.00%
16	Numerical	Enumeration 7 BEQ15	83	98.81%	10.91%	5.56%	3.81%
17	Matching	IPI 2022 Prorocentrum - diagnostic features	83	93.71%	10.00%	5.56%	5.79%
18	Matching	IPI 2022 Prorocentrum - identification of benthic species	83	84.52%	18.45%	5.56%	7.32%

Table 6: Facility index IPI2021 OT exercise

The OTGA facility index shows that the worst answered question in the test was Q11 (66.7%) a numerical question and the best Q15 (100%) another numerical question (Table 6).

The breakdown of scores per question can be found in Annex XVI of the annex report.
Q1 was based on Alexandrium taxonomic terminology using a line diagram of *Alexandrium* in dorsal, ventral, apical and antapical views plus details of the Apical pore complex (APC) and sulcal area. The average score was 94.35% (table 6) and generally there was good consensus between analysts for all taxonomic characters. The posterior attachment pore (pap) answer was a little bit below the average (85%) compared to the other characters (92-100%) (Annex XVI).

Q2 continued around the *Alexandrium* theme and an image plate showing different *Alexandrium* species with their corresponding species name and images mainly taken in SEM show particular species' characters. The question was to choose whether the species depicted here belonged to a chain or a non-chain forming *Alexandrium* species. There were 11 images with 6 chain forming species and 4 non-chain forming. The answers were for the chain formers: *A. compressum, affine, fraterculus, tamiyavanichi, catenella & pacificum* and for the N-chain formers: *A.kutnerae, leei, insuetum, gaardnerae & tamarense*. The consensus was high among analysts for most species and the average was 92.8% (Table 6). The most erroneous answers for the chain formers was *A.pacificum* with only 85% correct and for the N-chain formers *A.kutnerae* 54% correct only, well below the average for the question (Annex XVI).

Q3 showed an image plate depicting a number of naked dinoflagellates belonging to the gimnodiniales. The gimnodiniales are a group of organisms particularly difficult to identify and one special diagnostic feature for this group is the 'acrobase'. The 'acrobase' or apical groove has a different shape and size depending on the genus and these can be recognised at least to genus level. The analysts were given a list of descriptions for the 'acrobase' that they had to match to the respective genera. The average score for this question was 87.6% slightly lower than the average (Table 6). The 'acrobase' for most genera was actually answered without problems, except for *Karlodinium*. The model answer was given as 'apical groove is straight' (53) but it was argued that 'apical groove descends dorsally' (11) could also apply here. The advisory group decided that this answer should be also given as correct. So, Annex XVI shows the upgraded results and the final marks have been upgraded for these 11 analysts.

Q4 & 5 were two multiple choice questions with built-in penalty percentages for each erroneous answer. Q4 is an image depicting a *Dinophysis* cell with two flagella. There were 5 options to this

answer and the correct one was 'planozygote' which is effectively a mobile diploid zygote hatched from a cyst. In this stage, the cell bears two longitudinal flagella whereas vegetative cells have only one. Planozygote will in time divide two produce two daughter cells with just one longitudinal flagellum each. 8 analysts chose 'vegetative cell' which is incorrect. In Q5, the image depicts again two *Dinophysis* cells fusing during sexual reproduction. When the pair are exactly the same size this is called 'isogamy' which was the right answer. The average for Q4 was 89% compared to 97% for Q5.

Q6 & 7 were drag and drop questions based on benthic dinoflagellates taxonomic nomenclature. Q6 depicted both SEM images and a line diagram of *Gambierdiscus* in dorsal and ventral view and Q7 of *Ostreopsis*. There are two modern conflicting alternatives for the kofoidean tabulation of benthic armoured dinoflagellates. One version uses the convention by Fraga + Besada (4' 6" 5"' 2"") and the other uses Faust, Chinain and Litaker (3' 7" 5"' 2"" 1p). The idea is that analysts are aware of this dichotomy and were able to differentiate between these different patterns.

Analysts had no issues answering Q6 (following Faust, Chinain and Litaker), there were only a few mistakes in relation to the 2"" where 10 analysts chose the Sp plate instead. In Q7 (following Fraga + Besada), analysts had more problems. 57 analysts (68%) were correct choosing the 4' tabulation whereas the other 27 analysts chose the 3' tabulation. This meant a domino effect to the epithecal tabulation with errors and carrying over the whole plate pattern. Otherwise, there were no issues in relation to the hypothecal plate pattern. Therefore the average mark for Q6 was 98% to Q7 drop to 85% (Table 6).

Q8 shows a line diagram of a chain forming chaetoceros and analysts were asked to name the different diagnostic characters for these diatoms. The average score was 97.6% and there were no clear difficulties with the terminology.

The multiple choice question 9 depicted an image of a Thalassiosira chain and analysts were asked which statements were true for the family Thalassiosiraceae. The average score was 89%. There were 4 correct answers to this question (a, c, e & h). One answer (e) 'cell wall have a marginal ring of smaller labiate processes' was only given as correct by 53 analysts (63%) compared to 88 to 96% for the other 3. 10 analysts ticked option b) 'they are unipolar centrics' and 6 further option f) 'internal foramina external criba' which were wrong and deducted 20% of the score for these analysts.

Q10 was another multiple choice question following on the Thalassiosira theme and asking about which species in the images belonged to the order thalassiosirales. There were 4 correct answers, and the average score was high (92.9%) and no particular difficulties here.

Q11-12-15 &16 were numerical type questions and only required of analysts to count the cells visible in the images provided of dinoflagellates and diatoms. Q11 was a chain of *Thalassiosira* cells. The consensus answer was 9 +/- 1 cell, 56 analysts were within the model response given for the question, with 28 analysts giving other answers. 24 analysts gave 13 cells as their answer showing that there are differences in counting chain forming diatoms. There were no such issues with Q12 (*Asterionellopsis glacialis* chain), Q15 (*Paralia sulcata*) with a perfect score or Q16 (*Scrippsiella*). The average score for Q11 was 67%, the worst for all the questions, 98.8% (Q12), 100% (Q15) and 98.8% (Q16) (Table 6 and Annex XVI).

Q13 & Q14 are two drag and drop questions where objects must be placed in the right place holders. Q13 asked analysts to match the toxin compound to the right toxin producing dinoflagellate. There were 10 images and 10 draggable items. The average score was high 98.7% and most answers correct, the only difficulty was matching spirolides to *A.ostenfeldii* in this group (See Annex XVI) for all answers. Q14 follows Q13 on toxic producing dinoflagellates and asks analysts to place the species in the right order the peridiniales or the goniaulacales based on one character that differentiates both genera, the 'x plate'. Analysts did not find any difficulties answering this question with high average score of 96.7% and a small number of incorrect answers generally around *A.languida* and *A.spinosusm* which are 'incertae sedae' and not placed in any particular order at present, however, based on the 'x plate' which both have, should have been included in the Peridiniales group for this question.

Finally, Q17 & 18 were built around the *Prorocentrum* genera. Q17 was a matching question and analysts were asked to choose the character shown in the photos. The average score was 93.7% in Q17 (Table 6) and the only major issue was with number 6: 'Areolae' which was given correct by 62 people with 18 choosing 'thecal plate'. For Q18, analysts were asked to identify the species depicted in the images. Generally there were no major issues for the majority of the images, however, *P.sipadanensis* was confused for *P.lima* by 12 analysts. *P.concavum* for *P.leve* by 14 and *P.hoffmanianum* for various different species by 23 analysts (See annex XVI). The average score was 84.5% for this question (Table 6).