United Nations Educational, Scientific and Cultural Organization

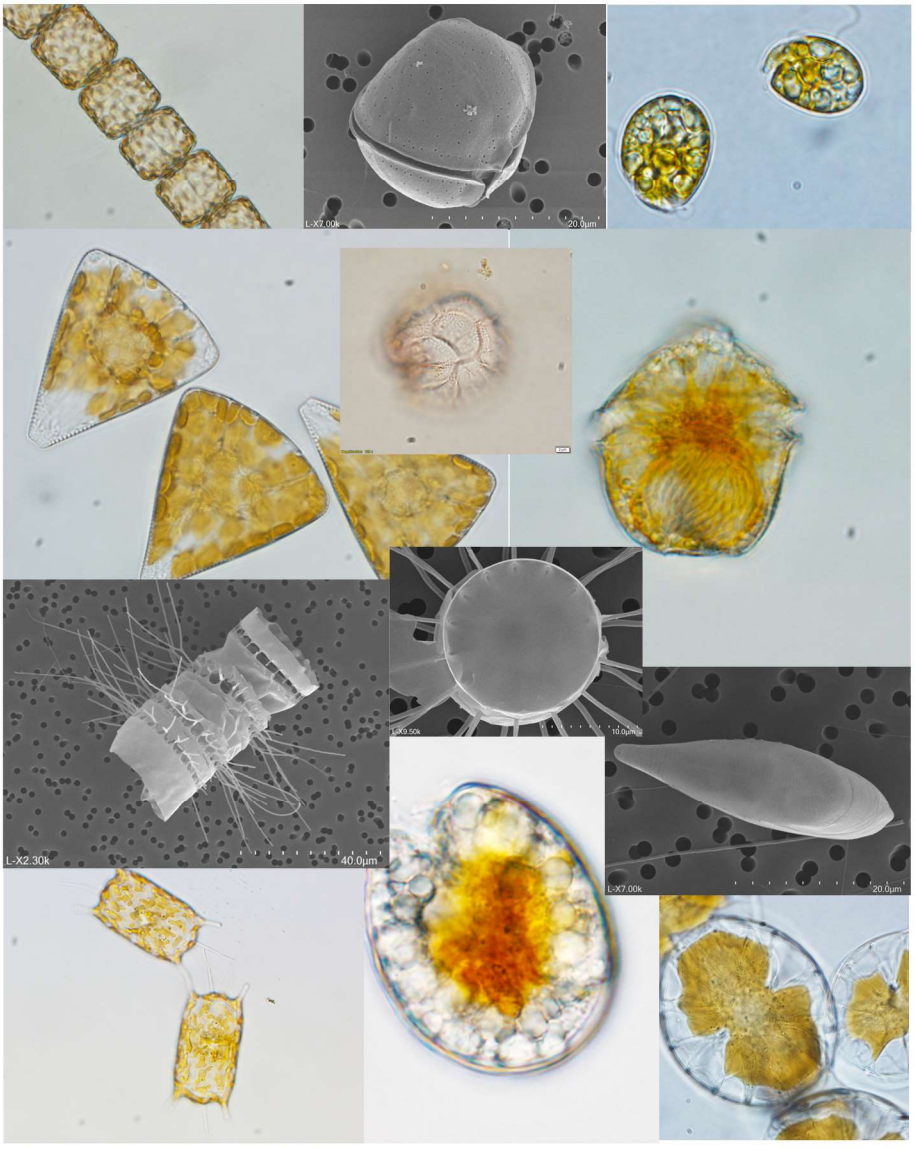
Intergovernmental Oceanographic Commission

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 2023 report



Rafael Salas[1] & Jacob Larsen[2]

[1] Observatorio Canario de Algas Nocivas (OCHABS), Calle Miramar 121, 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 2023, 79 analysts across 43 laboratories around the world participated in the IPI2023 exercise. 76 analysts returned sample results and 73 completed the Oceanteacher online test. 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.

- The samples were spiked with nine species; five diatoms and four dinoflagellates.

- The dinoflagellates were ***Prorocentrum emarginatum*** Y.Fukuyo, 1981, ***Gonyaulax hyalina*** Ostenfeld & Schmidt, 1901, ***Coolia monotis*** Meunier, 1919 and ***Amphidinium carterae*** Hulburt, 1957.

- The Diatom species were ***Amphiprora hyalina*** Greville, 1865, ***Trieres mobiliensis*** M.P.Ashworth & E.C.Theriot, 2013, ***Lauderia annulata*** Cleve, 1873, ***Bacteriastrum*** G. Shadbolt, 1854, and ***Licmophora*** C.A. Agardh, 1827.

- The cell counts of ***Prorocentrum emarginatum*** were deemed null and void for the purpose of this intercomparison due to homogenization issues. The species were embedded in an organic matrix and did not mix properly.

- 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.

- There were a very small number of warning and action signals across measurands for the quantification results. 14 Red flags (2.3%), 22 (3.7%) yellow flags and 4 (0.7%) non-detection flags (Grey triangles) from 600 results is evidence of good performance overall.

- 6 analysts did not pass the test from 75 returned results. Analysts 5, 118 and 124 failed 3 out of 8 results, analysts 114 and 173 4 out of 4 and analyst 138 5 and 3. 53 analysts had all the measurands (8) within the tolerance limits, 13 analysts had one failed measurand and 3 analysts two.

- The hardest identification for the participants was *Gonyaulax hyalina* with 22 incorrect and 1 non-detected flag. There were 4 non-detections in total, a very small number, *A.carterae* wasn't detected by 3 analysts.

- There was no difficulty overall with identification for any of the species, even with *G.hyalina* at low abundances in the samples, was identified by most participants.

- In 2023, most analysts passed the qualitative test except for one analyst (181) with 2 incorrect and one non-detected flag. 53 analysts identified correctly all measurands (8). 13 analysts identified incorrectly 1 measurand and 3 analysts 2 measurands.

- Overall, from 600 possible correct identifications, there were a total of 562 correct answers at least to genus level (93.7%), 34 incorrect identifications (5.7%) and 4 non-detections (0.7%).

- There were 73 attempts at the OceanTeacher assessment, the median overall grade was 90.3%. 60.8% of analysts performed above the proficiency threshold of 90% and 27.0% of all analysts between 80-90%. 5.4% above 70% and another 5.4% below 70% requiring improvement.

- The OTGA facility index shows that the worst answered question in the test was Q15 (65.75%) a numerical question and the best Q13 (98.51%) a matching 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.

Nine species were used in the IPI2023 exercise. The samples were spiked with nine species; five diatoms and four dinoflagellates. The dinoflagellates were *Prorocentrum emarginatum* Y.Fukuyo, 1981, *Gonyaulax hyalina* Ostenfeld & Schmidt, 1901, *Coolia monotis* Meunier, 1919 and *Amphidinium carterae* Hulburt, 1957. The Diatom species were *Amphiprora hyalina* Greville, 1865, *Trieres mobiliensis* M.P.Ashworth & E.C.Theriot, 2013, *Lauderia annulata* Cleve, 1873, *Bacteriastrum* G. Shadbolt, 1854, and *Licmophora* C.A. Agardh, 1827.

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 2023, 79 analysts across 43 laboratories around the world participated in the IPI2023 exercise. 76 analysts returned sample results and 73 completed the Oceanteacher online test. European countries including the UK accounted for 72% of the total participation, 6% came from South America, 10% from African countries, 9% from Oceania and 3% from Asia. (Figure 1). 19 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.

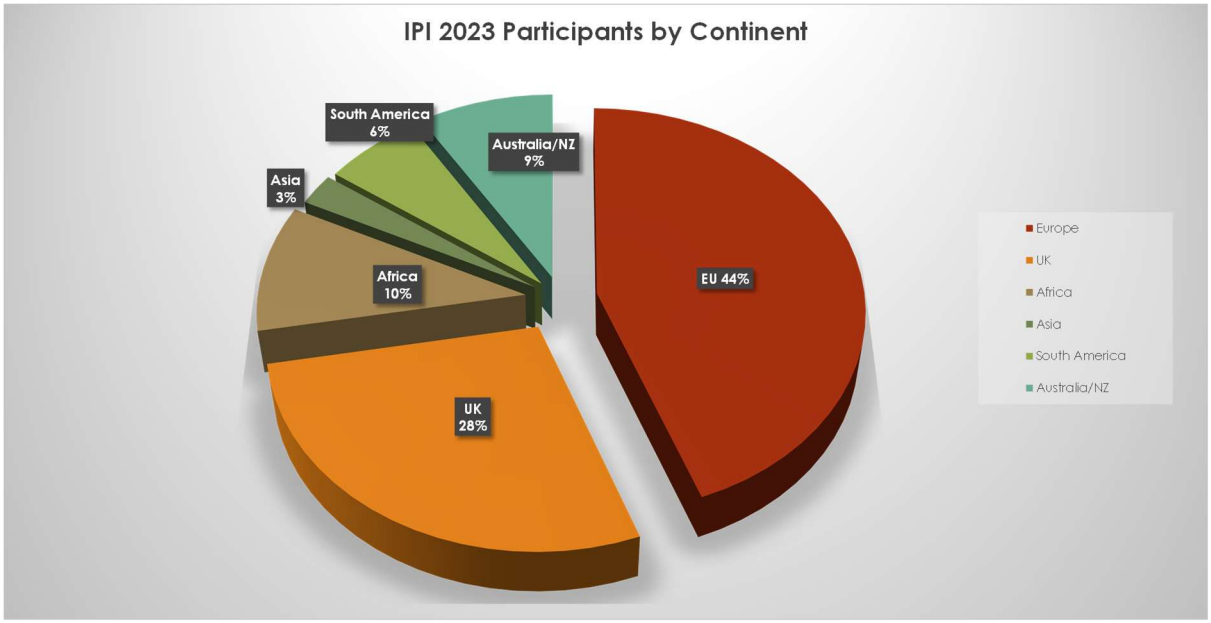


Figure 1: Participants by continent IPI2023

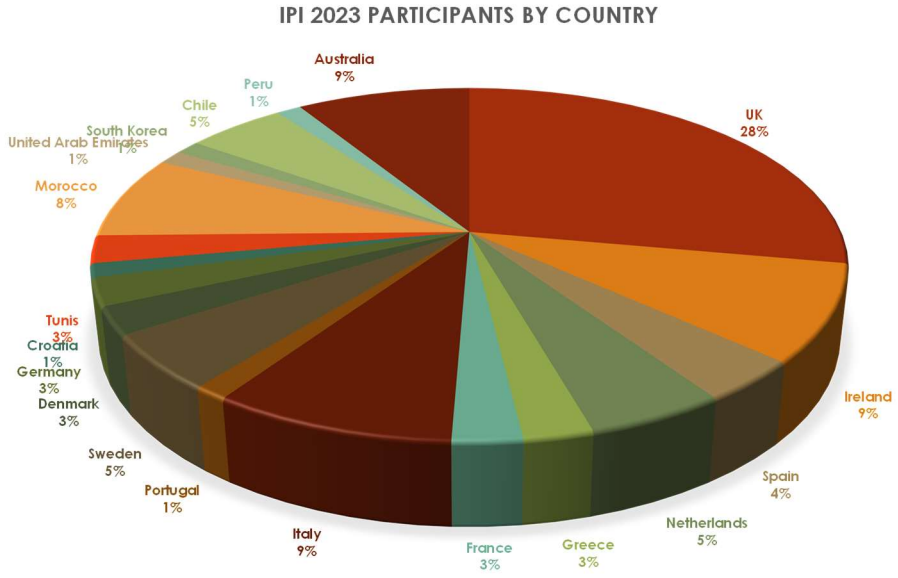


Figure 2: Participants by country IPI 2023

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).

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 Taliarte pier, Gran Canaria, Spain and it was filtered through 47mm GF/C Whatmann filters (Whatmann™, Kent, UK) and autoclaved (Systec V100, Wettenberg, Germany) and preserved using neutral Lugol's iodine solution (Clin-tech, Dublin, Ireland).

The materials were produced from several 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 (IPI2023 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. 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. 79 analysts in 43 laboratories were sent a total of 316 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 Microsoft Excel.

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 followed 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 20 questions (see Annex XVI). Question types used in the quiz were 'matching type' (Q4-7-8-11-13-14-16) which have dropdown menus including a selection of answers that analysts must choose from, 'multiple choice' (Q17-19-20) 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 (Q3-6-9-12-15) where analysts had to count the cells in the videos provided and 'drag and drop' types (Q1-2-10) 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, if participants don't press submit.

4. Results

4.1 Homogeneity and stability study

The homogeneity and stability test in 2023 included 8 measurands (Table 1) and all of them except for *C.monotis* 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.

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
Amphidinium carterae	no outliers found	Not OK	Not OK	Ok	Ok	Ok
Amphiprora hyalina	no outliers found	Ok	Ok	Ok	Ok	Ok
Bacteriastrum furcatum	no outliers found	Ok	Ok	Ok	Ok	Ok
Coolia monotis	no outliers found	Not OK	Not OK	Not OK	Not OK	Ok
Gonyaulax hyalina	no outliers found	Ok	Ok	Ok	Ok	Ok
Lauderia annulata	no outliers found	Not OK	Not OK	Ok	Ok	Ok
Licmophora gracilis	no outliers found	Not OK	Not OK	Ok	Not OK	Ok
Prorocentrum emarginatum	no outliers found	Ok	Not OK	Ok	Ok	Ok
Trieres Mobiliensis	no outliers found	Not OK	Ok	Ok	Ok	Ok

Table 1: IPI2023 Homogeneity and stability results according to ISO13528:2015

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 (Annex VI). 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 *Amphiprora, Bacteriastrum, G.hyalina* and *Trieres obiliensis* (table 1). However, no significant heterogeneity was found according to the expanded criterion except for *Coolia monotis.*

The within sample difference of the homogeneity cell counts for *Coolia monotis* 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 *Coolia monotis*. Hence, the proficiency test items cannot be considered fully homogeneous but not significantly heterogeneous (Table 1) except for *Coolia monotis.*

As most analysts achieved good Z-scores for *Coolia monotis*, there was no need to disregard this results. 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. (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.

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

B)

Where;

u_x= Standard uncertainty of the assigned value,

s^*= robust standard deviation for the test

p= number of analysts

Species	A.hyalina	Licmophora	L.annulata	Bateriastrum	T.mobiliensis	A.carterae	C.monotis	P.emarginatum	G.hyalina
Robust mean x*	7836	16026	13819	11121	7447	12779	8165	238	506
Robust Stdev s*	1933	2010	3356	2482	2632	4477	2243	157	191
Standard Ux	279	290	484	358	380	660	324	27	28
n=	75	75	75	75	75	72	75	53	74
if Ux < 0.3xSTdev	580	603	1007	745	790	1343	673	47	57
then Ux is negligible	neg	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**	4936	13011	8785	7398	3499	6063	4800	2	220
*x *+1.5s**	10735	19040	18853	14843	11395	19494	11530	473	793
Homogeneity test	A.hyalina	Licmophora	L.annulata	Bateriastrum	T.mobiliensis	A.carterae	C.monotis	P.emarginatum	G.hyalina
Reference value mean	10396	17422	15948	10228	9608	20928	10786	838	604
Reference value stdev	419	1006	1511	260	1035	2373	948	101	94
	Comparison with assigned value								
	A.hyalina	Licmophora	L.annulata	Bateriastrum	T.mobiliensis	A.carterae	C.monotis	P.emarginatum	G.hyalina
*x *-X*	2560	1396	2129	893	2161	8149	2621	600	98
Uncertainty of diff.	395	410	685	507	537	933	458	38	39
2* Uncertainty of diff.	789	820	1370	1013	1074	1865	916	76	78
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 2023, 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 blank in the Z-scores graphs and they are reported as a +3SD score in the individual statements of performance for the test. All qualitative scores are included for the final evaluation of analysts.

There were a very small number of warning and action signals across measurands for the quantification results. 14 Red flags (2.3%), 22 (3.7%) yellow flags and 4 (0.7%) non-detection flags (Grey triangles) from 600 results is evidence of good performance overall.

6 analysts did not pass the test from 75 returned results. Analysts 5, 118 and 124 failed 3 out of 8 results, analysts 114 and 173, 4 out of 4 and analyst 138, 5 and 3. 53 analysts had all the measurands (8) within the tolerance limits, 13 analysts had one failed measurand and 3 analysts two.

Quantitatively, The measurand *G.hyalina* was the most difficult organism to count in the samples, with 6 red flags, but also one of the most difficult to identify in the samples (22 incorrect ids). It is possible that some *Coolia* cells were counted as *Gonyaulax* whose cell density was low in the samples. The robust average was approximately 500 cells/L or 10 cells in a 50 ml sample. The *Licmophora* count also caused quantification issues among analysts with 4 red and 2 yellow flags and *Bacteriastrum* 3 red flags.

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 a particular type species or group of species: 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 the identification and quantitation results must be correct to pass this test. 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 2023, all analysts passed the qualitative test except for one. 52 analysts identified correctly all measurands. 13 analysts identified incorrectly 1 measurand, 9 analysts 2 measurands and one analyst 3 measurands. There were 4 non-detections, 3 of them on *Amphidinium* and 35 incorrect identifications across the 8 measurands.

Dinoflagellates were generally more difficult to identify than diatoms in this test. The results show this trend with 27 incorrect flags for 3 dinoflagellate species compared to 8 incorrect flags for 5 diatoms. Also, 4 non-detections for dinoflagellates and none for diatoms (Table 3).

The hardest identification was *G.hyalina* with only 51 analysts identifying correctly to at least genus level. No special difficulties were found with the identification of the diatoms in the sample. All analysts detected all diatoms in the samples but not all were identified correctly.

Analyst code	Amphiprora	Licmophora	Lauderia	Bacteriastrum	Trieres	Amphidinium	Coolia	Gonyaulax	Overall Flag	Number of Measurands correct
2	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
3	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
4	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Incorrect	Pass	7
5	Correct	Correct	Correct	Correct	Correct	Correct	Incorrect	Incorrect	Pass	6
8	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
9	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
12	Correct	Correct	Correct	Correct	Correct	Correct	Incorrect	Correct	Pass	7
14	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
18	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
19	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
24	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Incorrect	Pass	7
25	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
27	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
31	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
33	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
35	Correct	Correct	Incorrect	Correct	Correct	Correct	Correct	Incorrect	Pass	6
38	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
41	Correct	Correct	Incorrect	Correct	Correct	Correct	Correct	Incorrect	Pass	6
42	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
43	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
44	Correct	Correct	Correct	Correct	Correct	nd	Correct	Correct	Pass	8
49	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
51	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
55	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
58	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
61	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
67	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
73	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
74	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Incorrect	Pass	7
76	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
81	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
82	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Incorrect	Pass	7
86	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
92	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Incorrect	Pass	7
93	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
96	Correct	Correct	Incorrect	Correct	Correct	Correct	Correct	Incorrect	Pass	6
97	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Incorrect	Pass	7
99	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
103	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Incorrect	Pass	7
110	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
Analyst code	Amphiprora	Licmophora	Lauderia	Bacteriastrum	Trieres	Amphidinium	Coolia	Gonyaulax	Overall Flag	Number of Measurands correct
112	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
114	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
116	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
118	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
124	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
126	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Incorrect	Pass	7
129	Correct	Correct	Correct	Correct	Correct	Incorrect	Correct	Incorrect	Pass	6
131	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
134	Correct	Correct	Correct	Correct	Correct	Correct	Incorrect	Incorrect	Pass	6
136	Correct	Correct	Correct	Correct	Correct	Correct	Correct	nd	Pass	7
138	Correct	Correct	Correct	Correct	Correct	nd	Correct	Incorrect	Pass	7
140	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
143	Correct	Correct	Incorrect	Correct	Correct	Correct	Correct	Incorrect	Pass	6
144	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
145	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
146	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
147	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
148	Correct	Correct	Incorrect	Correct	Correct	Correct	Correct	Incorrect	Pass	6
151	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
157	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
159	Incorrect	Correct	Correct	Correct	Correct	Correct	Correct	Incorrect	Pass	6
160	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
163	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
168	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
169	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
170	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Incorrect	Pass	7
171	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
173	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Incorrect	Pass	7
176	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
179	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
181	Correct	Correct	Incorrect	Correct	Correct	nd	Correct	Incorrect	Fail	5
182	Correct	Correct	Incorrect	Correct	Correct	Correct	Correct	Incorrect	Pass	8
183	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
185	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
189	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Pass	8
No results	0	0	0	0	0	0	0	0	**0**	
Non-Detected	0	0	0	0	0	3	0	1	**4**	
Total Correct	74	75	68	75	75	71	72	51	**561**	
Total Incorrect	1	0	7	0	0	1	3	23	**35**	
Total Results	**75**	**75**	**75**	**75**	**75**	**75**	**75**	**75**	**600**	

Table 3: Qualitative results IPI2023 by Analyst and Measurand. Not- detected (ND)

Most measurands were identified easily. *Licmophora, Bacteriastrum* and *Trieres* were correctly identified by all analysts. There was only 1 incorrect identification for *Amphiprora/Thalassiophysa* and 3 incorrect identifications for *Coolia.* However, *Lauderia* was incorrectly identified by 7 analysts and *G.hyalina* by 23 analysts. The most common incorrect identifications for *Gonyaulax hyalina* were *Lingulodinium* (9) and *Protoceratium* (6) and for *Lauderia* was *Detonula* (Table 4).

There was a ninth measurand in the sample, the dinoflagellate *Prorocentrum emarginatum* which was included in the samples at low cells densities. There were difficulties with the homogenisation of these species tending to bunch together in organic matrix and hyaline spheres in twos and fours, which made it very difficult to homogenise properly even after fixation. These species were disregarded for the purpose of this test both quantitatively and qualitatively even if they appear in the samples. These results have not been considered for this test.

Species	Identification
Amphiprora hyalina	74
Incorrect ID	1
Total analysts	75
Incorrect IDs	
Amphora lineolata	1
Total analysts	1

Species	Identification
Bacteriastrum furcatum	51
Bacteriastrum shadbolti	11
Bacteriastrum criophilum	5
Bacteriastrum tenue	1
Bacteriastrum elegans	5
Bacteriastrum sp.	2
Total analysts	75

Species	Identification
Coolia monotis	31
Coolia sp.	38
Coolia malayensis	2
Coolia tropicalis	1
Incorrect ID	3
Total analysts	75
Incorrect IDs	
Prorocentrum concavum	2
Alexandrium sp.	1
Total analysts	3

Species	Identification
Licmophora sp.	39
Licmophora abbreviata	25
Licmophora paradoxa	5
Licmophora gracilis	6
Total analysts	75

Species	Identification
Trieres mobiliensis	57
Odontella sp.	13
Odontella longicruris	3
Odontella aurita	2
Total analysts	75

Species	Identification
Gonyaulax hyalina	40
Gonyaulax sp.	11
non detected	1
Incorrect ID	23
Total analysts	75
Incorrect IDs	
Lingulodinium polyedrum	9
Protoceratium reticulatun	6
Gyrodinium instriatum	3
Gymno/Gyrodinium sp.	4
Alexandrium catanella	1
Total analysts	23

Species	Identification
Lauderia annulata	63
Lauderia sp.	2
Lauderia borealis	2
Lauderia pumila	1
Incorrect ID	7
Total analysts	75
Incorrect IDs	
Thalassiosira sp.	2
Detonula sp.	5
Total analysts	7

Species	Identification
Amphidinium carterae	56
Amphidinium sp.	15
non detected	3
Incorrect ID	1
Total analysts	75
Incorrect IDs	
Prorocentrum rhathymum	1
Total analysts	1

Table 4: Qualitative data by measurand.

Overall, from 600 possible correct identifications, there were a total of 561 correct answers at genus level (93.5%), 34 incorrect identifications (5.7%) and 4 non-detections (0.7%).. Four measurands were identified correctly to species level by most analysts, these were: *Amphiprora* (99%), *Amphidinium* (74.6%), *Trieres* (76%) and *Lauderia* (84%)) with 250 correct answers from a possible 300 correspond to 83.33% correct to highest taxon. *Bacteriastrum* and *Licmophora,* were difficult to bring to species level. For *Licmophora,* 39 analysts preferred to identify as 'sp' but there

was consensus among 25 analysts that should be *L.abbreviata*. For *Bacteriastrum,* the consensus was *B.furcatum* by 51 analysts and for Coolia 38 analysts preferred to identify to genus only. *Gonyaulax* was also identified to species level by 40 analysts with 11 going to genus level only.

4.7 Ocean Teacher 2023 online taxonomic assessment

The test itself consisted of 20 questions (see Annex XVI in the annex report) and annex XVII shows the overall results and grades of the participants. There were 73 attempts at the OceanTeacher assessment, the median overall grade was 90.3%. 60.8% of analysts performed above the proficiency threshold of 90% and 27.0% of all analysts between 80-90%. 5.4% above 70% and another 5.4% below 70% requiring improvement (Table 5).

Analyst code	%	Analyst code	%	Analyst code	%	Analyst code	%
14	100	116	97.3	160	90.8	4	84.2
112	100	146	97.3	38	90.7	41	82.8
171	100	110	97	181	90.4	182	82.6
176	100	96	96.9	82	90.3	44	82.5
19	99.6	2	96.8	Average	90.3	25	81.9
31	99.6	33	96.7	131	90.1	103	81.9
76	99.6	58	95.8	144	89.6	143	77.3
173	99.6	49	95.7	126	89	134	76.5
9	99.5	61	95.2	151	88.7	136	76.3
92	99.5	140	95.1	168	88.5	145	74.3
124	99	8	95	129	88.3	185	67.7
5	98.8	157	95	159	87.5	30	62.6
51	98.8	189	95	179	87.4	138	62
183	98.8	18	94.6	27	87.2	163	52.3
3	98.5	21	93.9	67	86.7		
86	98.2	35	93.2	148	85.8		
97	98	36	92.8	42	85.1		
147	97.8	155	92.8	74	85		
170	97.8	169	92.1	81	85		
114	97.3	118	90.8	24	84.8		

Table 5: Ocean Teacher IPPI2023 scores by analyst code

The OTGA facility index shows that the worst answered question in the test was Q15 (65.75%) a numerical question and the best Q13 (98.51%) a matching question (Table 6).

Q#	Question type	Question name	Attempts	Facility index	Standard deviation	Random guess score	Intended weight	Effective weight	Discrimination index	Discriminative efficiency
1	Drag and drop onto image	IPI 2023 Flagellates 1	73	85.75%	18.92%	10.00%	5.00%	4.74%	33.56%	34.63%
2	Drag and drop onto image	IPI 2023 Flagellates II	73	88.77%	13.33%		5.00%	4.53%	49.38%	50.99%
3	Numerical	IPI 2023 Numerical I	73	95.89%	19.99%	0.00%	5.00%	4.97%	34.89%	55.65%
4	Matching	IPI 2023 Diatoms Centric versus Pennate	73	97.15%	6.69%	50.00%	5.00%	3.55%	64.61%	72.73%
5	Matching	IPI 2023 Diatoms Radial-multipolar-araphid-raphid	73	87.44%	15.29%	25.00%	5.00%	5.25%	58.73%	59.83%
6	Numerical	IPI 2023 Numerical II	73	97.26%	16.44%	0.00%	5.00%	3.42%	17.01%	29.03%
7	Matching	IPI 2023 Diatoms colony formation types	73	92.81%	17.54%	12.50%	5.00%	5.76%	61.47%	70.67%
8	Matching	IPI 2023 Diatoms life cycle	73	91.23%	20.54%	20.00%	5.00%	5.70%	48.06%	55.57%
9	Numerical	IPI 2023 Numerical III	73	97.26%	16.44%	0.00%	5.00%	2.61%	6.30%	10.86%
10	Drag and drop onto image	IPI 2023 Life cycle Dinoflagellates	73	96.03%	10.10%	10.00%	5.00%	3.87%	48.58%	58.71%
11	Matching	IPI 2023 Dinoflagellates Complex organelles	73	93.74%	14.28%	14.29%	5.00%	4.38%	41.56%	48.07%
12	Numerical	IPI 2023 Numerical IV	73	95.89%	19.99%	0.00%	5.00%	3.91%	17.18%	26.35%
13	Matching	IPI 2023 Gonyaulacales Toxic	73	98.51%	2.96%	50.00%	5.00%	2.01%	47.17%	59.10%
14	Matching	IPI 2023 Intro Dinoflagellates	73	96.71%	12.48%	20.00%	5.00%	4.05%	41.22%	54.69%
15	Numerical	IPI 2023 Numerical V	73	65.75%	47.78%	0.00%	5.00%	8.97%	40.04%	53.39%
16	Matching	IPI 2023 Kofoidean tabulation	73	93.84%	17.20%	7.14%	5.00%	5.64%	59.98%	70.74%
17	Multiple choice	IPI 2023 Raphidophytes	73	83.29%	22.83%		5.00%	5.84%	43.57%	44.89%
18	Numerical	IPI 2023 Numerical VI	73	79.45%	40.68%	0.00%	5.00%	7.88%	36.73%	46.31%
19	Multiple choice	Protoperidinium 2	73	84.93%	36.02%	10.00%	5.00%	6.49%	24.44%	31.32%
20	Multiple choice	Protoperidinium 3	73	84.93%	36.02%	10.00%	5.00%	6.43%	23.64%	29.55%

Table 6: Facility index IPI2023 OT exercise

The breakdown of scores per question can be found in Annex XVI of the annex report.
Q1 + 2 were drag and drop questions using the same line diagram depicting flagellate groups. Analysts were asked in Q1, at which group these figures belong based on the flagellar arrangement. In Q2 analysts had to decide which of these groups have members able to produce toxins (See Annex XVI Q1 + 2).

Most analysts were able to distinguish the shapes of the different flagellate groups, especially dinoflagellates, euglenophytes, raphidophytes, prymnesiophytes and prasinophytes. Xantophytes and chrysophytes created more difficulties with 23 analysts (30.6%) thinking that xantophytes were chrysophytes and viceversa but also eustigmatophytes and raphidophytes. Both xantophytes and chrysophytes are very similar to each other, with a trailing flagellum with flagellar swelling and an anteriorly positioned flagellum with hairs, the only main difference is in their shape with Chrysophytes having a more triangular shape and more asymmetrical than xantophytes.

For Q2, most analysts recognised the which groups represent toxic species. Not all species within a group must produce toxins to be considered a toxic group. Dinoflagellates were identified correctly by all, but 5 and 10 analysts failed to recognise prymnesiophytes and raphidophytes as having toxic species within their groups.

Questions 3, 6, 9, 12, 15 and 18 were numerical type questions. In each a video clip shows a moving transect containing cells which must be counted. These were given as links to YouTube

to watch the video clips. The diatoms *Lauderia (*Q3), *Amphiprora* (Q6), *Licmophora* (Q15), *Trieres* (Q18) and the dinoflagellates *Coolia* (Q9) and *Amphidinium* (Q12) were used for this purpose. All of them were well resolved by the analysts except for Q15 *Licmophora* count. The scores for Q3, 6, 9 and 12 were in the high 90s (table 6) except for Q15 (65%) and Q18 (79.5%). It is interesting to note that all the species featured in the test were also part of the counting exercise in the samples. *Licmophora* as reported above in the quantification results was one of the diatoms with a larger out of specification results with 4 red and 3 yellow flags among participants. The difficulty with *Licmophora* a species easily recognizable must be attributed solely to the difficulty in counting cells that are dividing but continue to be attached to each other. However, in the case of *Trieres*, there were no difficulties with analysts results, no red or yellow flags for this count, so the issue lies directly on the difficulty of the transect as recorded. As cells divide, they stay in the same frustule and sometimes it is difficult to decide how many cells are to be counted.

Q4 and 5 were two matching questions where you must choose from a drop down list the right answer. Q4 asked analysts from a list of diatoms, which were considered to be 'centric' or 'pennates'. There were some difficulties with assigning *Mediopyxis* and *Helicotheca* to the centrics, 7 and 6 analysts were incorrect here but otherwise analysts did not find major difficulties here.

In Q5 using the same list of diatoms from the previous question we wanted to go a bit further and assign these centric and pennate diatoms evolutionary speaking into distinct groups 'radial or multipolar' centrics and 'raphid or araphid' pennates. There was a mistake in the test, where we considered 'Amphiprora' to be and 'araphid pennate' rather than 'raphid' which was wrong, but this now has been corrected in the OT results. Also, there was a query in relation to *Rhizosolenia* which cannot be considered 'multipolar centric' nor 'radial centric'. In this case, *Rhizosolenia* is a 'unipolar centric'. We consider that from an evolutionary point of view, these species did not reach the 'multipolar centric' stage and therefore we consider the other option to be the correct one.

Q7 and 8 continues with diatoms, both matching questions. Q7 refers to the way in which diatoms link in chains with each other. This character aids in the identification of diatoms and in the discrimination from other groups. This question was answered expertly by the analysts and results were around 90% of all analysts. In Q8 analysts are given terminology about diatoms life cycle and several answers related to these processes. The analysts were asked to match the

terminology to the undergoing process. As with the previous question, there were no difficulties for most analysts.

The second part of the test was based mostly on dinoflagellate questions. Q11 was based on their life cycle as a drag and drop questions with draggable items that had to be dropped in defined dropped zones. The diagram showed a simplified life cycle based on the reproduction stages of dinoflagellates. These can initiate the asexual or sexual cycle depending on their life condition. The arrows indicate the process the cells are undergoing depicted as green labels and the culmination of that process into a different form is depicted by the blue labels. Both, draggable labels were the answers to this question. There was some confusion between encystment and excystment and between ecdysal cyst and resting cyst, with various analysts mistaking both terms.

In Q12, we were back on evolutionary theory this time for dinoflagellates and the reason for the development of complex organelles like Ocellus or nematocysts, for example. Here, analysts were asked to match the description to the complex organelles. There were some levels of misunderstanding between nematocysts and pistons for example with several erroneous answers. Nematocysts are described as 'Extrusive harpoons' whereas pistons are 'a contractile organelle'. Also between cytostome and peduncle. The former is an invaginated membrane used in phagocytosis and the peduncle is an extendable feeding tentacle.

Q13 depicted a plate with images of species belonging to the order Gonyaulacales and analysts were asked to tell us whether the species were 'toxic' or 'non-toxic'. This was answered correctly by the majority of analysts, and they were able to differentiate between the toxic and non-toxic genera. There was a query in relation to *Goniodoma*. This genus is non-toxic according to the literature and there are no species within this group that produce toxins, however, according to both WoRMS and Algabase, *Goniodoma* can be considered synonymous with *Alexandrium*, so we decided to give this answer correct to all participants that chose 'toxic' as their answer.

In Q14 we are back with cell organelles evolution in dinoflagellates and this time we are looking at chloroplasts and pigmentation in dinoflagellates. We asked analysts based on the origin of the plastids, to match the plastid type to the species in the list. There were nearly perfect scores for this question.

In Q16, the formulas of the kofoidean tabulation of armoured dinoflagellates were given and analysts were asked to match the genus to the kofoidean formula. Most tabulations were recognised correctly by the analysts and there were only a small number of wrong answers for *Amphidoma* as *Protoceratium* or *Thecadinium* and *Gonyaulax* as *lingulodinium* or *Goniodoma.*

The remaining questions in the test Q17, 19 and 20 were multiple choice questions. Here analysts must check the right answer/s for a pool of choices. A wrong choice can cause penalty points to analysts. In Q17, analysts must choose the raphidophytes depicted in a plate of dinoflagellates. Images C, D, F and J are the right answers. Most analysts recognise *Chatonella*, *Fibrocapsa* and *Heterosigma akashiwo*, however *Haramonas dimorpha* is only correctly identified by 48 analysts (64%). Also, 5 analysts incorrectly checked *Amphidinium* and 4 analysts *Akashiwo sanguinea.*

In Q19 and Q20, there is only one true answer. Both depict a *Protoperidinium* species in light microscopy and in fluorescence microscopy using calcofluor staining to look at the plates. Most *Protoperidinium* species can be recognised by their ditinctive shape and size, but sometimes this is not enough to separate similar species. These can be distinguished looking at the shape and sides of the 1' plate (Ortho/Meta/Para) and the shape of the 2 intercalary plates (Quadra/Penta/Hexa), that is whether their paltes are four, five or six sided. Here, we give this information in the fluorescence images and using the right taxonomic keys, the right answer can be found.

In Q19, the right answer is *P.conicum*. It is Ortho (4) -Hexa (6) (1'-2a) and has that typical inverted 'V' shape suture in sulcal view extending from the APC to the cingulum. It can be confused with *P.leonis* also Ortho-Hexa but this species bear small spines in their antapical horns and no suture In Q20, the right answer is *P.divergens*. This can be confused with *P.crassipes* but the latter's right antapical horn is larger and longer than the left one, sort of assimetrical compared to *P.divergens*. Both are Meta-Quadra species with long apical and antapical horns. *P.depressum is* Ortho-Quadra, a different configuration, so it should not be confused with *P.divergens.*

Discussion:

Every year the IPI scheme attempts to develop a Proficiency testing exercise in marine phytoplankton enumeration and identification that assesses analysts fairly but robustly. The IPI program is possibly one of the most demanding, difficult, and thorough proficiency tests that

anyone can participate in. Each exercise is unique from the previous ones and requires the development and production of culture materials and the creation of a unique taxonomic examination through Ocean Teacher.

This year was no different, the test included a mixture of diatoms and dinoflagellates, with emphasis on toxic/harmful species but also experimenting with new species. Generally, is good to repeat some of the species from year to year, because it gives information about repeatability and reproducibility over time at different concentrations of similar or same species.

The main issues we found with our materials this year were in relation to the enumeration of the species more than their identification. All species in the samples were identified quite well by the analysts except for Gonyaulax hyalina, a somewhat lesser-known species of this genus. The shape is somewhat different to other Gonyaulax and it has no visible antapical spines. It is unusual and therefore complicated to identify correctly. To compound this, the cell concentration was low 10-20 cells /vial, so it was vital to get a good specimen to attempt to identify. Due to this difficulty, 22 analysts failed to detect this species in the sample and a further 6 analysts obtained red flags overestimating the count.

Apart from G.hyalina there were no significant issues with the identification of the other measurands. More problematic was the enumeration of two measurands, Bacteriastrum and Prorocentrum emarginatum.

The diatom was used for the first time in an IPI exercise, and it is similar in many aspects to Chaetoceros species, in that it is a chain former diatom with setae (spines) arising from the cell valves. The main difference with Chaetoceros is that in Bacteriastrum there are several spines arising from the valve face all around its circumference whereas in Chaetoceros there is just a pair of setae arising from opposite sides of the valve. This means that for Bacteriastrum, the diatom chain can be settled into the sedimentation chamber and seen 'standing up' rather than straight along its axis, creating issues when counting. There were 3 red flags and 4 yellow flags on this species count. To count a diatom chain when the only cell visible is the end of the chain without disturbing the sample is quite difficult. How many cells are in that chain? Most importantly, there are no guidelines in relation to issues like this. In the IPI we have introduced a few years ago, a counting guide to help with different problems that may arise during the exercise, but sometimes it is difficult without giving away too much information, that is why it would be important to

develop working groups among laboratories to reach a consensus on counting cell units with clear examples.

In relation to P.emarginatum, a benthic dinoflagellate, we encountered that when grown in culture it produced these hyaline spheres where pairs of cells (2,4,6,8….more) were dividing and developing inside this translucent egg. Also, we encountered that they produce an organic matrix that the cells were embedded into making it very difficult to homogenize but ultimately to count. Because of this clumping of cells in this organic matrix and its low cell concentration in the samples, we were unable to use the results from the participants.

The Ocean Teacher exercise was designed around the idea of evolutionary theory in phytoplankton and many questions on diatoms and dinoflagellates were probing in that direction. We also like to use images and videos for cell counting purposes to be able to corroborate or dispel ideas around cell counting and how do we go about it. This year, we used embedded video transects of species that were also found in the samples.

A sort of intercalibration counting exercise, where everyone must count the same number of cells in the video transects, with a series of rules. The results are certainly not discouraging, and many analysts do have similar counting strategies, which is good, however, not all the videos were successful. Two of the worst scores were numerical questions. Q15 and 18 specifically caused quite a bit of trouble. Q15 was the Licmophora cell count and Q18 Trieres. In theory, easy species to identify and count. This suggests that more discussions must be had around cell counting in diatoms, especially chain formers but also around life cycle stages in diatoms and how to count them.

Most analysts performed quite well in this assessment and certainly there is a high level of taxonomic expertise needed to perform well in these tests. Analysts in multiple choice questions seem to perform worse than in matching questions. For example, Q19 and Q20 on Protoperidinium identification, where there is only one right answer from multiple possible answers. Also, there are possible penalties from wrong choices, which doesn't happen in matching questions.

There was an error in Q5 where we entered the wrong answer for Amphiprora as correct, which was queried, upheld and changed in the final results. Also, in Q13, we had accepted Goniodoma

as toxic for its synonymity with Alexandrium in the database and also because we did not mention the IOC HAB list in the enunciation of this question. There was a query about Rhizosolenia in Q5 somewhat falling in between 'radial centric' and 'multipolar centric' and better described as 'unipolar centric'. We concluded that it hasn't evolved to be multi-polar centric and therefore the best answer was 'radial centric', even if not completely correct either.

It is important to reiterate how important it is the correct identification in these assessments. There was only one analyst that failed the identification part and six analysts the enumeration in this test, which suggest a good level of performance and competency overall. Analysts are tested in both aspects of this test, and both are in my view equally important. It is no good having good numbers for your measurands if you cannot recognize the measurands with certainty and viceversa. Non detections are the worst possible scenario for analysts as not only are the species not identified but also there are no enumeration results, that is no Z-scores. The Z-score is given as a +3 SD and the identification as ND.