United Nations Educational, Scientific and Cultural Organization

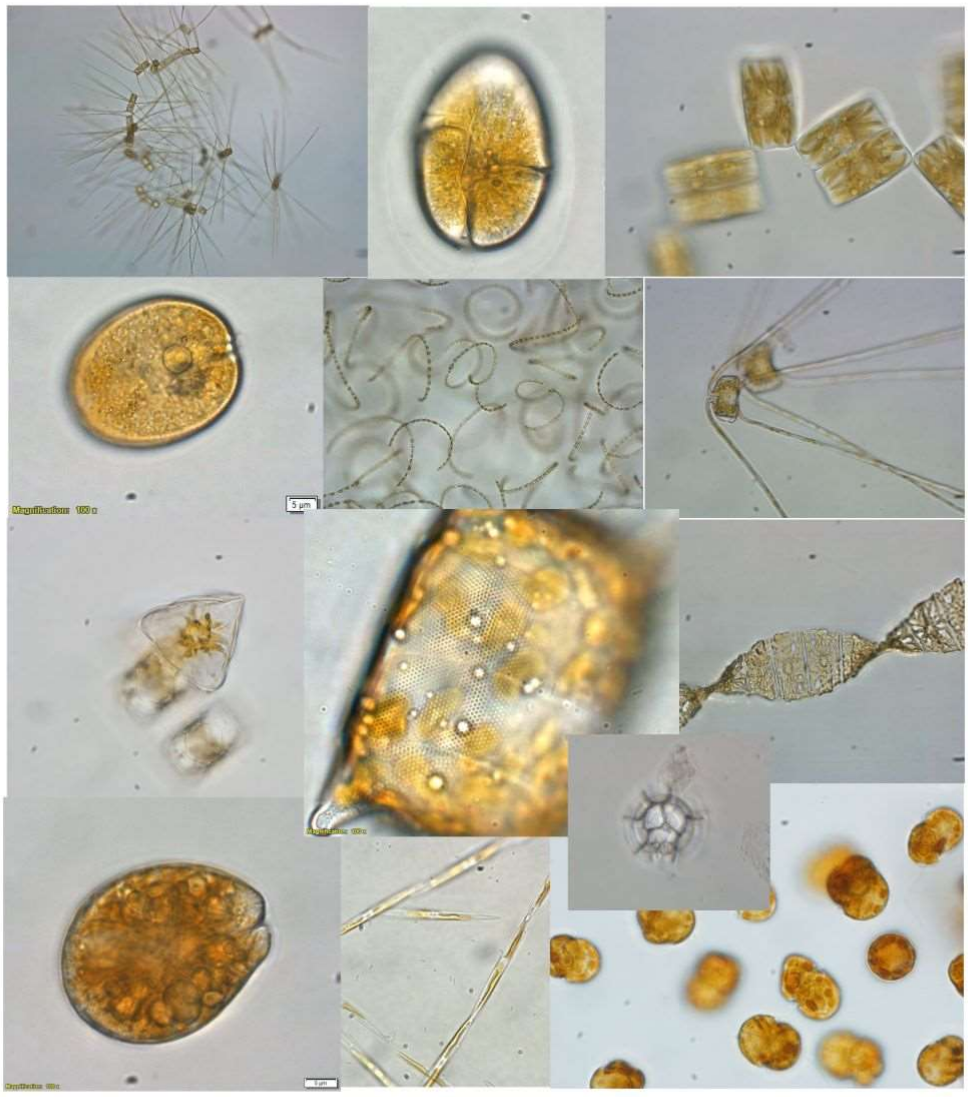
Intergovernmental Oceanographic Commission

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 2024 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 2024, 87 analysts across 51 laboratories around the world participated in the IPI2024 exercise. 86 analysts returned sample results and 84 completed the Ocean teacher taxonomic assessment. European countries accounted for 67% of the total participation 20% from the UK alone, 12% from African countries, 10% from Oceania, 8% from Asia and 3% came from South America.

- There were 8 new laboratories participating for the first time. Amanzi Biosecurity, Nelson Mandela University and SeeWise from South Africa, NOVA from the Faroe Islands, HYDRECO from France, IDAH from Romania, NIFS from South Korea and IZMIR from Turkey.

- There were ten measurands in the samples. The dinoflagellates were ***Ostreopsis cf. ovata*** Fukuyo, 1981, ***Prorocentrum cf. compressum*** (Bailey) T.H.Abé ex J.D.Dodge, 1975, ***Prorocentrum gracile*** F.Schütt, 1895, ***Levanderina fissa*** (Levander) Moestrup, Hakanen, Gert Hansen, Daugbjerg & M.Ellegaard, 2014 and ***Heterocapsa pseudotriquetra*** Iwataki, G.Hansen & Fukuyo, 2002.

- The Diatom species were ***Helicotheca tamesis*** (Shrubsole) M.Ricard, 1987, ***Lithodesmium undulatum*** Ehrenberg, 1839, ***Chaetoceros rostratus*** Ralfs, 1864, ***Pseudo-nitzschia delicatissima*** **group** (Cleve) Heiden, 1928, and ***Grammatophora marina*** (Lyngbye) Kützing, 1844.

- Grammatophora marina did not appear in all samples, therefore some analysts did not return results for this measurand. These results were not used for calculation purposes or to assess competency for these analysts. All the other results were used for statistical purposes

- 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.

- 5 measurands (*Helicotheca, Lithodesmium, Heterocapsa, P.gracile and Ostreopsis*) were deemed adequately homogeneous according to ISO13528:2022 criterion. The other 5 were not adequately

homogeneous but not significantly heterogeneous and it passed this criterion according to ISO13528:2022. All the measurands passed the expanded criterion for stability according to the same standard.

- There were a very small number of warning and action signals across measurands for the quantification results. 7 Red flags (0.83%), 27(3.2%) yellow flags and 23 (2.7%) non-detection flags (Grey triangles) from 844 results is evidence of good performance overall.

- 7 analysts did not pass the test from 86 returned results. Analysts 148 and 51 failed 5/10 results, analyst 86 6/10, analysts 10 and 75 failed 7/10 and analysts 29 and 99 6/9. 56 analysts had all the measurands (10) within the tolerance limits, 21 analysts had one failed measurand and 1 analyst two.

- Most analysts passed the qualitative test except for 8 analysts. 4 analysts had also failed the quantitative test (51/86/99/148) plus analysts 22/24/83/102. The hardest identification for the participants was *Levanderina fissa* with 16 incorrect and 10 non-detected flags.

- Overall, from approximately 860 possible correct identifications, there were a total of 694 correct answers at least to genus level (80.7 %), 43 incorrect identifications (5%) and 29 non-detections (3.4%) in total.

- There were 84 attempts at the OceanTeacher assessment, the median overall grade was 87.18%. 49.41% of analysts performed above the proficiency threshold of 90% and 30.58% of all analysts between 80-90%. 12.94% above 70% and another 5.88% below 70% requiring improvement.

- The OTGA facility index shows that the worst answered question in the test was Q20 (64.29%) a numerical question and the best Q12 (100.00%) also numerical.

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.

There were ten measurands in the samples. The dinoflagellates were *Ostreopsis cf. ovata* Fukuyo, 1981, *Prorocentrum cf. compressum* (Bailey) T.H.Abé ex J.D.Dodge, 1975, *Prorocentrum gracile* F.Schütt, 1895, *Levanderina fissa* (Levander) Moestrup, Hakanen, Gert Hansen, Daugbjerg & M.Ellegaard, 2014 and *Heterocapsa pseudotriquetra* Iwataki, G.Hansen & Fukuyo, 2002. The Diatom species were *Helicotheca tamesis* (Shrubsole) M.Ricard, 1987, *Lithodesmium undulatum* Ehrenberg, 1839, *Chaetoceros rostratus* Ralfs, 1864, *Pseudo-nitzschia delicatissima* group (Cleve) Heiden, 1928, and *Grammatophora marina* (Lyngbye) Kützing, 1844.

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 2024, 87 analysts across 51 laboratories around the world participated in the IPI2024 exercise. 86 analysts returned sample results and 84 completed the Ocean teacher taxonomic assessment. European countries accounted for 67% of the total participation 20% from the UK alone, 12% from African countries, 10% from Oceania, 8% from Asia and 3% came from South America (fig. 1).

There were 8 new laboratories participating for the first time. Amanzi Biosecurity, Nelson Mandela University and SeeWise from South Africa, NOVA from the Faroe Islands, HYDRECO from France, IDAH from Romania, NIFS from South Korea and IZMIR from Turkey.

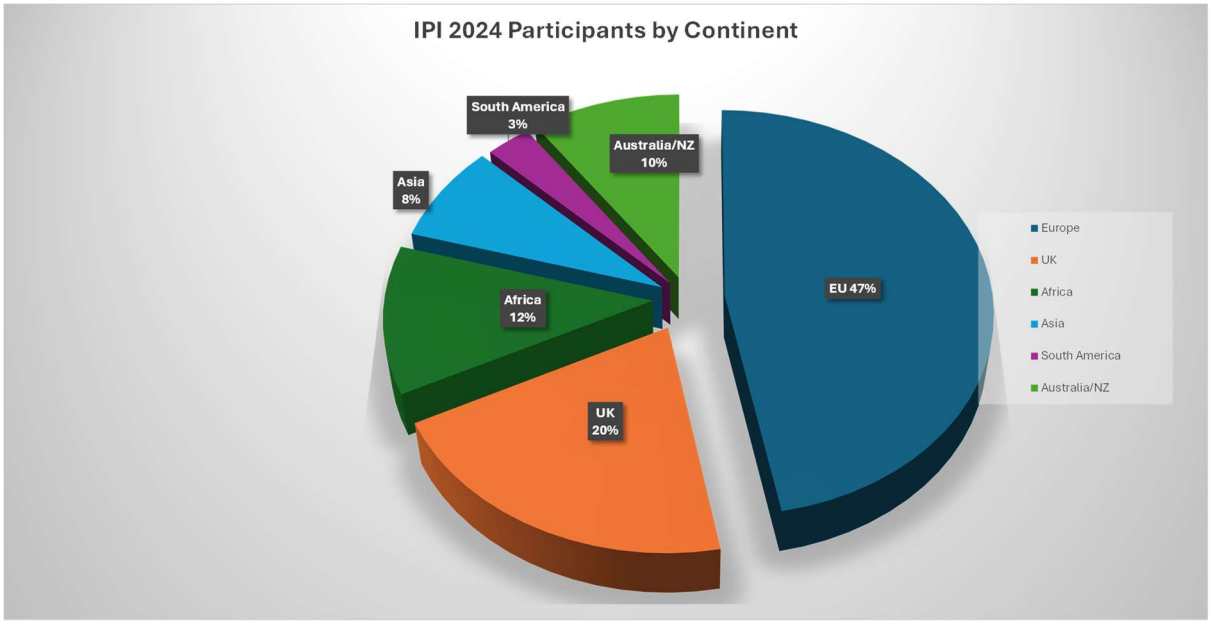


Figure 1: Participants by continent IPI2024

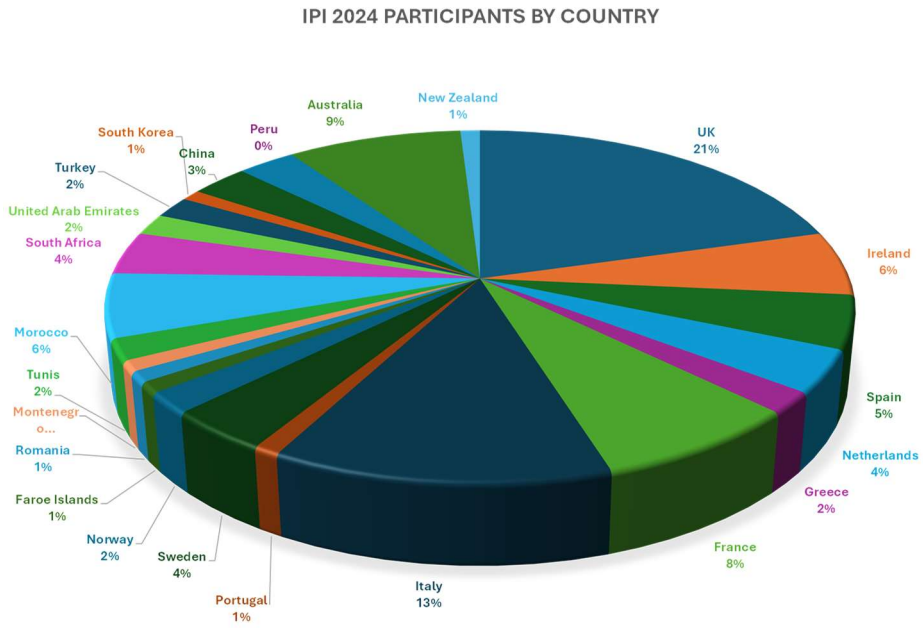


Figure 2: Participants by country IPI 2024

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-2024. The number of IPI participants has increased significantly since 2011 and the influence of the test has also widened to many regions across the globe (figure 2).
Pre-registration to the IPI intercomparison is through our dedicated website https://hab.ioc-unesco.org/ipi-home/ 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.

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 (IPI2024 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. 86 analysts in 51 laboratories were sent a total of 344 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, allows 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 were sent via e-mail to all registered participants including their unique identifiable laboratory and analyst code. Also, a counting guide was sent with the instructions to advise in the identification and counting of the species.

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 2024.7.30.0 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 13528:2022, 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 on 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 rafaelsalasipi@gmail.com

The test itself consisted of 20 questions (see Annex XVI). Question types used in the quiz were 'matching type' (Q3/7/11/15/19) which have dropdown menus including a selection of answers that analysts must choose from, 'multiple choice' (Q1/5/9/13/17) 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 (Q4/8/12/16/20) where analysts had to count the cells in the videos provided and 'drag and drop' types (Q2/6/10/14/18) 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 2024 included 10 measurands (Table 1) and all of them satisfied at least the ISO13528:2022 requirements and are not significantly heterogeneous, 5 of the measurands were found to be adequately homogeneous, these were *H.tamesis, H.pseudotriquetra, Lithodesmium, Ostreopsis and P.gracile.* Also, all materials passed the stability assessment according to the expanded criterion and only two failed the stability according to

ISO13528:2022. This means, as in previous years, that not all the materials are adequately homogeneous but also that they are not significantly heterogeneous.

Measurands	Cochran outliers	F-test	ISO 13528:2022 test for adequate homogeneity	ISO 13528:2022 - test for significant heterogeneity	Stability test ISO 13528:2022	Stability test ISO ISO 13528:2022 - expanded criterion	Harmonized Protocol / ISO 13528:2022 - expanded criterion
Chaetoceros rostratus	no outliers found	Not OK	Not OK	Ok	Ok	Ok	Ok
Grammatophora marina	outliers found: to id	Ok	Not OK	Ok	Not OK	Ok	Ok
Helicotheca tamesis	no outliers found	Ok	Ok	Ok	Ok	Ok	Ok
Heterocapsa pseudo-triquetra	no outliers found	Ok	Ok	Ok	Ok	Ok	Ok
Levanderina fussa	no outliers found	Ok	Not OK	Ok	Ok	Ok	Ok
Lithodesmium sp.	no outliers found	Ok	Ok	Ok	Ok	Ok	Ok
Ostreopsis sp.	no outliers found	Ok	Ok	Ok	Not OK	Ok	Ok
Prorocentrum compressum	no outliers found	Not OK	Not OK	Ok	Ok	Ok	Ok
Pseudo-nitzschia delicatissima group	no outliers found	Not OK	Not OK	Ok	Ok	Ok	Ok
Prorocentrum gracile	no outliers found	Ok	Ok	Ok	Ok	Ok	Ok

Table 1: IPI2024 Homogeneity and stability results according to ISO13528:2022

The procedure for a homogeneity and stability test is recorded in annex b of ISO13528:2022. 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 2024.7.30.0 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 ISO13528:2022 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 13528:2022, 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 five measurands failed the adequate homogeneity criterion except for *H.tamesis, H.pseudotriquetra, Lithodesmium, Ostreopsis and P.gracile* (table 1). However, no significant heterogeneity was found to any measurand according to the expanded criterion and only two failed the strictest criterion.

Some issues are reported in the homogeneity and stability test in relation to the *G.marina* cell count. Low cell densities of this measurand were introduced in the samples and while other chain forming diatom species tend to break down into smaller units, *G.marina* chain bonds are quite strong and the diatom chains do not breakdown easily. This made their homogeneity more difficult compounded by their low concentration, it caused that some samples did not contain this measurand.

Of 26 samples analysed for the homogeneity test, 8 samples did not contain any *G.marina* cells. The significance of this is that some analysts received samples that did not contain any cells of this species. 8 analysts did not find any cells in any of their 3 samples, and another 8 did not find any cells in 2 of 3 samples and a final 16 analysts did not find cells in 1 of the 3 samples. This left 53 analysts were able to find *G.marina* in all their samples. We were able to use the data of the 16 analysts that found cells in at least 2 samples, as statistically we could measure the mean of at least 2 samples, one of the requirements of the test. However, we were not able to use the data from the other 8 analysts that had counts in only 1 of 3 samples. This left 16 analysts in total outside of the *G.marina* count for statistical purposes, so their results are not applicable (n/a) for this measurand.

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 13528:2022 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 ISO13528:2022. 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

Ss= 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	L.undulatum	P.del.group	G.marina	H.tamesis	C.rostratus	O.cf ovata	H.pseudotriquetra	L.fissa	P.compressum	P.gracile
Robust mean x*	5298	18634	1106	10992	4988	2642	1957	232	911	7589
Robust Stdev s*	3103	6301	499	2318	2078	734	879	150	467	2722
Standard Ux	442	865	74	312	283	99	121	21	63	367
n=	77	83	71	86	84	85	83	76	85	86
if Ux < 0.3xSTdev	931	1890	150	695	623	220	264	45	140	817
then Ux is negligible	neg	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**	644	9182	358	7515	1871	1541	639	8	210	3506
*x *+1.5s**	9953	28086	1854	14469	8105	3742	3275	456	1612	11673
Homogeneity test	L.undulatum	P.del.group	G.marina	H.tamesis	C.rostratus	O.cf ovata	H.pseudotriquetra	L.fissa	P.compressum	P.gracile
Reference value mean	13380	14684	2360	15888	7732	5332	3916	868	1888	12064
Reference value stdev	1236	2979	997	516	1193	391	527	156	452	756
	Comparison with assigned value									
	L.undulatum	P.del.group	G.marina	H.tamesis	C.rostratus	O.cf ovata	H.pseudotriquetra	L.fissa	P.compressum	P.gracile
*x *-X*	8082	3950	1254	4896	2744	2690	1959	636	977	4475
Uncertainty of diff.	625	1223	105	442	401	141	170	30	90	519
2* Uncertainty of diff.	1250	2445	209	884	802	281	341	61	179	1038
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* (yellow values) is less than 0.3 times the standard deviation (green values) 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 ISO13528:2022 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 have been 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 2024.7.30.0. 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 2024, 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. The small red 'x' only applies to analysts that did not find G.*marina* cells in their samples and it means for the purpose of this intercomparison that is 'not applicable' to the analyst. All qualitative scores are also included for the final evaluation of analysts.

There were a very small number of warning and action signals across measurands for the quantification results. 7 Red flags (0.83%), 27(3.2%) yellow flags and 23 (2.7%) non-detection flags (blanks) from 844 results is evidence of good performance overall.

7 analysts did not pass the test from 85 returned results. Analysts 148 and 51 failed 5/10 results, analyst 86 6/10, analysts 10 and 75 failed 7/10 and analysts 29 and 99 6/9. 56 analysts had all the measurands (10) within the tolerance limits, 21 analysts had one failed measurand and 1 analyst two.

Most analysts passed the qualitative test except for 8 analysts. 4 analysts had also failed the quantitative test (51/86/99/148) plus analysts 22/24/83/102. The hardest identification for the participants was *Levanderina fissa* with 16 incorrect and 10 non-detected flags.

Quantitatively, The measurands with more failed results were *P.delicatissima* group with 7 yellow flags and the dinoflagellate *Ostreopsis sp.* with 8. The most difficult to detect species was *Levanderina fissa* with 10 non-detected flags, probably related to their low cell concentration in the samples, but also one of the most difficult to identify in the samples (16 incorrect ids).

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 methodological 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

The identification of measurands in the samples are given as 'correct', 'incorrect' and 'non-detected' flags in the statement of performance certificate of each analyst. This parameter is with the cell concentration calculations, the most important components of this test. Analysts must be able to recognize correctly all the species in the samples, at least to genus level.

At least 80% of the identification and quantitation results had to be correct to pass this test. In 2024, 10 measurands had to be identified in the samples. The only exception is for analysts that did not detect *G.marina* in their samples. These analysts were tested in 9 measurands only, so the pass mark for this small number of participants was reduced to 75%. The reason for this is that it is very possible that some samples did not contain any *G.marina* cells. This allows for all analysts to be able to pass the test with a maximum of two failed items.

A. Code	Lithodesmium undulatum	P-nitzschia delic group	Grammatophora marina	Helicotheca tamesis	Chaetoceros rostratus	Ostreopsis cf ovata	Heterocapsa pseudotriquetra	Levanderina fissa	Prorocentrum cf compressum	Prorocentrum gracile	Number of Measurands correct	Overall Flag
2	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
3	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
4	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Incorrect	Correct	Correct	9	PASS
5	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
6	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
7	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
8	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
9	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
10	Incorrect	Correct	Correct	Correct	Correct	Correct	Correct	nd	Correct	Correct	8	PASS
11	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
12	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
13	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
14	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
15	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
22	nd	Correct		Incorrect	Correct	Correct	Incorrect	Correct	Correct	Correct	7	FAIL
24	Incorrect	Correct	Correct	Incorrect	Correct	Correct	Correct	Incorrect	Correct	Correct	7	FAIL
25	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
26	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
28	Correct	Correct		Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
29	Correct	Correct		Correct	Correct	Correct	Correct	Incorrect	Correct	Correct	9	PASS
33	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
36	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
37	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
38	Incorrect	Correct	Correct	Incorrect	Correct	Correct	Correct	Correct	Correct	Correct	8	PASS
41	Correct	Correct	Correct	Incorrect	Correct	Correct	Correct	nd	Correct	Correct	8	PASS
42	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
43	Correct	Correct		Correct	Correct	Correct	nd	Correct	Correct	Correct	8	PASS
45	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Incorrect	Correct	Correct	9	PASS

A. Code	Lithodesmium undulatum	P-nitzschia delic group	Grammatophora marina	Helicotheca tamesis	Chaetoceros rostratus	Ostreopsis cf ovata	Heterocapsa pseudotriquetra	Levanderina fissa	Prorocentrum cf compressum	Prorocentrum gracile	Number of Measurands correct	Overall Flag
47	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Incorrect	Correct	Correct	9	PASS
51	nd	nd	Correct	Correct	nd	nd	nd	Incorrect	Correct	Correct	4	FAIL
53	Correct	Correct		Correct	Correct	Correct	Incorrect	Incorrect	Correct	Correct	7	PASS
55	nd	Correct	Correct	Correct	Correct	Correct	Correct	Incorrect	Correct	Correct	8	PASS
58	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
61	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
62	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
72	Correct	Correct	Correct	Correct	Incorrect	Correct	Correct	Correct	Correct	Correct	9	PASS
75	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
76	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
77	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Incorrect	Correct	Correct	9	PASS
83	Incorrect	Correct	Correct	Incorrect	Correct	Correct	Correct	Incorrect	Correct	Correct	7	FAIL
86	nd	Correct	Correct	Correct	Correct	Correct	Incorrect	Incorrect	Correct	Correct	7	FAIL
87	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
88	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
92	nd	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	9	PASS
97	Correct	Incorrect	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	9	PASS
99	nd	nd		Correct	Correct	Correct	nd	nd	Correct	Correct	5	FAIL
100	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
101	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
102	nd	Correct	Correct	Correct	Correct	Incorrect	Incorrect	Incorrect	Correct	Correct	6	FAIL
105	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
110	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
111	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
113	nd	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	9	PASS
114	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Incorrect	Correct	Correct	9	PASS
115	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS

A. Code	Lithodesmium undulatum	P-nitzschia delic group	Grammatophora marina	Helicotheca tamesis	Chaetoceros rostratus	Ostreopsis cf ovata	Heterocapsa pseudotriquetra	Levanderina fissa	Prorocentrum cf compressum	Prorocentrum gracile	Number of Measurands correct	Overall Flag
116	Correct	Correct		Incorrect	Correct	Correct	Correct	Correct	Correct	Correct	8	PASS
117	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
118	Correct	Correct	Correct	Correct	Correct	Correct	Correct	nd	Correct	Correct	9	PASS
122	Correct	Correct	Correct	Correct	Correct	Correct	Correct	nd	Correct	Correct	9	PASS
124	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
132	Correct	Correct		Incorrect	Correct	Correct	Correct	nd	Correct	Correct	8	PASS
133	Incorrect	Correct	Correct	Correct	Correct	Correct	Correct	Incorrect	Correct	Correct	8	PASS
134	Correct	Correct	Correct	Correct	nd	Correct	Correct	Correct	Correct	Correct	9	PASS
135	Correct	Correct	Correct	Correct	Correct	Correct	Correct	nd	Correct	Correct	9	PASS
136	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
141	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
142	Correct	Correct	Correct	Incorrect	Correct	Correct	Correct	Incorrect	Correct	Correct	8	PASS
143	Correct	Correct	Correct	Incorrect	Correct	Correct	Correct	nd	Correct	Correct	8	PASS
145	Correct	Correct	Correct	Correct	Incorrect	Correct	Correct	Correct	Correct	Correct	9	PASS
147	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
148	nd	nd	Correct	Correct	Correct	Incorrect	Incorrect	nd	nd	Correct	4	FAIL
149	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
151	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
152	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
153	Correct	Correct	Correct	Correct	Correct	Correct	Incorrect	nd	Correct	Correct	8	PASS
159	Correct	Correct	Incorrect	Correct	Incorrect	Correct	Correct	Correct	Correct	Correct	8	PASS
163	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
173	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
174	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Incorrect	Correct	Correct	9	PASS
182	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
183	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
184	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
185	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
193	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
194	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS
195	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	10	PASS

Table 3: Qualitative results IPI2024 by Analyst and Measurand. Nd= not detected

Most analysts passed the qualitative test except for 8 analysts. 4 analysts had also failed the quantitative test (51/86/99/148) plus analysts 22/24/83/102. 49 analysts identified correctly all measurands. 17 analysts identified incorrectly or did not detect 1 measurand, 12 analysts 2 measurands, 4 analysts 3 measurands and 4 analysts 4 or more measurands.

The hardest identification for the participants was *Levanderina fissa* with 16 incorrect and 10 non-detected flags. The next most difficult measurands were *Lithodesmium* with 9 nds and 5 incorrect flags, *Helicotheca* with 9 incorrect flags and *Heterocapsa* with 3nds and 6 incorrect flags.

The easiest measurands to identify were both *Prorocentrum gracile* (0 nds/0 incorrect) and *Prorocentrum cf. compressum* (1 nd/0 incorrect) and *Ostreopsis* (1 nd/2 incorrect) or the dinoflagellates and for the diatoms: *Grammatophora* (0 nd/1 incorrect), *Pseudo-nitzschia* (3 nds/1 incorrect) and *Chaetoceros* (2 nds/3 incorrect).

Species	ID
Lithodesmium sp.	48
Lithodesmium undulatum	18
Lithodesmium duckerae	4
Lithodesmium variabile	2
non detected	9
Incorrect id. Number	5
Total analysts	86
Incorrect ID names	
Lampriscus sp.	4
B.malleus	1
Total analysts	5

Species	ID
P. delicatissima group	63
P. seriata group	9
P. delicatissima	6
P. pungens	2
P. micropora	1
P. pseudodelicatissima	1
non detected	3
Incorrect id. Number	1
Total analysts	86
Incorrect ID names	
Synedropsis fragilis	1
Total analysts	1

Species	ID
Grammatophora sp.	51
Grammatophora marina	25
Grammatophora oceanica	1
Incorrect id. Number	1
Total analysts	78
Incorrect ID names	
Fragillaria sp.	1
Total analysts	1

Species	ID
Helicotheca tamesis	77
Incorrect id. Number	9
Total analysts	86
Incorrect ID names	
Mediopyxis helysia	9
Total analysts	9

Species	ID
Chaetoceros rostratus	42
Chaetoceros phaeoceros	19
Chaetoceros atlanticus	17
Chaetoceros dichaeta	1
Chaetoceros cerastoporus	1
Chaetoceros fallax	1
non detected	2
Incorrect id. Number	3
Total analysts	86
Incorrect ID names	
Attheya sp.	3
Total analysts	3

Species	ID
Ostreopsis cf ovata	49
Ostreopsis cf siamensis	15
Ostreopsis lenticularis	17
Ostreopsis marinus	3
non detected	1
Incorrect id. Number	1
Total analysts	86
Incorrect ID names	
P.panamense	1
Coolia sp.	1
Total analysts	2

Species	ID
Heterocapsa sp.	32
Aza/Heterocapsa sp.	13
Heterocapsa niei	10
Heterocapsa pseudotriquetra	19
Heterocapsa ovata	2
Heterocapsa triquetra	1
non detected	3
Incorrect id. Number	6
Total analysts	86
Incorrect ID names	
Azadinium spinosum	3
Amphidinium sp.	1
Karlodinium veneficum	2
Total analysts	6

Species	ID
Levanderina fissa	14
Gymnodinium sp.	33
Gymno/Gyrodinium sp.	12
G.chlorophorum	1
non detected	10
Incorrect id. Number	16
Total analysts	86
Incorrect IDs	
A.sanguinea	13
K.selliformis	1
K.mikimotoi	2
Total analysts	16

Species	ID
Prorocentrum cf compressum	66
Prorocentrum concavum	6
Prorocentrum maculosum	6
Prorocentrum rathymum	1
Prorocentrum lima	6
non detected	1
Incorrect id. Number	0
Total analysts	86
Incorrect ID names	
Total analysts	0

Species	ID
Prorocentrum gracile	52
Prorocentrum micans	32
Prorocentrum triestinum	1
Prorocentrum sigmoides	1
non detected	0
Incorrect id. Number	0
Total analysts	86
Incorrect ID names	
Total analysts	0

Table 4: IPI 2024 Qualitative data by measurand.

The identification of measurands to genus level is sufficient to achieve a correct flag for any given identification, but we always aim for analysts to identify to species level where possible as an incorrect species determination won't affect the result. Table 4 shows all the qualitative results of all the analysts for each measurand and gives information about which incorrect answers they selected for each measurand.

It is interesting to note that for example, for *P.gracile* analysts were in disagreement at species level. 52 analysts said it was *P.gracile* and another 32 said it was *P.micans*. The consensus was not strong enough at species level but perfect at genus level. For *H.tamesis*, 77 analysts (89.5%) identified correctly to species level, but 9 analysts identified incorrectly this species with *Mediopyxis helysia.* There were no other answers here. *P.compressum* also showed a high number of correct answers at species level with 66 analysts (76.7%) while other analysts chose different species names like for example *P.lima, P.concavum* and *P.maculosum*.

For other species, analysts were content to identify to genus level and not to go further. 51 analysts went for *Grammatophora sp.,* 63 for *P.delicatissima group*, 48 for *Lithodesmium sp.* and 32 for *Heterocapsa sp.*

In the identification of *Ostreopsis,* analysts found consensus around *O. cf ovata* with 49 analysts but other options were also supported: *O.lenticularis* (17) and *O.siamensis* (15) were also popular. Something similar to *Chaetoceros rostratus* (42), with *C.atlanticus* (17) and *C. phaeoceros* group (19) as other options chosen by analysts.

The identification of *Levanderina fissa* was the most difficult measurand to identify or detect in the samples. This species was not detected by 10 analysts probably because of its low cell concentration in the samples. Also, it was misidentified by 16 analysts, mostly with *Akashiwo sanguinea* (Table 4).

Overall, from approximately 860 possible correct identifications, there were a total of 694 correct answers at least to genus level (80.7 %), 43 incorrect identifications (5%) and 29 non-detections (3.4%) in total.

4.7 Ocean Teacher 2024 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 84 attempts at the OceanTeacher assessment, the median overall grade was 87.18%. 49.41% of analysts performed above the proficiency threshold of 90% and 30.58% of all analysts performed between 80-90%. 12.94% above 70% and another 5.88% below 70% requiring improvement (Table 5).

A. Code	%	A. Code	%	A. Code	%	A. Code	%
8	98.84	**117**	93.38	**15**	89.92	**163**	83.93
29	98.84	**147**	93.38	**133**	89.81	**36**	83.51
12	98.11	**7**	93.07	**142**	89.71	**173**	82.46
62	98.00	**182**	93.07	**24**	89.50	**3**	81.72
136	97.37	**152**	92.96	**28**	89.50	**41**	81.30
61	97.06	**92**	92.75	**77**	89.18	**55**	81.20
149	97.06	**195**	92.65	**100**	89.08	**153**	79.62
115	96.95	**25**	92.54	**42**	88.66	**185**	79.62
183	96.85	**47**	92.23	**88**	88.03	**97**	79.10
4	96.85	**122**	92.23	**184**	87.82	**13**	78.99
9	96.64	**124**	91.91	**10**	87.50	**114**	78.68
111	96.43	**151**	91.91	**33**	87.29	**198**	78.15
43	96.22	**194**	91.70	**Average**	**87.18**	**102**	77.21
116	95.59	**14**	91.28	**26**	86.03	**58**	76.89
145	95.59	**37**	91.18	**174**	86.03	**132**	76.68
118	95.59	**141**	91.18	**75**	85.82	**135**	75.53
143	95.48	**5**	91.07	**53**	85.61	**22**	74.37
72	94.96	**2**	90.76	**193**	85.61	**148**	69.64
134	94.54	**38**	90.23	**110**	85.40	**83**	68.38
6	93.70	**11**	90.13	**101**	84.98	**86**	67.86
76	93.38	**113**	90.13	**99**	84.14	**159**	64.50
						105	63.24

Table 5: Ocean Teacher IPI 2024 scores by analyst code

The OTGA facility index shows that the worst answered question in the test was Q20 (64.29%) a numerical question and the best Q12 (100.00%) also numerical (Table 6).

Q#	Question type	Question name	Attempts	Facility index	Standard deviation	Random guess score	Intended weight	Effective weight	Discrimination index	Discriminative efficiency
1	Multiple choice	IPI 2024 Benthic Gonyaulacales	84	89.76%	17.20%		5.56%	5.11%	18.07%	20.39%
2	Drag and drop onto image	IPI 2024 Amphidomataceae Plate	84	93.23%	7.01%		5.56%	4.27%	47.57%	50.70%
3	Matching	IPI 2024 Chetoceros identification image plate	84	95.45%	11.51%	6.25%	5.56%	4.45%	26.32%	32.39%
4	Numerical	IPI 2024 Numerical I C.curvisetus	84	84.52%	36.38%	0.00%	5.56%	7.76%	6.64%	8.36%
5	Multiple choice	IPI 2024 Dinophysiales plate	84	86.43%	15.80%		5.56%	6.89%	51.21%	54.46%
7	Matching	IPI 2024 Prorocentrum benthic vs planktonic	84	97.34%	6.81%	50.00%	5.56%	4.02%	43.03%	55.93%
8	Numerical	IPI 2024 Numerical II Lithodesmium	84	98.81%	10.91%	0.00%	5.56%	2.03%	-0.46%	-1.52%
9	Multiple choice	IPI 2024 Heterocapsa	84	78.97%	32.75%		5.56%	10.57%	50.91%	55.59%
11	Matching	IPI 2024 Diatoms Centric versus Pennate	84	97.77%	4.52%	50.00%	5.56%	3.22%	42.71%	51.58%
12	Numerical	IPI 2024 Numerical III Ostreopsis	84	100.00%	0.00%	0.00%	5.56%	0.00%		
14	Drag and drop onto image	IPI 2024 Azadinium ventral pore position	84	91.22%	16.78%	12.50%	5.56%	4.51%	12.10%	13.64%
15	Matching	IPI 2024 Diatoms colony formation types I	84	76.86%	14.27%	8.33%	5.56%	5.27%	29.12%	31.32%
16	Numerical	IPI 2024 Numerical VI Grammatophora	84	85.71%	35.20%	0.00%	5.56%	9.07%	22.74%	29.67%
17	Multiple choice	IPI 2024 Suessiales	84	78.57%	26.21%		5.56%	7.69%	27.56%	30.77%
18	Drag and drop onto image	IPI 2024 Alexandrium chain vs non-chain formers	84	93.51%	10.98%		5.56%	5.28%	43.82%	50.73%
19	Matching	IPI 2024 Protoperidinium identification	84	84.92%	22.20%	7.69%	5.56%	8.03%	45.13%	48.04%
20	Numerical	IPI 2024 Numerical V Pseudonitzschia	84	64.29%	48.20%	0.00%	5.56%	9.78%	4.65%	5.64%

Table 6: Facility index IPI2024 OT exercise

This test was divided into 20 questions with 4 different types used. 5 questions for each type. The types used were numerical, matching, multiple choice and drag & drop.

Questions 4, 8, 12, 16 and 20 were numerical type questions. In this test analysts had to watch a short video clip of a transect from a sedimentation chamber showing several cells, the analysts were given instructions on how to count the cells shown in the transects and to write down the number of cells they counted. Each question had a hyperlink to the video in YouTube. The diatoms *Chaetoceros curvisetus (*Q4), *Lithodesmium undulatum* (Q8), *Grammatophora marina* (Q16) and *Pseudo-nitzschia delicatissima* group (Q20) plus the dinoflagellate *Ostreopsis cf ovata* (Q12) were used for this purpose. All of them were well resolved by the analysts except for Q20 the *Pseudo-nitzschia delicatissima* group count. The scores for Q8 and 12 (98.81% and 100%) were nearly perfect for all analysts. Q4 and 16 (84.52% and 85.71%) were reasonable scores and Q20 (64.29%) was the worst score of all questions. Both, in the *Chaetoceros curvisetus (*Q4) and *Grammatophora marina* (Q16) count, there is evidence of bias in counting, in both instances there were a sizeable number of analysts under reporting cell numbers in chain forming diatoms. In the *Chaetoceros* case, the consensus counts 42 +/- 1 cell, however 13 analysts reported cells below 38 and in the *Grammatophora* count the consensus count was 76 +/- 2 cells, but 7 analysts reported 67, 7 cells less than the consensus count, suggesting a bias compared to the rest of analysts. In Q20, the problem is the opposite, with 27 analysts reporting overestimates (more than 100 cells counted) of the consensus count for *Pseudo-nitzschia* (94+/-5 cells). The tolerance was widened for this count because of the difficulty of counting this transect.

The Multiple-choice questions (Q1-5-9-13-17) are a type of question where analysts must tick the right answer/s from a list of choices given. Their difficulty lies in that a wrong choice penalise the

analysts with a percentage deduction. Q1 asked analysts to choose the benthic genera from a list provided of the order Gonyaulacales. This oder is known to contain both planktonic and benthic genera. Most analysts chose the genus *Coolia, Gambierdiscus* and *Ostreopsis* as benthic, but only 65 chose *Adenoides* a lesser know benthic genus. Also 14 analysts picked *Alexandrium* as benthic.

This is an issue that was raised with the technical Advisory group. The view by the group is that *Alexandrium* should be considered planktonic, most of the 31 species included in the genus are planktonic except for *A. hiranoi* that has some benthic characteristics and *A. Phylotypes*. Hoppenrath does include *A.hiranoi* in her benthic dinoflagellates book but also writes in the remarks section that *Alexandrium* are 'planktonic' and that *A.hiranoi* is and I quote 'Benthopelagic'. The *A.phylotypes* is included in the new version of her book. It says that *A.phylotypes* are epiphytic to macroalgae in Japan. These are exceptional types rather than the rule and *Alexandrium* cannot be considered at this stage a benthic genus. Also, cyst formation is not a character that defines what is a benthic genus, otherwise, most dinoflagellate genera could be considered benthic as they most have a benthic cyst stage.

Q5 asked analysts to choose with genera belongs to the dinophysaceae family. Most analysts correctly chose *Dinophysis/Citharistes/Histioneis/Ornithocercus* and *Parahistioneis.* As many as 42 also included *Sinophysis* into the family and 27 and 21 analysts included *Pseudophalacroma* and *Oxyphisis* which are incorrect answers.

I think that the question clearly states: 'Place the genera that can be assigned to the family with **'certainty'**. *Oxyphysis* is now *Phalacroma* and *Phalacroma* is in the Oxyphysiaceae family, *Synophysis* and *Pseudophalacroma* are 'Incertae sedis', that is they have and uncertain origin. It is a tricky question and, I think that some people have investigated the WoRMS database to get the answer to this question. However, it is actually correct in Algaebase and it is hinted in the question. The records in WoRMS cite Algaebase for their records but it is not completely up to date in some cases. It is true that Chomerat 2016, does leave *Synophysis* in the family even though he argues against it in his paper discussion on *Synophysis*. This genus which is benthic and not planktonic like the others cannot be placed with certainty in the family Dinophysaceae. The only phylogenetic data available from Gomez and Hoppenrath also hints at this. 'Incertae sedis' is the right answer here.

Q9 asks analysts which other dinoflagellate genera other than *Heterocapsa* have 'body scales'. The right answers were *Lepidodinium, Apmphidinium* and *Oxyrrhis*. It was a moderately well answered question with over 70 correct answers for *Apmphidinium* and *Oxyrrhis* and 61 only for *Lepidodinium*.

In Q13, analysts were asked which families within the Peridiniales order could be 'obligate autotrophic' from the list given. The correct answers given were *Scrippsiella/ Heterocapsa/ Calciodinellum/ Durinskia* and *Peridinium*. However, there is evidence that some members of the generally phototrophic genera of the Peridiniales can in exceptional circumstances use mixotrophy. Some analysts were not in agreement with the 'obligate autotrophy' of the genera.

There is no doubt that the 'obligate autotrophy' line in the question created some confusion and there is no doubt that there is some evidence of mixotrophy in a very reduced number of species in the genus named in the question. The Peridiniales are a large order of dinoflagellates that have a very clear heterotrophic component and therefore it would not be unusual that some of the photosynthetic cousins may have also the ability to be mixotrophic. The papers that have been used to back up the mixotrophy claim are generally laboratory experiments with cultures deprived of light and nutrients to induce heterotrophy. These genera are photosynthetic in 'normal conditions', perhaps the issue is more with the question itself and the way it was phrased. Therefore, the question was left out of the final score for the test.

The last of the multiple-choice questions Q17 asked analysts which of the following genera belongs to the suessiales order. The correct answers were: *Biecheleria, Pelagodinium, Symbiodinium* and *Biecheleriopsis*. Most analysts correctly checked *Biecheleria* and *Symbiodinium*, but only 53 and 52 analysts chose *Pelagodinium* or *Biecheleriopsis*. Some analysts said that *Pelagodinium* should be in the gymnodiniales and *Biecheleropsis* in the lophodiniales orders.

This is again a database issue, it is possible that analysts looking at the WoRMS database got it wrong and those looking into Algaebase got it correct, because the information is not up to date in WoRMS. The papers are Siano et al. 2010 for *Pelagodinium*, a new genus where *Gymnodinium beii* was transferred to *pelagodinium* and the suessiales order. The other paper is Moestrup et al. 2009 on *Biecheleriopsis*. Both papers are associated with these records. Both Genera are order suessiales.

In 'matching' type questions (Q3-7-11-15-19) analysts must choose an answer from a list of options in a drop-down menu, The list generally contains more species names than the number

of images to make it a bit more difficult. In Q3, analysts were given an image plate showing several *Chaetoceros* species and were asked to identify them from the drop-down menu. All analysts performed well and identify correctly most species. There were only a small number of incorrect answers between species that are quite similar. *Chaetoceros lorenzianus* (image 9 a-b) is generally confused with *C.decipiens* (image 11 a-b), the difference between them being the point of fusion of the 'setae'. There were 7 incorrect answers. Also, *C.rostratus* and *C.atlanticus* were confused by 5 analysts. *C.atlanticus* (image 1a-b-c) has straight chains, the setae are in the apical axis and they cross away from the valve faces, whereas in *C.rostratus* , the chains are not necessarily straight, the setae are not in the apical axis and they cross closer to the valve face. These are the main distinguishing features for these two species.

In Q7, analysts were asked to choose which of the following *Prorocentrum* species were benthic or planktonic and most responses were correct for this question. Only *P.rathymum* had more erroneous responses than the others with 6 analysts choosing planktonic rather than benthic. Q11 was a similar question to Q7, in this instance we asked analysts to tell us which diatoms in the plate image shown were centric or pennate diatoms. There was an error in this question, where we erroneously placed *Grammatophora* as 'centric' which is incorrect. This error was amended and scores updated. There were no problems with this question and most analysts were correct with their answers. The most difficult item was *Grammatophora* which 13 analysts placed in the 'centric' group.

For Q15, analysts were shown an image plate of chain forming diatoms and were asked to match each image to the description that best fits how the chains link together. There was a problem with this question as it doesn't allow two answers to be exactly the same, and we repeated 'ribbons' and 'stellate chains' twice for *Sinedropsis* and *Thalassionema* for stellate chains and *Helicotheca* and *Lithodesmium* for ribbons. The problem is caused by the programme which it only allows the answers in a particular order, otherwise it gives the answer as incorrect. This issued was corrected and the scores updated to reflect these changes. Nonetheless, there were some difficulties with this question, the most obvious one is the descriptions for *Lauderia* and *Guinardia*. *Lauderia* chains link with each other by short processes while *Guinardia* joins their chains by abbuting valve faces. Analysts switched these two species, and it was mostly incorrectly answered except for 15 and 18 analysts respectively. The other difficult description was the answer for *Actinoptychus* which chains link by a ring of labiate processes. Only 44 analysts correctly answered this species.

Q19 the last of the matching type questions showed images of 6 *Protoperidinium* sp. These species can only be characterised with certainty to species level by the shape of the 1' apical plate and 2a intercalary plate (if 3 intercalary plates are present). Other characters can also be used in conjunction with these. The image plates showed these features to allow the analysts to make the right decisions. The question was answered reasonably well with over 65-70 analysts correct per species and only a small number of mistakes. The most common ones were *P.divergens* confused with *P.crassipes/curtipes* types. *P.divergens* has a Meta/Quadra arrangement and have sulcal lists and diverging spines, longer than wider in size and shape, it can be differentiated from *P.crassipes* which is wider than longer. *P.conicum* and *P.leonis* are always confused because they are both Ortho/Hexa arrangements and the only visible difference is that *P.conicum* has this inverted 'V' shape suture in its ventral side running down from the apical point (see image 4 in annex report XVI) and *P.leonis* doesn't. *P.leonis* have small antapical horn types. Also. *P.thorianum* with *P.puntulatum* both have reticulated armoured plates but *P.thorianum* has only 2 intercalary plates.

The other question 'type' used in this test was the 'drag and drop' type question (Q2-6-10-14-18). In these questions analysts must drag images to the right place holders and drop them there. We had several technical issues with Q6 and Q10, where there were more dragging objects than place holders for the objects. This caused objects to jump out of their placings. We decided to write off these two questions for that reason, the analysts scores reflect this update.

Q2 asked analysts which Amphidomataceae species of this family were considered toxic/non-toxic or unknown. They had to place a red 'x' for toxic a green one for non-toxic and a blue one for unknown. There were no serious difficulties with this question and analysts placed most answers in their correct places. The most split decision was *Az. caudatum var. caudatum* where 44 analysts chose 'non-toxic' and 40 chose 'unknown'. *Az.perforatum* 67 non-toxic and 17 unknown was the next most difficult decision. Both are non-toxic. The species that were discovered and described from SEM stubs do not have a corresponding live image and therefore no molecular or toxicity data is available, these are the only unknown ones.

In Q6 we introduced a key for Heterocapsa species. This is an expanded key from Iwataki 2008 with the inclusion of newly described species. This key must be worked out from the top and make your way down to make sense of it. In this question several objects were to be deposited in the same place holder. This caused some technical issues which were compounded by one mistake, where two species were supposed to be placed under the key 'large epitheca-small

hypotheca' and '>15µm size', the species *H.lanceolata* and *H.arctica* fall under this characteristics but only one place was available for both species, the dropping zone for *H.lanceolata* was mistakenly placed somewhere else. Besides that, there weren't many difficulties using the key. The full answer to this question is in the annex XVI of the annex report. The main contentions are *H.iwatakii* which is placed under 'large epitheca-small hypotheca' and '<15µm size' even though the actual drawing doesn't really shows this feature well, the paper describes the species as Large epitheca-small hypotheca type and it is therefore included in this section. Other issues arise from not following the key correctly, so that species like *H.borneoensis* are placed in 'elongated nucleus' rather than in 'more than one pyrenoid anterior to the nucleus' (see key in the annex report annex XVI).

There were similar issues with place holders in Q10 and a decision was made no to compute these two questions for the final scores. The scores were updated successfully excluding the two drag and drop questions. Q10 was for analysts to link phycotoxin families to species. Analysts were asked to place the species to the family of biotoxins they belong to. In this case several species could belong to the same phycotoxin family or just one or two. The main concerns with this question arose around 'Ichthyotoxic' species and 'Reactive Oxygen Species' and what is the difference between these two. The answer is that ROS species are also Ichthyotoxic but may or may not produce 'Ichthyotoxins' and their mode of action, that is the way they affect fish can not be fully understood or attributed to the production of one or several toxin compounds but rather by an oxidative reaction usually caused by the release of reactive oxygen compounds in the water. It is true that at present this fish kills are not fully understood and the question intended to separate species known for causing fish kills due to a particular known toxin with a particular mode of action and toxicology associated to it and those that harm fish by other means.
This would separate the raphidophytes from other ichthyotoxin producing species like *Karlodinium, Karenia, Pfiesteria* etc.

The family amphidomataceae needs electron microscopy to identify with certainty to species level. One of the main identification features using electron microscopy is the location of the ventral pore (v.p.) on the cell. It is not the only diagnostic feature but an important one if working with *Azadinium* and *Amphidoma* cells. The image in Q14 shows 8 images of *Azadinium* species that have the v.p. in the Po plate right hand side, so they require other features to separate them, so a description was added to the bottom of the image to help designating the species into their right place holders. Only one holder per object works well for these questions. All analysts

performed well in Q14 and there were only issues with *Az. perfusorium* and *Az. perforatum*. The former has large 1a and 3a intercalary plates and the latter has thecal pores in the Po plate, which is sufficient to clarify their identify.

The last drag and drop question (Q18) showed an image plate of several *Alexadrium* species and analysts were asked to place the objects onto the species that were chain-forming and non-chain-forming. Scores were good for this question and there were a small number of erroneous answers for two chain forming species: *A. pacificum* and *A.compressum* with 13 and 7 wrong answers respectively. The most controversial was *A.tamarense* which is a non-chain former species that can at times appear in short chains of 2s and sometimes 4 cells with 59 analysts said it was a non-chain former to 25 analysts saying it is a chain former. Due to this fact we introduced in the question the ability to form 'long chains'. Normally, these species don't appear to form chains readily and we don't consider them from this point of view as chain formers. The ability of dinoflagellates to form short chains when dividing would make most species to be chain forming, which is not the case.

5.0 Discussion:

The IPI program is possibly one of the most demanding, difficult, and thorough proficiency tests that anyone can partake, in the world. Each exercise is unique and requires the development and production of culture materials and the creation of a taxonomic examination through Ocean Teacher. Our intention is to develop a Proficiency testing exercise in marine phytoplankton enumeration and identification that assesses analysts fairly but robustly.

In relation to the production of materials for the test, this year we learned that the diatom *Grammatophora marina* did not homogenised well because their chains do not tend to break down easily into smaller units and our cell density at the time of preparing the materials was not sufficiently high. This meant that some samples did not contain any *Grammatophora* chains. Also, we had some problems with *Lithodesmium* which in contrast to *Grammatophora* its chains were shorter, and they also broke into single cells, although easily identifiable because of its characteristic triangular shape in valve view.

Also, among the dinoflagellates *Levanderina fissa* was the most difficult species to identify in the samples. We have included *Levanderina* in a previous exercise in 2018 and we found similar issues

with its identification. In 2018, there were several analysts that identified *Levanderina* erroneously as *Akashiwo sanguinea.* In 2024, we had a similar pattern of misidentifications for this species. It is interesting to note that it does not happen the other way around. In 2017, we included *Akashiwo sanguinea* in our materials and analysts had no difficulty identifying it. So, it is more likely to identify incorrectly *Levanderina* as *Akashiwo* than the other way around. *Gymnodinium/Gyrodinium* were accepted as correct answers for *Levanderina*. There were also some issues with the identification of *Helicotheca tamesis* which was mistaken by *Mediopyxis helysisa* by several analysts. Both are very similar species.

The other interesting result was the identification of *Prorocentrum gracile* to species level. 52 analysts identified the species as *P.gracile* while 32 identified them as *P.micans*. This was a surprising result taking into consideration that both are species that are cosmopolitan and found all over the world and that we all consider to a certain extent as easy species to identify. This suggests that their identification is not as simple as suggested and that the different varieties and morphotypes may account for the split in this identification. We should attempt to define these species better to avoid these unexpected identification conflicts in the future.

In relation to the quantification of the measurands in the samples, the only species with a higher number of out of specification results were *Pseudo-nitzschia* and *Ostreopsis* with 7 and 8 yellow flags respectively. Both species were included in the Ocean Teacher exercise in Q12 and Q20 as numerical questions. A video clip showing a transect of short duration for analysts to enumerate the cells in the transect, a sort of repeatability study where all analysts are looking at the same number of cells. In OT, for Q12, most analysts performed well, so the issue lies with the ability to identify the species in samples rather than counting. For *Pseudo-nitzschia* the issue is with counting rather than identifying them. This is clear from the results in Q20 of the OT exercise, where 30 analysts were outside of the tolerance for that cell count which was +/- 10 cells. There was not real consensus for this count.

The Ocean Teacher 2024 exercise was challenging this year, and the results were compounded by some technical issues on our side. There were some glaring errors like in Q11 *Grammatophora* was entered as a 'centric' diatom instead of a pennate diatom and in Q15 with the repetition of 'ribbons' and 'stellate chains' included twice in the question requires that there are entered in the right order to be computed as correct. There were also technical issues with Q6 and Q10 in drag and drop type questions where more than one object has the same dropping zone. Both

questions had to be excluded from the final score unfortunately. As we probe the capabilities of the Moodle software in Ocean Teacher, we are finding some inconsistencies with question behaviour and some limitations as to what we are capable to do.

There were other issues of a taxonomical nature that had to be answered in relation to several queries that were resolved fairly by the Advisory Expert Committee. These issues were explained in the results section. Q13 was excluded also from the final score because how the question was phrased and that it was interpreted in a way that was not intended.

Overall, the final scores and grades were not dissimilar to previous years, with a slight increase on failed identification and enumeration results which resulted in an increase on unsuccessful attempts at the test.