UNESCO
United Nations
Educational, Scientific and
Cultural Organization

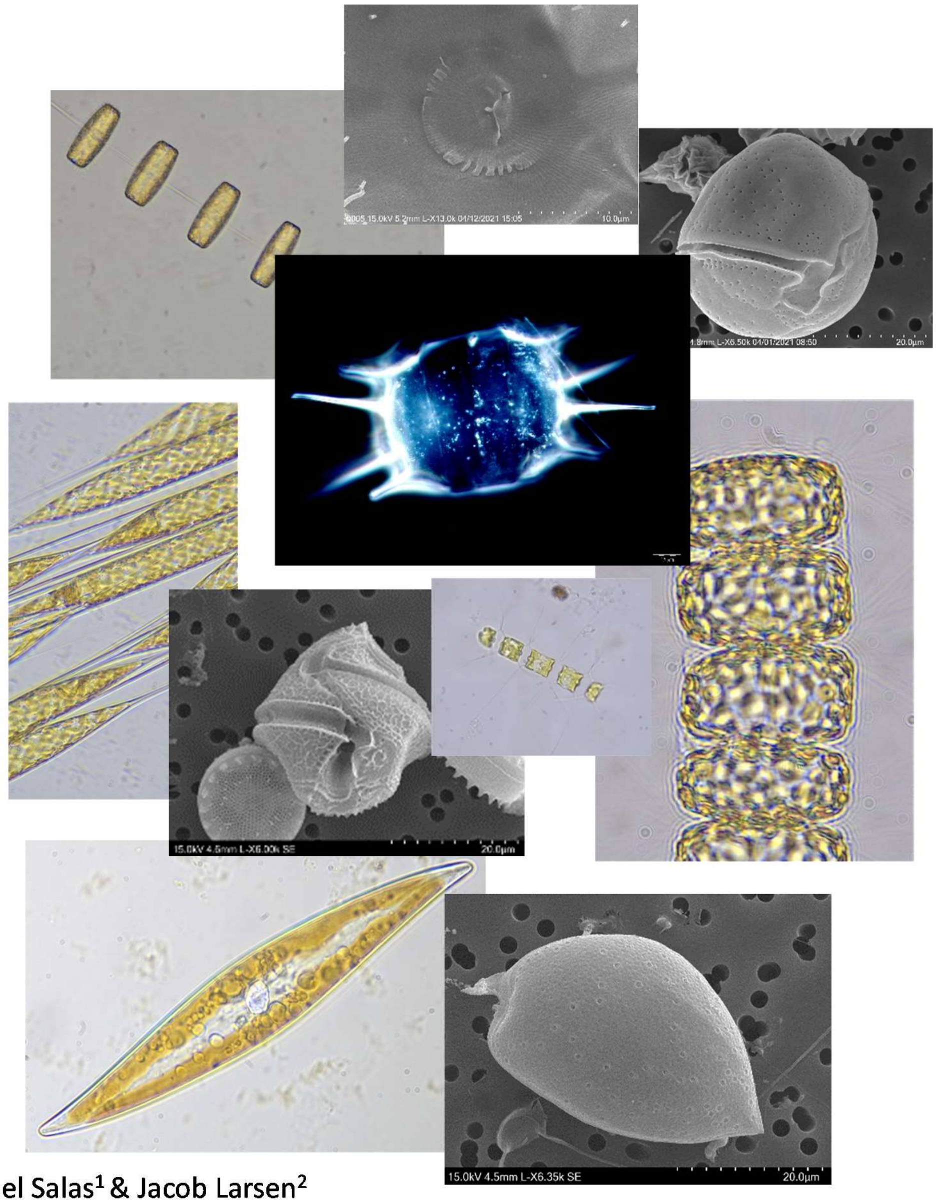
Intergovernmental
Oceanographic
Commission

IPI
INTERNATIONAL PHYTOPLANKTON
INTERCOMPARISON

ULPGC
Universidad de
Las Palmas de
Gran Canaria

OCHABs
Observatorio Canario de algas nocivas

# INTERNATIONAL PHYTOPLANKTON INTERCOMPARISON (IPI) Proficiency testing in the abundance and composition of marine microalgae 2021 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 2021, 124 analysts across 59 laboratories around the world participated in the IPI exercise. European countries accounted for 78% of the total participation, 7% came from South America, 7% from African countries, 4% from Oceania and 4% from Asia.

- 14 species were used in total but only 10 species were inoculated per sample. There were four dinoflagellates and ten diatoms in the samples distributed in a batch system.

- The dinoflagellates were ***Alexandrium minutum*** *(K.Hirasaka)* Halim, 1960**,** ***Prorocentrum micans*** Ehrenberg, 1834, ***Gonyaulax spinifera*** *(Claparède & Lachmann) Diesing, 1866* and ***Coolia monotis*** Meunier, 1919.

- The Diatoms species were ***Bacillaria paxillifer*** (O.F.Müller) T.Marsson, 1901, ***Chaetoceros didymus*** Ehrenberg, 1845, ***Coscinodiscus centralis*** Ehrenberg, 1844, ***Lauderia annulata*** Cleve, 1873, ***Odontella aurita*** (Lyngbye) C.Agardh, 1832, ***Pleurosigma*** W. Smith, 1852, ***Pseudo-nitzschia delicatissima*** **complex** (Cleve) Heiden, 1928, ***Rhizosolenia setigera*** Brightwell, 1858, ***Thalassiosira rotula/gravida*** Meunier, 1910, ***Trieres mobiliensis*** (J.W.Bailey) Ashworth & E.C.Theriot in Ashworth, Nakov & E.C.Theiriot, 2013

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

- All measurands passed the expanded criterion for homogeneity and stability according to ISO13528:2015 except for *C.centralis* and *O.aurita* which we discuss in the main body of the report.

- There were a very small number of warning and action signals across measurands. 10 Red flags (1.2%), 31 (2.5%) yellow flags and 29 (2.3%) non-detection flags (Grey) from 1220 results is evidence of good performance overall.

- 21 analysts weren't successful at the overall test. 9 analysts failed the quantitation, with 4 analysts just below the requirement with three failed test items (70%) and 4 analysts with 4 failed items (60%) need some improvement. One analyst failed 6 out 10 items (40%) score requires training and improvement for the next round.

- The other 12 analysts failed the qualitative test with 10 failing 3 identifications and 2 analysts failing 4 identifications. There were 3 analysts that failed the test quantitatively and qualitatively.

- The hardest species to recognize in this test was *Coolia monotis* which was not detected by 16 analysts and erroneously classified by 44 analysts, one third of the total. 32 analysts confused this species with *Alexandrium ostenfeldii* or *tamarense.*

- The most undetected species in the samples was also *Coolia monotis* which had a relatively low cell density. 16 analysts did not detect this organism compared with 5 analysts for *A.minutum* or 3 for *G.spinifera.* Generally, dinoflagellates were harder to identify than diatoms. 24 non-detections on 4 dinoflagellates compared to 6 non-detections on 14 diatoms

- Overall, from 1220 possible correct identifications, there were a total of 784 correct answers at species level (64%) and 1024 correct answers at genus level that is 84% correct, 171 (14%) incorrect identifications.

- There were 113 attempts at the OTGA HAB asessement. The median overall grade was 91.3%. 57.5% of analysts performed above the proficiency threshold of 90% and 27.4% of all analysts between 80-90%. 7.9% above 70% and another 7.9% below 70% requiring improvement.

- The OTGA facility index shows that the worst answered question in the test was Q10 (61%) a numerical question and the best Q1(98%) a question about *Pseudo-nitzschia* terminology, this was closely matched by the percentage of correct identifications for Q3 through Q9 on species identification of that same genus.

## 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.

14 species were used in the IPI2021 samples, but only 10 species were inoculated per sample. There were four dinoflagellates and ten diatoms in the samples distributed in a batch system. The dinoflagellates were *Alexandrium minutum* (K.Hirasaka) Halim, 1960, *Prorocentrum micans* Ehrenberg, 1834, *Gonyaulax spinifera* (Claparède & Lachmann) Diesing, 1866 and *Coolia monotis* Meunier, 1919. The Diatoms species were *Bacillaria paxillifer* (O.F.Müller) T.Marsson, 1901, *Chaetoceros didymus* Ehrenberg, 1845, *Coscinodiscus centralis* Ehrenberg, 1844, *Lauderia annulata* Cleve, 1873, *Odontella aurita* (Lyngbye) C.Agardh, 1832, *Pleurosigma* W. Smith, 1852, *Pseudo-nitzschia delicatissima complex* (Cleve) Heiden, 1928, *Rhizosolenia setigera* Brightwell, 1858, *Thalassiosira rotula/gravida* Meunier, 1910, *Trieres mobiliensis* (J.W.Bailey) Ashworth & E.C.Theriot in Ashworth, Nakov & E.C.Theiriot, 2013.

From 2021 to 2025, the IPI programme 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 2021, 124 analysts in 59 laboratories from across the world participated in the IPI exercise. European countries accounted for 78% of the total participation, 7% from South America, 7 % from African countries, 4% from Oceania and 4% from Asia (Figure 1). 22 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.

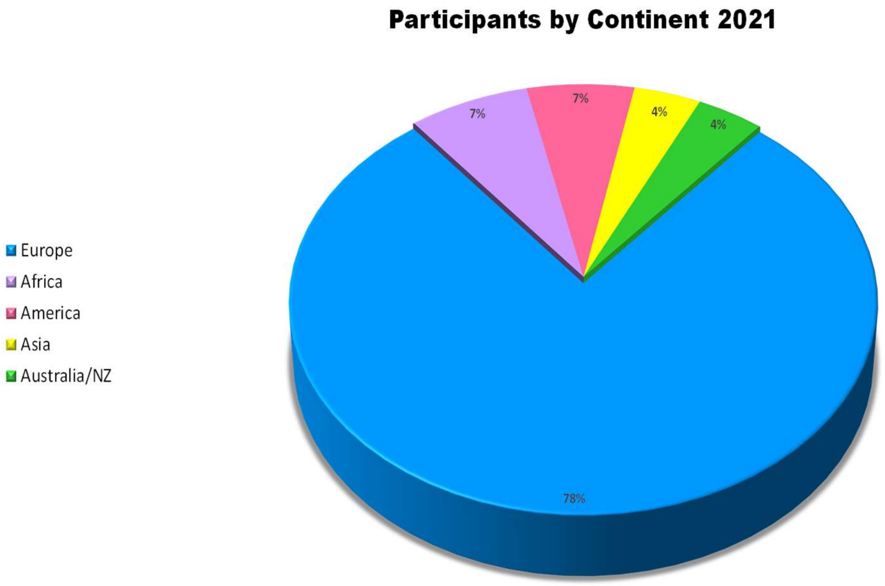


**Figure 1**: Participants by continent IPI2021

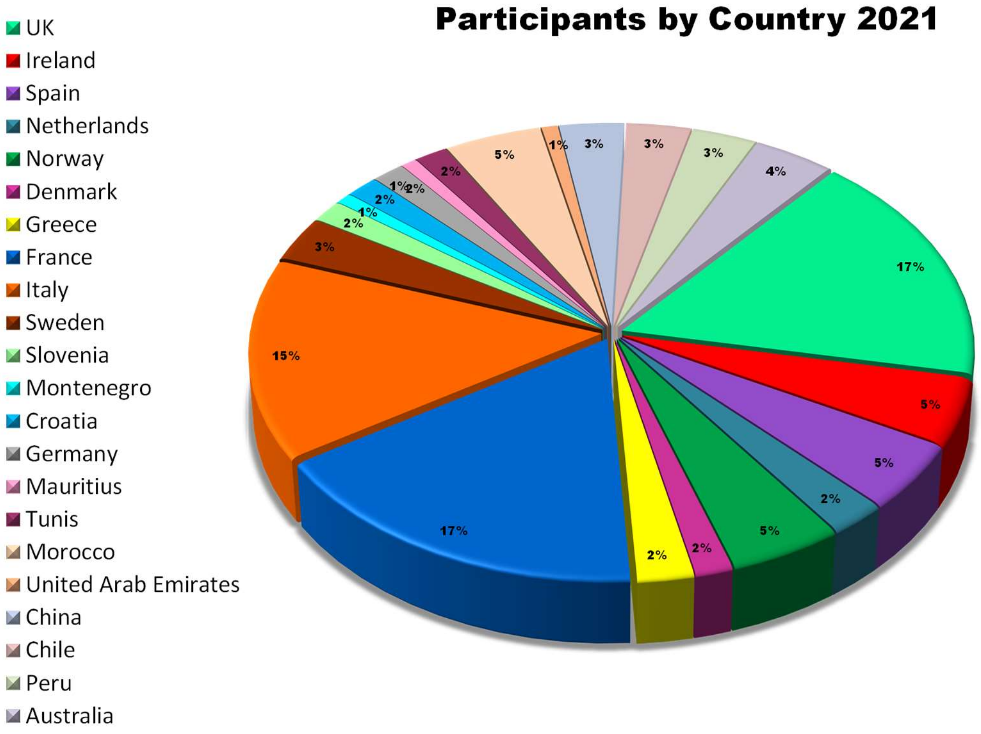


**Figure 2**: Participants by country IPI 2021

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-2021. 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). This year we reached the highest number of analysts (124) and the largest number of laboratories (59).

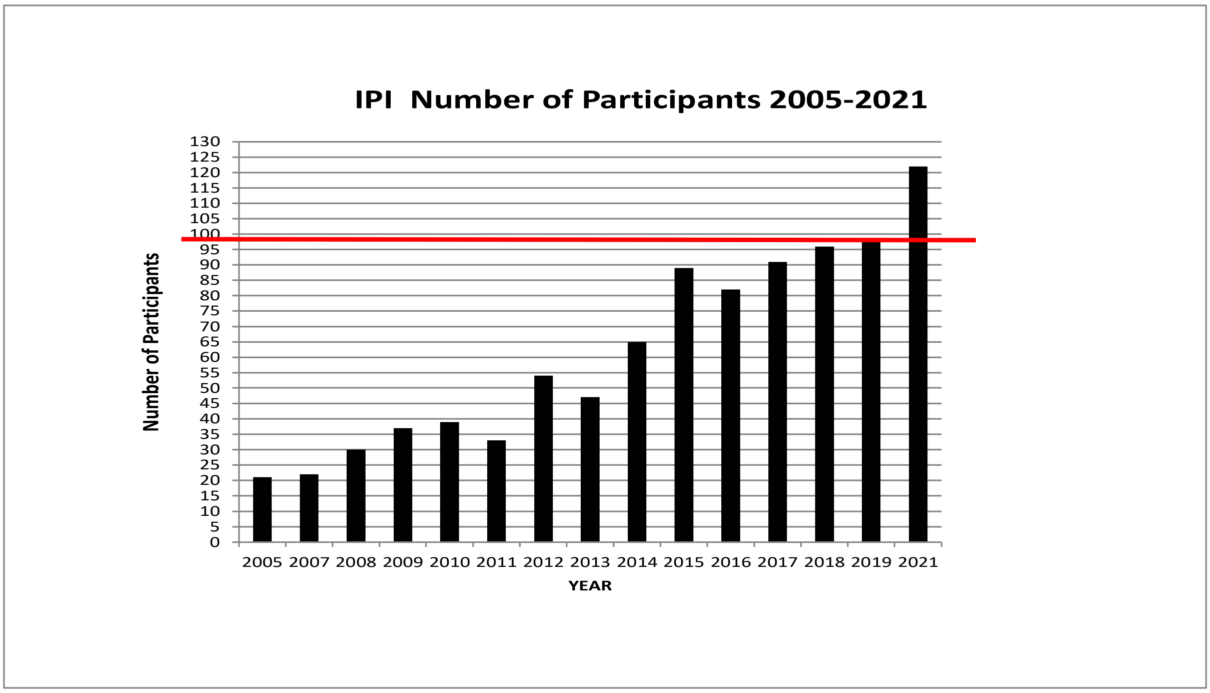


**Figure 3:** IPI participation since 2005

Many laboratories participate on a regular basis and several analysts have more than 14 result contributions since 2005. In 2021, it is the first time we have two Chinese laboratories participating in the scheme (Figure 2).

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

## 3. Materials and Methods

### 3.1 Sample preparation, homogenization and inoculation

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

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

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

Each ampoule was labeled with a sequential number and each box of ampoules was also labeled to differentiate sample sets produced from different working stocks (IPI2021 batches #1.1, #1.2, #1.3, #1.4, #2.1, #2.2, #2.3, #2.4) 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 West of Ireland. Most species were identified through light microscopy techniques using an inverted microscope Olympus IX-51 and a compound research Olympus microscope 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 1152 ampoules were produced for this study. Each participant was sent a set of four replicates. 124 analysts in 59 laboratories were sent a total of 496 ampoules. Each sample set consisted of a padded brown envelope containing 4 ampoules, 4 x 50 ml skirted centrifuge tubes and 4 plastic droppers.

### 3.3 Cell concentrations

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

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

### 3.4 Sample randomization

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

### 3.5 Forms and instructions

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

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

At the end of the exercise and with the publication of this report, analysts will be issued with a statement of performance certificate (Annex V in separate annex report) which is tailored specifically for each test. This is an important document for auditing purposes and ongoing competency.

### 3.6 Statistical analysis

Statistical analysis was carried out using PROlab Plus version 2021.7.22.0 dedicated software for the statistical analysis of intercalibration and proficiency testing exercises from Quodata, and Microsoft office Excel 2016.

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

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

### 3.7 IPI Ocean teacher online taxonomic assessment

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

This year, participants were sent information from ioc.training@unesco.org to register to the OTGA website. The preparatory phase consisted of an online quiz made available on the IOC/OceanTeacher e-Learning Platform, direct link https://classroom.oceanteacher.org/course/view.php?id=730.

In order, to access the quiz, participants had to create an account on OceanTeacher (www.oceanteacher.org), not later than 8 October 2021 (Central European time). Once they received confirmation of their account, each participant then was able to enrol to the course (link above using the enrolment key IPI#730). Participants that already have an account on OT were able, instead, to enrol directly using the link and enrolment key to the course/quiz as from 11 October 2021 onwards. 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/event/2991. 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 (this means these are 2 different accounts and as such usernames and passwords are not necessarily the same).

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 XV). Question types used in the quiz were; 'matching type' (Q1, 2, 11) which have dropdown menus including a selection of answers which analysts must choose from, 'multiple choice' (Q 3-9, 12-13, 15, 18) where the participant must fill in the right option from those given, 'numerical' (Q10, 14, 17, 20) where analysts had to watch a video clip of a transect containing cells to be counted and 'drag and drop' types (Q16, 19) 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. In multiple choice type questions, we introduced penalties for wrong answers, where an incorrect choice incurs a percentage deduction. The amount of this deduction depends on the number of possible answers and ranges from 5% to 25% per wrong answer.

The online quiz can only be submitted once. After submission, no changes can be made. However, analysts can login and out as many times as they wish throughout the allocated time periods and make changes. The changes are saved and can be accessed at a later stage, as long as participants don't press submit.

## 4. Results

### 4.1 Homogeneity and stability study

The 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 2021.7.22.0 and the reports for homogeneity and stability are given separately for each measurand. The top of the report gives you information on the measurand, mean and analytical standard deviation for the homogeneity analysis and the homogeneity and stability mean comparison in the stability analysis. The reports, also show the target standard deviation for each measurand, which in this case was calculated manually using the consensus results of the participants and taking into consideration the heterogeneity of the samples, as will be explained later.

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

According to ISO 13528:2015, the heterogeneity standard deviation (s(sample)) between the proficiency test items should not exceed 30 % of the standard deviation for the proficiency assessment. If the homogeneity test fails, the heterogeneity standard deviation is then, taken into consideration, when calculating the standard deviation for the measurand. The consensus values new heterogeneity standard deviation (STD) was used for all measurands as most items failed the adequate homogeneity criterion except for *A.minutum*, *P.micans*, *Pseudo-nitzschia delicatissima* and *Thalassiosira rotula/Gravida* (table 1). However, no significant heterogeneity was found according to the expanded criterion except for *C. centralis* and *O.aurita*.

In the *C. centralis* case, low cell densities close to the limit of detection creates large variance in the precision of the measurements as the difference between seeing one cell or no cells is a 100% difference. Also, homogenization of large organisms is more challenging as the species tend to settle down straightaway after movement stops. This is an issue that requires using this type of species carefully in future exercises.

In the *O.aurita* cell counts, something altogether different happened. Some of the *O.aurita* cells did not break down in smaller chains after homogenization, creating samples with large variance within portions of the same sample. If you look at the homogeneity test for *O.aurita* in Annex VI Pgs 34-35, you will find that the data exhibit significant heterogeneity, this can be gleamed from the graphical representation of *O.aurita* results from the analysts in Annex XIII pg 96. In this graph, analysts found big differences between samples. The sample difference between two aliquots of the same sample meant that homogenization was not easily achieved for this species, as a chain former, if the chains were large and did not break down enough, one sample could end up with one large chain and the other with smaller amounts. This is, demonstrated by the analysts results which show the largest variance between replicates for these species (see Annex XIII: Graphical summary of results in the annex report for *O.aurita*.

Hence, the proficiency test items cannot be considered fully homogeneous but not significantly heterogeneous (Table 1) except for *O.aurita* and *C.centralis*.

| ISO13528 | 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 |
|---|---|---|---|---|---|---|
| Alexandrium minutum | no outliers found | Ok | Ok | Ok | Ok | Ok |
| Bacillaria paraxillifer | no outliers found | Ok | Not OK | Ok | Ok | Ok |
| Chaetoceros didymus batch 1 | no outliers found | Not OK | Not OK | Ok | Ok | Ok |
| Chaetoceros didymus batch 2 | no outliers found | Not OK | Not OK | Ok | Ok | Ok |
| Coolia monotis | no outliers found | Ok | Not OK | Ok | Ok | Ok |
| Coscinodiscus centralis | no outliers found | Not OK | Not OK | Not OK | Not OK | Not OK |
| Gonyaulax spinifera | no outliers found | Not OK | Not OK | Ok | Ok | Ok |
| Lauderia annulata | no outliers found | Not OK | Not OK | Ok | Ok | Ok |
| Odontella aurita | no outliers found | Not OK | Not OK | Not OK | Not OK | Ok |
| Pleurosigma-Gyrosigma | no outliers found | Not OK | Not OK | Ok | Ok | Ok |
| Prorocentrum micans | no outliers found | Ok | Ok | Ok | Ok | Ok |
| Pseudo-nitzschia delicatissima | no outliers found | Ok | Ok | Ok | Ok | Ok |
| Rhizosolenia setigera batch 1 | no outliers found | Not OK | Not OK | Ok | Not OK | Ok |
| Rhizosolenia setigera batch 2 | no outliers found | Ok | Not OK | Ok | Ok | Ok |
| Thalassiosira rotula/gravida | no outliers found | Ok | Ok | Ok | Not OK | Ok |
| Trieres Mobiliensis | no outliers found | Not OK | Not OK | Ok | Ok | Ok |

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

As all analysts achieved good Z-scores for *O.aurita* within 2SD of the test, there is no need to change anything there and we can continue using this data in the test and in the final certification for analysts.

In the case of *C.centralis,* we must be careful with the data. It is possible that some samples contained no cells, the homogeneity test showed that when you divide a sample in two halves, one may contain no cells, while the other may contain all of them. This happened in one homogeneity test sample of 26 analysed. In the case of analysts, if they only analysed one half of the sample, there was a possibility that some samples would not contain any cells at all or too many.

Said that, this didn't affect too many analysts, most analysts manage to spot several cells in the samples and return an average for this measurand but 3 analysts failed to detect the cells (analysts 56, 33 and 129) and 2 others (1 and 17) gave an inflated result, most likely as a consequence of this heterogeneity. The results on this measurand for these analysts, therefore should be forfeited in this case as *C.centralis* are a large diatom and it would be hard to miss in the samples.

Again, *Cocinodiscus* in general are large heavy diatoms that will deposit very quickly after homogenization stops and will create this type of difficulty. I would recommend for further exercises that analysts analyse both halves of the sample and return an average to make sure there are not issues with lab homogenization.

In relation to the stability test, most items were considered stabled according to the expanded criterion (table 1), except for *C.centralis*, this again is an effect of the difficulty homogenizing these species given their cell size and compounded by their low cell density values for an otherwise, rather easy diatom to identify and hard to miss given their size

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

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

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

### 4.3 Analysts' Data

The full table of participants' results can be found in Annex VII pgs 50-56 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;

$\sigma_{r1}$ =the new SD for the homogeneity test

$\sigma_r$ =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;

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

B)

Where;

$u_x$= Standard uncertainty of the assigned value,

$s^*$= robust standard deviation for the test

$p$= number of analysts

| Species | A.minutum | P.micans | G.spinifera | C.monotis | R.setigera #1 | O.aurita | C.didymus #1 | Thalassiosira gravida/rotula |
|---|---|---|---|---|---|---|---|---|
| Robust mean x* | 4270 | 3066 | 1888 | 1136 | 3638 | 1332 | 14192 | 8985 |
| Robust Stdev s* | 1274 | 842 | 725 | 371 | 591 | 538 | 7482 | 1474 |
| Standard Ux | 147 | 95 | 82 | 45 | 95 | 86 | 1207 | 238 |
| n= | 118 | 123 | 121 | 107 | 61 | 61 | 60 | 60 |
| if Ux < 0.3xSTdev | 382 | 253 | 218 | 111 | 177 | 161 | 2245 | 442 |
| then Ux is negligible | neg | neg | neg | neg | neg | neg | neg | neg |
| **The equation is satisfied in all cases** | | | | | | | | |

| Species | B.paxillifer | R.setigera #2 | T.mobiliensis | C.didymus #2 | L.annulata | P.delicatissima group | Pleurosigma/ Gyrosigma | C.centralis |
|---|---|---|---|---|---|---|---|---|
| Robust mean x* | 92319 | 1703 | 19306 | 7950 | 11722 | 92233 | 1101 | 333 |
| Robust Stdev s* | 19363 | 466 | 3483 | 3698 | 2901 | 28968 | 258 | 158 |
| Standard Ux | 3125 | 74 | 553 | 592 | 461 | 4599 | 41 | 26 |
| n= | 60 | 62 | 62 | 61 | 62 | 62 | 62 | 58 |
| if Ux < 0.3xSTdev | 5809 | 140 | 1045 | 1109 | 870 | 8690 | 77 | 47 |
| then Ux is negligible | neg | neg | neg | neg | neg | neg | neg | neg |
| **The equation is satisfied in all cases** | | | | | | | | |

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

The performance statistics for the exercise have been calculated using ProLab Plus Version 2021.7.22.0. The summary table of all the Z-scores can be found in Annex IX of the annex report. The performance statistics (Annex X) 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/Huber), 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.

### 4.4.1 Z-scores

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

There were a very small number of warning and action signals across measurands. 10 Red flags (1.2%), 31 (2.5%) yellow flags and 29 (2.3%) non-detection flags (Grey) from 1220 results is evidence of good performance overall.

Nine analysts failed the quantitative part of the test (see annex X). Four analysts (70%) are just below the requirement with three failed test items and 4 analysts (60%) failed 4 items need some improvement. One analyst (40%) score failed 6 out 10 items require training and improvement for the next round.
85 analysts had all the measurands (10) within the tolerance limits, 19 analysts had one fail measurand and 9 analysts two.

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

The chart of RLP against RSZ (Annex XIV in the annex report) for all measurands combine all the analysts results for the 10 measurands in one combined Z-score which indicates how close or far the analyst is to the consensus robust average and standard deviation. The X axis shows the RSZ or the Rescaled sum of all your Z-scores for the test and the y axis shows the RLP or relative Lab performance, which indicates the length of the combined standard deviations. In other words, systematic laboratory bias. 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 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 plots of repeatability standard deviations or Lischer plots (Annex XV in the annex report) are something similar but measurand by measurand, instead of all combined. Here, you may be able to glean other problems more specific to the identification and counting of certain species. Perhaps, a tendency to underestimate particular species or group of species or have a particular difficulty with dinoflagellates and not diatoms for example. These graphs will show how you did compared to everyone else in a very interactive way.

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

Analyst performance on the correct composition of species in the samples was generally quite good (Table 3). To pass the qualitative test, analysts must identify correctly at least 80% of the measurands, that is at least 8 of the species in the samples. In 2021, 12 analysts failed the qualitative test. 9 analysts identified incorrectly 3 measurands and 3 analysts up to 4 measurands. Of these analysts, 3 also failed to pass the 80% Z-scores. A non-detection (ND) is also a fail flag.

*Prorocentrum micans* was the easiest to recognize of the dinoflagellates. Most analysts (121) identified to species level (table 4). *Gonyaulax spinifera* was recognized to species level by 78 analysts with a further 20 identifying to genus level, that is an 80% correct identification rate.

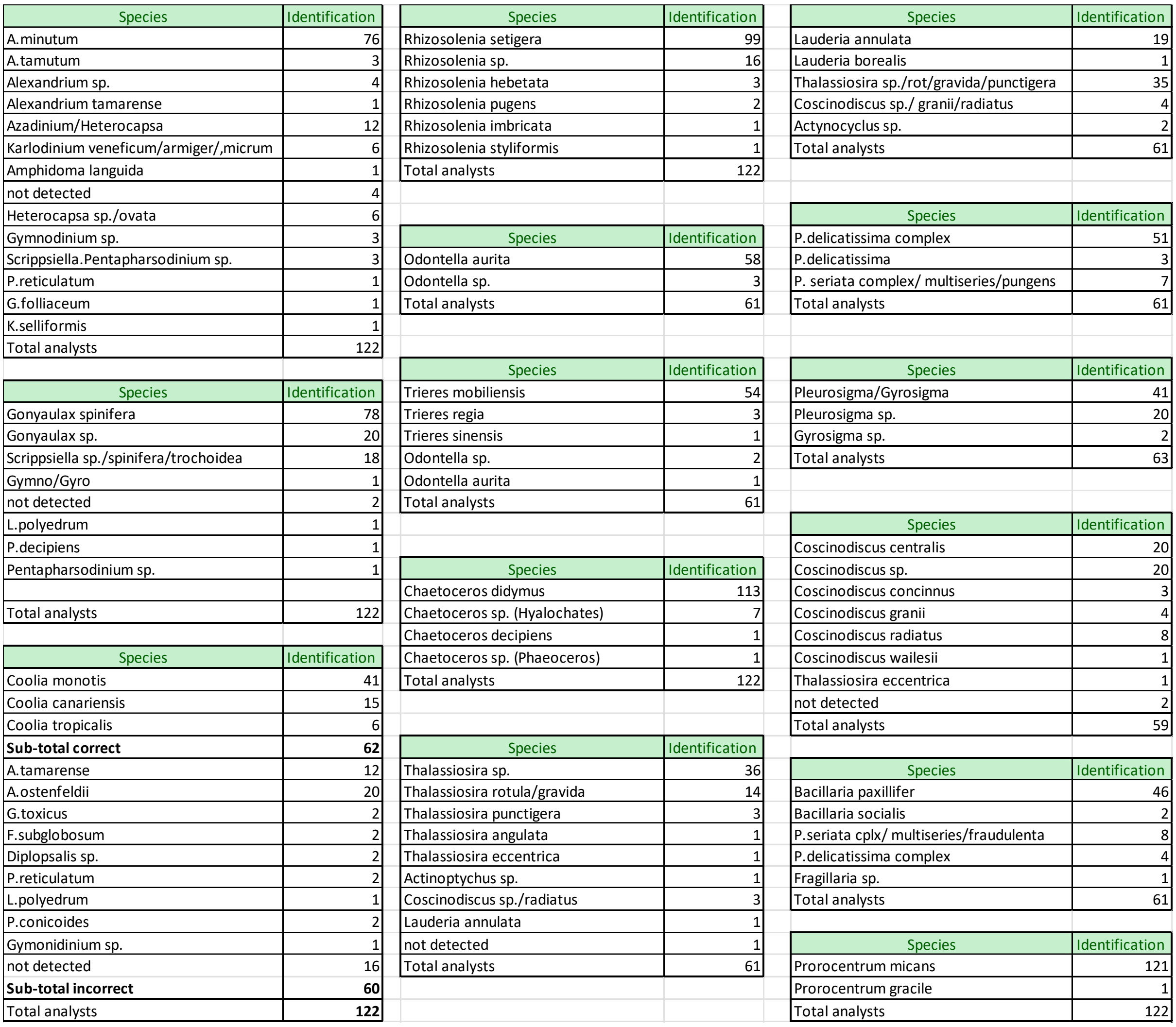
*Scrippsiella* was used by a further 18 analysts as the most popular incorrect answer to this identification.

| A. Code | *AMIN* | *PMIC* | *GSPIN* | *CMON* | *RSET* | *CDIDY* | *OAUR* | *THALA* | *BPAXI* | *PLEU* | A. Code | *AMIN* | *PMIC* | *GSPIN* | *CMON* | *RSET* | *CDIDY* | *OAUR* | *THALA* | *BPAXI* | *CCENT* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 120 | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 16 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 5 | ✓ | ✓ | ✓ | ND | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 50 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 87 | ✓ | ✓ | X | ND | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 79 | ✓ | ✓ | ✓ | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 45 | X | ✓ | X | ✓ | ✓ | ✓ | ✓ | ✓ | X | ✓ | 90 | ✓ | ✓ | X | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 40 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 117 | ✓ | ✓ | ✓ | ND | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 118 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 77 | ✓ | ✓ | ✓ | ND | ✓ | ✓ | ✓ | ✓ | X | ✓ |
| 8 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 115 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 76 | ✓ | ✓ | ✓ | X | ✓ | ✓ | ✓ | ✓ | X | ✓ | 65 | ND | ✓ | ND | ND | ✓ | ✓ | ✓ | ✓ | X | ✓ |
| 14 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 82 | X | ✓ | ✓ | X | ✓ | ✓ | ✓ | ✓ | X | ✓ |
| 134 | X | ✓ | ✓ | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 78 | ✓ | ✓ | ✓ | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 53 | ✓ | ✓ | ✓ | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 55 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 29 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 89 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 126 | ✓ | ✓ | X | X | ✓ | ✓ | ✓ | ✓ | X | ✓ | 88 | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 105 | ✓ | ✓ | ✓ | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 70 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 9 | ✓ | ✓ | ✓ | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 35 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 4 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 59 | ✓ | ✓ | X | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 103 | ND | ✓ | ND | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 93 | ✓ | ✓ | ✓ | X | ✓ | ✓ | ✓ | ✓ | X | ✓ |
| 10 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 42 | X | ✓ | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 128 | ✓ | ✓ | ✓ | ND | ✓ | ✓ | ✓ | ✓ | X | ✓ | 24 | ✓ | ✓ | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 84 | ✓ | ✓ | ✓ | ND | ✓ | ✓ | ✓ | ✓ | X | ✓ | 17 | ✓ | ✓ | X | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 104 | ✓ | ✓ | ✓ | ND | ✓ | ✓ | ✓ | ✓ | X | ✓ | 81 | ND | ✓ | ✓ | X | ✓ | ✓ | ✓ | ✓ | X | ✓ |
| 71 | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 72 | ✓ | ✓ | ✓ | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 110 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 7 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 54 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 6 | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 111 | ✓ | ✓ | ✓ | X | ✓ | ✓ | ✓ | ✓ | X | ✓ | 56 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 85 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 131 | ✓ | ✓ | ✓ | ND | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 15 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 133 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 119 | ✓ | ✓ | X | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 32 | X | ✓ | ✓ | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 109 | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 127 | X | ✓ | ✓ | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 26 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |  |  |  |  |  |  |  |  |  |  |  |
| 125 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |  |  |  |  |  |  |  |  |  |  |  |
| 38 | X | ✓ | X | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |  |  |  |  |  |  |  |  |  |  |  |

| A. Code | AMIN | PMIC | GSPIN | CMON | RSET | CDIDY | TMOB | LAUD | PDEL | CCEN | A. Code | AMIN | PMIC | GSPIN | CMON | RSET | CDIDY | TMOB | LAUD | PDEL | PLEU |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 49 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 13 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | X | ✓ |
| 75 | ✓ | ✓ | ✓ | X | ✓ | ✓ | ✓ | ✓ | X | ✓ | 27 | X | ✓ | X | ND | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 83 | X | ✓ | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 2 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 94 | ✓ | ✓ | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 41 | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 98 | ✓ | ✓ | X | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 96 | X | ✓ | ✓ | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 11 | ✓ | ✓ | ✓ | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 62 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 48 | X | ✓ | ✓ | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 114 | ✓ | ✓ | ✓ | X | ✓ | ✓ | ✓ | ✓ | X | ✓ |
| 116 | ✓ | ✓ | ✓ | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 19 | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | X | ✓ |
| 22 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 58 | ✓ | ✓ | ✓ | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 97 | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 91 | X | ✓ | X | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 51 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 36 | ✓ | ✓ | ✓ | ND | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 33 | ✓ | ✓ | X | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 25 | X | ✓ | ✓ | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 69 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 44 | X | ✓ | ✓ | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 129 | ✓ | ✓ | ✓ | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 34 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 61 | ✓ | ✓ | ✓ | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 68 | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 60 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 21 | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 102 | X | ✓ | X | ND | ✓ | ✓ | ✓ | ✓ | X | ✓ | 23 | ✓ | ✓ | ✓ | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 20 | ✓ | ✓ | ✓ | X | ✓ | ✓ | ✓ | ✓ | X | ✓ | 63 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 80 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 3 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 66 | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 64 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 95 | X | ✓ | ✓ | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 122 | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 108 | ✓ | ✓ | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 112 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 57 | X | ✓ | ✓ | ND | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 113 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 18 | ✓ | ✓ | ✓ | ND | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 46 | ND | ✓ | ✓ | ✓ | ✓ | X | ✓ | ✓ | ✓ | ✓ |
| 43 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 73 | ✓ | ✓ | ✓ | ND | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 47 | X | ✓ | X | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 67 | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 74 | X | ✓ | X | ND | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 100 | X | ✓ | ✓ | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 132 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 92 | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 1 | X | ✓ | X | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 86 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 12 | X | ✓ | ✓ | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 124 | ✓ | ✓ | ✓ | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
|  |  |  |  |  |  |  |  |  |  |  | 123 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

Table 3: Qualitative results IPI2021 by Analyst and Measurand. Not- detected (ND); Correct ( ✓); Incorrect (**x**)

For *Alexandrium minutum* the rate was similar to Gonyaulax with 76 analysts identifying to species level and a further 8 analysts to genus level, the rate was slightly lower at 69% correct identification. The hardest species to recognize in this test was *Coolia monotis* which was not detected by 16 analysts and erroneously classified by 44, one third of the total. 32 analysts confused this species with *Alexandrium ostenfeldii* or *tamarense.*



| Species | Identification |
|---|---|
| A.minutum | 76 |
| A.tamutum | 3 |
| Alexandrium sp. | 4 |
| Alexandrium tamarense | 1 |
| Azadinium/Heterocapsa | 12 |
| Karlodinium veneficum/armiger/,micrum | 6 |
| Amphidoma languida | 1 |
| not detected | 4 |
| Heterocapsa sp./ovata | 6 |
| Gymnodinium sp. | 3 |
| Scrippsiella.Pentapharsodinium sp. | 3 |
| P.reticulatum | 1 |
| G.folliaceum | 1 |
| K.selliformis | 1 |
| Total analysts | 122 |

| Species | Identification |
|---|---|
| Gonyaulax spinifera | 78 |
| Gonyaulax sp. | 20 |
| Scrippsiella sp./spinifera/trochoidea | 18 |
| Gymno/Gyro | 1 |
| not detected | 2 |
| L.polyedrum | 1 |
| P.decipiens | 1 |
| Pentapharsodinium sp. | 1 |
| | |
| Total analysts | 122 |

| Species | Identification |
|---|---|
| Coolia monotis | 41 |
| Coolia canariensis | 15 |
| Coolia tropicalis | 6 |
| **Sub-total correct** | **62** |
| A.tamarense | 12 |
| A.ostenfeldii | 20 |
| G.toxicus | 2 |
| F.subglobosum | 2 |
| Diplopsalis sp. | 2 |
| P.reticulatum | 2 |
| L.polyedrum | 1 |
| P.conicoides | 2 |
| Gymonidinium sp. | 1 |
| not detected | 16 |
| **Sub-total incorrect** | **60** |
| Total analysts | **122** |

| Species | Identification |
|---|---|
| Rhizosolenia setigera | 99 |
| Rhizosolenia sp. | 16 |
| Rhizosolenia hebetata | 3 |
| Rhizosolenia pugens | 2 |
| Rhizosolenia imbricata | 1 |
| Rhizosolenia styliformis | 1 |
| Total analysts | 122 |

| Species | Identification |
|---|---|
| Odontella aurita | 58 |
| Odontella sp. | 3 |
| Total analysts | 61 |

| Species | Identification |
|---|---|
| Trieres mobiliensis | 54 |
| Trieres regia | 3 |
| Trieres sinensis | 1 |
| Odontella sp. | 2 |
| Odontella aurita | 1 |
| Total analysts | 61 |

| Species | Identification |
|---|---|
| Chaetoceros didymus | 113 |
| Chaetoceros sp. (Hyalochates) | 7 |
| Chaetoceros decipiens | 1 |
| Chaetoceros sp. (Phaeoceros) | 1 |
| Total analysts | 122 |

| Species | Identification |
|---|---|
| Thalassiosira sp. | 36 |
| Thalassiosira rotula/gravida | 14 |
| Thalassiosira punctigera | 3 |
| Thalassiosira angulata | 1 |
| Thalassiosira eccentrica | 1 |
| Actinoptychus sp. | 1 |
| Coscinodiscus sp./radiatus | 3 |
| Lauderia annulata | 1 |
| not detected | 1 |
| Total analysts | 61 |

| Species | Identification |
|---|---|
| Lauderia annulata | 19 |
| Lauderia borealis | 1 |
| Thalassiosira sp./rot/gravida/punctigera | 35 |
| Coscinodiscus sp./ granii/radiatus | 4 |
| Actynocyclus sp. | 2 |
| Total analysts | 61 |

| Species | Identification |
|---|---|
| P.delicatissima complex | 51 |
| P.delicatissima | 3 |
| P. seriata complex/ multiseries/pungens | 7 |
| Total analysts | 61 |

| Species | Identification |
|---|---|
| Pleurosigma/Gyrosigma | 41 |
| Pleurosigma sp. | 20 |
| Gyrosigma sp. | 2 |
| Total analysts | 63 |

| Species | Identification |
|---|---|
| Coscinodiscus centralis | 20 |
| Coscinodiscus sp. | 20 |
| Coscinodiscus concinnus | 3 |
| Coscinodiscus granii | 4 |
| Coscinodiscus radiatus | 8 |
| Coscinodiscus wailesii | 1 |
| Thalassiosira eccentrica | 1 |
| not detected | 2 |
| Total analysts | 59 |

| Species | Identification |
|---|---|
| Bacillaria paxillifer | 46 |
| Bacillaria socialis | 2 |
| P.seriata cplx/ multiseries/fraudulenta | 8 |
| P.delicatissima complex | 4 |
| Fragillaria sp. | 1 |
| Total analysts | 61 |

| Species | Identification |
|---|---|
| Prorocentrum micans | 121 |
| Prorocentrum gracile | 1 |
| Total analysts | 122 |

Table 4: Qualitative data by measurand.

Diatoms depended on batch assignation, except for *Chaetoceros didymus* and R*hizosolenia setigera* which were found in both batches but at different abundances. Diatoms generally did not create any problems among participants and there were no differences in performance between batch #1 or #2 participants. All analysts were able to identify correctly these species except for a small number of participants (Table 4).

There is a doubt about the full identity of *Chaetoceros didymus* where some analysts said that it could be *Chaetoceros protuberans*. This could be the case and we can not fully ascertain this at this point. The light microscopy images we have point to the likelihood and probability that it is indeed *C.protuberans*. Either way, this doesn't deflect from the fact that all identifications were correct or the most correct they could be, in this instance as both species are extremely similar and as there was no option for *Chaetoceros protuberans* in the list of species.

The only diatoms that caused some concern among participants were *Lauderia annulata* and *Bacillaria paxillifer*. The case for *Lauderia*, is that the species did not preserve well in the samples and chains of this diatom broke down into single cells that could not be discerned with authority by analysts, as there was not enough detail for identification. Most analysts decided to identify as *Thalassiosira*. Here, the case is made for giving these identifications as correct for the purpose of this intercomparison as these cells weren't of good enough quality for this identification.

For *Bacillaria paxillifer*, 13 analysts out of 61 incorrectly identified these species. It is clear why this mistake can be made. The single cells can look very similar to *Pseudo-nitzschia* although they can still be recognized by looking at the ends of the cells which are very different. Also, the *Bacillaria* chains are quite distinct to *Pseudo-nitzschia* chains. In any case, these were the most difficult sample identifications for diatoms.

The most undetected species in the samples was also *Coolia monotis* which had a relatively low cell density. 16 analysts did not detect this organism compared with 5 analysts for *A.minutum* or 3 for *G.spinifera.* Generally, dinoflagellates were harder to identify than diatoms. 24 non-detections on 4 dinoflagellates compared to 6 non-detections on 14 diatoms

Overall, from 1220 possible correct identifications, there were a total of 784 correct answers at species level (64%) and 1024 correct answers at genus level that is 84% correct, 171 (14%) incorrect identifications.

## 4.7 Ocean Teacher 2021 online taxonomic assessment

The test itself consisted of 20 questions (see Annex XVI in the annex report) and annex XVII show the overall results and grades of the participants. There were 113 attempts at the OTGA HAB asessement, the median overall grade was 91.3%. 57.5% of analysts performed above the proficiency threshold of 90% and 27.4% of all analysts between 80-90%. 7.9% above 70% and another 7.9% below 70% requiring improvement (Table 5).

| A. Code | % | A. Code | % | A. Code | % | A. Code | % | A. Code | % |
|---|---|---|---|---|---|---|---|---|---|
| 15 | 100 | 85 | 97.3 | 8 | 92.5 | 75 | 88.7 | 100 | 82 |
| 2 | 99.7 | 19 | 97.1 | 35 | 92.5 | Overall | 88.7 | 133 | 82 |
| 16 | 99.7 | 111 | 96.7 | 49 | 92.5 | 123 | 88.3 | 108 | 81.2 |
| 62 | 99.7 | 71 | 96.5 | 53 | 92.3 | 41 | 87.9 | 57 | 80.9 |
| 117 | 99.7 | 72 | 95.5 | 94 | 92.2 | 24 | 87.6 | 98 | 79.6 |
| 127 | 99.7 | 68 | 95.1 | 7 | 91.9 | 4 | 87.2 | 1 | 77.3 |
| 131 | 99.7 | 26 | 95 | 34 | 91.9 | 92 | 87 | 46 | 76.1 |
| 51 | 99.6 | 48 | 94.7 | 65 | 91.8 | 119 | 87 | 67 | 76 |
| 50 | 99.4 | 89 | 94.7 | 79 | 91.4 | 84 | 86.2 | 21 | 75.7 |
| 13 | 99.3 | 78 | 94.6 | 83 | 91.4 | 25 | 85.8 | 102 | 75.7 |
| 29 | 99.3 | 96 | 94.6 | 63 | 91.3 | 23 | 85.7 | 45 | 73.2 |
| 43 | 99.3 | 124 | 94.6 | 125 | 91.3 | 66 | 85.5 | 126 | 72.4 |
| 47 | 99.3 | 10 | 94.3 | 116 | 91.2 | 76 | 84.9 | 87 | 70.4 |
| 109 | 99.3 | 38 | 94.3 | 20 | 91 | 113 | 84.7 | 61 | 69.9 |
| 122 | 99 | 115 | 94.3 | 14 | 90.9 | 93 | 84.5 | 128 | 69.9 |
| 55 | 98.9 | 120 | 94.2 | 58 | 90.9 | 40 | 84.4 | 132 | 69.1 |
| 80 | 98.9 | 77 | 94.1 | 18 | 90.7 | 95 | 84.1 | 74 | 66.8 |
| 64 | 98.8 | 103 | 94 | 88 | 90.6 | 110 | 84 | 104 | 66.6 |
| 118 | 98.8 | 91 | 93.8 | 82 | 90.3 | 129 | 83.8 | 114 | 63.4 |
| 9 | 98.4 | 5 | 93.4 | 11 | 89.8 | 70 | 83.6 | 12 | 62.7 |
| 56 | 98.3 | 86 | 93.4 | 134 | 89.4 | 3 | 82.3 | 112 | 61.6 |
| 22 | 98.1 | 27 | 93.2 | 97 | 89.2 | 36 | 82.3 | 33 | 57 |
| 60 | 97.3 | 59 | 92.8 | 42 | 88.9 | 32 | 82 | | |

Table 5: Ocean Teacher scores by analyst code

The OTGA facility index shows that the worst answered question in the test was Q10 (61%) a numerical question and the best Q18 (100%) a multiple choice question on *Akashiwo sanguinea* (Table 6).

| Q# | Question type | Question name | Attempts | Facility index | Standard deviation | Intended weight | Effective weight |
|---|---|---|---|---|---|---|---|
| 1 | Matching | IPI2021 Diatoms identification Plate | 113 | 92.40% | 8.30% | 5.00% | 2.97% |
| 2 | Matching | Identification of Pseudo-nitzschia, terminology | 113 | 98.08% | 7.64% | 5.00% | 2.67% |
| 3 | Multiple choice | Pseudo-nitzschia sp.1 | 113 | 94.69% | 22.52% | 5.00% | 5.10% |
| 4 | Multiple choice | Pseudo-nitzschia sp.2 | 113 | 89.38% | 30.95% | 5.00% | 6.42% |
| 5 | Multiple choice | Pseudo-nitzschia sp.3 | 113 | 96.46% | 18.56% | 5.00% | 3.73% |
| 6 | Multiple choice | Pseudo-nitzschia sp.4 | 113 | 85.84% | 35.02% | 5.00% | 7.00% |
| 7 | Multiple choice | Pseudo-nitzschia sp.5 | 113 | 92.04% | 27.20% | 5.00% | 6.78% |
| 8 | Multiple choice | Pseudo-nitzschia sp.6 | 113 | 96.46% | 18.56% | 5.00% | 4.68% |
| 9 | Multiple choice | Pseudo-nitzschia sp.7 | 113 | 93.81% | 24.21% | 5.00% | 5.04% |
| 10 | Numerical | IPI 2021 Diatom chains cell counting | 113 | 61.06% | 48.98% | 5.00% | 6.04% |
| 11 | Matching | ICHA2018 Tripos | 113 | 90.44% | 17.95% | 5.00% | 4.80% |
| 12 | Multiple choice | Dinoflagellate terminology IPI16 | 113 | 79.29% | 30.19% | 5.00% | 7.28% |
| 13 | Multiple choice | ICHA2018 Ocelloid | 113 | 93.58% | 15.23% | 5.00% | 3.51% |
| 14 | Numerical | IPI2021 Diatoms cell counting 2 | 113 | 86.73% | 34.08% | 5.00% | 6.42% |
| 15 | Multiple choice | IPI21 Family Coscinodiscaceae | 113 | 80.62% | 24.65% | 5.00% | 6.92% |
| 16 | Drag and drop onto image | IPI2021 Amphidomataceae Plate | 113 | 87.89% | 14.99% | 5.00% | 3.27% |
| 17 | Numerical | IPI2021 Diatoms counting 3 | 113 | 80.53% | 39.77% | 5.00% | 5.44% |
| 18 | Multiple choice | ICHA2018 Akashiwo | 113 | 100.00% | 0.00% | 5.00% | 0.00% |
| 19 | Drag and drop onto image | IPI Azadinium ventral pore position | 113 | 90.15% | 19.37% | 5.00% | 5.39% |
| 20 | Numerical | IPI2021 Armoured dinoflagellate cell counting | 113 | 84.07% | 36.76% | 5.00% | 6.53% |

Table 6: Facility index IPI2021 OT exercise

In Q1 the classification of species from a list of potential candidates scored highly (Annex XVI for full breakdown of results). Most species were easily recognized, except for one species:

*Coscinodiscus centralis* which was also found in the samples, had only 66 correct answers to species level, with 26 *C.radiatus* and 14 *C.granii* responses. *Lauderia annulata* with 92 correct responses suggest that most analysts would have recognised this species in the samples if they had seen them in chains. Although 15 of them would have probably identify them incorrectly as *Detonula.*

Questions 2 through to 9 were dedicated to the *Pseudo-nitzschia* genus and their responses were above 90% for all the questions except for Q6 with a 85% correct index. The right answer was *P.pungens* but 16 analysts chose a different species here, 6 and 4 analysts chose *P.brasiliana* and *P.obtusa* (See details in Annex XVI in the annex report).

Q10 one of several numerical questions with a recorded video transect caused trouble to quite a few analysts (44), for many this was not a counting problem but rather an interpretation error of what was to be counted and what was to be left out of the count. At least, this was true for 20 of those analysts that were out by 3-4 cells of the consensus count. However, the other 24 analysts gave very different answers (See Annex XVI).

Q11 did not cause serious difficulties to the analysts except for the *T.macroceros/T.massiliensis* duet which caught a few analysts. *T.macroceros* was confused with *T.massiliensis,* while *T.massiliensis* was confused with both *T.macroceros* but also *T.longipes.*

In Q12, about the taxonomy of the Suessiaceae, a group of dinoflagellates in between armoured and naked, again there were a good level of proficiency among analysts (79%), but not as high as with diatom terminology (Q2) 98%. Albeit, a lesser known group of dinoflagellates, it caused analysts issues with the terminology. There were four correct responses to choose from and most analysts got the ‘Elongated Apical Vesicles’ , ‘Suessiaceae’ , and ‘x plate’ answers. However, the ‘Latitudinal series’ answer was only correct for 77 analysts. Many (21) gave ‘Amphiesmal vesicles’ as incorrect answer and 13 chose longitudinal series rather than latitudinal.

There were no issues with Q13 on dinoflagellates containing ocellus, a light sensing organelle. Most analysts recognised the two species with ocellus (images C (113) and F (111), however 16 people chose image A and 8 people chose image D which caused a 25% penalty for each wrong answer.

Q14 was another numerical question and most analysts (98) achieved consensus around the 58 to 62 cells in the video clip showing cells of *Bacillaria paxillifer*. We build a tolerance of +/- 2 cells into the model answer. However, another 15 analysts did not give the correct answer.

Q15 was a taxonomic question about the Coscinodiscaceae family. Analysts gave a good performance in this question, there were five right answers and most of them had high scores, except for 'cingulum has open bands' with only 71 correct answers.

Q16 and Q19 on *Azadinium* species were answered correctly by most people. In Q16, we asked analysts to choose between toxic, non-toxic or unknown and most analysts returned the right answer for the 16 species. A small number of discrepancies were found in relation to image 13 *A.zhuanum* (32 unknown, 6 toxic) and image 14 *A.concinnum* (83 unknown, 1 toxic) where both are non-toxic. In Q19, the ventral pore in the right hand side of the Apical Pore Complex in *Azadinium* species is a diagnostic feature for many species which have it. Together with other characters you can establish them to species level. Most analysts achieved a high score here with no real issues among analysts.

Q17 and 20 were numerical questions. Q17 was a chain of *Thalassiosira* and Q20 *Alexandrium* cells. In Q17, a tolerance of +/-1 was given however, there were 22 non-consensus answers. For Q20, a dinoflagellate in single cells and a tolerance of +/-1 cell there were 18 non-consensus answers.

## 5. Discussion

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 programme ProLab Plus to calculate the descriptive statistics for the test and the performance characteristics including the graphical representation of all the results.

### Homogeneity and stability test

The homogeneity and stability test in 2021 included 16 measurands (Table 1) and most of them except for *C.centralis* and *O.aurita* satisfied at least the ISO13528:2015 requirements for significant heterogeneity which allows the standard deviation to be greater than 30%. Also, most materials passed the stability assessment according to the expanded criterion except for *Coscinodiscus centralis*. This means, as in previous years that the materials are not adequately homogeneous but not significantly hetereogeneous, except for the two measurands above.

The stability of the materials is fine for all measurands and these can be used over time, the only failed item (table 1) was *Coscinodiscus* but as we have shown this is really related to homogenization issues rather than with the degradation of the material itself over time. For this reason, we are confident that the materials which were analysed for the homogeneity test in October 2021 and stability in December 2021, around the same time as these materials were analysed by all the participating laboratories around the world, are kept stabled and with a high degree of homogeneity.

In order, to avoid bias among laboratories, ISO 17043 gives some solutions to the materials lack of sufficient homogeneity. Essentially, it uses the analysts data and the homogeneity data to establish a new standard deviation for each test item. This is done through an iterative process (Annex VIII) where the standard deviation from the analysts and that of the homogeneity and stability test generates a new standard deviation for the test.

It is this new standard deviation that it is used to establish the confidence limits for the test (table 2). In this way, the test takes into account the heterogeneity of the samples. In practical terms this widens the confidence limits for each measurand.

### Calculation of performance statistics

The consensus values from the participants + the 'between samples standard deviation' from the homogeneity test were used to calculate the performance statistics for the test. These values are derived by iterative calculation using the new modified averages and standard deviations until the process converges (Annex VIII). This method deals with outliers in the dataset and missing values.

These assigned values were then used to calculate the Z-scores (Annex IX). Laboratory bias assumes a normal distribution of the data across zero and any results outside the warning signal (+/-2SD) or action signal (+/-3SD) would suggest an out of specification result. The results show that Z-scores are generally within the requirement for the test for most analysts with a small number of warning and action signals. A warning signal is a result between +/- 2 and +/- 3SD of zero and an action signal is a result outside +/-3SD. Two warning signals in consecutive intercomparisons give rise to an action signal. An action signal signifies that an investigation of the causes by the laboratory should be carried out.

There were a very small number of warning and action signals across measurands. 10 Red flags (1.2%), 31 (2.5%) yellow flags and 29 (2.3%) non-detection flags (Grey) from 1220 results is evidence of good performance overall. This compares with 18 (1.8%) and 13 (1.4%) red flags, 23 (2.34%) and 31 (3.26%) yellow flags, 12 (1.22%) and 22 (2.3%) non-detection flags in previous rounds. This suggest good performance and ongoing competency for most analysts over time with a small number of analysts that must improve their performance.

Nine analysts failed the quantitative test (see annex X). Four analysts (70%) are just below the requirement with three failed test items and 4 analysts (60%) failed 4 items need some improvement. One analyst (40%) score failed 6 out 10 items require training and improvement for the next round.

Quantitatively speaking, no measurand was more difficult to quantify than the rest. The largest amount of red flags (Annex X) were 2 for *Gonyaulax* and *Thalassiosira*, otherwise there were only 1 or none for the rest. The largest amount of yellow flags were 4 for *P.micans*, *A.minutum*, *T.mobiliensis* and 3 for *L.annulata*, *G.spinifera* and *Pseudo-nitzschia delicatissima group*.

The chart of RLP against RSZ (Annex XIII) 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.

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 graphically shown as Lischer plots in annex XV are in essence a representation of the RLP vs RSZ plot for individual components. Here you can visualize individual Z-scores per measurand against the other participants. 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.

Qualitative sample data

At least 80% of identification results must be correct to pass the test in conjunction with 80% of your quantitative results.  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 to genus level for all species.

Dinoflagellates were generally more difficult to identify than diatoms in this test. The results show this trend with approximately 121 incorrect flags for 3 dinoflagellate species compared to approximately 21 incorrect flags for all the diatoms (9) excluding *Lauderia annulata* which was not taken into account for this test.

The biggest problem was found with *Coolia* which many analysts failed to detect and also with *A.minutum* which is small and hard to identify species. However, diatoms did not pose any difficulty and many analysts were correct to species level for most species. Interestingly, the largest cell *Coscinodiscus* was the hardest to identify to species level and at least another 4 species were considered by the analysts for this measurand.

Chain formers like *Bacillaria* or *Pseudo-nitzschia* were described without trouble but a small number of laboratories confused *Bacillaria* with *Pseudo-nitzschia seriata group* members. These two species were inoculated in high densities and as we have expected that analysts would have to use a transect cell count rather than a whole chamber, we thought there would be significant variability in the results, however, the results show (Annex X: Z-scores) that only one analyst failed *Bacillaria* and 4 analysts failed the *Pseudo-nitzschia* demonstrating that transect counts, although there are not the preferred counting strategy, they do work for large cell densities in samples and the means are not significantly different across laboratories.

There were a number of different challenges in the samples, we mentioned *Lauderia annulata*, a diatom easily enough to identify in samples when in chains but more difficult if found in single cells. Only 21 analysts manage to identify correctly this species. However in the Oceanteacher test, question 1, image C, 92 analysts identified correctly this species from the image suggesting that the difficulty with samples was their bad condition and not the inability of the analysts to identify it. Therefore, we decided to waive the identification only, to all analysts within that batch.

There were other difficulties with *Coscinodisucs centralis* and *Odontella aurita*. In the case of *C.centralis,* cells are large and heavy and in this test in low cell densities. This created issues while homogenizing the samples, as these heavy diatoms tend to sediment quite quickly and it is possible that large variations between sample portions occur. This was exacerbated by the low cell density in the samples. Basically, analysts could have samples with all the cells in one portion of their 50 ml sample and samples with no cells. For this reason, and also because *Coscinodiscus* is so large, it would be difficult to miss if it was there, we waived the scores of five analysts for this measurand as there is a possibility that the cells weren't there.

With *O.aurita* we had a slightly different issue, *O.aurita* is not as large and heavy as *C.centralis,* however, there are chain formers and not solitary and the chains in this species did not break down into perfectly small group of cells but rather some large chains stayed together while others broke down into single cells. This caused issues that we picked up in the homogeneity and stability test and caused the repeatability within samples to be very large between same sample portions, which translated into a significant hererogeneity for this species. Interestingly, this same effect was found in the analysts data as the graphical representation of the *O.aurita* data suggests (Annex XIII). As the data of all analysts for the *O.aurita* count was basically within the Standard deviation for Proficiency Assessment (SDPA), there was no need for discarding this data for the exercise but just to be aware that there is an effect also with this measurand. These examples show that analyzing the second 25ml sample could ameliorate some of these problems and a suggestion as an improvement for next year's intercomparison.

I would like to insist to laboratories and individual analysts that in order to achieve a good performance in this test, that samples from the ampoules must be aliquot into your 50 ml sterilin tube first and that you are not supposed to aliquot the ampoule directly into your sedimentation chamber, because this nullifies the Utermöhl approach, which requires of a proper homogenization of the sample by inversion and the random effect of sedimentation into the chamber. This approach is not only wrong, but it also creates issues when trying to calculate the cell concentration as it was clear from the results. The initial sample is never 3ml but 50ml.

Ocean Teacher online taxonomic assessment IPI2021

There were 113 attempts at the OTGA HAB asessement, the median overall grade was above the proficiency threshold of 90% at 91.3%. 57.5% of analysts performed above that threshold. 27.4% of all analysts between 80-90% which is regarded as good scores. 7.9% above 70% which are regarded as acceptable and another 7.9% below 70% requiring improvement (Table 4).

Compared this to the 2019 exercise where, 74% were proficient, 21% good and 5% acceptable it suggests that the 2021 exercise was slightly more challenging for analysts.
For questions Q2 to Q9 on *Pseudo-nitzschia,* analysts did not find great difficulties in finding the right answer for most questions. The most challenging of these, appear to be the *P.pungens* identification in Q6, which was erroneously indicated as *P.brasiliana* and *P.obtusa* by several analysts. *P.obtusa* is generally considered similar to *P.seriata* in relation to the number of rows of poroids and lacking a central interspace, the bands are definitely different in *P.obtusa.* Yes, it is true that *P.pungens* share some of those characters but there are clear differences in shape and the number of striae and interstriae. *P.brasiliana* has very rounded ends, the interstriae are equidistant to the fibulae and it has 2-3 rows of poroids to *P.pungens* two rows. Many differences indeed.

In Q10, analysts found difficulties interpreting what should have been counted in a video clip showing a transect of *O.aurita.* The difficulty arose, because the transect showed transversal and length wise lines that belong to the Sedgewick-Rafter (SR) sedimentation chamber. Some analysts used the SR line to the right running up and down the transect as the limit for their count and they missed 4 cells that were to the right of this line but visible in the video field of view. The question was clear as to what had to be counted and this error can not be considered as a failure to count properly, but about understanding what was asked of the analysts.

On the genus *Tripos* (Q11) some analysts erroneously identified *T.macroceros* as *T.massiliensis* and viceversa. The difference between those two species is quite subtle. In *T.macroceros,* the antapical horns extend downwards first before doing a U-turn and growing on the apical direction, whereas in *T.massiliensis* one horn at least doesn't extend antapically but rather goes on a sideway trajectory before turning apically while the other is similar to *T.macroceros.*

Q13 depicted some dinoflagellates where participants were asked to choose the images that represented dinoflagellates bearing an Ocellus. The ocellus is a complex ultrastructural organelle typical of Warnowiids (*Nematodinium* and *Warnowia*). These were depicted in image C and image F, the correct answers. 16 analysts (14%) also chose image A (*Cochlodinium*) and 8 analysts (7%) chose image D (Polykrikos), both these genera are not ocellus bearing species.

Q14 showed how difficult sometimes is to determine what can be considered as 'one cell' as it is gleamed from the counting of *Bacillaria* cells in this video clip. This shows how complicated it is to count accurately, even when all analysts are counting the same thing. In this instance a tolerance of +/- 2 cells were allowed to take into consideration this issue. However, 15 analysts counted differently to the rest. The determination of what should be considered a 'cell' is not something that is learned or taught, because, we all consider that we know how to count and something that we take for granted, nonetheless, a few rules regarding cell counting would not go amiss when dealing with biological subjects that are constantly reproducing and changing shape and are tri-dimensional in nature. Q17 and Q20 also, shows that these differences are not found exclusively in counting cells in chain former species but also how to count solitary cells. This can affect enumeration results in real samples.