



Direct Toxicity Assessment Proficiency Scheme (DTAPS)

Round 3 Report

2007/2008



Provider: Environment Agency, Ecosystems Science.
Report Author: Dean Leverett (Environment Agency)

Executive Summary

The Environment Agency's Direct Toxicity Assessment Proficiency Scheme (DTAPS) has been implemented to support the Analytical Quality Control (AQC) components of the Monitoring Certification Scheme (MCERTS) for Direct Toxicity Assessment (DTA). Participation is invited from laboratories who are, or wish to become, involved in DTA testing for regulatory purposes.

Five DTA methods were proficiency tested in Round 3 of the DTAPS (2007-2008). The numbers of participants in the scheme across all test methods ranged from 5 (Oyster Embryo-Larval Development test) to 9 (*Daphnia magna* Immobilisation & *Tisbe battagliai* lethality tests).

The assessment comprised the 'spot' testing of centrally prepared reference toxicant solutions of unknown concentration and repeat testing of internally prepared reference toxicant solutions.

Laboratory performance in terms of proximity to the applied accuracy targets for each test method (using a Z-score assessment procedure) was generally good for this Round and ranged from 39% (Marine Algae Inhibition of Growth test) to 96% (*Tisbe battagliai* lethality test) compliance (Z-score = -2 to 2) across all participants.

Internal laboratory performance in terms of compliance with an applied precision target (using a comparison of variance approach) was also good (83-100% compliance) for all the methods except for the Oyster Embryo-Larval Development test with which a relatively low level of overall laboratory compliance was demonstrated (25%).

Three laboratories participating in Round 3 of the scheme complied with all the provisionally applied accuracy and precision targets for all the test methods they undertook, although none of these laboratories undertook the Oyster Embryo-Larval Development or Marine Algae Inhibition of Growth tests.

Contents

Executive Summary	Page 2
1. Introduction	Page 5
2. General	Page 6
3. Performance Criteria	Page 7
3.1. Accuracy	Page 7
3.2. Precision	Page 7
4. Test Methods	Page 9
5. Proficiency Testing Samples	Page 10
6. Internal Laboratory Reference Testing	Page 11
7. Data Treatment	Page 12
8. Participation	Page 13
9. Results	
9.1 General	Page 14
9.2 Accuracy Assessments	
9.2.1 <i>Daphnia magna</i> immobilisation test	Page 14
9.2.2 Freshwater Algae inhibition of growth test	Page 17
9.2.3 <i>Tisbe battagliai</i> lethality test	Page 19
9.2.4 Oyster Embryo-Larval (OEL) development test	Page 21
9.2.5 Marine Algae inhibition of growth test	Page 24
9.2.6 Summary	Page 27
9.3 Precision Assessments	
9.3.1 <i>Daphnia magna</i> immobilisation test	Page 33
9.3.2 Freshwater Algae inhibition of growth test	Page 34
9.3.3 <i>Tisbe battagliai</i> lethality test	Page 35
9.3.4 Oyster Embryo-Larval (OEL) development test	Page 36
9.3.5 Marine Algae inhibition of growth test	Page 37
9.3.6 Summary	Page 38
10. Update of DTAPS Performance Criteria	Page 42
11. Comparison of Statistical Methods used to Derive EC50 Values	Page 44
12. References	Page 46

1. Introduction

The Environment Agency's Direct Toxicity Assessment Proficiency Scheme (DTAPS) has been implemented to support the Analytical Quality Control (AQC) components of the Monitoring Certification Scheme (MCERTS) for Direct Toxicity Assessment (DTA). Participation is invited from laboratories who are, or wish to become, involved in DTA testing for regulatory purposes.

The scheme has evolved from previous Environment Agency research (Refs 1, 2, 3, 6), and has been implemented following the insertion of DTA as a prescribed hazard assessment tool within the Integrated Pollution prevention and Control (IPPC) regulations (for England and Wales) and the extension of the MCERTS scheme to encompass DTA activities. The scheme's design both builds on and extends the format and outcomes of the previous research into DTA proficiency testing, with an ultimate goal of providing robust technical performance assessments for laboratories undertaking the DTA methods for regulatory purposes.

2. General

The general conduct and operation of the DTAPS is described in the Environment Agency DTAPS Quality Manual (DTAPS 1.0) and Environment Agency DTAPS Operational Manual (DTAPS 3.0) (Refs 9, 10).

The provider for the scheme is the Environment Agency's Ecosystem Science Team.

The scheme is governed according to the International Laboratory Accreditation Cooperation (ILAC) guidelines for the operation of proficiency testing schemes (Ref 17) and is accredited by the United Kingdom Accreditation Service (UKAS) for proficiency testing.

In this Round of the scheme (2007/08), the provider undertook the preparation and distribution of centrally prepared reference toxicant solutions (zinc sulphate), collated both internal and external laboratory data returns, and assessed and reported on external laboratory performance.

Registered participant laboratories were requested to test two centrally prepared reference toxicant solutions (DTAPS A & B) of unknown concentration using the DTA methods they had opted to perform. In addition, participants were asked to provide internal reference testing data for a minimum of six repeat tests (performed within a defined period) for each DTA method.

Inter-laboratory accuracy assessments comprised two Z-scores calculated for each laboratory/ method combination based on each laboratory's submitted EC50 values for DTAPS A and DTAPS B. Intra-laboratory assessments comprised a single Z-score calculated from each laboratory's mean internal EC50 value.

Intra-laboratory precision assessments were achieved by calculating the variance of each laboratory's internal EC50 values and comparing to a target variance using a Chi-squared test (0.05 significance level).

3. Performance Criteria

3.1 Accuracy

In the context of the DTAPS, accuracy is defined as the degree of proximity of a laboratory result to the specified target reference value.

Accuracy assessments were undertaken using a Z-score methodology which incorporated a reference value (EC50 to zinc sulphate) and standard deviation target value for each test method. The reference values and standard deviation target values were derived from previous Rounds of the DTAPS.

Z-scores were calculated for each proficiency testing sample (DTAPS A & B) and for the mean of a each laboratory's internal data returns for each test method according to the formula;

$$Z = [\text{Laboratory Result} - \text{Ref. Value}] / [\text{Standard Deviation Target}]$$

The reference values and standard deviation target values for each test method (as applied in Round 3) are given in Table 1 below.

Table 1. DTAPS Accuracy Performance Criteria

Test Method	Reference Value (mg/L Zn)	Standard Deviation Target	Compliant Z-score Range
<i>Daphnia magna</i> 48hr immobilisation	1.47	0.51	-2 to 2
<i>Tisbe battagliai</i> 48hr mortality	0.55	0.23	-2 to 2
Oyster embryo-larval development	0.14	0.06	-2 to 2
Freshwater algae inhibition of growth	0.07	0.03	-2 to 2
Marine algae inhibition of growth	0.1	0.03	-3 to 3*

* A wider range of proficiency testing results were accepted as compliant for the Marine algae test in acknowledgement of the continuing developmental nature of quality control criteria for this test method.

3.2 Precision

In the context of the DTAPS, precision is defined as the degree of agreement between the results derived from a set of repeated measurements (tests).

In Round 3 of the DTAPS, the variance of each laboratory's internal data submissions was compared with a provisional (independently derived) target variance (σ^2) for each test method using a Chi-squared test;

$$\chi^2 = [S^2 / \sigma^2] \times [n-1]$$

The estimated χ^2 value was then compared to a critical χ^2 value ($p=0.05$).

An acceptable level of variance is indicated by an estimated χ^2 value which is less than the critical χ^2 value. That is, the variability between repeat tests is less than expected based on previous DTAPS data.

The target variance values applied in Round 3 for each test method is shown in Table 2 below.

Table 2. DTAPS Precision Performance Criteria

Test Method	Target Variance (σ) *
<i>Daphnia magna</i> 48hr immobilisation	0.03
<i>Tisbe battagliai</i> 48hr mortality	0.03
Oyster embryo-larval development	0.03
Freshwater algae inhibition of growth	0.03
Marine algae inhibition of growth	0.04

* Variance of Log Base 10 EC50 values.

4. Test Methods

Five ecotoxicity tests have been assessed within Round 3 of the DTAPS;

48 hour *Daphnia magna* immobilisation test;
72 hour Freshwater alga inhibition of growth test;
48 hour *Tisbe battagliai* lethality test;
24 hour Oyster Embryo-Larval Development test;
72 hour Marine alga inhibition of growth test.

These comprise the tests prescribed by the Environment Agency for Direct Toxicity Assessments under the IPPC regulatory regime, and are the DTA tests for which laboratories can be gain MCERTS approval.

5. Proficiency Testing Samples

The proficiency testing samples, DTAPS A and B, comprised solutions of zinc sulphate ($\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$) in Reverse Osmosis (RO) water, which had been acidified with concentrated hydrochloric acid to ensure stability over the duration of each Round (approx. 3 months). The concentration of zinc in each sample was unknown to participants. Each batch of DTAPS A and B was provided to participants as a 500mL sub-sample in a single inert plastic bottle. Further sub-samples were available to participants on request.

The degree of acidification and choice of RO water as a dilution media were applied to enable the investigation of laboratory proficiency in the adjustment of sample water quality parameters such as pH, total hardness and salinity. The modification of the physico-chemical parameters of industrial effluent samples can be critical in the application of DTA tests undertaken within a regulatory context and it is therefore considered important to include such procedures within the proficiency testing regime.

Each batch of proficiency testing sample was analysed chemically for zinc content by the provider (using the Environment Agency National Laboratory Service) prior to distribution and at 2 weekly intervals throughout the duration of each Round of the scheme. This ensured the verification of nominal zinc concentrations and demonstrated the homogeneity and stability of each batch of sample over the duration of each Round.

In Round 3 of the DTAPS (2007/08), the proficiency testing samples were prepared and distributed by the provider. DTAPS Sample A was prepared at a nominal zinc concentration of 6 mg/L and was acidified to a pH of between 2 and 3 prior to distribution to participants. Participating laboratories were instructed to test the DTAPS A sample as received. pH adjustment of this sample (either as a whole or at specific test dilutions) was required for all test methods. In addition, higher test concentrations required the adjustment of salinity or total hardness depending on the test method applied.

DTAPS Sample B was prepared at a nominal zinc concentration of 600 mg/L and also acidified to a pH of between 2 and 3 before distribution. Participants were instructed to dilute this sample to one-hundredth of its initial strength before testing. The initial dilution provided an identical sample to DTAPS A but removed the pH adjustment necessity.

6. Internal Laboratory Reference Testing

In addition to the results of testing with the distributed proficiency samples DTAPS A and B, participant laboratories were also required to submit a minimum of six results (EC50 value as mg/L zinc) for repeat reference tests undertaken using internally prepared zinc sulphate standards, for the period 13th February 2007 to 8th February 2008.

The procedures and concentrations used in the preparation of internal zinc standards were left to each participant. Some laboratories opted to prepare a high concentration (100-1000 mg/L as zinc) standard which was retained and diluted for use in DTA tests as and when required, while others prepared a new standard (from solid zinc sulphate) on each testing occasion.

The assessment of performance across each participant's internal laboratory reference testing therefore contains an element of potential variability beyond that measured using the distributed proficiency samples since any error (random or systematic) associated with the internal preparation of zinc standards has not been controlled or standardised (by the scheme). This potential for additional error is, however, balanced to some degree since each test will have been undertaken using a known concentration of reference toxicant solution. Thus, it is possible to select an optimal series of test concentrations to enable the derivation of the expected internal laboratory EC50 value for zinc.

7. Data Treatment

In both Rounds, participants were instructed to report an EC50 value (% v/v) for both DTAPS proficiency samples (DTAPS A & B) for each test method they had elected to undertake. EC50 calculations on tests undertaken with DTAPS Sample B were required to account for the initial dilution of the sample (1/100). The EC50 values reported by participants were converted to a mg/L value (as zinc) by the provider, using the mean measured zinc concentrations of the distributed samples.

A mean internal EC50 value (mg/L zinc) was also calculated for each participant for each method, based on their internal reference test submissions.

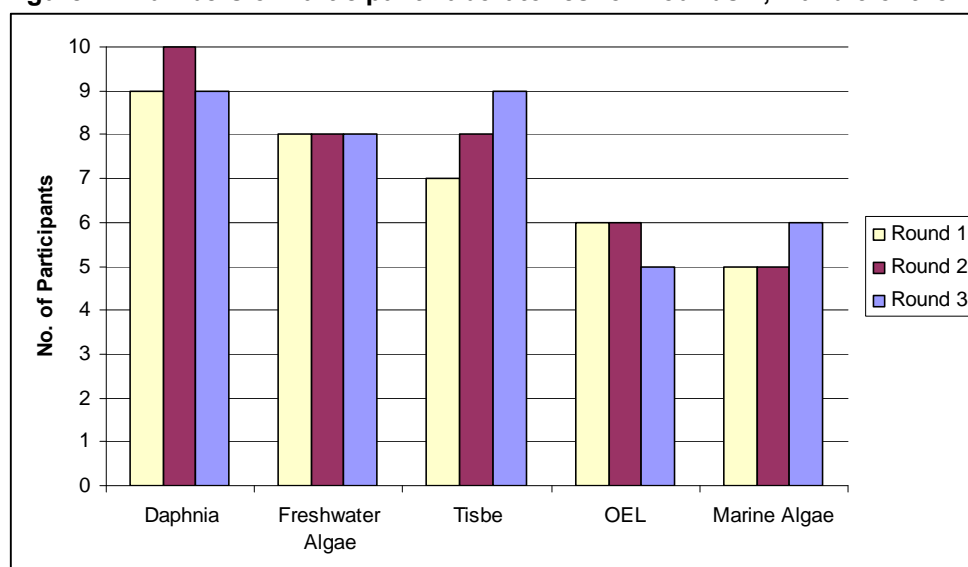
Each participant's internal reference test submissions were also used to calculate an internal variance value for precision assessment.

8. Participation

The scheme attracted the participation of the vast majority of the UK contract ecotoxicological testing laboratories. A small number of specialist/ overseas laboratories also participated in some of the methods in order to ascertain their relative performance for reasons other than for MCERTS approval. The total numbers of participants for each test method across the previous and current Round of the scheme is shown in Figure 1.

The freshwater methods (*Daphnia* immobilisation & Freshwater Algae inhibition of growth) are traditionally used for basic chemical hazard assessments and are generally more widely applied across laboratories than the marine methods. It is therefore encouraging to see the *Tisbe* lethality test attracting an increasing number of participants. In Round 3, the participants for the *Tisbe* tests equalled those for the *Daphnia* immobilisation test (previously the most subscribed test) and these two methods attracted the most participants in Round 3 of the scheme. The increase in numbers for the *Tisbe* test is probably in part due to the higher level of estuarine DTA regulatory work being undertaken by the participant laboratories compared to freshwater work (although this has not caused a reciprocal drop in subscription for the freshwater assays). The Marine Algae inhibition of growth test also saw an increase in participation compared to previous Rounds. In Round 3, the Oyster Embryo-Larval Development test was the least subscribed method.

Figure 1. Numbers of Participant Laboratories for Rounds 1, 2 and 3 of the DTAPS.



There was also a substantial improvement with the level of data submission for participant laboratories especially with regard to internal data returns for Round 3 of the scheme. Only a single laboratory (No. 4) submitted proficiency testing data (DTAPS A and B) with no accompanying internal data (for *Tisbe* and OEL), while one participant (No. 11) submitted insufficient internal data ($n < 6$) for two test methods (*Daphnia* and OEL). No assessment of internal accuracy or precision was made for Laboratory No.4, but were carried out for Laboratory No.11 since there was considered to be sufficient data to in both tests to conduct the analysis ($n=5$). Both the algal inhibition of growth test methods yielded full data set for all participants across in Round 3 of the scheme.

9. Results

9.1 General

All of the data submitted by participants in Round 3 was suitable for undertaking the proficiency assessments for accuracy and precision as outlined in Section 3.

9.2 Accuracy Assessments

9.2.1 *Daphnia magna* 48 hour immobilisation test

Inter-laboratory Z-scores for unknown samples (DTAPS A & B)

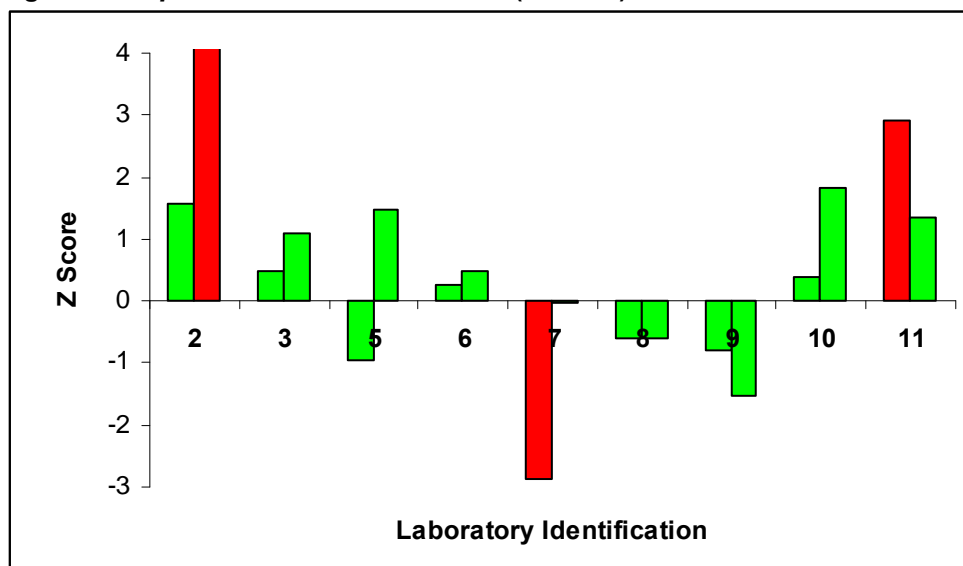
The results of the Round 3 accuracy (Z-score) assessments for the *Daphnia* test are presented in Table 3 and Figure 2, below.

Table 3. *Daphnia* EC50 Values and Z-scores for Round 3 (2007/08) of the DTAPS.

Laboratory Identification	DTAPS A EC50 (mg/L Zn)*	DTAPS A Z-score	DTAPS B EC50 (mg/L Zn)*	DTAPS B Z-score
2	2.28	1.58	8.97	14.7
3	1.72	0.48	2.03	1.09
5	0.985	-0.97	2.21	1.46
6	1.61	0.27	1.72	0.49
7	0.0130	-2.86	1.46	-0.03
8	1.17	-0.59	1.17	-0.59
9	1.06	-0.81	0.69	-1.52
10	1.67	0.38	2.40	1.82
11	2.95	2.91	2.15	1.34
Consensus				
Mean	1.50	0.0433	2.53	2.08
Range	0.013-2.95	-2.86-2.91	0.69-8.97	-1.52-14.7
Standard Deviation	0.84	NA	2.48	NA
% RSD (% CV)	56	NA	98	NA

* As reported by participant.

Figure 2. *Daphnia* Z-scores for Round 3 (2007/08) of the DTAPS.



For each Laboratory: 1st Bar = DTAPS A; 2nd Bar = DTAPS B.
Green denotes compliance with provisional performance criteria ($Z = -2$ to 2); Red denotes non-compliance.

Compliance with the provisional DTAPS performance criteria for accuracy was generally good for the *Daphnia* immobilisation test in Round 3 of the DTAPS.

Nine laboratories participated and six obtained compliant Z-scores for both DTAPS A and B. Laboratories 7 and 11 generated a non-compliant Z-score for DTAPS A while only Laboratory 2 generated a Z-score for DTAPS B which was outside the -2 to 2 range. None of the participant laboratories produced non-compliant Z-scores for both of the proficiency testing samples.

The Round 3 performance statistics for inter-laboratory accuracy represent a similar level of overall compliance with the applied performance criteria as seen in Round 2 although the result submitted for DTAPS B by Laboratory No.2 demonstrated a much more extreme deviation from the applied reference value (1.47 mg/L Zn) than any of the results submitted in Round 2 of the scheme.

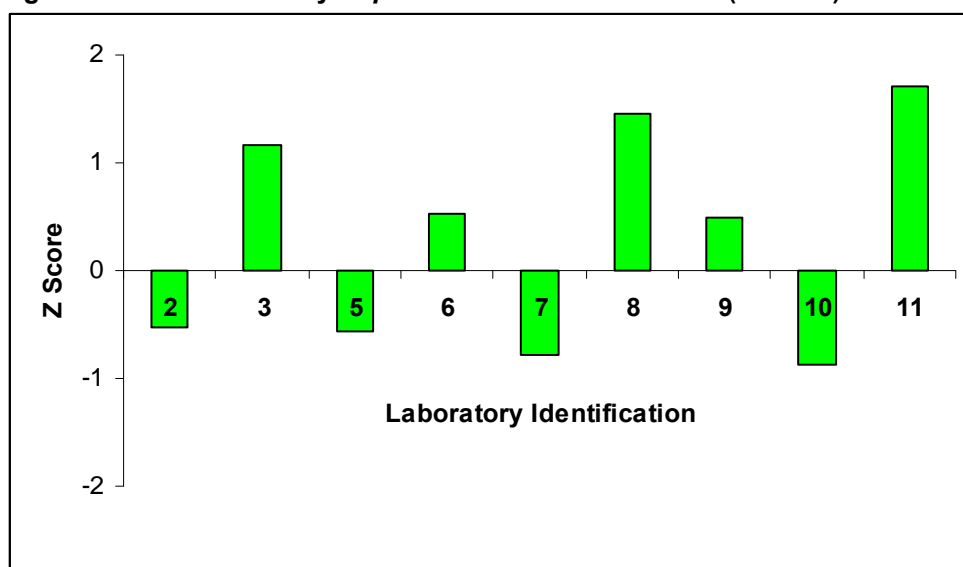
Intra-Laboratory Z-scores (Mean Internal Reference Standard Result)

In addition to the Z-scores calculated for each laboratory for the DTAPS Sample A and B (unknown concentrations), an additional Z-score was generated for each laboratory based on the mean of their internal reference tests submissions (EC50 to zinc (mg/L)).

Table 4. *Daphnia* Mean internal EC50 Values and Z-scores for Round 3 (2007/08) of the DTAPS.

Lab ID	Mean Internal EC50 (mg/L Zn)	Z-score	n
2	1.20	-0.53	6
3	2.06	1.16	6
5	1.18	-0.57	6
6	1.74	0.52	6
7	1.07	-0.79	6
8	2.21	1.46	6
9	1.73	0.50	6
10	1.03	-0.87	6
11	2.34	1.71	5
Consensus			
Mean	1.62	0.29	
Range	1.03-2.34	-0.87-1.71	
Std Dev.	0.51	NA	
% RSD (% CV)	32	NA	

Figure 3. Intra-Laboratory *Daphnia* Z-scores for Round 3 (2007/08) of the DTAPS.



Green denotes compliance with provisional performance criteria ($Z = -2$ to 2); Red denotes non-compliance.

The mean internal EC50 value for all participants in Round 3 of the scheme complied with the standard deviation target applied around the applied reference value of 1.47 mg/L zinc.

Compliance of mean internal EC50 values was much improved from Round 2 of the scheme in which four laboratories generated Z-scores which were outside the -2 to 2 range.

9.2.2 Freshwater Algae 72 hour inhibition of growth test

Inter-laboratory Z scores for unknown samples (DTAPS A & B)

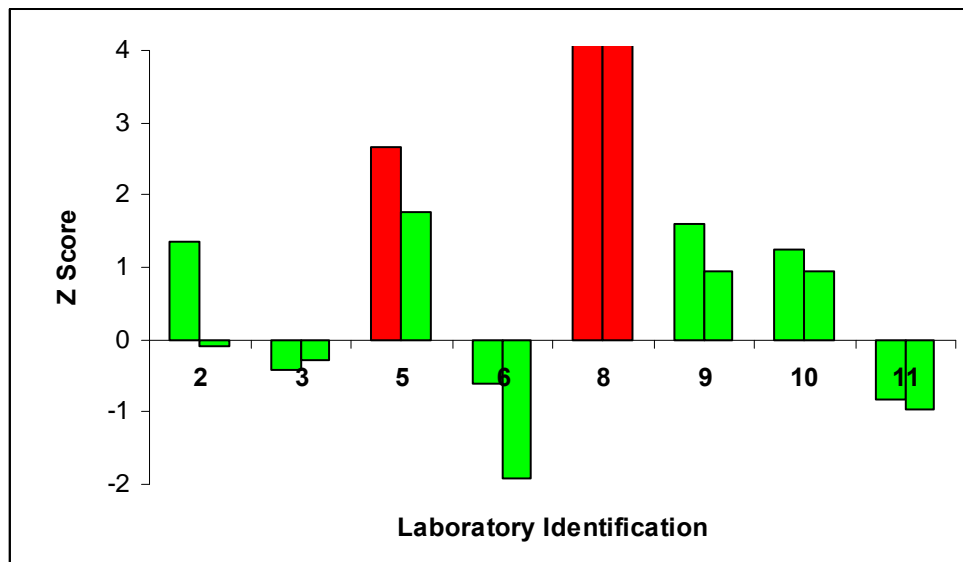
The results of the Round 3 accuracy (Z-score) assessments for the Freshwater Algae test are presented in Table 5 and Figure 4, below.

Table 5. Freshwater Algae EC50 Values and Z-scores for Round 3 (2007/08) of the DTAPS.

Laboratory Identification	DTAPS A EC50 (mg/L Zn)*	DTAPS A Z-score	DTAPS B EC50 (mg/L Zn)*	DTAPS B Z-score
2	0.111	1.36	0.0676	-0.08
3	0.0572	-0.43	0.0615	-0.28
5	0.150	2.67	0.123	1.76
6	0.0517	-0.61	0.0123	-1.92
8	1.01	31.3	0.596	17.5
9	0.118	1.60	0.0983	0.94
10	0.108	1.25	0.0983	0.94
11	0.0450	-0.83	0.0408	-0.97
Consensus				
Mean	0.206	4.54	0.137	2.24
Range	0.045-1.01	-0.83-31.3	0.012-0.6	-1.92-17.5
Standard Deviation	0.327	NA	0.189	NA
% RSD (% CV)	158	NA	137	NA

* As reported by participant.

Figure 4. Freshwater Algae Z-scores for Round 3 (2007/08) of the DTAPS.



For each Laboratory: 1st Bar = DTAPS A; 2nd Bar = DTAPS B.

Green denotes compliance with provisional performance criteria ($Z = -2$ to 2); Red denotes non-compliance.

Compliance with the provisional DTAPS performance criteria for accuracy was generally good for the Freshwater Algae growth test, with similar overall results being generated in Round 3 as in the developmental Rounds (1 and 2) of the DTAPS.

Eight laboratories participated and six obtained compliant Z-scores for both DTAPS A and B. Laboratory 8 was the only laboratory to generate non-compliant Z-scores (for both DTAPS A and B) and these were extreme compared to the other laboratory's results (Z-score = 31.3 for DTAPS A and 17.5 for DTAPS B).

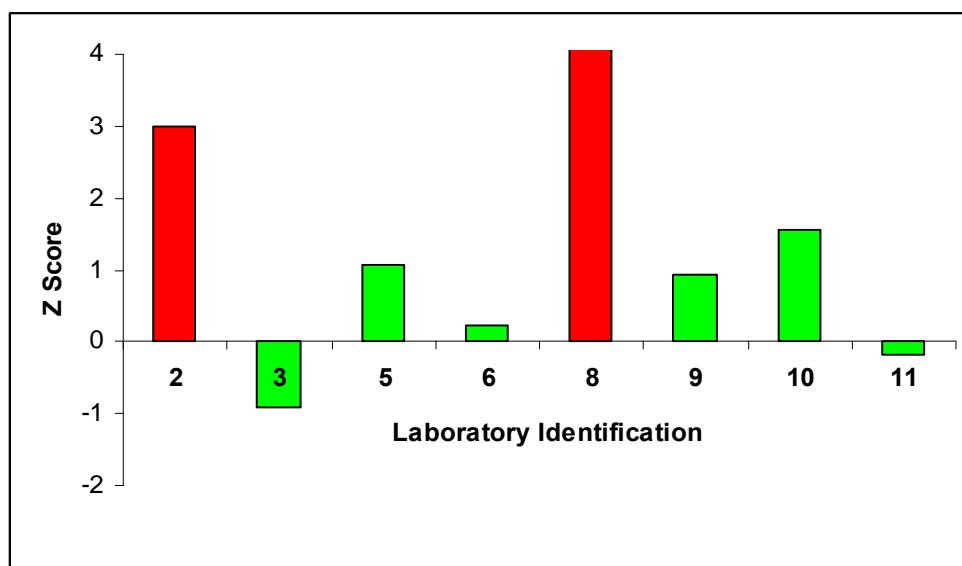
Intra-Laboratory Z-scores (Mean Internal Reference Standard Result)

In addition to the Z-scores calculated for each laboratory for the DTAPS Sample A and B (unknown concentrations), an additional Z-score was generated for each laboratory based on the mean of their internal reference tests submissions (EC50 to zinc (mg/L)).

Table 6. Freshwater Algae Mean internal EC50 Values and Z-scores for Round 3 (2007/08) of the DTAPS.

Lab ID	Mean Internal EC50 (mg/L Zn)	Z-score	n
2	0.16	3.00	6
3	0.04	-0.92	6
5	0.10	1.06	6
6	0.08	0.22	6
8	0.67	19.9	6
9	0.10	0.94	6
10	0.12	1.56	6
11	0.07	-0.17	6
Consensus			
Mean	0.168	3.20	
Range	0.04-0.67	-0.92-19.9	
Std Dev.	0.206	NA	
% RSD (% CV)	123	NA	

Figure 5. Intra-Laboratory Freshwater Algae Z-scores for Round 3 (2007/08) of the DTAPS.



Green denotes compliance with provisional performance criteria ($Z = -2$ to 2); Red denotes non-compliance.

The mean internal EC50 value for six laboratories (of 8) participating in Round 3 of the scheme complied with the standard deviation target applied around the reference value of 0.07 mg/L zinc. Of the two participants generating non-compliant results, Laboratory No.2 generated a marginally non-compliant Z-Score ($z=3$) while the mean internal EC50 for Laboratory No.8 deviated significantly from the accuracy target (Z-score = 19.9).

Compliance of mean internal EC50 values with the applied reference value in Round 3 followed a similar pattern to the previous Round, although in Round 2 only one participant failed to comply with the internal accuracy target.

9.2.3 *Tisbe battagliai* 48 hour lethality test

Inter-laboratory Z-scores for unknown samples (DTAPS A & B)

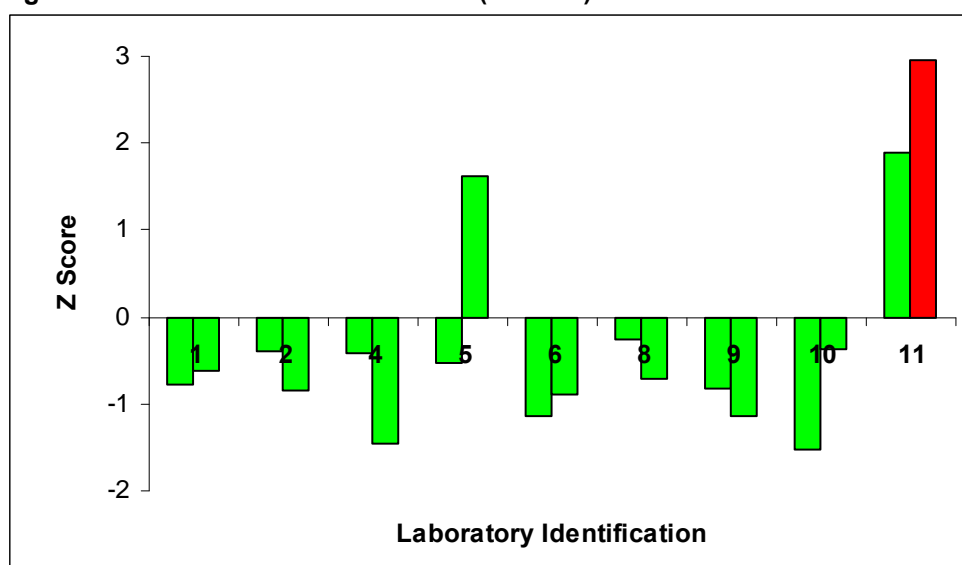
The results of the Round 3 accuracy (Z-score) assessments for the *Tisbe* test are presented in Table 7 and Figure 6 below.

Table 7. *Tisbe* EC50 Values and Z-scores for Round 3 (2007/08) of the DTAPS.

Laboratory Identification	DTAPS A EC50 (mg/L Zn)*	DTAPS A Z-score	DTAPS B EC50 (mg/L Zn)*	DTAPS B Z-score
1	0.360	-0.78	0.406	-0.63
2	0.461	-0.39	0.356	-0.84
4	0.456	-0.41	0.215	-1.46
5	0.429	-0.52	0.922	1.62
6	0.287	-1.15	0.342	-0.90
8	0.491	-0.26	0.387	-0.71
9	0.359	-0.83	0.291	-1.13
10	0.199	-1.53	0.467	-0.36
11	0.984	1.89	1.23	2.95
Consensus				
Mean	0.447	-0.44	0.513	-0.16
Range	0.199-0.984	-1.53-1.89	0.215-1.23	-1.46-2.95
Standard Deviation	0.222	NA	0.34	NA
% RSD (% CV)	50	NA	65	NA

* As reported by participant.

Figure 6. *Tisbe* Z-scores for Round 3 (2007/08) of the DTAPS.



For each Laboratory: 1st Bar = DTAPS A; 2nd Bar = DTAPS B. Green denotes compliance with provisional performance criteria (Z = -2 to 2); Red denotes non-compliance.

Compliance with the applied DTAPS performance criteria for accuracy was excellent for the *Tisbe* lethality test, with only a single DTAPS B result falling outside the standard deviation accuracy target range (Z-score -2 to 2).

Nine laboratories participated in Round 3 of the scheme and eight produced Z-scores for both proficiency testing samples which were within the target error range. Laboratory No.11 generated a non-compliant result for DTAPS B only (Z-score = 2.95). The majority of participants generated EC50 values (mg/L Zn) which were less than the specified reference value (0.55 mg/L Zn) derived using the data submitted for this test in Rounds 1 and 2 of this scheme.

The results in Round 3 were very similar to those generated in Round 2 for the *Tisbe* test in which a single participant also failed the accuracy performance criteria for DTAPS Sample B.

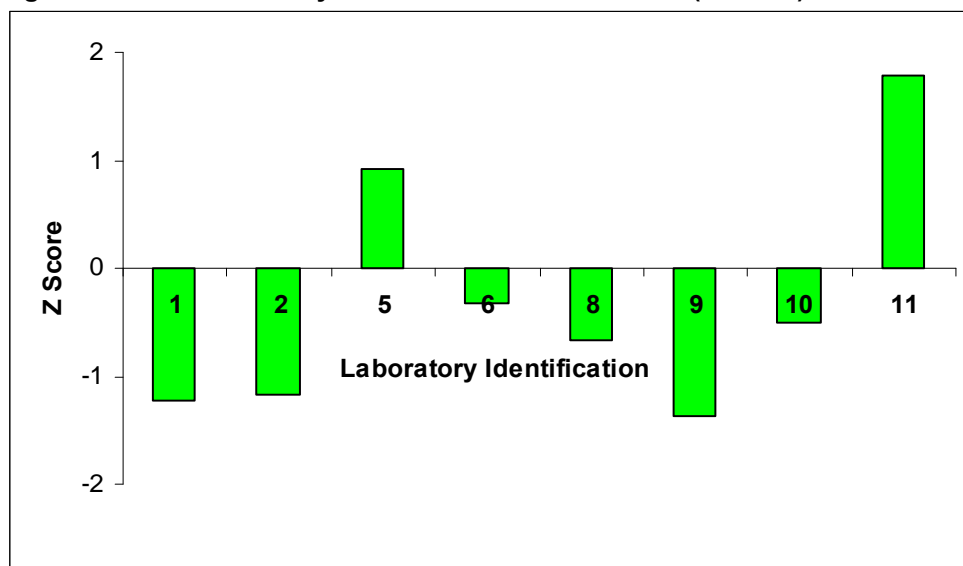
Intra-Laboratory Z-scores (Mean Internal Reference Standard Result)

In addition to the Z-scores calculated for each laboratory for the DTAPS Sample A and B (unknown concentrations), an additional Z-score was generated for each laboratory based on the mean of their internal reference tests submissions (EC50 to zinc (mg/L)).

Table 8. *Tisbe* Mean internal EC50 Values and Z-scores for Round 3 (2007/08) of the DTAPS.

Lab ID	Mean Internal EC50 (mg/L Zn)	Z-score	n
1	0.27	-1.23	6
2	0.28	-1.18	6
4	No Results Submitted	Not calculable	0
5	0.76	0.91	6
6	0.47	-0.33	6
8	0.40	-0.66	6
9	0.23	-1.37	6
10	0.43	-0.51	6
11	0.96	1.78	6
Consensus			
Mean	0.48	-0.32	
Range	0.23-0.96	-1.37-1.78	
Std Dev.	0.258	NA	
% RSD (% CV)	54	NA	

Figure 7. Intra-Laboratory *Tisbe* Z-scores for Round 3 (2007/08) of the DTAPS.



Green denotes compliance with provisional performance criteria ($Z = -2$ to 2); Red denotes non-compliance.

The mean internal EC50 value for all participants submitting a full data set ($n \geq 6$) in Round 3 of the scheme complied with the standard deviation target applied around the reference value of 0.55 mg/L zinc. Laboratory 4 did not submit any internal reference test results.

As with the testing of the distributed proficiency testing samples (DTAPS A and B) the majority of participants generated mean internal EC50s which were less than the reference value.

There remains an excellent level of compliance with the applied reference value among DTAPS participants for repeated internal reference tests and this is the only DTAPS method not displaying a single non-compliant result across all three Rounds of the scheme.

9.2.4 Oyster Embryo-Larval (OEL) 24 hour development test

Inter-laboratory Z-scores for unknown samples (DTAPS A & B)

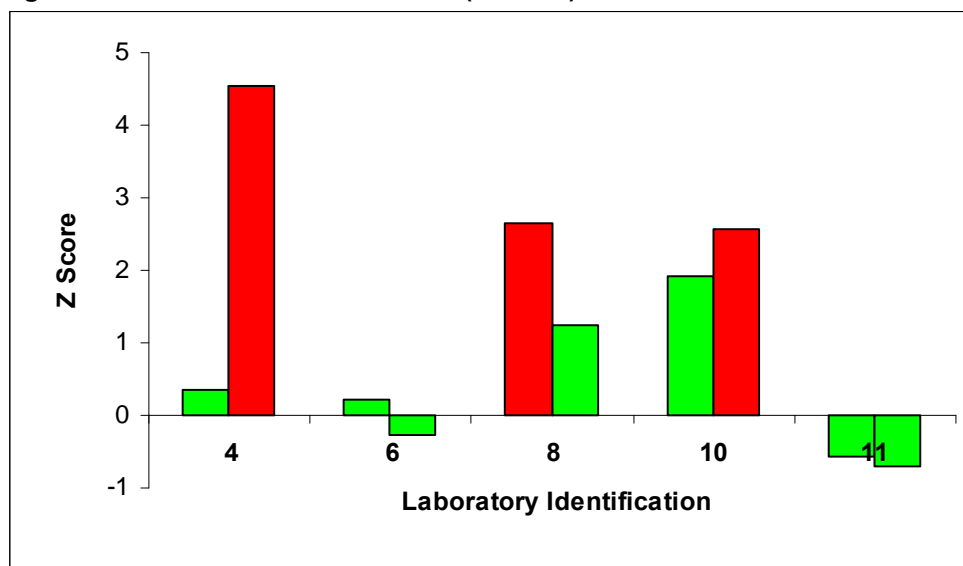
The results of the Round 3 accuracy (Z-score) assessments for the OEL test are presented in Table 9 and Figures 8 below.

Table 9. OEL EC50 Values and Z-scores for Round 3 (2007/08) of the DTAPS.

Laboratory Identification	DTAPS A EC50 (mg/L Zn)*	DTAPS A Z-score	DTAPS B EC50 (mg/L Zn)*	DTAPS B Z-score
4	0.161	0.34	0.412	4.53
6	0.153	0.22	0.123	-0.28
8	0.298	2.64	0.215	1.25
10	0.255	1.91	0.295	2.58
11	0.106	-0.56	0.098	-0.69
Consensus				
Mean	0.195	0.91	0.229	1.48
Range	0.106-0.298	-0.56-2.64	0.098-0.412	-0.69-4.53
Standard Deviation	0.079	NA	0.129	NA
% RSD (% CV)	41	NA	56	NA

* As reported by participant.

Figure 8. OEL Z-scores for Round 3 (2007/08) of the DTAPS.



For each Laboratory: 1st Bar = DTAPS A; 2nd Bar = DTAPS B.
Green denotes compliance with provisional performance criteria ($Z = -2$ to 2); Red denotes non-compliance.

Round 3 of the DTAPS appeared to show a decline in laboratory performance (in terms of accuracy) for the OEL test compared with the results obtained in Rounds 1 and 2. Six laboratories participated in both developmental Rounds of the scheme and only a single non-compliant result was generated. In Round 3, three non-compliant results were generated from only five participants, and one of these (Laboratory 4) submitted a DTAPS B result which was almost 3 times the specified reference value (Z -Score = 4.53).

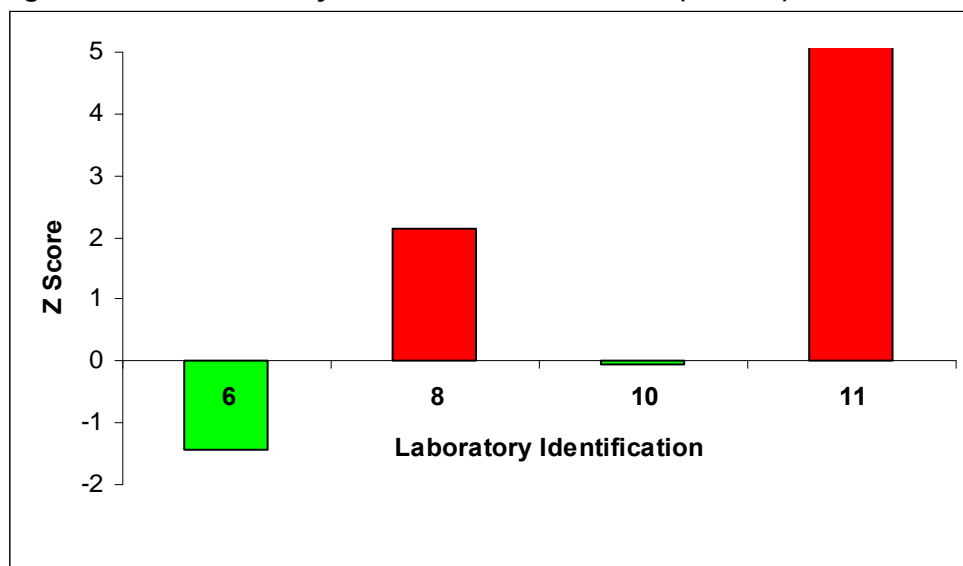
Intra-Laboratory Z scores (Mean Internal Reference Standard Result)

In addition to the Z-scores calculated for each laboratory for the DTAPS Sample A and B (unknown concentrations), an additional Z-score was generated for each laboratory based on the mean of their internal reference tests submissions (EC50 to zinc (mg/L)).

Table 10. OEL Mean internal EC50 Values and Z-scores for Round 3 (2007/08) of the DTAPS.

Lab ID	Mean Internal EC50 (mg/L Zn)	Z-score	n
4	No results submitted	Not calculable	0
6	0.06	-1.42	6
8	0.27	2.14	6
10	0.14	-0.06	6
11	0.68	9.0	5
Consensus			
Mean	0.29	2.42	
Range	0.06-0.68	-1.42-9.0	
Std Dev.	0.28	NA	
% RSD (% CV)	96	NA	

Figure 9. Intra-Laboratory OEL Z-scores for Round 3 (2007/08) of the DTAPS.



Green denotes compliance with provisional performance criteria (Z = -2 to 2); Red denotes non-compliance.

Only half the participants submitting internal reference data in Round 3 of the scheme complied with the standard deviation target applied around the reference value of 0.14 mg/L zinc. Laboratory No. 8 marginally exceeded the target ($z = 2.14$) while Laboratory 11 generated a mean internal EC50 value which was almost five times the specified reference value ($z = 9.0$). Laboratory No. 4 did not submit any internal reference testing data. Both laboratories which complied with the standard deviation target generated mean EC50 values which were less than the reference value, while those failing to comply both generated values in excess of the reference value.

While the results of the Round 3 internal assessment suggest a continued worsening of the internal accuracy of participants seen between Rounds 1 and 2 of the scheme (0 and 40% participant non-compliance), this is in part due to the reduction in participants returning internal data sets for this test method in Round 3 (from 5 to 4). The actual number of non-compliant laboratories (2) remains the same as in Round 2. The two laboratories failing the applied target criteria in Round 2 only marginally exceeded the Z-score range as did Laboratory No. 8 in Round 3. The extreme deviation of the mean internal EC50 from the reference value generated by Laboratory No. 11 is perhaps of most concern though this is likely to be caused by a laboratory specific issue rather than any generic problems with the internal application of the test methodology.

9.2.5 Marine Algae inhibition of growth test

Inter-laboratory Z-scores for unknown samples (DTAPS A & B)

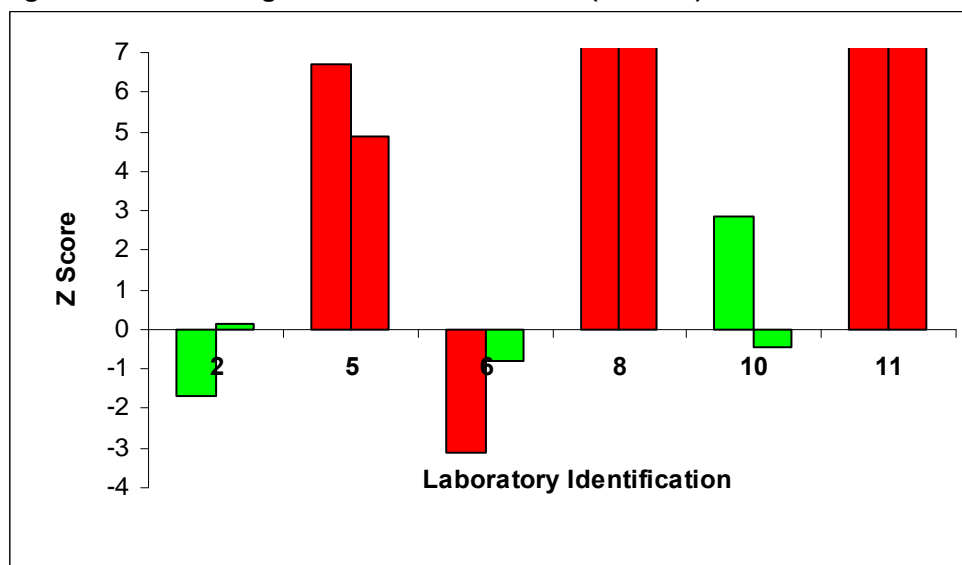
The results of the Round 3 accuracy (Z-score) assessments for the OEL test are presented in Table 11 and Figure 10 below.

Table 11. Marine Algae EC50 Values and Z-scores for Round 3 (2007/08) of the DTAPS.

Laboratory Identification	DTAPS A EC50 (mg/L Zn)*	DTAPS A Z-score	DTAPS B EC50 (mg/L Zn)*	DTAPS B Z-score
2	0.049	-1.69	0.104	0.15
5	0.301	6.69	0.246	4.86
6	0.007	-3.09	0.077	-0.77
8	0.648	18.3	0.799	23.3
10	0.186	2.88	0.086	-0.47
11	2.01	63.7	2.08	65.9
Consensus				
Mean	0.534	14.5	0.565	15.5
Range	0.007-2.01	-3.09-63.7	0.077-2.08	-0.77-65.9
Standard Deviation	0.759	NA	0.791	NA
% RSD (% CV)	142	NA	140	NA

* As reported by participant.

Figure 10. Marine Algae Z-scores for Round 3 (2007/08) of the DTAPS.



For each Laboratory: 1st Bar = DTAPS A; 2nd Bar = DTAPS B. Green denotes compliance with provisional performance criteria (Z = -3 to 3); Red denotes non-compliance.

In the developmental Rounds of the DTAPS a reference value was not specified for the Marine Algae test owing to the paucity of previous (pre-DTAPS) data available on which to base a value. A reference value was subsequently derived from the data submitted for Rounds 1 and 2 of the scheme which was based on the application of the method using a test and culture media containing a minimum concentration of EDTA, and reporting results calculated using biomass data.

Compliance with the applied accuracy (standard deviation) target around the reference value was generally poor despite the application of a wider compliant Z-score range (-3 to 3) than the other DTA test methods (-2 to 2). Two of the six participants generated compliant Z-scores for both DTAPS Samples A and B (Laboratories Nos. 2 and 10) while Laboratory No. 6 generated a compliant Z-score for DTAPS B only (accompanied by a marginal failure for DTAPS A). The other three participants all submitted results (EC50s to zinc) which were considerably higher (approximately 2.5 – 21 times) than the reference value.

This may suggest that the applied reference value (which although derived from previous DTAPS data is still based on significantly smaller data set than the other methods) may not accurately reflect the actual level of toxicity generally obtained for the zinc reference toxicant. It is clear from the results that there may, in fact, be two sub-sets of laboratories, the first (Nos. 2, 6 and 10) generally performing within the level of performance applied here and the second (Nos. 5, 8 and 11) which are all operating at a level of zinc toxicity to marine algae which is much lower than expected. This may indicate a bias or systematic difference in the performance of the test method between the two sub-sets of laboratories which may require further standardisation. The publication of a new SCA guideline for this method (due before Round 4 of the DTAPS) should assist with identifying and/or resolving the difference in levels of zinc toxicity between participant laboratories. The fact that most of the non-compliant results are significantly higher than the reference value (0.1 mg/L Zn) implicates the use of test media containing high levels of EDTA which can chelate metals (such as zinc) in solution and prevent them from being available to the test organism under exposure.

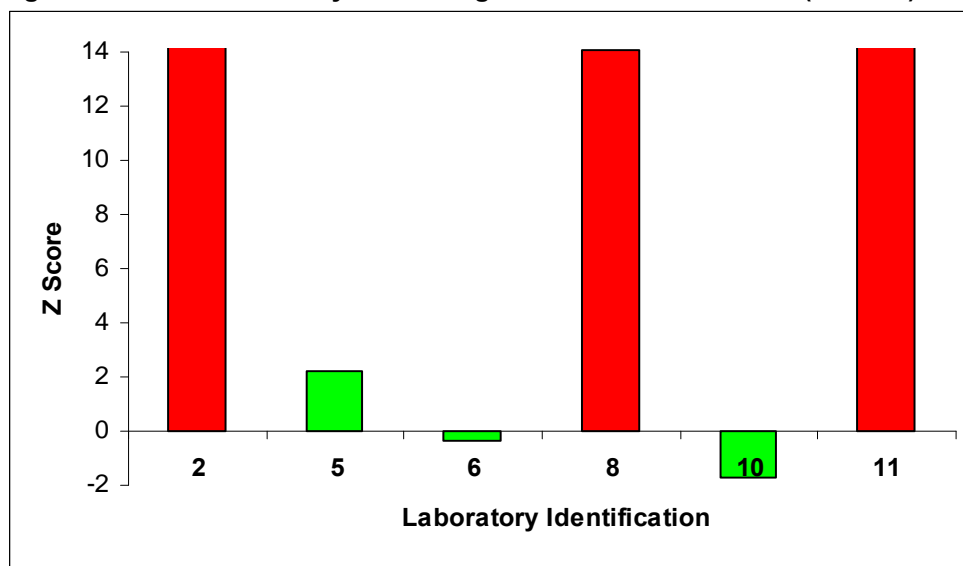
Intra-Laboratory Z scores (Mean Internal Reference Standard Result)

In addition to the Z-scores calculated for each laboratory for the DTAPS Sample A and B (unknown concentrations), an additional Z-score was generated for each laboratory based on the mean of their internal reference tests submissions (EC50 to zinc (mg/L)).

Table 12. Marine Algae Mean internal EC50 Values and Z-scores for Round 3 (2007/08) of the DTAPS.

Lab ID	Mean Internal EC50 (mg/L Zn)	Z-score	n
2	2.15	68.3	6
5	0.17	2.22	6
6	0.09	-0.39	6
8	0.52	14.1	6
10	0.05	-1.72	6
11	2.10	66.6	6
Consensus			
Mean	0.85	24.9	
Range	0.05-2.15	-1.72-68.3	
Std Dev.	1.00	NA	
% RSD (% CV)	119	NA	

Figure 11. Intra-Laboratory Marine Algae Z-scores for Round 3 (2007/08) of the DTAPS.



Green denotes compliance with provisional performance criteria (Z = -2 to 2); Red denotes non-compliance.

All the participants submitted a full internal data set (n=6) for the Marine algal test in Round 3 of the scheme. As with the testing of the distributed proficiency samples (DTAPS A and B) the overall level of compliance with the applied reference value/ standard deviation target was generally poor. Three laboratories (Nos. 5, 6 and 10) complied with the accuracy target for internal performance, while the other three did not. All those failing to comply with the target Z-score range (Nos. 2, 8 and 11) generated mean internal EC50 values which were significantly greater than the applied reference value of 0.1 mg/L zinc.

The results were also similar to the external accuracy assessments in demonstrating two sub-sets of laboratories with respect to the general level of zinc toxicity measured by the assay, although interestingly the split of laboratories was not identical. Laboratory No. 5 generated higher than expected EC50 values for DTAPS A and B (0.2-0.3 mg/L Zn) but a mean internal EC50 value which was within the expected range (0.17 mg/L Zn). The reverse situation was apparent for Laboratory No. 2 which submitted compliant results for DTAPS A and B but generated a very high mean internal EC50 for the test (2.15 mg/L Zn). Laboratory Nos. 8 and 11 demonstrated consistently higher than expected EC50 values for both internal and external accuracy assessments.

9.2.6 Summary

This part of the assessment is concerned primarily with the proximity of laboratory test values to the reference values applied for each test method.

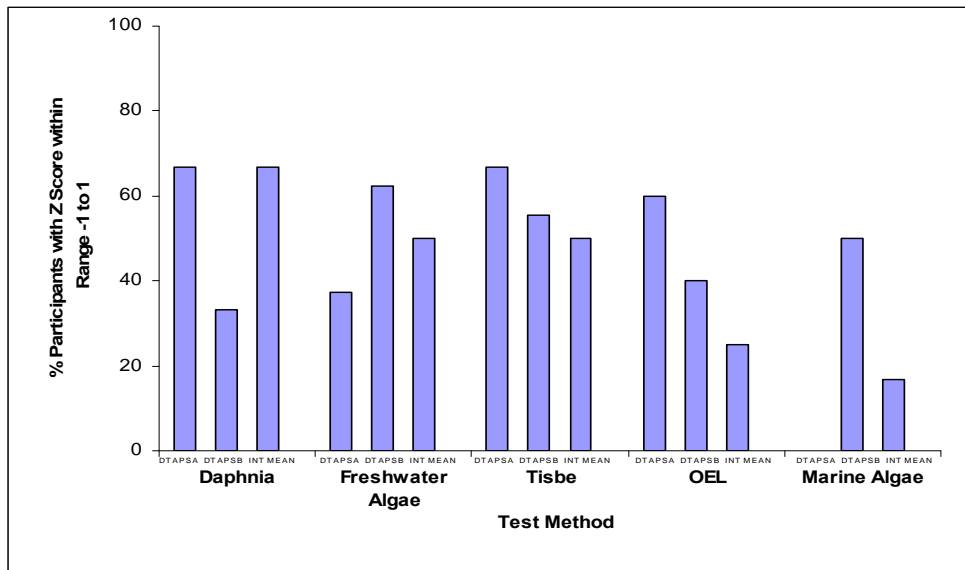
The reference values and error targets applied for each test method appear to have been achievable for most participating laboratories for both the proficiency samples and internal reference testing results. There were, however, some notable exceptions of laboratories being significantly without the specified Z-score range for one or two of the methods.

The overall summary accuracy statistics for the developmental Rounds of the scheme are given in Table 13 and Figures 12, 13 and 14.

Table13. Summary of Test Method Consensus Accuracy Performance Statistics for Round 3 of the DTAPS.

Method	Result Type	Mean EC50 (mg/L Zn)	Standard Deviation of EC50 values	% RSD of EC50 values	n	Number of Laboratories in Z-score Range				Mean Z-score	Z-score Span	Overall Z-score Span
						-1 to 1 (optimal performance)	-2 to 2	-3 to 3	<-3 or >3			
<i>Daphnia</i>	DTAPS A	1.50	0.84	56	9	6	1	2	0	0.04	5.77	17.6
	DTAPS B	2.53	2.48	98	9	3	5	0	1	2.08	16.2	
	Internal Laboratory Mean	1.62	0.51	32	9	6	3	0	0	0.29	2.58	
Freshwater Algae	DTAPS A	0.206	0.327	158	8	3	3	1	1	4.54	32.1	33.2
	DTAPS B	0.137	0.189	137	8	5	2	0	1	2.24	19.4	
	Internal Laboratory Mean	0.168	0.206	123	8	4	2	1	1	3.20	20.8	
<i>Tisbe</i>	DTAPS A	0.447	0.222	50	9	6	3	0	0	-0.44	3.42	4.48
	DTAPS B	0.513	0.34	65	9	5	3	1	0	-0.16	4.41	
	Internal Laboratory Mean	0.48	0.258	54	8	4	4	0	0	-0.32	3.15	
OEL	DTAPS A	0.195	0.079	41	5	3	1	1	0	0.91	3.2	10.4
	DTAPS B	0.229	0.129	56	5	2	1	1	1	1.48	5.22	
	Internal Laboratory Mean	0.29	0.28	96	4	1	1	1	1	2.42	10.4	
Marine Algae	DTAPS A	0.534	0.759	142	6	0	1	1	4	14.5	66.8	71.4
	DTAPS B	0.565	0.791	140	6	3	0	0	3	15.5	66.7	
	Internal Laboratory Mean	0.85	1.00	119	6	1	1	1	3	24.9	70	

Figure 12. Percentage of participants achieving optimal performance in Round 3 of the DTAPS.



The measurement of the percentage of participants achieving optimal performance for each Z-score statistic (Z-score = -1 to 1), suggests that overall the DTA methods produced similar levels of accuracy (40-70%), with the possible exception of the Marine Algae test.

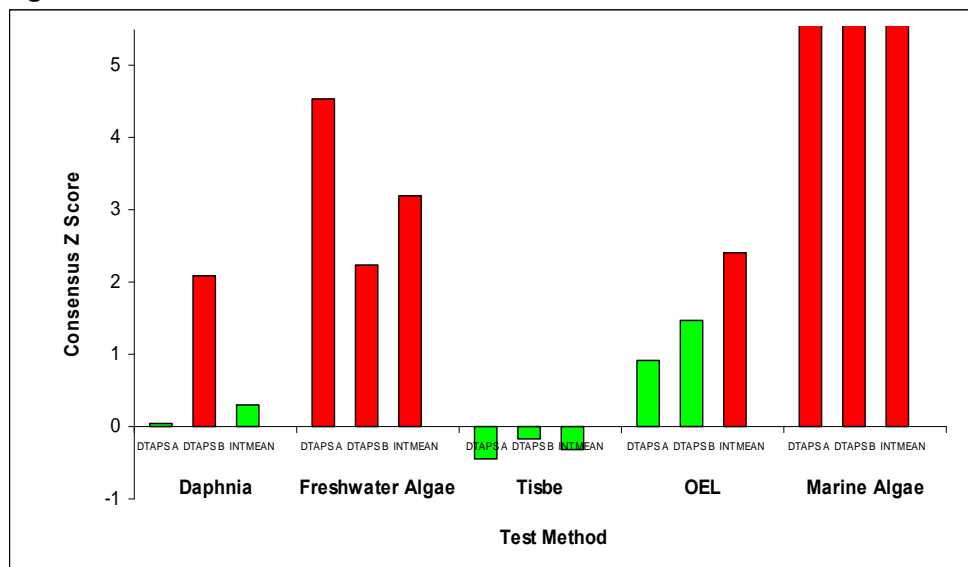
The *Daphnia* test displayed an identical level of accuracy for tests undertaken with DTAPS Sample A and internally but the results for DTAPS Sample B were approximately half as accurate. DTAPS Sample B is diluted 1:100 prior to use in DTAPS tests and it is possible that this process may infer some additional error in the analysis. This is not supported, however, by the results of the other test methods with DTAPS Sample B. With the exception of DTAPS Sample B, the results for the *Daphnia* test display an improvement in accuracy (measured as % participants with optimal performance) over Rounds 1 and 2 of the scheme.

The Freshwater Algae test was most accurate under testing with DTAPS Sample B and least accurate using DTAPS Sample A, while the *Tisbe* and OEL tests were most accurate using DTAPS Sample A and least accurate under internal testing conditions. The Freshwater Algae test displayed slightly improved accuracy (measured as % participants with optimal performance) over Rounds 1 and 2. The *Tisbe* and OEL tests displayed slightly reduced accuracy compared with Round 2 of the scheme.

The Marine Algae test showed the least accuracy (across all methods) under testing with DTAPS Sample A (no laboratories achieving optimal performance). The value for internal accuracy was also the lowest for any of the test methods (17% optimal). The DTAPS Sample B value was, however, comparable to the other DTA assays. A new SCA method for the Marine Algae test is currently under review and should be published prior to Round 4 of the DTAPS. This should assist in improving the accuracy of this methodology.

Consensus Z-scores for each method (i.e. the Z-score of the mean EC50 value obtained for each 'set' of data) can also be used to give an overall indication of the consensus proximity to the prescribed reference value for each method (Figure 13). Consensus Z-scores which exceed the specified Z-score limits (-2 to 2) for a particular method indicate poor inter and/ or intra-laboratory accuracy (relative to the reference values). There remains a strong reliance on the robustness of the reference values applied within such accuracy comparisons. The use of reference values which are unrepresentative of the consensus situation will produce similar results to those produced by poor relative accuracy.

Figure 13. Consensus Z-scores for DTAPS Test Methods for Round 3 of the DTAPS.



Green denotes compliance with provisional performance criteria ($Z = -2$ to 2); Red denotes non-compliance.

Only the *Tisbe* method generated consensus Z-Scores which were within the compliant range for both unknown concentration proficiency testing samples (DTAPS A and B) and internal reference testing (all consensus Z-Scores $= -2$ to 2), and overall displayed slightly improved performance over Rounds 1 and 2 of the scheme.

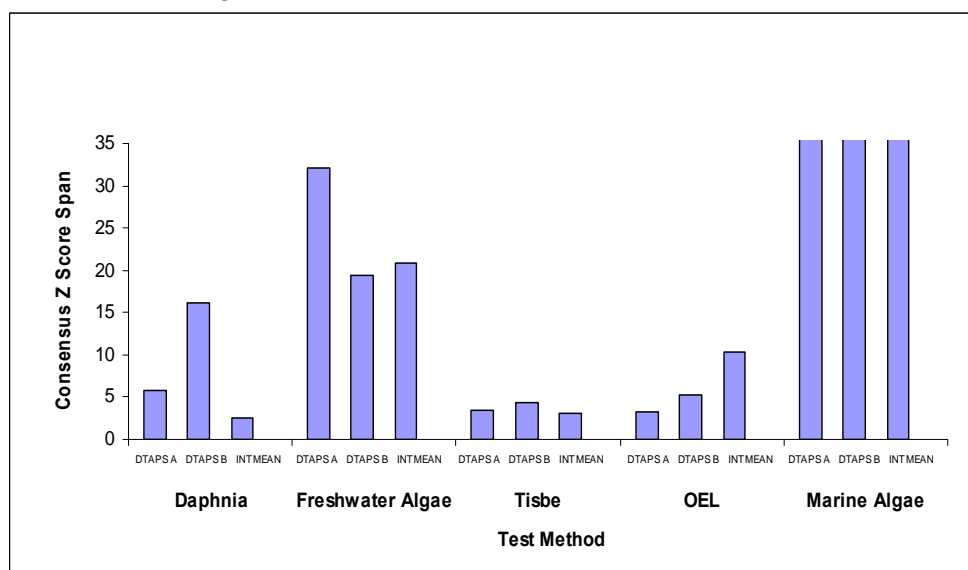
The *Daphnia* test showed improved accuracy for DTAPS Sample A and internal results compared to Rounds 1 and 2 but, as with the percentage of optimal Z-Score measurement, displayed a reduction in accuracy measured by consensus Z-Score for DTAPS Sample B (which was marginally outside the acceptable range at $z=2.08$).

The OEL test displayed a reduction in accuracy against the developmental Rounds which was most marked in the assessment of internal results (Consensus-Z Score= 2.42), although this is likely to be the combined effect of a single participant's highly non-compliant results and the relatively low number of participants.

The algal methods appear from this assessment to be the least accurate tests. The Freshwater Algal test was also the least accurate test method (as measured by consensus Z-Score) in Round 2 of the scheme and there appears to have been little improvement in Round 3. No assessment was possible in the developmental Rounds for the Marine Algae test but the very low level of accurate performance (Consensus Z-Score) measured in Round 3 provides some cause for concern. It is clear that there remains some development required with this test method to gain control over the accuracy of this test, both across and within laboratories. As described in Section 9.2.5 there may be a systematic difference in the performance of the method between two sub-sets of participant laboratories. Those failing to meet the criteria applied in Round 3 all generated EC50 values which were significantly in excess of the reference value (0.1 mg/L Zn) and these therefore have a large effect on the calculation of the consensus Z-Score (and Z-Score Span – see below). The removal of the non-compliant values provides consensus Z-Scores of -1.9 , -1.09 and 0.11 respectively for DTAPS A, B and the internal mean results.

The overall Z-score span for each method also provides an indication of the total inter and intra-laboratory deviation around the provisional reference values (Figure 14).

Figure 14. Consensus Z-score Spans for DTAPS Test Methods in Round 3 of the DTAPS.



The Z-score Span provides a similar summary of test method accuracy as seen for the consensus Z-Score assessments. Again the *Daphnia* test shows a significant reduction in accuracy for DTAPS B only and the algal growth methods are by far the least accurate methods.

While the level of accuracy for each test method was generally good across participant laboratories, the assessments are, as stated previously, only useful if the reference values applied are close to optimal for each method. A highly 'erroneous' reference value (i.e. one with a large differential from the 'true' value) will obviously invalidate any conclusions inferred from assessments made using it, but since the 'true' value is not one which can be derived for ecotoxicology tests, the reference values used must be estimated using the data which is available. This may be particularly relevant for the Marine Algae test where there remains some doubt over the reference value used to approximate optimum performance with regard to accuracy. The methods and procedures used to derive updated Z-score reference values (along with other performance criteria) are discussed in Section 10.3.

Table 14 summarises the results of statistical comparisons between the different sets of EC50 results for Round 3 of the scheme. This has been undertaken to investigate any differences between approaches for testing using DTAPS A or B, or between inter and intra-laboratory testing, that may have generated a 'systematic' bias in the operational aspects of the scheme. The individual EC50 values (mg/L Zn) have been used for these comparisons to remove the influence of the reference values and error targets.

Table 14. Statistical Comparisons of DTAPS Round 3 Accuracy Data Sets.

Test Method	DTAPS A Versus DTAPS B (2 tailed t test, 0.05)	DTAPS A + B Versus Internal Mean (Mann-Whitney U test, 0.05)
<i>Daphnia</i>	No significant difference	No significant difference
Freshwater Algae	No significant difference	No significant difference
<i>Tisbe</i>	No significant difference	No significant difference
OEL	No significant difference	No significant difference
Marine Algae	No significant difference	No significant difference

DTAPS Sample A was tested as received which necessarily required some modification of the pH of exposure dilutions, especially at higher sample concentrations. DTAPS Sample B was diluted 100 times before testing, removing the need for pH adjustment, although other physico-chemical parameters did need modification (e.g. total hardness, salinity) depending on the media used to dilute the DTAPS sample (i.e. RO/distilled water or test media). Despite the differences in the way that the DTAPS Sample A and B inter-laboratory accuracy assessments were conducted there were no indications of significant statistical differences between the sets of EC50 results obtained.

Differences between the total data sets for inter (DTAPS Samples A and B) and intra-laboratory assessments might be expected because of the additional variability inferred by the application of a range of different internal procedures for preparing zinc solutions, the increased precision inferred by working with a test material of known concentration and toxicity (relative to tests with solutions of unknown concentration) or the need for water quality modification in the external samples. No significant statistical differences were found between the sets of intra and inter-laboratory EC50 data for any of the test methods, however, suggesting that the procedural differences between the intra and inter-laboratory elements of the scheme are unlikely to have generated a systematic bias in the DTAPS proficiency testing process.

9.3 Precision Assessments

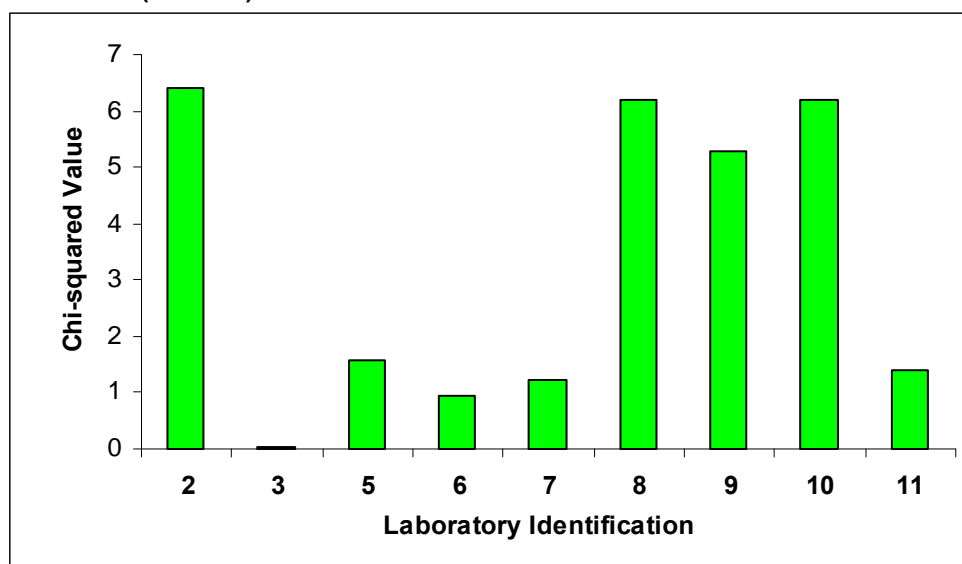
9.3.1 *Daphnia magna* 48 hour immobilisation test

The results of the Round 3 precision assessments for the *Daphnia* test are presented in Table 15 and Figure 15 below.

Table 15. *Daphnia magna* immobilisation test - Chi-squared values for Round 3 (2007/08) of the DTAPS.

Laboratory ID	χ^2 value	n	Critical χ^2 value (0.05)
2	6.40	6	11.07 (5 dof)
3	0.04	6	11.07 (5 dof)
5	1.59	6	11.07 (5 dof)
6	0.93	6	11.07 (5 dof)
7	1.24	6	11.07 (5 dof)
8	6.21	6	11.07 (5 dof)
9	5.28	6	11.07 (5 dof)
10	6.19	6	11.07 (5 dof)
11	1.41	5	9.49 (4 dof)

Figure 15. *Daphnia magna* immobilisation test – Chi-squared values for Round 3 (2007/08) of the DTAPS.



As in Round 2 of the DTAPS, all of the participant laboratories in Round 3 complied with the precision target (variance) applied for the *Daphnia* test. There was a general reduction in the levels of precision achieved by laboratories for this method between Rounds 2 and 3 but since all participants performed well within the level of precision expected (based on previous data) there is no immediate cause for concern regarding the overall level of precision.

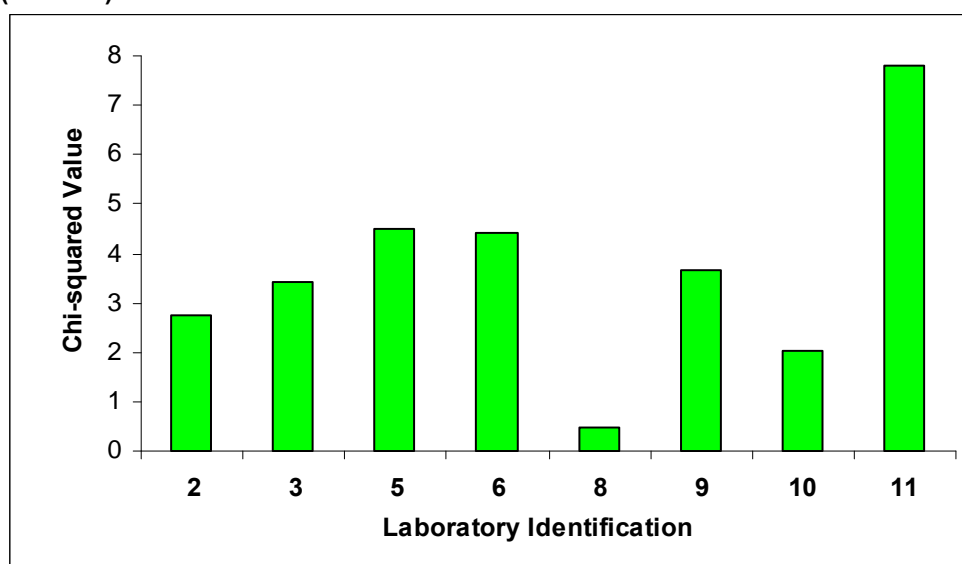
9.3.2 Freshwater Algae 72 hour inhibition of growth test

The results of the Round 3 precision assessments for the Freshwater Algae test are presented in Table 16 and Figure 16 below.

Table 16. Freshwater Algae inhibition of growth test – Chi-squared values for Round 3 (2007/08) of the DTAPS.

Laboratory ID	χ^2 value	n	Critical χ^2 value (0.05)
2	2.74	6	11.07 (5 dof)
3	3.42	6	11.07 (5 dof)
5	4.49	6	11.07 (5 dof)
6	4.41	6	11.07 (5 dof)
8	0.47	6	11.07 (5 dof)
9	3.68	6	11.07 (5 dof)
10	2.01	6	11.07 (5 dof)
11	7.80	6	11.07 (5 dof)

Figure 16. Freshwater Algae inhibition of growth test - Chi-squared values for Round 3 (2007/08) of the DTAPS.



Green denotes compliance with provisional performance criteria (Target Variance = 0.1 [Log Base 10 EC50 values]); Red denotes non-compliance.

The Freshwater Algae precision data generated a similar pattern of results as observed with the *Daphnia* test with no participants failing to meet the applied target. Overall, the degree of precision obtained within laboratories for the Freshwater Algae test was slightly less than in Round 2 of the scheme although, again as with the *Daphnia* test, all performed within the expected level of precision for the assay, based on a Chi-squared comparison of variances.

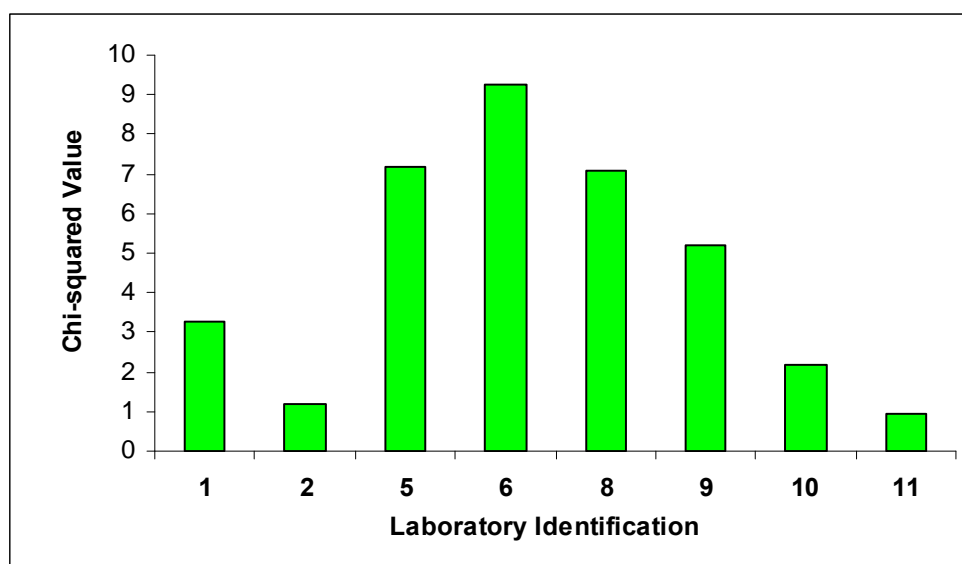
9.3.3 *Tisbe battagliai* 48 hour lethality test

The results of the Round 3 precision assessments for the *Tisbe* test are presented in Table 17 and Figure 17 below.

Table 17. *Tisbe battagliai* lethality test - Chi-squared values for Round 3 (2007/08) of the DTAPS.

Laboratory ID	χ^2 value	n	Critical χ^2 value (0.05)
1	3.27	6	11.07 (5 dof)
2	1.20	6	11.07 (5 dof)
4	Not calculable	0	Not calculable
5	7.17	6	11.07 (5 dof)
6	9.27	6	11.07 (5 dof)
8	7.08	6	11.07 (5 dof)
9	5.20	6	11.07 (5 dof)
10	2.18	6	11.07 (5 dof)
11	0.96	6	11.07 (5 dof)

Figure 17. *Tisbe battagliai* lethality test - Chi-squared values for Round 3 (2007/08) of the DTAPS.



Green denotes compliance with provisional performance criteria (Target Variance = 0.1 [Log Base 10 EC50 values]); Red denotes non-compliance.

The overall intra-laboratory precision of participants undertaking the *Tisbe* test was good in Round 3 with no incidences of failure to meet the target precision thresholds for variance. As with the *Daphnia* and Freshwater algae tests, there appears to be a general reduction in the overall level of precision achieved within individual laboratories between Rounds 2 and 3 which, if sustained could lead to some participants generating non-compliant precision values in Round 4 of the scheme. At this time, however, all participant laboratories are operating within the expected level of precision for this test.

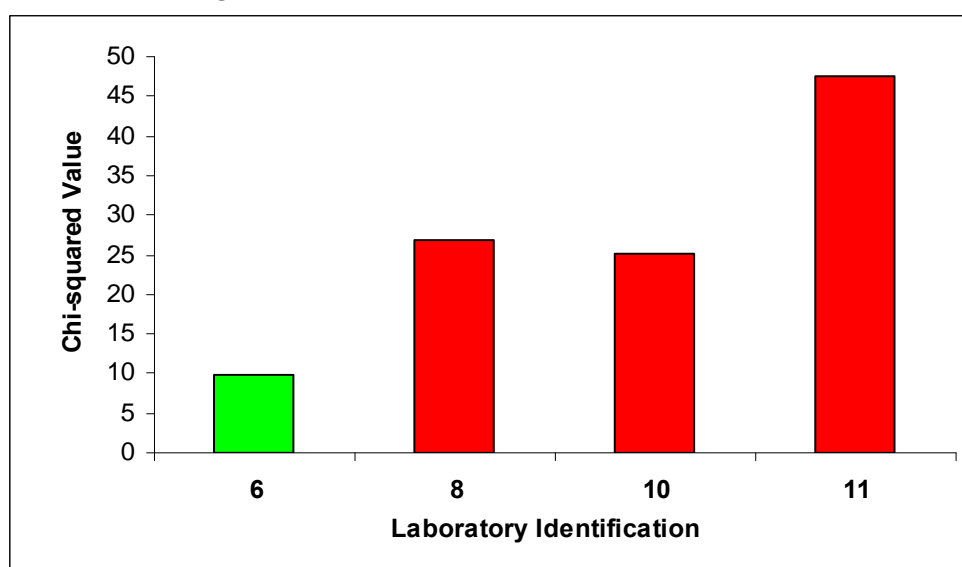
9.3.4 Oyster Embryo-Larval (OEL) 24 hour development test

The results of the Round 3 precision assessments for the OEL test are presented in Table 18 and Figure 18 below.

Table 18. OEL development test - Chi-squared values for Round 3 (2007/08) of the DTAPS.

Laboratory ID	χ^2 value	n	Critical χ^2 value (0.05)
4	Not calculable	0	Not calculable
6	9.93	6	11.07 (5 dof)
8	26.87	6	11.07 (5 dof)
10	25.16	6	11.07 (5 dof)
11	47.47	5	9.49 (4 dof)

Figure 18. OEL development test - Chi-squared values for Round 3 (2007/08) of the DTAPS.



Green denotes compliance with provisional performance criteria (Target Variance = 0.03 [Log Base 10 EC50 values]); Red denotes non-compliance.

The intra-laboratory performance for precision in the OEL test was generally poor in Round 3 of the scheme and displayed are marked reduction over Round 2. Only a single laboratory (No.6) achieved the applied precision target (Ch-squared comparison of variances). The other three participants all generated Chi-squared values which were considerably in excess of the relevant critical values, and such imprecision over repeated tests provides some considerable cause for concern over the ability of laboratories to maintain control over the process of undertaking this method. The fact that most participants met identical criteria in Round 2 of the scheme suggests that applied criteria are probably optimal but that (at least in three laboratories) there has been a degradation in the level of precision achieved. Two of the participants failing to achieve the precision target in Round 3 (Nos. 8 and 11) additionally failed the accuracy target as applied to repeated internal tests which may indicate overall analytical control over the OEL testing process has somehow been compromised within these laboratories.

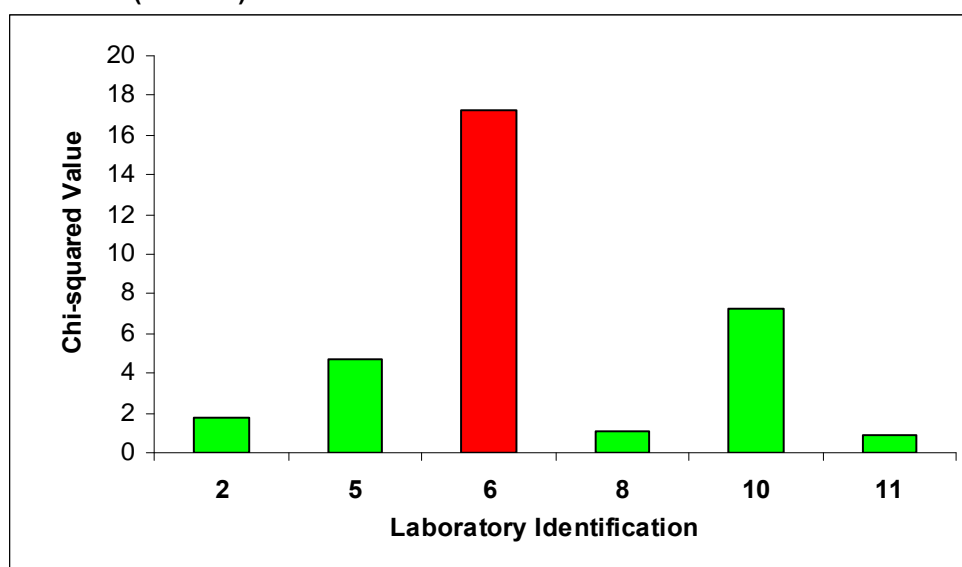
9.3.5 Marine Algae 72 hour inhibition of growth test

The results of the Round 3 precision assessments for the Marine algae test are presented in Table 19 and Figure 19 below.

Table 19. Marine Algae inhibition of growth test – Chi-squared values for Round 3 (2007/08) of the DTAPS.

Laboratory ID	χ^2 value	n	Critical χ^2 value (0.05)
2	1.79	6	11.07 (5 dof)
5	4.70	6	11.07 (5 dof)
6	17.22	6	11.07 (5 dof)
8	1.06	6	11.07 (5 dof)
10	7.30	6	11.07 (5 dof)
11	0.92	6	11.07 (5 dof)

Figure 19. Marine Algae inhibition of growth test – Chi-squared values for Round 3 (2007/08) of the DTAPS.



Green denotes compliance with provisional performance criteria (Target Variance = 0.03 [Log Base 10 EC50 values]); Red denotes non-compliance.

The overall results for Round 3 of the scheme show that precision in the Marine Algae tests is generally under control within laboratories. A single laboratory (No.6) failed the applied precision target but all the remaining participants were well within the level of precision expected for this test method.

9.3.6 Summary

This element of the assessments is concerned primarily with the spread of results from repeated tests undertaken by individual laboratories using internally prepared reference toxicant solutions, as assessed using the comparison of a laboratory's variance with a prescribed target variance.

With the exception of the OEL test, it would appear that the overall level of precision obtained within laboratories for the DTA test is good (based on what would be expected from previous data) and that any non-compliances with accuracy targets (Section 9.2) is more likely to be caused by the systematic bias of a laboratory's process with a specific test method than from a lack of repeatability. There may be both bias and repeatability issues with the OEL test.

The variability of repeated internal laboratory results have been summarised (Table 20 and Figures 20-21) in order to provide an indication of the relative precision of each DTAPS method. These comparisons have been described using the performance statistics (variance) used within the DTAPS for laboratory assessments and also using % Relative Standard Deviation to act as a check on the variance comparisons.

Table 20. Summary of Intra-Laboratory Test Method Consensus Precision Performance Statistics for Round 3 of the DTAPS.

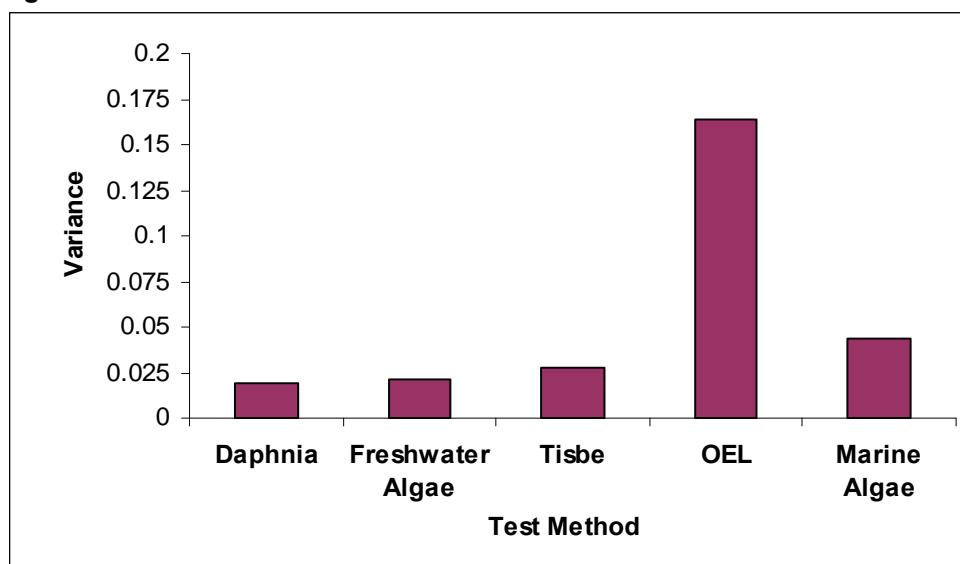
Method	Consensus Mean EC50 (mg/L Zn)	Consensus Variance*	Consensus Standard Deviation**	Consensus % RSD***	Number of Laboratories
<i>Daphnia</i>	1.62	0.0195	0.466	29	9
Freshwater Algae	0.168	0.0218	0.0395	24	8
<i>Tisbe</i>	0.48	0.0273	0.156	33	8
OEL	0.29	0.164	0.275	95	4
Marine Algae	0.85	0.044	0.206	24	6

* Mean of internal laboratory variances (Log Base 10 EC50 values).

** Mean of internal laboratory Standard Deviations.

*** (Consensus SD / Consensus Mean EC50) * 100.

Figure 20. Consensus Variance for Test Methods in Round 3 of the DTAPS.



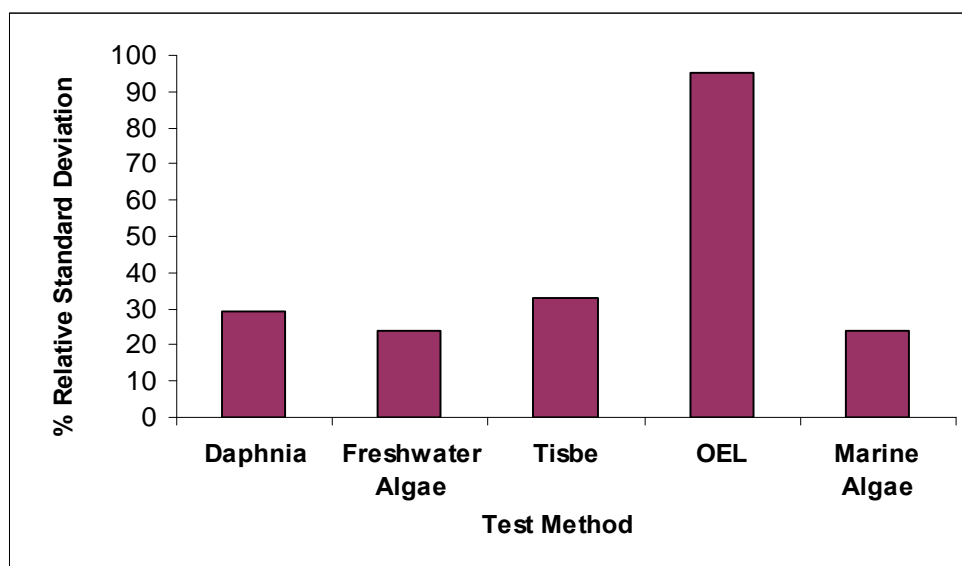
Log Base 10 EC50 values.

The consensus variances for three of the methods (*Daphnia*, Freshwater Algae & *Tisbe*) were less than the target variance value (0.03) suggesting that the overall level of variability was under good control for these tests, and this is also supported by the individual laboratory precision assessments for these methods. The level of precision for these methods appears roughly stable with no significant degradation or improvement in precision between Rounds 2 and 3.

The Marine Algae test generated a consensus variance value which was marginally above the target variance value for this method (0.04) and indicates that there may be some issues with overall precision for this method. Individual laboratory assessments suggest, however, that this high value is the result of a single laboratory's internal variance being significantly greater than the other five participants. There also appears to be some degradation in performance since Round 2 of the scheme but this is also likely to be caused by a single participant's internal results.

As suggested by the individual laboratory assessments for the OEL test, the consensus variance for this method is well beyond the applied variance target of 0.03 and indicates a considerable degradation in control over the variability of test results since Round 2 of the scheme.

Figure 21. Consensus Intra-Laboratory % Relative Standard Deviation (%RSD) for Test Methods in Round 3 of the DTAPS.



In general, use of the relative standard deviation (%RSD) of internal EC50 values supports the results of the variance assessment although the Marine Algae test appears to display a similar level of precision as the *Daphnia*, Freshwater Algae and *Tisbe* tests.

The OEL method is again highlighted as having a low level of overall precision in Round 3 of the scheme and shows a significant degradation in precision between Rounds 2 and 3.

Whilst the external (inter-laboratory) assessments were not primarily concerned with the variability of repeated results, it is possible to calculate similar consensus variability values for the results (EC50s) of testing using the distributed DTAPS (A and B) samples. These are presented in Table 21. These results take into account the facts that the participants did not know the concentrations of the samples being tested and that water quality adjustments may have been required in each test.

It would be expected that the variability of results for each test method would be greater *between* laboratories than *within* them.

A comparison of the intra and inter-laboratory variability for each method is provided in Figure 22.

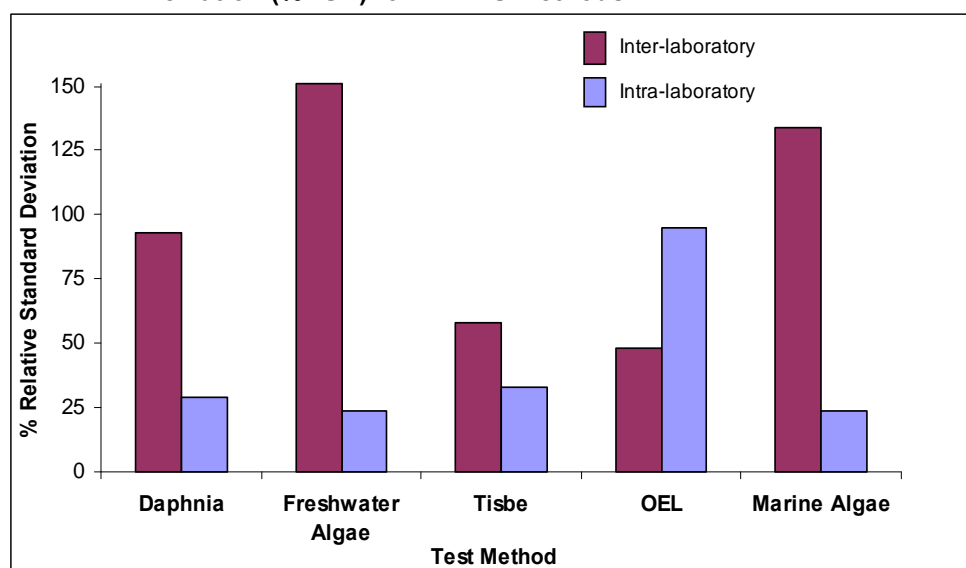
Table 21. Summary of Inter-Laboratory Test Method Consensus Precision Performance Statistics for Round 3 of the DTAPS.

Method	Consensus Mean (DTAPS A & B EC50 values)	Consensus Standard Deviation (DTAPS A & B EC50 values)*	Consensus %RSD (DTAPS A & B EC50 values)**	n
<i>Daphnia</i>	2.02	1.87	93	18
Freshwater Algae	0.172	0.26	151	16
<i>Tisbe</i>	0.48	0.278	58	18
OEL	0.212	0.102	48	10
Marine Algae	0.55	0.739	134	12

* Standard Deviation of all reported DTAPS A and B EC50 values (as mg/L Zn).

** (Consensus SD / Consensus Mean EC50) * 100.

Figure 22. Comparison of Inter and Intra-Laboratory Consensus Relative Standard Deviation (%RSD) for DTAPS Methods.



Comparison of intra and inter-laboratory precision (using relative standard deviation) clearly highlights the expected increase in control over repeated tests within laboratories as compared to between laboratories for all but the OEL test. The greatest differences are seen in the algal growth tests which is perhaps unsurprising given the potential for systematic differences between laboratories observed in inter-laboratory data sets for these two methods. There does not appear to be any significant improvements or reductions in external precision for any of the methods (again with the exception of the OEL test) between Rounds 2 and 3 of the scheme.

Unusually for such assessments, the OEL test has maintained (or slightly improved) external precision while a showing significant degradation in intra-laboratory precision to the extent that the %RSD value is approximately half that expressed for external precision. While this may be an effect of relatively low numbers of intra-laboratory participants for the OEL test (n=4), it adds strong evidence (as outlined earlier in this report) regarding the need to improve internal variability control for this method.

10. Update of DTAPS Performance Criteria

Performance criteria (reference and target values) are applied in each Round of the DTAPS which are derived from the full DTAPS data set for each test method, including all submitted internal and external EC50 (mg/L Zn) values. This assists in ensuring that the criteria values against which participants are assessed are as representative of optimal performance as possible.

The performance criteria values applied in Round 3 of the scheme were derived from all of the data submitted in Rounds 1 and 2. Updated values for use in Round 4 (08/09) of the scheme have been derived as outlined below.

The new (Round 3) data set for each test method was initially compared with each existing DTAPS data set using a Mann Whitney U test of differences (95% Confidence Limits). If the new and existing data exhibited no significant statistical difference the existing data was used to calculate outlier limits to apply to the new data set and, following exclusion of outliers, the new data was pooled with the existing data set to derive new performance criteria values. If a significant statistical difference was found between the new and existing data sets for a method, the latest data (Round 3) was considered to represent current optimal performance and was used in derive direct outlier limits to apply to the new data and the new data set (minus outliers) was used in isolation of the existing data to update performance criteria values.

Each test method data set (either existing or new) was then tested for normal distribution using an Anderson-Darling normality test. Non-normal data sets were then transformed by applying a Log Base 10 transformation to each individual data point (EC50 as mg/L Zn). The arithmetic mean and standard deviation (untransformed or Log Base 10 values) were then calculated for each data set. Values within each new (Round 3) test method data set which lay outside of the range of the mean ± 2 x the standard residual were excluded as atypical (outlying) values. The remainder was considered to represent optimal performance. If greater than 25% of the new data set was excluded using this procedure, the new data set was considered invalid and rejected. The existing performance criteria (derived from Rounds 1 and 2 and applied in Round 3) were then retained.

Table 22. New DTAPS Test Method Data Sets (Round 3) and Numbers of Outlier Values.

Test Method	<i>n</i> (Round 3 Data Set)	Number of Outliers (%)	Resultant <i>n</i> (Full Valid Data Set)	DTAPS Rounds contributing to the Full Valid Data Set
<i>Daphnia</i>	71	10	179	1, 2 and 3
Freshwater Algae*	64	5	61	3 only
<i>Tisbe</i>	66	3	64	3 only
OEL	33	33	56	1 and 2
Marine Algae*	48	10	43	3 only

The updated values were then derived from the full 'valid' data set as outlined in Table 23.

Table 23. Source of Updated DTAPS Performance Criteria.

Z Score Reference Value	Consensus Mean of Full 'Valid' Data Set
Z Score 'Standard Deviation' Value	Consensus Standard Deviation of Full 'Valid' Data Set
Target Variance Value	Consensus Variance of Full 'Valid' Data Set

In order to utilise the entire valid DTAPS data set in updating the values, and therefore maximise statistical viability, the updated performance values will reflect ‘consensus’ (i.e. inter-laboratory) performance. While DTAPS precision (variance) targets will therefore reflect the amount of variation expected across laboratories, participants who achieve them will be demonstrating that their internal precision is as least as controlled as that expected between laboratories. This provides a measure of ‘allowance’ around the internal targets and ensures that only those participants with a significant lack of internal precision (i.e. worse than expected across a number of laboratories) are highlighted for investigation.

Table 24. Updated DTAPS Performance Criteria.

Test Method	Accuracy			Precision
	Reference Value (mg/L Zn) <small>(Previous Reference Value)</small>	Standard Deviation Target (mg/L Zn) <small>(Previous SD Target)</small>	Compliant Z Score Range	Variance Target* <small>(Previous Variance Target)</small>
Daphnia	1.53	0.53		-2 to 2
	(1.47)	(0.51)	(0.03)	
Freshwater Algae	0.16	0.19	0.12	
	(0.07)	(0.03)	(0.03)	
Tisbe	0.48	0.29	0.06	
	(0.55)	(0.23)	(0.03)	
OEL	0.14	0.06	0.03	
	(0.14)	(0.06)	(0.03)	
Marine Algae	0.86	0.93	0.36	
	(0.1)	(0.03)	(0.04)	

* Log Base 10 EC50 values.

The DTAPS performance values are updated following each DTAPS Round to ensure that they are relevant to the optimal levels of performance achievable for each test method and should therefore represent a gradual improvement in performance across successive DTAPS Rounds. In general, it is not recommended that performance values be updated to reflect a reduction in the level of performance (accuracy and precision) and this could promote a gradual deterioration in the level of control applied to the test processes.

The new accuracy targets represent a larger allowance than previously applied in all cases except the OEL test which remains unchanged. It is therefore recommended that the previous standard deviation targets (applied around the updated reference value) should be retained for Round 4 of the scheme. Similarly the updated variance (precision) targets also represent a reduction in the level of precision expected for three of the tests methods (Freshwater Algae, *Tisbe* and Marine Algae) while the *Daphnia* and OEL tests remain the same. It is recommended that the variance targets applied in Round 3 be retained for the Freshwater Algae and *Tisbe* tests. The significant increase in the reference value, and widening of the accuracy and precision targets, for the Marine Algae test has been caused as a direct result of the existence of two sub-sets of laboratories with regard to zinc toxicity using this method (one group showing considerably less toxicity than the other). There remains some investigative work to be undertaken to ascertain the reason for this (probably systematic) difference between laboratories. To allow for this, it is recommended that the values are updated for this test method for the next Round, despite representing a fall in overall performance. The performance values can then be re-aligned following the next Round of the scheme.

11. Comparison of Statistical Methods used to Derive EC50 Values

In addition to the above data submissions, participant laboratories were also requested to calculate EC50 values using the example test data provided in each SCA Test Method Guideline. This process was undertaken to allow a comparison of results derived using different statistical software packages as applied to identical raw test data.

Laboratories No. 4, 7, 6 (algae only) & 11 failed to submit a demonstration of their EC50 calculation methods using example data from the SCA methods.

Tables 25 and 26 display each participant's calculated EC50 value for each test method alongside the guideline EC50 value given in each SCA Test Method.

Table 25. Comparison of calculated EC50 values using identical raw data for the *Daphnia*, *Tisbe* & OEL tests.

Laboratory ID	SCA Method Guideline	Scenario 1		Scenario 2		Scenario 3	
		Lab Value (EC50)	Guideline Value (EC50)	Lab Value (EC50)	Guideline Value (EC50)	Lab Value (EC50)	Guideline Value (EC50)
1	<i>Tisbe</i>	3.26	3.3	4.68	4.8	3.23	3.2
2	<i>Daphnia</i>	3.26	3.3	4.82	4.8	3.18	3.2
	<i>Tisbe</i>	3.26	3.3	4.82	4.8	3.18	3.2
3	<i>Daphnia</i>	3.52	3.3	4.77	4.8	3.18	3.2
5	<i>Daphnia</i>	3.26	3.3	4.82	4.8	3.18	3.2
	<i>Tisbe</i>	3.26	3.3	4.82	4.8	3.18	3.2
6	<i>Tisbe</i>	3.26	3.3	-	-	-	-
	<i>Daphnia</i>	3.26	3.3	-	-	-	-
	OEL	0.66	0.67	-	-	-	-
8	<i>Daphnia</i>	3.26	3.3	-	-	-	-
	<i>Tisbe</i>	3.26	3.3	-	-	-	-
	OEL	0.67	0.66	-	-	-	-
9	<i>Daphnia</i>	3.3	3.3	4.84	4.8	3.18	3.2
	<i>Tisbe</i>	3.3	3.3	4.84	4.8	3.18	3.2
10	<i>Daphnia</i>	3.26	3.3	4.82	4.8	-	-
	<i>Tisbe</i>	3.26	3.3	4.82	4.8	-	-
	OEL	0.66	0.66	0.501	0.52	-	-

Table 26. Comparison of calculated EC50 values using identical raw data for the Algal Inhibition of Growth tests.

Laboratory ID	Lab Value (EC50)	Guideline Value (EC50 by linear interpolation)
2	0.122*	0.11
3	0.09*	0.11
5	0.122*	0.11
8	0.089**	0.11
9	0.1***	0.11
10	0.11*	0.11

*Toxcalc linear interpolation

** Toxcalc Maximum-Likelihood Probit

***Safestat

The comparisons outlined above clearly demonstrate that those laboratories participating in this exercise produced EC50 values which were all very close to the guideline values provided in the SCA Method Guidelines.

The calculation of test endpoint values was highlighted in the developmental Rounds (2005-2007) of the scheme as an area of DTA test method performance which might be subject to significant variation amongst laboratories undertaking the DTA test methods. There was some considerable variation between EC50 values (for the DTAPS Samples) derived by participants and those derived by the provider (using the participants raw data) in Rounds 1 and 2, however, comparisons of values derived from identical data across all participants in Round 3 does not appear to show a similar level of variation. In part this is probably because no interpretation of statistical outputs was required in Round 3 (e.g. decisions on which particular statistical procedure to apply) and the comparisons have been undertaken on data which is designed to provide the optimal EC50 output. This may therefore provide an erroneously positive view of the agreement between EC50 derivation at different laboratories. The level of agreement may, however, simply have improved following the publication of a full set of SCA methods which provide clear guidance on the application of different statistical models to derive an EC50 value.

The testing of effluent samples under the regulatory DTA regime (PPC) usually involves the use of a No Observed Effect Concentration (NOEC) to express the hazard posed by an industrial effluent, rather than an EC50. The EC50 value of reference tests, although useful in assessing overall laboratory capability with regard to precision and accuracy, may not be the optimal statistic to provide a demonstration of the potential effect on risk assessments of the application of different statistical procedures/ proficiencies. Beyond the DTAPS itself, it is evident from the provider's ongoing review of the regulatory DTA process that some laboratories fail to derive a LOEC in the DTA tests they undertake, usually as a result of applying a truncated range of the lower (<10% v/v effluent) test concentrations. This can provide considerable issues in hazard assessments based on LOECs and it might therefore be of more benefit to base any further investigations with regard to statistical outputs in later Rounds of the scheme on the LOEC value. For example, participants should be able to derive a LOEC for the tests undertaken with the DTAPS samples.

12. References

- 1. Environment Agency (1996) – Performance Standards for Ecotoxicity Tests; SR 4166/1.**
- 2. Environment Agency (1996) – Implementation of Performance Standards of Ecotoxicological Methods; E18.**
- 3. Environment Agency (1999) – A Proposed Scheme to Ensure the Quality of Data Generated by Laboratories Undertaking Regulatory Ecotoxicological Testing; P166.**
- 4. Environment Agency (1999) – Report on a Ring Test of the 48 Hour *Tisbe battagliai* Lethality Test; E90.**
- 5. International Organization for Standardisation [ISO] / International Electrotechnical Commission [IEC] (1999) – Proficiency Testing by Inter-laboratory Comparisons, Part 1: Development and Operation of Proficiency Testing Schemes & Part 2: Selection and use of Proficiency Testing Schemes by Laboratory Accreditation Bodies; ISO/IEC Guide 43-1 & 2.**
- 6. Environment Agency (2001) – Assessment of the Options for a Regulatory Ecotoxicology Testing Quality Scheme (RETQS); P426.**
- 7. Environment Agency (2001) – Ecotoxicity Test Methods for Effluent and Receiving Water Assessment – Comprehensive Guidance; EP236.**
- 8. Environment Agency (2001) – Ecotoxicity Test Methods for Effluent and Receiving Water Assessment – Supplementary Advice to International Test Guidelines; EP236.**
- 9. Environment Agency (2005-2007) – Direct Toxicity Assessment Proficiency Scheme (DTAPS) Quality Manual; Issues 1-3.**
- 10. Environment Agency (2005-2007) – Direct Toxicity Assessment Proficiency Scheme (DTAPS) Operational Manual; Issues 1-6.**
- 11. Environment Agency (2006) – Monitoring Certification Scheme (MCERTS) Performance Standard for Laboratories undertaking Direct Toxicity Assessment of Effluents; Version 1.**
- 12. Environment Agency (2006) – Monitoring Certification Scheme (MCERTS) for Laboratories undertaking Direct Toxicity Assessment of Effluents; Laboratory Assessment; Version 1.**
- 13. Aquacheck Limited (2006) – Assessment of the Accuracy of Results from the Direct Toxicity Assessment Proficiency Scheme (DTAPS).**

- 14. Environment Agency (2006 & 2007) – Standing Committee of Analysts (SCA), The Direct Toxicity Assessment of Aqueous Environmental Samples using the Oyster (*Crassostrea gigas*) Embryo-Larval Development Test; Blue Book Nos. 203 & 209.**
- 15. Environment Agency (2006 & 2007) - Standing Committee of Analysts (SCA), The Direct Toxicity Assessment of Aqueous Environmental Samples using the Marine Copepod *Tisbe battagliai* Lethality Test; Blue Book Nos. 205 & 210.**
- 16. Environment Agency (2007) - Standing Committee of Analysts (SCA), The Direct Toxicity Assessment of Aqueous Environmental Samples using the Juvenile *Daphnia magna* Immobilisation Test; Blue Book No. 208.**
- 17. International Laboratory Accreditation Cooperation [ILAC] (2007) – Guidelines for the Requirements for the Competence of Providers of Proficiency Testing Schemes; ILAC-G13:07/2007.**
- 18. Environment Agency (2008) – Standing Committee of Analysts (SCA), The Direct Toxicity Assessment of Aqueous Environmental Samples using the *Pseudokirchneriella subcapitata* Freshwater Algal Growth Inhibition Test, Blue Book No. 219.**
- 19. Environment Agency (2008) – Direct Toxicity Assessment Proficiency Scheme (DTAPS) Developmental Phase Report.**

Appendix 1

DTAPS Participants - Compliance with Performance Criteria for Round 3 of the scheme

Laboratory Identifications (Round 3)		Method	DTAPS A	DTAPS B	Internal Mean	Variance
1	<i>Daphnia</i>		NP	NP	NP	NP
	Freshwater Algae		NP	NP	NP	NP
	<i>Tisbe</i>		Y	Y	Y	Y
	OEL		NP	NP	NP	NP
	Marine Algae		NP	NP	NP	NP
2	<i>Daphnia</i>		Y	N	Y	Y
	Freshwater Algae		Y	Y	N	Y
	<i>Tisbe</i>		Y	Y	Y	Y
	OEL		NP	NP	NP	NP
	Marine Algae		Y	Y	N	Y
3	<i>Daphnia</i>		Y	Y	Y	Y
	Freshwater Algae		Y	Y	Y	Y
	<i>Tisbe</i>		NP	NP	NP	NP
	OEL		NP	NP	NP	NP
	Marine Algae		NP	NP	NP	NP
4	<i>Daphnia</i>		NP	NP	NP	NP
	Freshwater Algae		NP	NP	NP	NP
	<i>Tisbe</i>		Y	Y	NP	NP
	OEL		Y	N	NP	NP
	Marine Algae		NP	NP	NP	NP
5	<i>Daphnia</i>		Y	Y	Y	Y
	Freshwater Algae		N	Y	Y	Y
	<i>Tisbe</i>		Y	Y	Y	Y
	OEL		NP	NP	NP	NP
	Marine Algae		N	N	Y	Y
6	<i>Daphnia</i>		Y	Y	Y	Y
	Freshwater Algae		Y	Y	Y	Y
	<i>Tisbe</i>		Y	Y	Y	Y
	OEL		Y	Y	Y	Y
	Marine Algae		N	Y	Y	N
7	<i>Daphnia</i>		N	Y	Y	Y
	Freshwater Algae		NP	NP	NP	NP
	<i>Tisbe</i>		NP	NP	NP	NP
	OEL		NP	NP	NP	NP
	Marine Algae		NP	NP	NP	NP

Appendix 1 (continued)

DTAPS Participants - Compliance with Performance Criteria for Round 3 of the scheme.

Laboratory Identifications (Round 3)		Method	DTAPS A	DTAPS B	Internal Mean	Variance
8	<i>Daphnia</i>		Y	Y	Y	Y
	Freshwater Algae		N	N	N	Y
	<i>Tisbe</i>		Y	Y	Y	Y
	OEL		N	Y	N	N
	Marine Algae		N	N	N	Y
9	<i>Daphnia</i>		Y	Y	Y	Y
	Freshwater Algae		Y	Y	Y	Y
	<i>Tisbe</i>		Y	Y	Y	Y
	OEL		NP	NP	NP	NP
	Marine Algae		NP	NP	NP	NP
10	<i>Daphnia</i>		Y	Y	Y	Y
	Freshwater Algae		Y	Y	Y	Y
	<i>Tisbe</i>		Y	Y	Y	Y
	OEL		Y	N	Y	N
	Marine Algae		N	Y	Y	Y
11	<i>Daphnia</i>		N	Y	Y	Y
	Freshwater Algae		Y	Y	Y	Y
	<i>Tisbe</i>		Y	N	Y	Y
	OEL		Y	Y	N	N
	Marine Algae		N	N	N	Y

Y = 'Compliant' Result

N = 'Non-compliant' Result

NP = Either did not participate or did not submit required data