Direct Toxicity Assessment Proficiency Scheme (DTAPS)

Round 4 Report

2008/2009

4095

Provider: Environment Agency, Ecosystems Science
Red Kite House, Howbery Park
Wallingford, Oxon OX10 8BD
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 4 of the DTAPS (2008-2009). The numbers of registered participants in the scheme across all test methods ranged from 5 (Oyster Embryo-Larval Development & Freshwater Algae Inhibition of Growth tests) to 7 (*Tisbe battagliai* lethality test).
The assessment comprised the 'spot' testing of centrally prepared reference toxicant solutions of unknown concentration and repeat testing of internally prepared reference toxicant solutions.
Round 4 laboratory performance in terms of accuracy (using a Z-score assessment procedure) was generally very good for the invertebrate DTA tests (62.5 to 100% compliance (Z-score = -2 to 2)) but poor for the algal growth tests (0 to 33% compliance (Z-score = -2 to 2) across all participants.
Internal laboratory performance for Round 4 in terms of compliance with an applied precision target (using a comparison of variance approach) was generally good (50-100% compliance) for all of the test methods.

Contents

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 (2008/09), the provider collated both internal and external laboratory data returns, and assessed and reported on internal and external laboratory performance. The preparation and distribution of centrally prepared reference toxicant solutions (zinc sulphate) was undertaken by LGC Ltd under contract to the provider.
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).

The proficiency testing samples (DTAPS A & B) were distributed to all registered participants by LGC Ltd in the week commencing 29th September 2008. The samples were tested by participants using the methods they had opted to participate in over the period 6th October 2008 to 30th January 2009. All participant data was submitted to the Environment Agency by the 13th February 2009. Data processing and analysis of submitted data by the provider was completed by 30th April 2009.

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 = [Laboratory Result – Ref. Value] / [Standard Deviation Target]

The reference values and standard deviation target values for each test method (as applied in Round 4) 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
Daphnia magna 48hr immobilisation	1.53	0.51	-2 to 2
Tisbe battagliai 48hr mortality	0.48	0.23	
Oyster embryo-larval development	0.14	0.06	
Freshwater algae inhibition of growth	0.15	0.03	
Marine algae inhibition of growth	0.1	0.03	

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 4 of the DTAPS, the variance of each laboratory's internal data submissions was compared with a target variance (σ^2), derived from previous Rounds of the DTAPS, 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 4 for each test method is shown in Table 2 below.

Table 2. DTAPS Precision Performance Criteria

Test Method	Target Variance (σ) *
Daphnia magna 48hr immobilisation	0.03
Tisbe battagliai 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 4 of the DTAPS;

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

5. Proficiency Testing Samples

The proficiency testing samples, DTAPS A and B, comprised solutions of zinc sulphate ($ZnSO_4.7H_2O$) 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-monthly intervals throughout the duration of the testing phase. This ensured the verification of nominal zinc concentrations and demonstrated the homogeneity and stability of each batch of sample over the duration of the Round.

In Round 4 of the DTAPS (2008/09), the proficiency testing samples were prepared and distributed by LGC Ltd under contract to the provider. DTAPS Sample A was prepared at a nominal zinc concentration of 7.5 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 750 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.

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

While the scheme attracted the participation of the all of the UK contract ecotoxicological testing laboratories currently actively involved in undertaking DTA effluent testing, a number of such facilities previously involved in the DTAPS declined to participate in Round 4 of the scheme. In addition, a number of laboratories continuing to participate in the scheme decided to reduce the range of test methods undertaken. The overall result is an appreciable fall in the numbers of participants for the DTAPS, particularly in the freshwater methods, compared to previous Rounds of the scheme.
The total numbers of participants for each test method across the previous and current Rounds 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. The majority of chemical industries required to undertake regulatory DTA assessments are, however, situated in estuaries and therefore discharge effluent to a saline environment, and the volume of freshwater DTA regulatory work available to the testing facilities is therefore much lower than DTA assessments on discharges to estuaries. This has inevitably led to a number of facilities focussing their DTA services on the marine test methods, and this is reflected in the significant drop in DTAPS participants for the *Daphnia* immobilisation and Freshwater Algae inhibition of growth tests.
The *Tisbe* lethality test also saw a drop in participation between Round 3 and 4, although this continued to represent the most subscribed test in Round 4 of the scheme. Participation in the Oyster Embryo-Larval Development and Marine Algae inhibition of growth tests remained the same as in Round 3 of the scheme.
The Oyster Embryo-Larval Development and Freshwater Algae inhibition of growth tests were the least subscribed methods in Round 4 of the scheme.

Figure 1. Numbers of Participant Laboratories for Rounds 1 - 4 of the DTAPS.

The quality of data submissions from participant laboratories remained high in Round 4 of the scheme and only a single laboratory (No.7) failing to submit internal data returns along with their proficiency testing results (DTAPS A and B). One participant (Laboratory No.1) also submitted insufficient internal data (n=<6) for a single test method. No assessment of internal accuracy or precision was made for Laboratory No. 7, but *were* carried out for Laboratory No.1 since there was considered to be sufficient data to conduct the analysis (n=5). In addition, a single registered participant failed to submit any data for Round 4 of the scheme (Laboratory No.9) and another (Laboratory No.3) did not submit any data for two of the four methods for which they had registered.

9. Results

9.1 General

All of the data submitted by participants in Round 4 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)

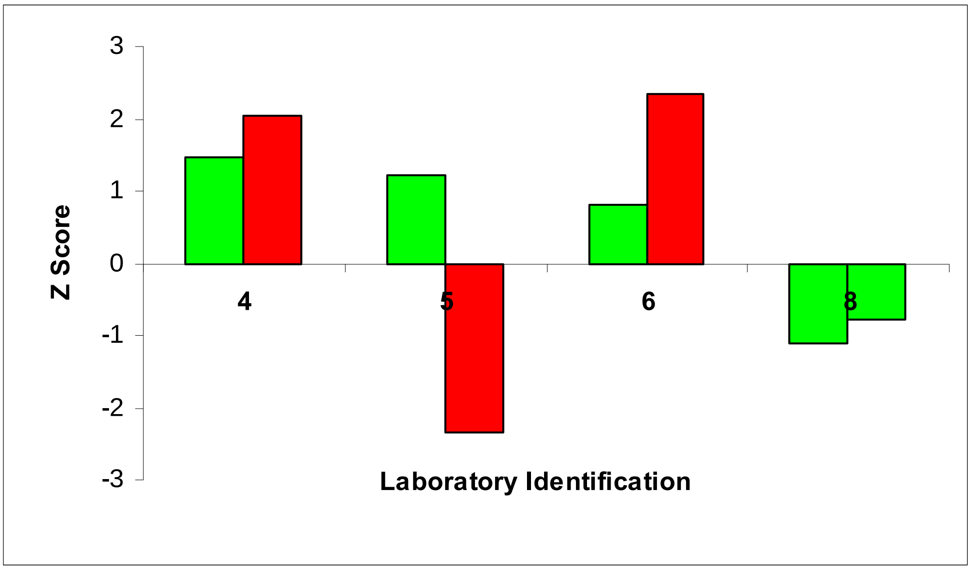
The results of the Round 4 accuracy (Z-score) assessments for the *Daphnia* test are presented in Table 3 and Figure 2, below.

Table 3. *Daphnia* EC50 Values and Z-scores for Round 4 (2008/09) 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	2.28	1.47	2.56	2.03
3	No result submitted	NA	No result submitted	NA
5	2.15	1.21	0.33	-2.35
6	1.94	0.80	2.72	2.33
8	0.97	-1.10	1.13	-0.78
9	No result submitted	NA	No result submitted	NA
Consensus				
Mean	1.84	0.595	1.69	0.308
Range (update with lab 5)	0.97-2.28	-1.10-1.47	0.33-2.72	-2.35-2.33
Standard Deviation	0.593	NA	1.15	NA
% RSD (% CV)	32.3	NA	68.4	NA

** As reported by participant.*

Figure 2. *Daphnia* Z-scores for Round 4 (2008/09) 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 4 of the DTAPS.

Four laboratories submitted results and one obtained compliant Z-scores for both DTAPS A and B. None of the participants generated a non-compliant Z-score for DTAPS A, while three generated marginally non-compliant Z-scores for DTAPS B.

While the number of non-compliances for the inter-laboratory accuracy assessments for the *Daphnia* test were similar in this Round to the results obtained in Round 3 of the scheme, the significant reduction in the total participant numbers means that the overall level of performance is lower than that demonstrated previously. Five (of nine) laboratories obtained compliant Z-scores for both DTAPS samples (A & B) in Round 3 while only a single laboratory (of four) managed to achieve this level of performance in Round 4, although all of the Round 4 non-compliances were very marginal exceedances of the specified Z-score range.

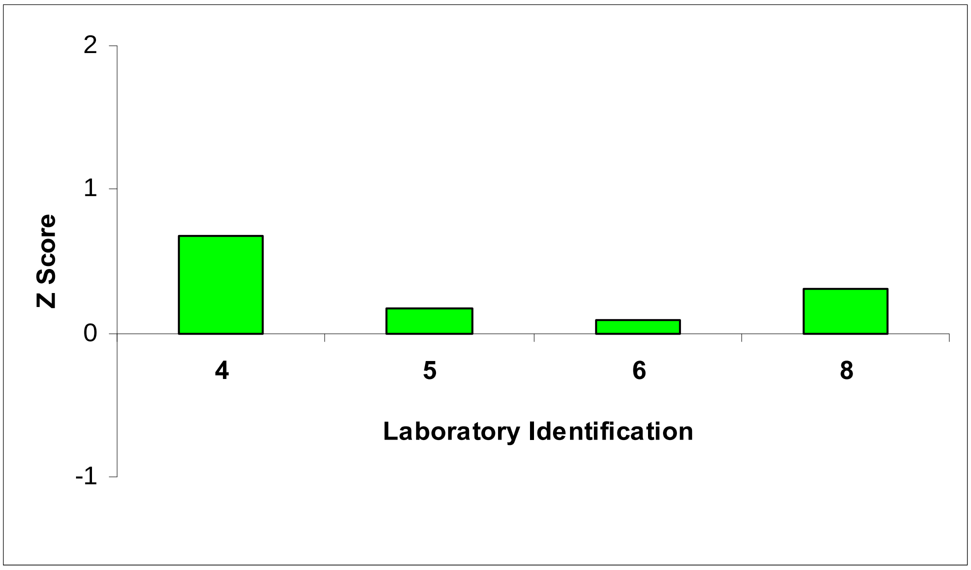
Intra-Laboratory Z-scores (Mean Internal Reference Standard Result)

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 4 (2008/09) of the DTAPS.

Lab ID	Mean Internal EC50 (mg/L Zn)	Z-score	n
4	1.88	0.68	6
3	No results submitted	NA	0
5	1.62	0.17	6
6	1.58	0.09	6
8	1.69	0.31	6
9	No results submitted	NA	0
Consensus			
Mean	1.69	0.313	
Range	1.58-1.88	0.09-0.68	
Std Dev.	0.133	NA	
% RSD (% CV)	7.86	NA	

Figure 3. Intra-Laboratory *Daphnia* Z-scores for Round 4 (2008/09) 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 four participants in Round 4 of the scheme complied with the standard deviation target applied around the applied reference value (1.53 mg/L zinc), and all laboratories generated results which were slightly above this reference value.

Compliance of mean internal EC50 values with the applied accuracy target was very similar to Round 3 of the scheme in which nine laboratories all generated compliant Z-scores, although unlike the current Round, the range of results above and below the specified reference value were roughly equal.

9.2.2 Freshwater Algae 72 hour inhibition of growth test

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

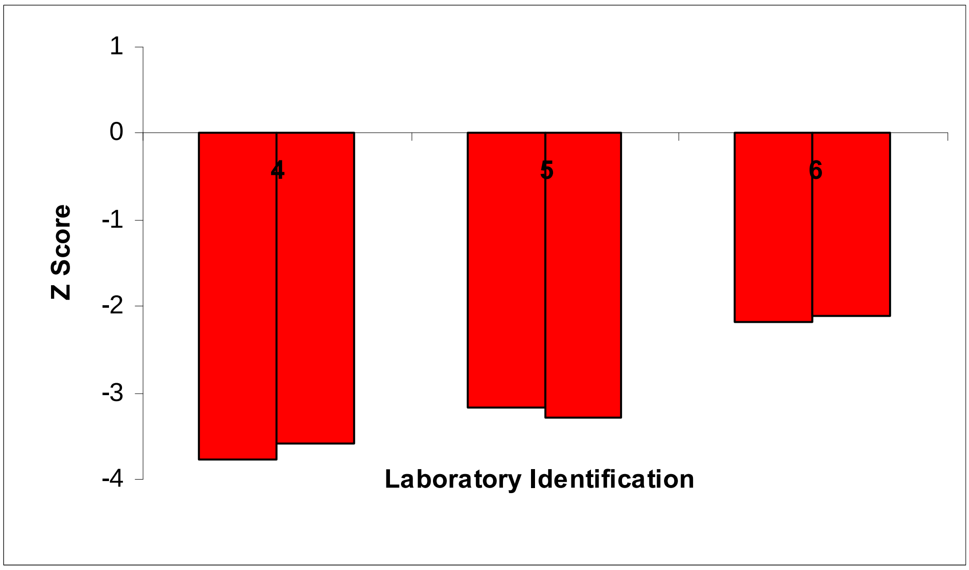
The results of the Round 4 accuracy (Z-score) assessments for the Freshwater Algae test are presented in Table 5 and Figure 4, below.

Table 5. Freshwater Algae EC50 Values and Z-scores for Round 4 (2008/09) 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
3	No result submitted	NA	No result submitted	NA
4	0.037	-3.77	0.042	-3.59
5	0.055	-3.17	0.051	-3.29
6	0.085	-2.18	0.087	-2.11
9	No result submitted	NA	No result submitted	NA
Consensus				
Mean	0.059	-3.04	0.06	-2.30
Range	0.037-0.085	-3.77 to -2.18	0.042-0.087	-3.59 to -2.11
Standard Deviation	0.024	NA	0.024	NA
% RSD (% CV)	41.1	NA	39.7	NA

** As reported by participant.*

Figure 4. Freshwater Algae Z-scores for Round 4 (2008/09) 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 poor for the Freshwater Algae growth test in this Round, in considerable contrast to the previous Round in which only two laboratories generated non-compliant results (from eight that submitted results). This suggests a significant decline in the level of performance (as expressed by accuracy assessments) displayed by laboratories undertaking this test.

Only three laboratories submitted results (of five registered to participate for this test method), and all obtained non-compliant Z-scores for both DTAPS A and B, all of which were significantly less than the applied target reference value.

The target reference value for the Freshwater Algae growth test was, after analysis of the Round 3 DTAPS data set, increased from 0.07 to 0.15 mg/L Zn for Round 4 of the scheme. All three laboratories submitting data for Round 4 would have complied with the target value (0.07 mg/L Zn) previously applied. This suggests that the increased reference value may not represent the current level of optimal performance that can be achieved by laboratories undertaking this test, despite representing the consensus level of performance achieved by participants over the previous Rounds of the scheme. It is also possible that a systematic bias has somehow been introduced into the prescribed methodology for this test (Ref 18) between Rounds 3 and 4, and has resulted in all three participants generating lower than expected EC50 values.

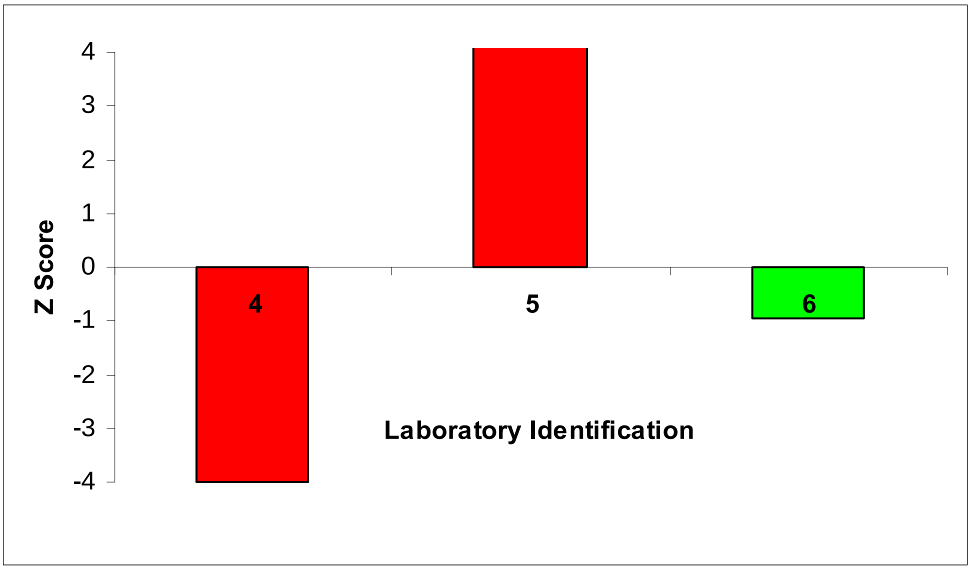
Intra-Laboratory Z-scores (Mean Internal Reference Standard Result)

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 4 (2008/09) of the DTAPS.

Lab ID	Mean Internal EC50 (mg/L Zn)	Z-score	n
3	No results submitted	NA	0
4	0.03	-4.00	7
5	0.81	22.06	6
6	0.12	-0.94	6
9	No results submitted	NA	0
Consensus			
Mean	0.48	-0.32	
Range	0.03-0.81	-4-22.06	
Std Dev.	0.26	NA	
% RSD (% CV)	54.2	NA	

Figure 5. Intra-Laboratory Freshwater Algae Z-scores for Round 4 (2008/09) of the DTAPS.



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

The mean internal EC50 value for two (of three) laboratories submitting results in Round 4 of the scheme did not comply with the standard deviation target applied around the reference value of 0.15 mg/L zinc. Of these, Laboratory No.4 generated a mean EC50 value that was much lower than the applied reference value (Z-score = -4.0), and supports the results obtained for the DTAPS A and B samples, while Laboratory No.5 generated a value which was significantly greater than the reference value (Z-score = 22.06), suggesting a different issue than that highlighted by the DTAPS A and B results. The result for the compliant laboratory was marginally lower than the applied reference value.

Two laboratories (of 8) also failed to comply with the applied reference value in Round 3 of the scheme, although both of these participants generated results which were higher than the reference value applied in that Round (and nearer the value applied in Round 4).

9.2.3 *Tisbe battagliai* 48 hour lethality test

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

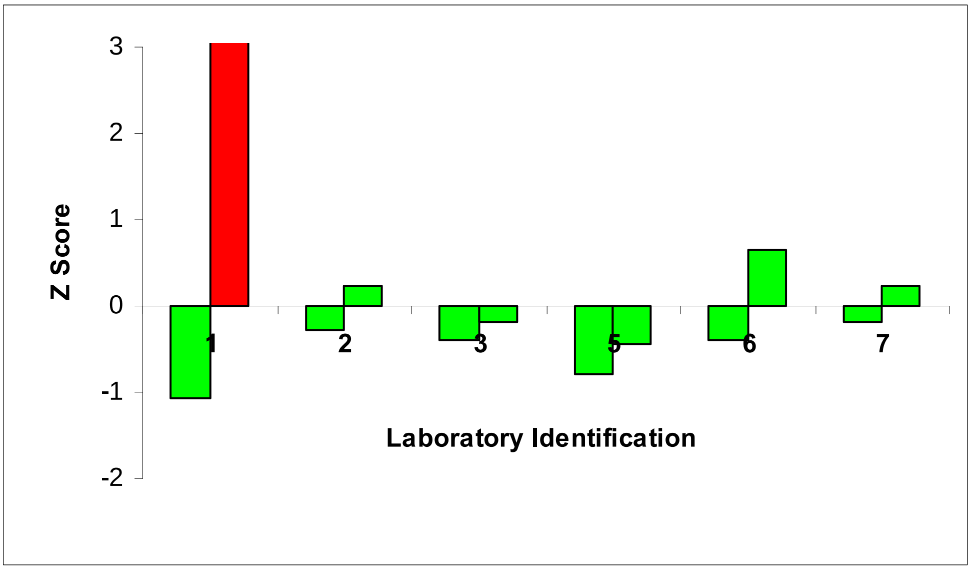
The results of the Round 4 accuracy (Z-score) assessments for the *Tisbe* test are presented in Table 7 and Figure 6 below.

Table 7. *Tisbe* EC50 Values and Z-scores for Round 4 (2008/09) 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	0.24	-1.06	23.8	102
2	0.41	-0.29	0.54	0.24
3	0.39	-0.39	0.44	-0.18
5	0.29	-0.80	0.38	-0.45
6	0.39	-0.39	0.63	0.64
7	0.44	-0.19	0.55	0.24
9	No result submitted	NA	No result submitted	NA
Consensus				
Mean	0.36	-0.52	4.39	17.1
Range	0.24-0.44	-1.06 to -0.19	0.38-23.8	-0.45-102
Standard Deviation	0.077	NA	9.51	NA
% RSD (% CV)	21.5	NA	217	NA

As reported by participant.

Figure 6. *Tisbe* Z-scores for Round 4 (2008/09) 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).

Laboratory No.1 generated a non-compliant result for DTAPS B only (Z-score = 102), although it was suspected (but not confirmed) that this result was generated owing to an error either in the preparation of the DTAPS B sample or in calculation of the EC50 (i.e. 1/100 dilution not performed or accounted for in result calculation). The majority of participants generated EC50 values (mg/L Zn) which were less than the specified reference value (0.48 mg/L Zn) derived using the data submitted for this test in Rounds 1 - 3 of this scheme.

The results in Round 4 were very similar to those generated in Round 3 for the *Tisbe* test in which a single participant also failed the accuracy performance criteria for DTAPS Sample B.

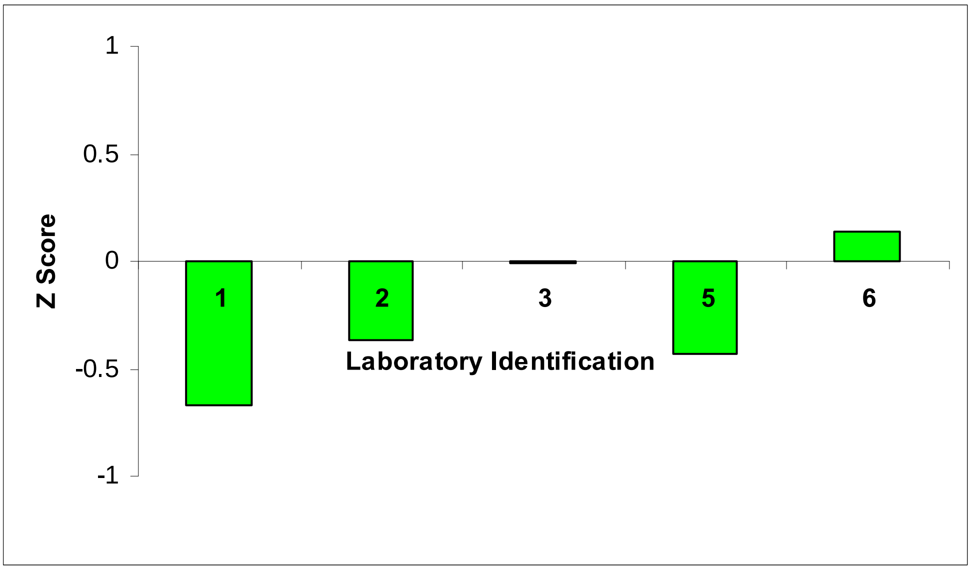
Intra-Laboratory Z-scores (Mean Internal Reference Standard Result)

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 4 (2008/09) of the DTAPS.

Laboratory ID	Mean Internal EC50 (mg/L Zn)	Z-score	n
1	0.33	-0.67	5
2	0.39	-0.37	7
3	0.48	-0.01	6
5	0.38	-0.43	7
6	0.51	0.14	6
7	No results submitted	NA	0
9	No results submitted	NA	0
Consensus			
Mean	0.42	-0.27	
Range	0.33-0.51	-0.67-0.14	
Std Dev.	0.075	NA	
% RSD (% CV)	17.9	NA	

Figure 7. Intra-Laboratory *Tisbe* Z-scores for Round 4 (2008/09) 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=>6) in Round 4 of the scheme complied with the standard deviation target applied around the reference value of 0.48 mg/L zinc. Laboratories No. 7 and 9 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 of 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 four 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)

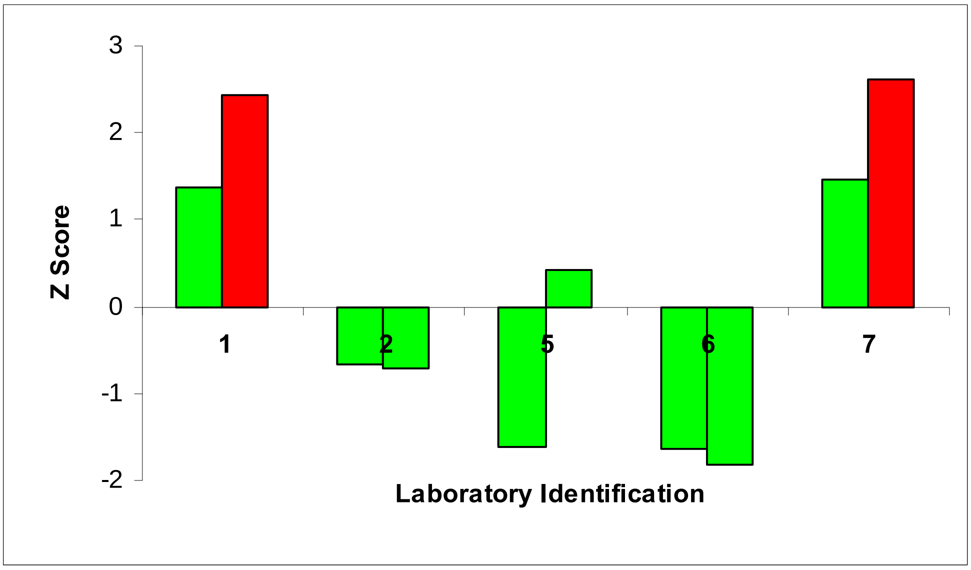
The results of the Round 4 accuracy (Z-score) assessments for the OEL test are presented in Table 9 and Figures 8 below.

Table 9. OEL EC50 Values and Z-scores for Round 4 (2008/09) 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	0.22	1.37	0.29	2.44
2	0.1	-0.67	0.098	-0.70
5	0.044	-1.61	0.17	0.43
6	0.042	-1.63	0.03	-1.83
7	0.23	1.47	0.3	2.61
Consensus				
Mean	0.13	-0.21	0.18	0.59
Range	0.042-0.23	-1.63-1.47	0.03-0.3	-1.83-2.61
Standard Deviation	0.092	NA	0.12	NA
% RSD (% CV)	72.6	NA	66.5	NA

As reported by participant.

Figure 8. OEL Z-scores for Round 4 (2008/09) 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 4 of the DTAPS showed an improvement in laboratory performance (in terms of accuracy) for the OEL test compared with the results obtained in Round 3, and appears to suggest that the decline in performance seen in the Round 3 results (as compared to previous Rounds) has not worsened. In Round 3, three non-compliant results were generated from five participants, with one laboratory submitting a DTAPS B result which was almost 3 times the specified reference value (Z-Score = 4.53). Round 4 again attracted five participants and while two non-compliant results (both for DTAPS B) were generated, both were marginal (Z-score = 2-3).

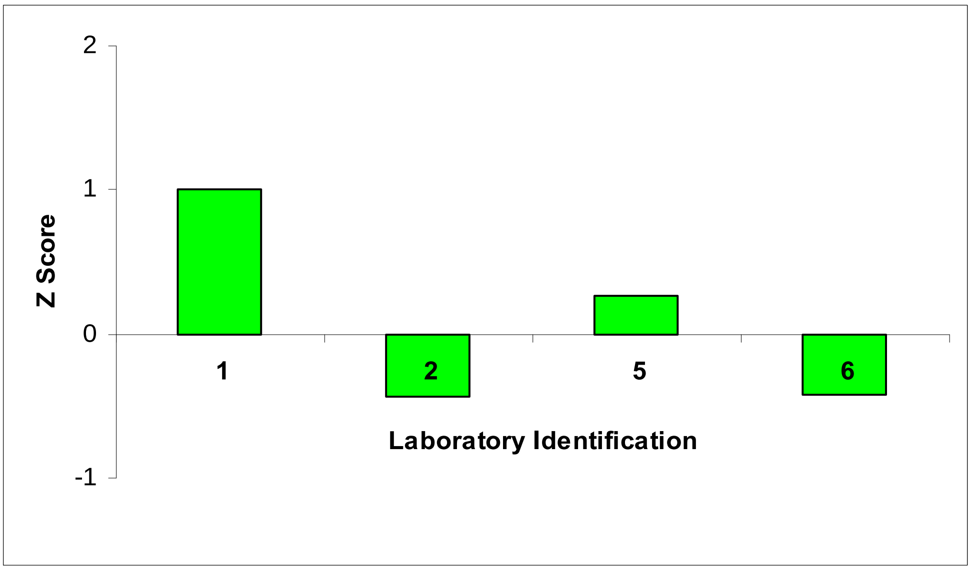
Intra-Laboratory Z scores (Mean Internal Reference Standard Result)

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 4 (2008/09) of the DTAPS.

Lab Identification	Mean Internal EC50 (mg/L Zn)	Z-score	n
1	0.20	1.01	6
2	0.11	-0.44	6
5	0.16	0.26	7
6	0.11	-0.42	6
7	No results submitted	NA	0
Consensus			
Mean	0.15	0.10	
Range	0.11-0.20	-0.44-1.01	
Std Dev.	0.044	NA	
% RSD (% CV)	30.1	NA	

Figure 9. Intra-Laboratory OEL Z-scores for Round 4 (2008/09) of the DTAPS.



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

All of the participants submitting internal reference data in Round 4 of the scheme for the OEL test complied with the standard deviation target applied around the reference value of 0.14 mg/L zinc. Laboratory No. 7 did not submit any internal reference testing data. Two laboratories generated mean EC50 values which were less than the reference value (Z-score = -0.44 to -0.42), while the other two generated mean EC50 values above the reference value (Z-score = 0.26 to 1.01).

The results of the Round 4 internal assessment suggest a considerable improvement on previous Rounds and the continued worsening of the internal accuracy of participants seen over Rounds 2 and 3 of the scheme (40% and 50% participant non-compliance respectively), has not been substantiated in this Round. Levels of compliance have returned to those seen in the first Round of the scheme (100%) although there were 50% more participants in this element of the scheme in Round 1 than in this current Round. It appears that those laboratories which had previously exhibited problems in achieving the performance criteria applied for this test have successfully addressed any deficiencies in performance for this Round (or have not participated).

9.2.5 Marine Algae inhibition of growth test

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

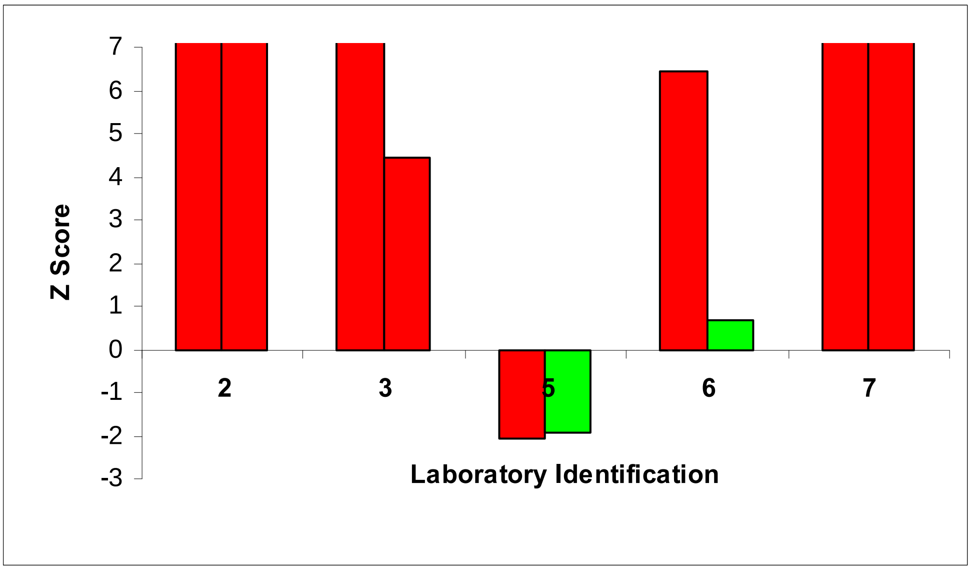
The results of the Round 4 accuracy (Z-score) assessments for the OEL test are presented in Table 11 and Figure 10 below.

Table 11. Marine Algae EC50 Values and Z-scores for Round 4 (2008/09) 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	>1.65	>51.7	>1.66	>52.0
3	2.18	69.2	0.23	4.46
5	0.038	-2.06	0.042	-1.93
6	0.29	6.42	0.12	0.69
7	0.75	21.7	6.19	203
9	No result submitted	NA	No result submitted	NA
Consensus				
Mean	0.98	29.4	1.65	51.6
Range	0.038-2.18	-2.06-69.2	0.042-6.19	-1.93-203
Standard Deviation	0.91	NA	0.12	NA
% RSD (% CV)	92.6	NA	66.5	NA

** As reported by participant. 'Greater than' values treated as calculated values for Consensus statistic calculations.*

Figure 10. Marine Algae Z-scores for Round 4 (2008/09) 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 in Round 3, despite the application of a wider compliant Z-score range (-3 to 3) than the other DTA test methods (-2 to 2). In fact, two sub-sets of laboratories appeared to be operating, the first generally achieving the accuracy performance criteria and the second operating at a level of zinc toxicity to marine algae which was much lower than expected, based on the performance criteria applied in the DTAPS. The use of test media containing high levels of EDTA which can chelate metals (such as zinc) in solution and prevent their bioavailability to the algae under exposure was implicated as a likely candidate for introducing systematic bias in the results of those laboratories generating EC50 values to zinc for marine algae which were significantly greater than the applied reference value of 0.1 mg/L Zn. The inclusion of the results from those laboratories generating significantly higher EC50 values, and implicated as using an incorrect test media for DTA testing, in the process of updating the reference value and proficiency testing target criteria for use in Round 4, would have undoubtedly increased the reference value to a value somewhere between the two groups of laboratories, and one likely to be unachievable for all, based on previous performance. It was therefore decided to retain the reference value and target criteria from Round 3 in order to reflect the level of performance expected using the correct media.

Despite the introduction of a Standing Committee of Analysts (SCA) guideline for conducting the marine algae inhibition of growth test which specifies the use of media containing a reduced concentration of EDTA, and is prescribed for use within this scheme, it appears that a number of laboratories have continued to use a media containing a higher than prescribed concentration of EDTA, and have hence generated EC50 values that are significantly higher than the applied reference value of 0.1 mg/L Zn.

No laboratories achieved compliant results for DTAPS Sample A, although Laboratory No. 5 only very marginally exceeded the standard deviation target around the reference value (Z-score = 2.03). Laboratories No. 5 and 6 achieved compliant Z-scores for DTAPS Sample B, and these represented the only compliant results returned by participants for this element of the assessment in Round 4 of the scheme.

The results appear to suggest that Laboratories 5 and 6 applied the correct (low EDTA) media in their tests (although Laboratory 6 still achieved a significantly higher than expected non-compliant value for DTAPS Sample A), while the remaining participants appear to have persisted with a media inferring lower toxicity of the test items to the exposed algae.

Further attempts should be made to ensure that laboratories participating in the next Round of the DTAPS apply the prescribed methodology (Ref 20), and in particular use a test medium containing a lowered concentration of EDTA for use with this test. Failure to fully standardise the procedure for this test are likely to render further assessments of laboratory accuracy (as compared to prescribed performance values) inconclusive and make inaccurate performance among participants difficult to resolve amongst the inter-laboratory variability caused by the use of differing test media.

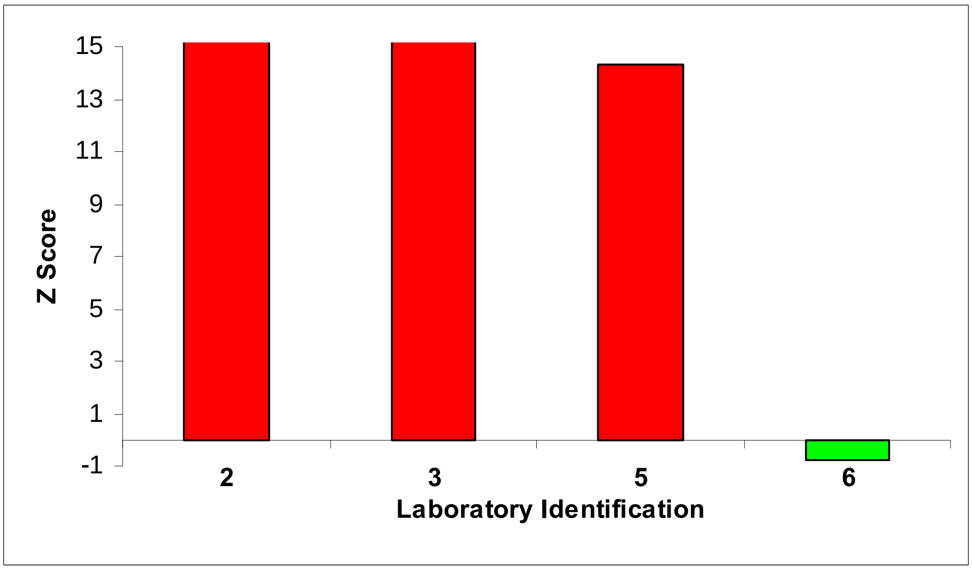
Intra-Laboratory Z scores (Mean Internal Reference Standard Result)

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 4 (2008/09) of the DTAPS.

Lab ID	Mean Internal EC50 (mg/L Zn)	Z-score	n
2	1.66	52.1	10
3	2.16	68.7	6
5	0.53	14.3	7
6	0.08	-0.75	6
7	No results submitted	NA	0
9	No results submitted	NA	0
Consensus			
Mean	1.11	33.6	
Range	0.08-2.16	-0.75-68.7	
Std Dev.	0.97	NA	
% RSD (% CV)	87.3	NA	

Figure 11. Intra-Laboratory Marine Algae Z-scores for Round 4 (2008/09) of the DTAPS.



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

Four of six participants submitted a full internal data set (n=>6) for the Marine algal test in Round 4 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. 2, 3 and 5) failed to comply with the accuracy target for internal performance, while Laboratory No 6 performed well within the applied standard deviation target around the reference value. All those failing to comply with the target Z-score range generated mean internal EC50 values which were significantly greater than the applied reference value of 0.1 mg/L Zn.

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. Laboratory No. 6 performed optimally overall, managing to obtain compliant Z-scores for DTAPS Sample B and mean internal EC50. Laboratory No. 5 also achieved a compliant result for DTAPS Sample B, but failed to generate results for DTAP Sample A or internal mean EC50 which were within the specified range, although the degree of non-compliance was considerably less (especially for DTAPS Sample A) than the other non-compliant participants.

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.

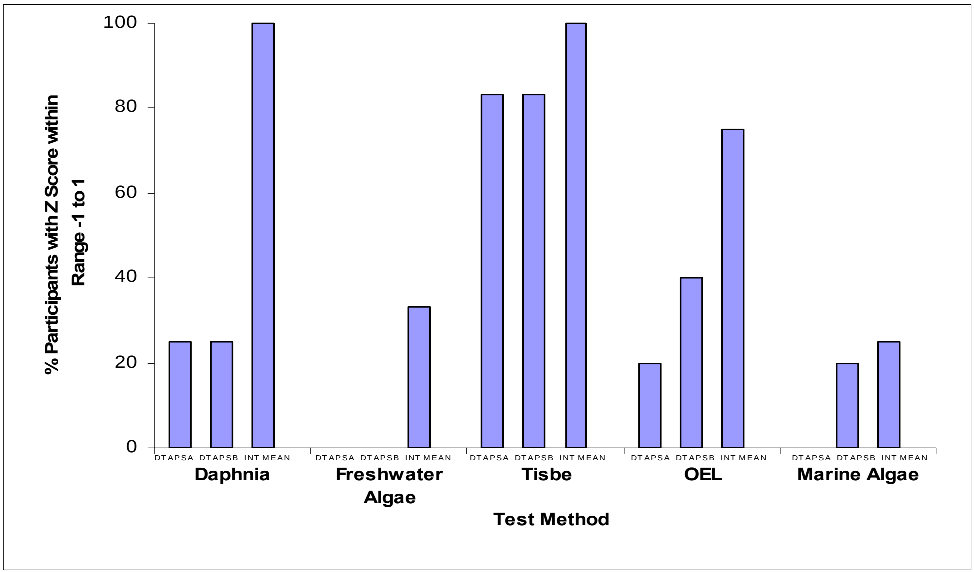
The reference values and error targets applied for the invertebrate test methods appear to have been achievable for most participating laboratories for both the proficiency samples and internal reference testing results. The results for the algal tests were much less encouraging, however, and while there remain some issues to be resolved by some laboratories with regard to test medium used for the marine algae inhibition of growth test, the apparent decline in performance (accuracy) with the freshwater algae inhibition of growth test may be a greater cause for concern, and should be closely monitored in future Rounds of the DTAPS.

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 4 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			
Daphnia	DTAPS A	1.84	0.59	32.3	4	1	3	0	0	0.6	2.57	4.68
	DTAPS B	1.69	1.15	68.4	4	1	0	3	0	0.31	4.68	
	Internal Laboratory Mean	1.69	0.13	7.86	4	4	0	0	0	0.31	0.77	
Freshwater Algae	DTAPS A	0.059	0.024	41.1	3	0	0	1	2	-3.04	5.95	26.1
	DTAPS B	0.06	0.024	39.7	3	0	0	1	2	-2.3	5.7	
	Internal Laboratory Mean	0.48	0.26	54.2	3	1	0	0	2	-0.32	26.1	
Tisbe	DTAPS A	0.36	0.077	21.5	6	5	1	0	0	-0.52	1.25	102
	DTAPS B	4.39	9.51	217	6	5	0	0	1	17.1	102	
	Internal Laboratory Mean	0.42	0.075	17.9	5	5	0	0	0	-0.27	0.81	
OEL	DTAPS A	0.13	0.092	72.6	5	1	4	0	0	-0.21	3.1	4.44
	DTAPS B	0.18	0.12	66.5	5	2	1	2	0	0.59	4.44	
	Internal Laboratory Mean	0.15	0.044	30.1	4	3	1	0	0	0.1	1.45	
Marine Algae	DTAPS A	0.98	0.91	92.6	5	0	0	1	4	29.4	71.3	205
	DTAPS B	1.65	0.12	66.5	5	1	1	0	3	51.6	205	
	Internal Laboratory Mean	1.11	0.97	87.3	4	1	0	0	3	33.6	69.5	

Figure 12. Percentage of participants achieving optimal performance in Round 4 of the DTAPS.



The percentage of participants achieving optimal performance for each Z-score statistic (Z-score = -1 to 1), suggests that the levels of accuracy produced by the DTA methods in Round 4 of the DTAPS were highly variable (0-100%).

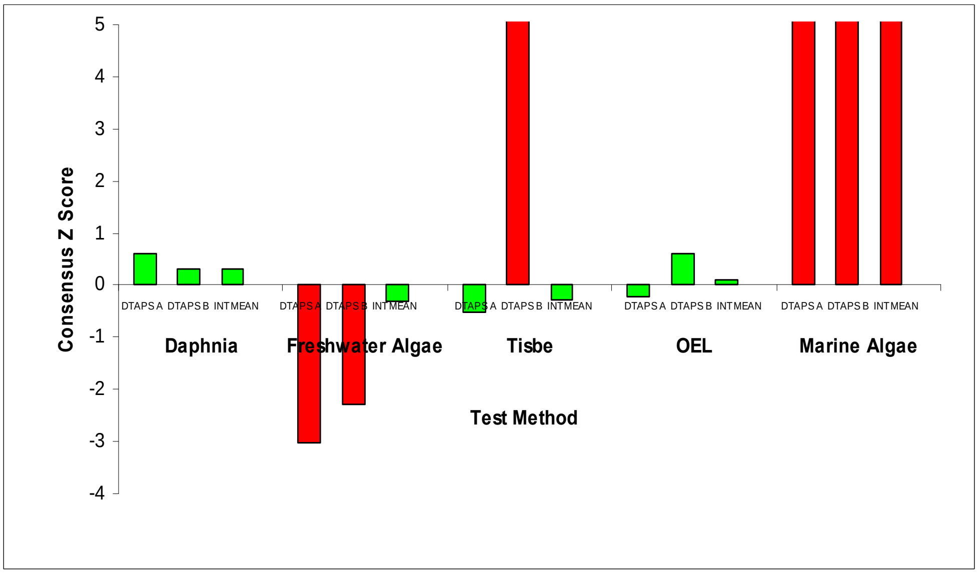
The *Daphnia* test displayed an identical level of accuracy for tests undertaken with DTAPS Samples A and B (25%) but the level of accuracy for these proficiency testing samples was significantly less than that obtained from internal reference testing (100% optimal). Overall, the results for the *Daphnia* test display a decline in accuracy (measured as % participants with optimal performance) for tests undertaken using the proficiency testing samples DTAPS A and B compared to Round 3, while a significant improvement has been observed in the accuracy achieved by repeated internal reference testing (from 65 to 100% optimal performance) which in Round 4 represented (along with the *Tisbe* test) the most accurate test under internal testing conditions.

Compared to Round 3, a significant decline in accuracy (measured as % participants with optimal performance) was observed for the Freshwater Algae test with no participants achieving optimal accuracy in tests with the proficiency testing samples. Similarly, no participants achieved optimal (Z-score = -1 to 1) accuracy in the Marine algae test using DTAPS Sample A. As also shown in the individual laboratory accuracy assessments, the algal inhibition of growth tests clearly represented the least accurate tests in Round 4 of the scheme. While this was also true of the Marine Algae test in the previous Round of the scheme, and is a likely symptom of continuing issues with the media used for testing, the decline in accuracy for the Freshwater Algae test may be a greater cause for concern if substantiated in subsequent Rounds of the scheme. It may, however, be that the reduction in accuracy as expressed as percentage optimal performance may simply be an artefact of the drop in participant numbers for this test and/or the significant change in reference value between Rounds 3 and 4.

The *Tisbe* test continues to improve in performance from Round to Round and now represents the most accurate across all testing conditions achieving >80% optimal performance for both proficiency testing samples and 100% for internal reference testing (along with the *Daphnia* test). The OEL test showed a decline in the level of accuracy (measured as % participants with optimal performance) for the proficiency testing samples DTAPS A and B compared to Round 3, while internal reference testing accuracy was considerably improved.

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 4 of the DTAPS.



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

Only the *Daphnia* and OEL methods 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 both displayed improved performance compared to Round 3 of the scheme (as measured by consensus Z-score).

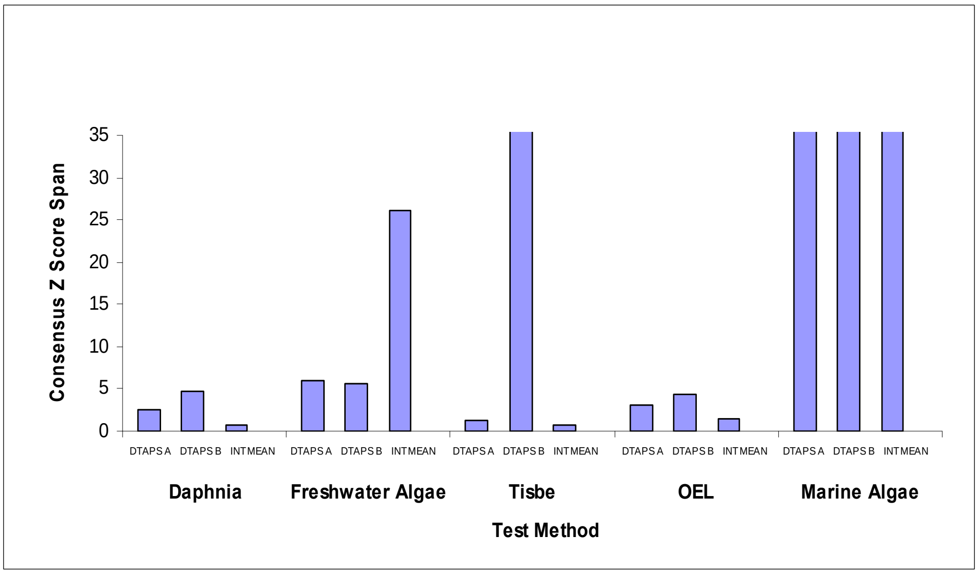
The *Tisbe* test showed similar accuracy for DTAPS Sample A and internal results compared to Round 3 but displayed a reduction in accuracy measured by consensus Z-Score for DTAPS Sample B, although this was heavily influenced by the very high EC50 value for this proficiency sample generated by Laboratory No.1.

As with Round 3 of the scheme, the algal methods appear from this assessment to be the least accurate tests. There appears to have been some improvement in accuracy for the Freshwater Algal test compared to Round 3 (as measured by consensus Z-Score) when both proficiency testing samples DTAPS A and B, and internal reference testing produced non-compliant consensus Z-scores. The consensus Z-score generated under internal reference testing conditions is optimal for Round 4, and although the consensus Z-scores for DTAPS A and B remain without the specified range, the extent of exceedance has lessened somewhat. As with the % optimal performance results, this could be an effect of low participant numbers (and therefore some cancelling out of extreme values above and below the reference) and increased reference value (it is noted that all three of the consensus Z-scores now represent consensus EC50 values which are less than the reference value whereas in Round 3 they were all above it).

The Marine Algae test continues to reflect the highly inaccurate performance also measured in Round 3. It is clear that there remains some development required to gain control over the accuracy of this test, both across and within laboratories. As described in Section 9.2.5 a systematic difference in the performance of the method between laboratories remains apparent (probably owing to the differences in test medium applied). Sub-optimal performance by individual laboratories cannot be resolved against such a background of extreme variability. Despite the publication of a standard guideline (Ref. 20) between Rounds 3 and 4 of the DTAPS, it appears that this has not been adopted and/or incorporated into internal procedures at all laboratories undertaking DTA assessments. It is vital that these guidelines are applied across all laboratories undertaking DTA testing in order to bring the overall accuracy of this test method under control, enabling the identification of individual laboratories which are not performing within specification. This is especially important since the Marine Algae test is (along with the *Tisbe* test) the most frequently applied method for regulatory DTA testing.

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 4 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 *Tisbe* test shows a significant reduction in accuracy for DTAPS B only, and the algal growth methods are by far the least accurate methods overall.

Such consensus assessments are only meaningful if the reference values applied in deriving Z-scores 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. 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 are available (i.e. the full DTAPS data set for previous Rounds). This may be particularly relevant for the Marine Algae test for which accuracy assessments suggest a significant lack of performance across participant laboratories. This may be caused by the application of a reference value which is specifically relevant to the methodology applied for DTA rather than other forms of ecotoxicological testing. It could be argued that the reference value and target criteria should be based on the level of performance achieved by the majority of laboratories regardless of doubts over the application of the prescribed methodology. The use of a 'low EDTA' test medium in Marine Algae tests is, however, specifically designed to address issues of potential metal toxicity in effluent samples, and it therefore seems sensible that any reference value employed in this scheme should reflect what is required 'in practice' in tests with real effluent samples. The suspected lack of application of the specified medium within the DTAPS suggests that they are also unlikely to be applied in the regulatory testing of DTA samples, and this could result in false negative results being obtained for effluent samples containing metals, and hence a lack of adequate protection for the receiving environment.

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 4 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)
Daphnia	No significant difference*	No significant difference
Freshwater Algae	No significant difference*	No significant difference
Tisbe	No significant difference**	No significant difference
OEL	No significant difference*	No significant difference
Marine Algae	No significant difference*	No significant difference

* Homoscedastic t test (equal variances)
** Heteroscedastic t test (unequal variances)

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 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

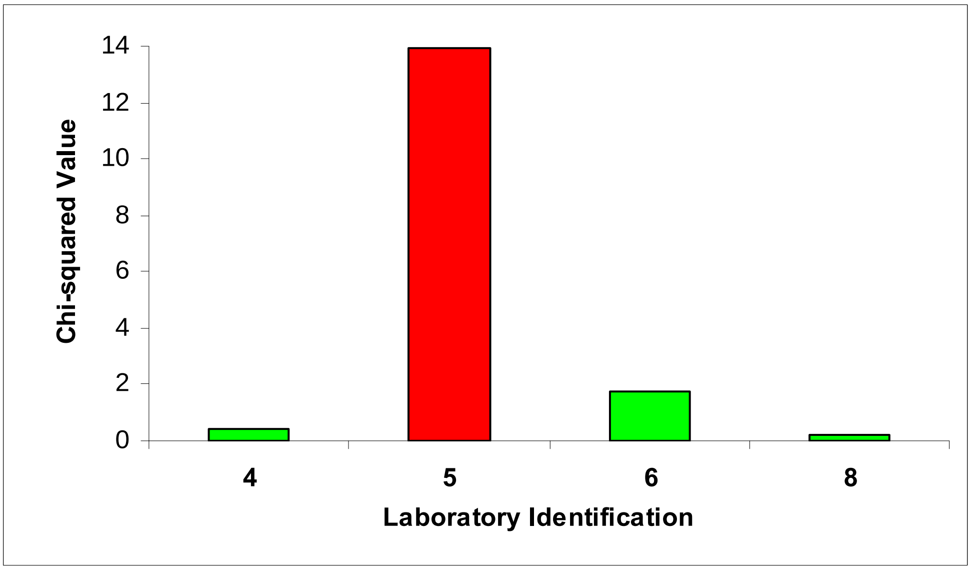
9.3.1 *Daphnia magna* 48 hour immobilisation test

The results of the Round 4 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 4 (2008/09) of the DTAPS.

Laboratory ID	X^2 value	n	Critical χ^2value (0.05)
4	0.45	6	11.07 (5 dof)
5	13.9	6	11.07 (5 dof)
6	1.72	6	11.07 (5 dof)
8	0.20	6	11.07 (5 dof)
9	No results submitted	0	NA

Figure 15. *Daphnia magna* immobilisation test – Chi-squared values for Round 4 (2008/09) of the DTAPS.



All of the participant laboratories in the previous Round of the DTAPS complied with the precision target (variance) applied for the *Daphnia* test and, while there was a general reduction in intra-laboratory precision achieved by laboratories for this method between Rounds 2 and 3, it appeared that all participants were able to adequately control the variability of their testing process, at least within the limits applied in the DTAPS.

Compliance with the applied precision target in Round 4 remained excellent for three of the four participants submitting data, however, Laboratory No. 5 generated a significant exceedance of the variance target which suggests a deterioration of control over the testing process within that facility.

It would be reasonably expected that the degree of precision within laboratories be improved over successive Rounds of the DTAPS as the DTA methods become more standardised and familiar within each laboratory. While this appears to be the case for those participants generating compliant variance statistics in this Round, where all achieved Chi-square values comparable to the optimally performing participants in Round 3, it only serves to further highlight the lack of precision exhibited by Laboratory No.5.

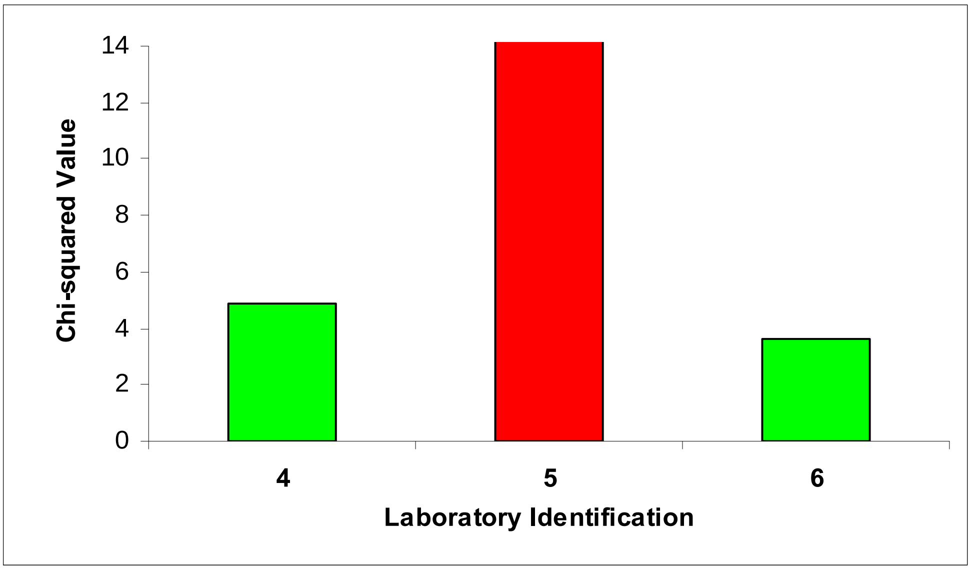
9.3.2 Freshwater Algae 72 hour inhibition of growth test

The results of the Round 4 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 4 (2008/09) of the DTAPS.

Laboratory ID	χ^2 value	n	Critical χ^2value (0.05)
4	4.87	7	12.59 (6 dof)
5	121.7	6	11.07 (5 dof)
6	3.60	6	11.07 (5 dof)
9	No results submitted	0	NA

Figure 16. Freshwater Algae inhibition of growth test - Chi-squared values for Round 4 (2008/09) 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. Of the three laboratories submitting internal reference testing data, two met the applied precision target and one exceeded the target by a significant margin. The laboratory failing to comply with the Chi-squared target (Laboratory No.5) also generated a non-compliant precision statistic for the *Daphnia* test, suggesting overall control over repeated DTA testing (at least for the freshwater methods) may be have been compromised at this facility. Overall, the degree of precision obtained within laboratories for the Freshwater Algae test was less than in Round 3 of the scheme when all eight laboratories submitting data conformed with the applied precision target.

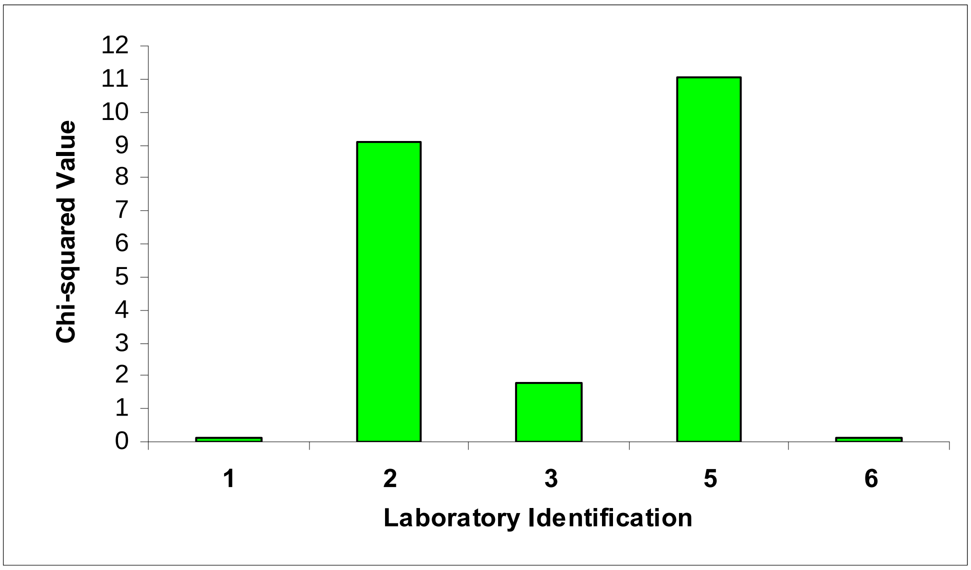
9.3.3 *Tisbe battagliai* 48 hour lethality test

The results of the Round 4 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 4 (2008/09) of the DTAPS.

Laboratory ID	χ^2 value	n	Critical χ^2value (0.05)
1	0.13	5	9.49 (4 dof)
2	9.06	7	12.59 (6 dof)
3	1.78	6	11.07 (5 dof)
5	11.03	7	12.59 (6 dof)
6	0.14	6	11.07 (6 dof)
7	No results submitted	0	NA
9	No results submitted	0	NA

Figure 17. *Tisbe battagliai* lethality test - Chi-squared values for Round 4 (2008/09) 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 remained good in Round 4 with no incidences of failure to meet the target precision thresholds for variance. Overall, the level of control over internal variability appears to have improved compared to Round 3, with three of the five laboratories generating a Chi-squared value of less than 2. Two laboratories (Nos. 2 and 5) generated precision variables which, while remaining within the limits specified in the scheme, were significantly higher than the other participants. Laboratory No. 5 demonstrated the least control over precision using the *Tisbe* test (Chi-squared=11.03 against a critical Chi-squared value of 12.59), which, when considered alongside their non-compliant precision results for the Freshwater DTA methods (see above) may signal a suggest a general issue with repeatability at this facility.

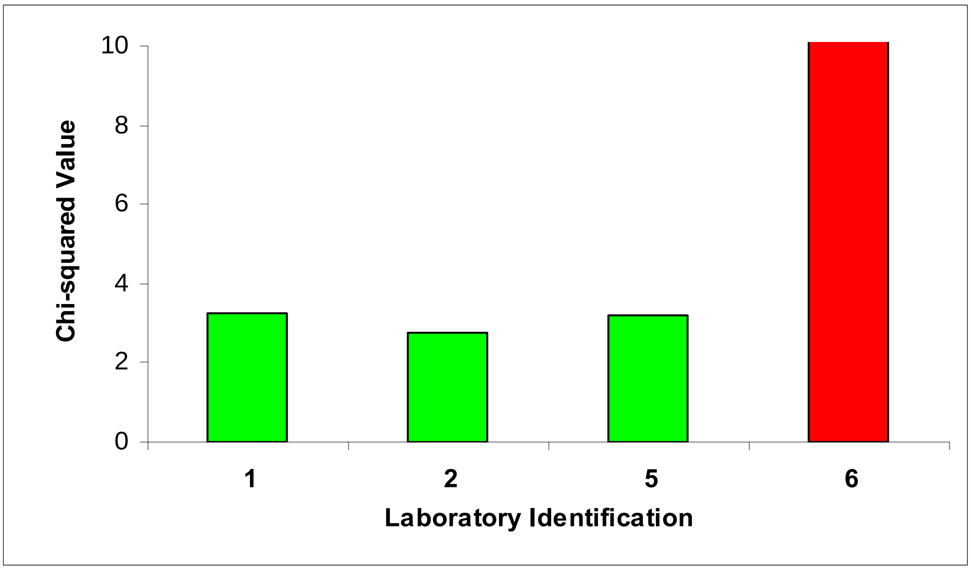
9.3.4 Oyster Embryo-Larval (OEL) 24 hour development test

The results of the Round 4 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 4 (2008/09) of the DTAPS.

Laboratory ID	χ^2 value	n	Critical χ^2value (0.05)
1	3.26	6	11.07 (5 dof)
2	2.77	6	11.07 (5 dof)
5	3.21	7	12.59 (6 dof)
6	90.7	6	11.07 (5 dof)
7	No results submitted	0	NA

Figure 18. OEL development test - Chi-squared values for Round 4 (2008/09) 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 much improved compared to Round 3 of the scheme when three of four participants failed to meet the specified variance target. Only a single laboratory (No.6) exceeded the applied precision target (Chi-squared comparison of variances). The other three participants all generated Chi-squared values which were well within the relevant critical values.

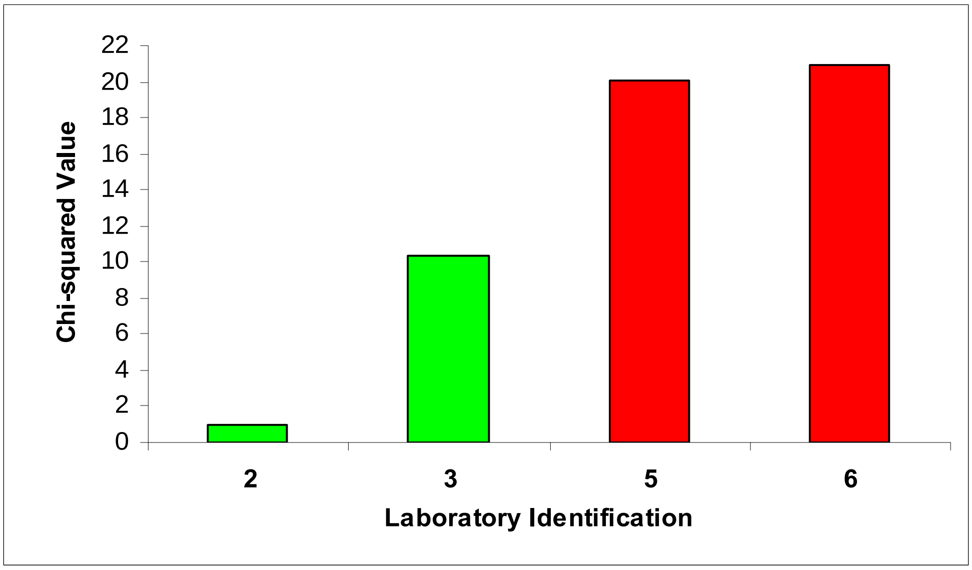
9.3.5 Marine Algae 72 hour inhibition of growth test

The results of the Round 4 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 4 (2008/09) of the DTAPS.

Laboratory ID	χ^2 value	n	Critical χ^2 value (0.05)
2	0.94	10	16.92 (9 dof)
3	10.4	6	11.07 (5 dof)
5	20.1	7	12.59 (6 dof)
6	20.9	6	11.07 (5 dof)
7	No results submitted	0	NA
9	No results submitted	0	NA

Figure 19. Marine Algae inhibition of growth test – Chi-squared values for Round 4 (2008/09) 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 4 of the scheme show that precision in the Marine Algae tests has generally worsened since Round 3 of the scheme. In the previous Round only one of six participants failed the applied precision target, while in this Round half of the participants have exceeded their critical Chi-squared values by some margin. Laboratory No.5 also exceeded the precision targets for the *Daphnia* and Freshwater algae tests, while Laboratory No.6 also failed to meet the targets applied for the Oyster Embryo-Larval Development test.

Given the issues that remain with the Marine Algae inhibition of growth test with respect to the optimal test medium to employ for DTA testing, it may be that the apparent lack of precision observed here is caused by undertaking repeat reference tests using media of differing composition.

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 by the comparison of a laboratory's variance with a prescribed target variance.

It would appear that the general level of precision obtained within laboratories for the DTA tests 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.

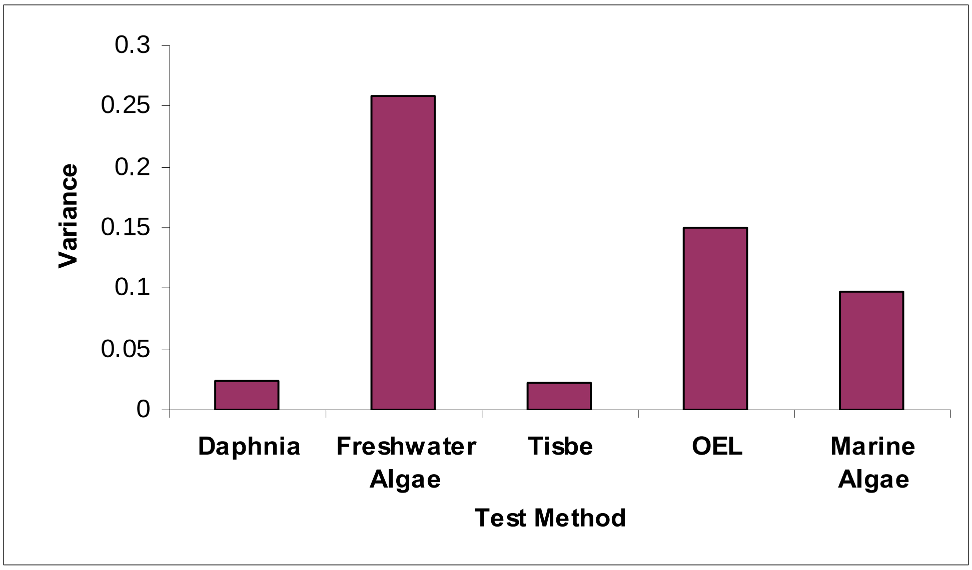
The variability of repeated internal laboratory results have been summarised (Table 20 and Figures 20-21) in order to provide an indication of the relative precision of each DTAPS method. These comparisons have been described using the performance statistics (variance) used within the DTAPS for laboratory 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 4 of the DTAPS.

Method	Consensus Mean EC50 (mg/L Zn)	Consensus Variance*	Consensus Standard Deviation**	Consensus % RSD***	Number of Laboratories
Daphnia	1.69	0.0243	0.437	25.9	4
Freshwater Algae	0.48	0.259	0.293	61	3
Tisbe	0.42	0.0226	0.101	24	5
OEL	0.15	0.15	0.0635	42.3	4
Marine Algae	1.11	0.098	0.481	43.3	4

** 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 4 of the DTAPS.

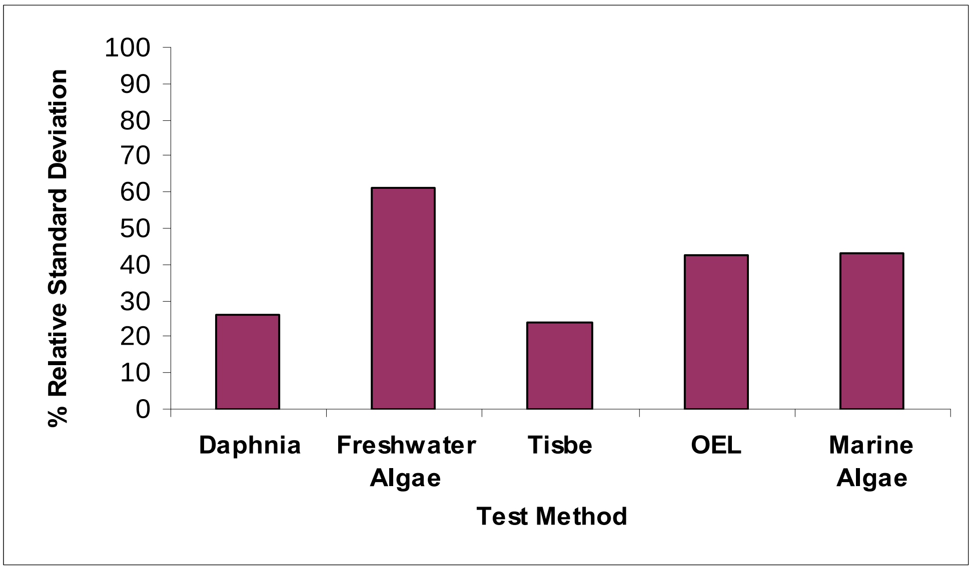


Log Base 10 EC50 values.

The consensus variances for two of the methods (*Daphnia* & *Tisbe*) were less than the target variance value applied for these methods (0.03) suggesting that the overall level of variability was under good control for these tests. The level of precision for these methods appears roughly stable with no significant degradation or improvement in precision between Rounds 3 and 4.

The consensus variances for the other three methods are all in excess of their applied target variances. The Marine Algae test generated a consensus variance value which was more than double the applied target variance for this method (0.04), while the OEL and Freshwater Algae tests generated variances which were 5 and almost 10 times the target value, respectively. While the consensus variance for the OEL test is in fact a very slight improvement on the Round 3 value, the results for the algal growth tests represent a significant degradation in overall performance since Round 3. However, the results are likely to be heavily influenced by a combination of the significant fall in participant numbers for these tests and the very high variance demonstrated by one or two laboratories in each group. The ongoing issues regarding the optimal test medium composition to be used in the Marine Algal test are also likely to have a significant bearing on the overall resulting variability for this test method.

Figure 21. Consensus Intra-Laboratory % Relative Standard Deviation (%RSD) for Test Methods in Round 4 of the DTAPS.



In general, a comparison of the relative standard deviations (%RSD) of internal EC50 values supports the results of the variance assessment.

The Freshwater Algae method is again highlighted as having the lowest level of overall precision in Round 4 of the scheme and shows a significant degradation in precision between Rounds 3 and 4.

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.

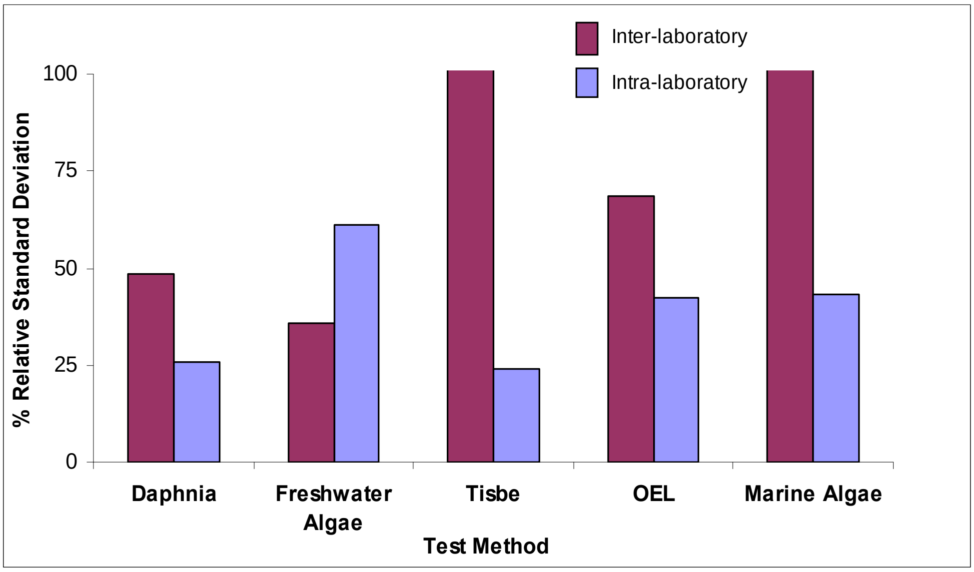
A comparison of the intra and inter-laboratory variability for each method is provided in Figure 22.

Table 21. Summary of Inter-Laboratory Test Method Consensus Precision Performance Statistics for Round 4 of the DTAPS.

Method	Consensus Mean (mg/L Zn) (DTAPS A & B EC50 values)	Consensus Standard Deviation (DTAPS A & B EC50 values)*	Consensus %RSD (DTAPS A & B EC50 values)**	n
Daphnia	1.76	0.852	48.4	8
Freshwater Algae	0.06	0.0215	35.8	6
Tisbe	2.38	6.78	285	12
OEL	0.15	0.103	68.7	10
Marine Algae	1.32	1.88	142	10

** 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 for Round 4 of the DTAPS.



Comparison of intra and inter-laboratory precision (using relative standard deviation) clearly highlights the expected increased level of process control over repeated tests within laboratories as compared to between laboratories, for all but the Freshwater Algae test. The Freshwater Algae intra-laboratory result is heavily influenced by a single laboratory's poor control over repeated internal testing (Laboratory No.5) combined with overall low participant numbers (n=3), and so should be taken as a further demonstration of the deterioration of precision in a single facility rather than an issue with the reproducibility of the overall testing process.

The greatest difference between intra and inter laboratory precision is seen in the *Tisbe* test, although again a single very high result for DTAPS B (Laboratory No.1) has a significant influence and an assessment excluding this result would reduce the inter-laboratory RSD to around 26%. This would suggest that (with a notable exception) the level of external variability achievable for the *Tisbe* test is close to that obtained for repeat tests within laboratories.

10. Update of DTAPS Performance Criteria

The performance criteria (reference and target values) applied in each Round of the DTAPS 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 4 of the scheme were derived from all of the data submitted in Rounds 1, 2 and 3. Updated values for use in Round 5 (09/10) of the scheme have been derived as outlined below.

The new (Round 4) 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 4) 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 (existing and/or new) was tested for normality using an Anderson-Darling 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 4) 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, and the existing performance criteria (derived from Rounds 1, 2 and 3, and applied in Round 4) were then retained.

Table 22. New DTAPS Test Method Data Sets (Round 4) and Numbers of Outlier Values.

Test Method	*n* (Round 4 Data Set)	% Outliers	Resultant *n* (Full Valid Data Set)	DTAPS Rounds contributing to the Full Valid Data Set
Daphnia	32	6	209	1-4
Freshwater Algae	25	32	61	3
Tisbe	43	7	104	3-4
OEL	35	17	85	1,2 & 4
Marine Algae	39	13	77	3-4

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 will 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) (Previous Reference Value)	Standard Deviation Target (mg/L Zn) (Previous SD Target)	Compliant Z Score Range	Variance Target* (Previous Variance Target)
Daphnia	1.57 (1.53)	0.54 (0.51)	-2 to 2	0.03 (0.03)
Freshwater Algae	0.16 (0.16)	0.19 (0.03)		0.12 (0.03)
Tisbe	0.47 (0.48)	0.24 (0.23)		0.04 (0.03)
OEL	0.15 (0.14)	0.06 (0.06)		0.03 (0.03)
Marine Algae	1.09 (0.1)	1.10 (0.03)		0.34 (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 5 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 4 be retained for the Algae and *Tisbe* tests.

The continued increase in the reference value, and widening of the accuracy and precision targets, for the Marine Algae test from Rounds 1 to 4 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). Investigative work undertaken following Round 3 of the scheme suggested that this systematic difference between laboratories was likely to be caused by the use of differing test media, particularly with regard to the concentration of EDTA available to chelate dissolved metal salts within the test system. Since the toxicant used within the DTAPS is a metal salt (zinc sulphate), the use of a media containing a high concentration of EDTA will chelate (and make unavailable to the growing algae) a higher proportion of the zinc toxicant than a media containing less EDTA, and hence the resultant EC50 will be increased to a degree that relates directly to the amount of chelation potential inherent in the media.

Prior to the commencement of Round 4 of the scheme, a new Standing Committee of Analysts guideline for the Marine Algal Inhibition of Growth test was published (for use in DTA assessments and the DTAPS) and it was hoped that the specification in this methodology of a ‘low EDTA’ test (and culture) medium would result in all participants in Round 4 of the scheme exhibiting results that were nearer the reference value (0.1 mg/L Zn) applied for the Marine Algal test. Following the completion of Round 4, a further investigation was carried out, taking into account the Round 4 submissions. An examination of the marine algal data included in this report shows that, despite the prescription of specific methodology, there remain some laboratories which appear to have continued to apply a media containing a higher than specified concentration of EDTA (e.g. Laboratories No. 2 & No. 3), and at least one participant (Laboratory No. 5) appears to have applied a ‘low’ EDTA media for tests with the proficiency testing samples (DTAPS A & B) but a ‘high’ EDTA media for internal reference testing. Furthermore, consultation with participants has suggested that some have experienced difficulties in culturing *Skeletonema*, and have experienced increased incidences of test failures (e.g. no or very low control growth) using the ‘low EDTA’ medium specified in the SCA guideline. Since zinc is added to the specified medium as an essential nutrient but has not been reduced in line with the reduction in EDTA, it is suspected that the reduced concentration of EDTA has resulted in a greater amount of ‘free’ zinc to be available to the growing algae in culture/ controls and therefore zinc toxicity may be occurring even before additional zinc is added in the form of the reference toxicant. This is supported by one or two EC50 results generated in DTAPS Round 4 which were well *below* the specified reference value of 0.1 mg/L Zn (e.g. Laboratory No.5). Clearly, there remains some additional work to be undertaken to ensure that the test medium applied in this method is balanced, both in terms of chelating potential and zinc concentration.

11. 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.

20. Environment Agency (2009) – Standing Committee of Analysts (SCA), The Direct Toxicity Assessment of Aqueous Environmental Samples using the *Skeletonema costatum M*arine Algal Growth Inhibition Test, Blue Book No. 225.

Appendix 1
DTAPS Participants - Compliance with Performance Criteria for Round 4 of the scheme

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

Appendix 1 (continued)
DTAPS Participants - Compliance with Performance Criteria for Round 4 of the scheme.

Laboratory Identifications (Round 3)	Method	DTAPS A	DTAPS B	Internal Mean	Variance
8	***Daphnia***	Y	Y	Y	Y
	Freshwater Algae	NP	NP	NP	NP
	Tisbe	NP	NP	NP	NP
	OEL	NP	NP	NP	NP
	Marine Algae	NP	NP	NP	NP
9	***Daphnia***	NP	NP	NP	NP
	Freshwater Algae	NP	NP	NP	NP
	Tisbe	NP	NP	NP	NP
	OEL	NP	NP	NP	NP
	Marine Algae	NP	NP	NP	NP

Y = 'Compliant' Result
N = 'Non-compliant' Result
NP = Either did not participate or did not submit required data