Results of Proficiency Test
Bisphenol A EN71-10/11
October 2020

Organized by: Institute for Interlaboratory Studies
Spijkenisse, the Netherlands

Author: ing. C.M. Nijssen-Wester
Correctors: ing. A.S. Noordman-de Neef & ing. R.J. Starink

Report: iis20V06 == Revised ==

January 2021

SUMMARY OF CHANGES

This revised report replaces the original report iis20V06 of January 2021.

It was discovered that the test result of laboratory 2108 was not corrected for the fact that the surface was taken as two-sided. This means that in order to compare this test result with the group (one-sided surface) this test result should have been multiplied by two. Since one test result on a small group of results can be of influence, iis decided to perform the statistical evaluation with the revised test result. Therefore, the result table and graphs as well as performance z-scores have been revised. Although it had a major influence on the z-score for laboratory 2108, it did not have a large effect on the z-scores of all other participants.

The following pages in this report have been revised:

- Paragraph 4.1: added text on page 9 (page 8 in the original report)
- Paragraph 4.2: revised table 4 on page 9 (page 8 in the original report)
- Paragraph 4.3: revised table 5 on page 10 (page 9 in the original report)
- Appendix 1: revised one test result and statistical evaluation including z-scores and graphs on page 12 (page 11 in the original report)
- Appendix 2: revised analytical details for lab 2108 on page 13 (page 12 in the original report)

CONTENTS

1 INTRODUCTION

Toy safety is the practice of ensuring that toys, especially those made for children, are safe, usually through the application of safety standards. In many countries toys must be able to pass safety tests in order to be sold. Many regions model their safety standards on the EU's EN71 standard. In Europe toys must meet the criteria set by the 2009 EC Toy Safety Directive 2009/48/EC, last updated on 18th of November 2019.

Migration of Bisphenol A (BPA) is described in EN 71-9 (Requirements), EN 71-10 (Sample preparation and extraction) and EN 71-11 (Methods of Analysis). The maximum specific limit as described in EN 71-9 is 0.1 mg/L aqueous substrate (or simulant). The European Union has further restricted this limit when it comes to toys. EU directive 2017/898 of 24 May 2017 amending Appendix C to Annex II to Directive 2009/48/EC as regards BPA describes a maximum specific migration limit of 0.04 mg/L aqueous substrate (or simulant). This has been implemented from November 26, 2018 in its member states.

Since 2017 the Institute for Interlaboratory Studies (iis) organizes a proficiency scheme for the determination of migratable Bisphenol A by EN71-10/11 every year. During the annual proficiency testing program 2020/2021 it was decided to continue the proficiency test for the determination of migratable Bisphenol A by EN71-10/11.
In this interlaboratory study 24 laboratories in 15 different countries registered for participation. See appendix 3 for the number of participants per country. In this report the test results of this proficiency test are presented and discussed. This report is also electronically available through the iis website www.iisnl.com.

2 SET UP

The Institute for Interlaboratory Studies (iis) in Spijkenisse, the Netherlands, was the organizer of this proficiency test (PT). Sample analyzes for fit-for-use and homogeneity testing were subcontracted to an ISO/IEC17025 accredited laboratory. It was decided to send one strip of white thermal paper labelled #20695 positive on Bisphenol A.
Furthermore, a number of test conditions (sample size, simulant, exposure temperature, exposure time and rotation speed) were prescribed. The participants were requested to report rounded and unrounded test results. The unrounded test results were preferably used for statistical evaluation.

2.1 QUALITY SYSTEM

The Institute for Interlaboratory Studies in Spijkenisse, the Netherlands, has implemented a quality system based on ISO/IEC17043:2010. This ensures strict adherence to protocols for sample preparation and statistical evaluation and 100% confidentiality of participant’s data. Feedback from the participants on the reported data is encouraged and customer’s satisfaction is measured on a regular basis by sending out questionnaires.

2.2 PROTOCOL

The protocol followed in the organization of this proficiency test was the one as described for proficiency testing in the report 'iis Interlaboratory Studies: Protocol for the Organisation, Statistics and Evaluation' of June 2018 (iis-protocol, version 3.5). This protocol is electronically available through the iis website www.iisnl.com, from the FAQ page.

2.3 CONFIDENTIALITY STATEMENT

All data presented in this report must be regarded as confidential and for use by the participating companies only. Disclosure of the information in this report is only allowed by means of the entire report. Use of the contents of this report for third parties is only allowed by written permission of the Institute for Interlaboratory Studies. Disclosure of the identity of one or more of the participating companies will be done only after receipt of a written agreement of the companies involved.

2.4 SAMPLES

A batch of white thermal paper positive on BPA was selected. From this batch 50 paperstrips of approximately 5x6 cm were wrapped in Aluminum foil and labelled #20695.
The homogeneity of the subsamples was checked by the determination of Bisphenol A content by EN71-10 on 8 stratified stratified randomly selected subsamples.

	BPA in mg/L
Sample #20695-1	4.043
Sample #20695-2	3.693
Sample #20695-3	3.784
Sample #20695-4	3.801
Sample #20695-5	3.877
Sample #20695-6	3.727
Sample #20695-7	4.091
Sample #20695-8	4.118

Table 1: homogeneity test results of subsamples #20695

From the above test results the repeatability was calculated and compared with 0.3 times the reproducibility of the reference method in agreement with the procedure of ISO13528, Annex B2 in the next table.

	BPA in mg/L
r (observed)	0.474
reference method	Horwitz
0.3 * R (reference method)	0.426

Table 2: evaluation of the repeatability of subsamples #20695

The calculated repeatability was in agreement with 0.3 times the reproducibility of the reference method. Therefore, homogeneity of the subsamples was assumed.

To each of the participating laboratories one sample labelled #20695 was sent on September 23, 2020.

2.5 ANALYZES

The participants were requested to determine on sample #20695 Bisphenol A in aqueous simulant using the prescribed test conditions (see table 3).
It was also requested to report if the laboratory was accredited for this determination. It was advised to keep the thermal paper stored dark, dry and cool and packed until the start of the test. It was also advised not to touch the sample with bare hands.

Sample size	cut the sample at **width=5cm/length=6cm** (surface area is: 5x6=30 cm^2)*
Simulant	deionized water
Simulant volume	as per method used
Exposure temperature	20°C
Exposure time	1 hour
Rotation speed	60 r/min

Table 3: prescribed test conditions for sample #20695
*) see also paragraph 4.1

It was explicitly requested to treat the sample as if it was a routine sample and to report the test results using the indicated units on the report form and not to round the test results but report as much significant figures as possible. It was also requested not to report ‘less than’ test results, which are above the detection limit, because such test results cannot be used for meaningful statistical evaluations.

To get comparable test results, a detailed report form and a letter of instructions are prepared. On the report form the reporting units are given as well as the reference test methods (when applicable) that will be used during the evaluation. The detailed report form and the letter of instructions are both made available on the data entry portal www.kpmd.co.uk/sgs-iis-cts/. The participating laboratories are also requested to confirm the sample receipt on this data entry portal. The letter of instructions can also be downloaded from the iis website www.iisnl.com.

3 RESULTS

During five weeks after sample dispatch, the test results of the individual laboratories were gathered via the data entry portal www.kpmd.co.uk/sgs-iis-cts/. The reported test results are tabulated per determination in appendix 1 of this report. The laboratories are presented by their code numbers.

Directly after the deadline, a reminder was sent to those laboratories that had not reported test results at that moment. Shortly after the deadline, the available test results were screened for suspect data. A test result was called suspect in case the Huber Elimination Rule (a robust outlier test) found it to be an outlier. The laboratories that produced these

suspect data were asked to check the reported test results (no reanalysis). Additional or corrected test results are used for data analysis and the original reported test results placed under 'Remarks' in the result tables in appendix 1. Test results that came in after the deadline were not taken into account in this screening for suspect data and thus these participants were not requested for checks.

3.1 STATISTICS

The protocol followed in the organisation of this proficiency test was the one as described for proficiency testing in the report 'iis Interlaboratory Studies: Protocol for the Organisation, Statistics and Evaluation' of June 2018 (iis-protocol, version 3.5).
For the statistical evaluation the *unrounded* (when available) figures were used instead of the rounded test results. Test results reported as '<…' or '>…" were not used in the statistical evaluation.
First, the normality of the distribution of the various data sets per determination was checked by means of the Lilliefors-test, a variant of the Kolmogorov-Smirnov test and by the calculation of skewness and kurtosis. Evaluation of the three normality indicators in combination with the visual evaluation of the graphic Kernel density plot, lead to judgement of the normality being either 'unknown', 'OK', 'suspect' or 'not OK'. After removal of outliers, this check was repeated. If a data set does not have a normal distribution, the (results of the) statistical evaluation should be used with due care.

According to ISO5725 the original test results per determination were submitted to Dixon's, Grubbs' or Rosner's outlier tests. Outliers are marked by D(0.01) for the Dixon's test, by G(0.01) or DG(0.01) for the Grubbs' test and by R(0.01) for the Rosner's test. Stragglers are marked by D(0.05) for the Dixon's test, by G(0.05) or DG(0.05) for the Grubbs' test and by R(0.05) for the Rosner's test. Both outliers and stragglers were not included in the calculations of averages and standard deviations.

For each assigned value the uncertainty was determined in accordance with ISO13528. Subsequently the calculated uncertainty was evaluated against the respective requirement based on the target reproducibility in accordance with ISO13528. In this PT, the criterion of ISO13528, paragraph 9.2.1, was met for all evaluated tests, therefore, the uncertainty of all assigned values may be negligible and need not be included in the PT report.

Finally, the reproducibilities were calculated from the standard deviations by multiplying these with a factor of 2.8.

3.2 GRAPHICS

In order to visualize the data against the reproducibilities from literature, Gauss plots were made, using the sorted data for one determination (see appendix 1). On the Y-axis the reported test results are plotted. The corresponding laboratory numbers are on the X-axis. The straight horizontal line presents the consensus value (a trimmed mean). The four striped lines, parallel to the consensus value line, are the +3s, +2s, -2s and -3s target reproducibility limits of the selected reference test method. Outliers and other data, which were excluded from the calculations, are represented as a cross. Accepted data are represented as a triangle.

Furthermore, Kernel Density Graphs were made. The Kernel Density Graph is a method for producing a smooth density approximation to a set of data that avoids some problems associated with histograms. Also, a normal Gauss curve was projected over the Kernel Density Graph for reference.

3.3 Z-SCORES

To evaluate the performance of the participating laboratories the z-scores were calculated. As it was decided to evaluate the performance of the participants in this proficiency test (PT) against the literature requirements, the z-scores were calculated using a target standard deviation. This results in an evaluation independent of the variation in this interlaboratory study.

The target standard deviation was calculated from the literature reproducibility by division with 2.8. In case no literature reproducibility was available, other target values were used, like Horwitz or an estimated reproducibility based on former iis proficiency tests.

When a laboratory did use a test method with a reproducibility that is significantly different from the reproducibility of the reference test method used in this report, it is strongly advised to recalculate the z-score, while using the reproducibility of the actual test method used, this in order to evaluate whether the reported test results is fit-for-use.

The z-scores were calculated in according to:

$z_{(target)}$ = (test result - average of PT) / target standard deviation

The $z_{(target)}$ scores are listed in the test result tables in appendix 1.

Absolute values for z<2 are very common and absolute values for z>3 are very rare.
The usual interpretation of z-scores is as follows:

	|z|	< 1	good
1 <	|z|	< 2	satisfactory
2 <	|z|	< 3	questionable
3 <	|z|		unsatisfactory

4 EVALUATION

In this interlaboratory study some problems were encountered with the dispatch of the samples due to the COVID-19 pandemic. Therefore, the reporting time on the data entry portal was extended with one week. One participant reported test results after the PT was closed and two participants did not report any test results.
Finally, 22 participants reported 21 numerical test results. No outliers were observed. In proficiency studies, outlier percentages of 3% - 7.5% are quite normal.

The original data set proved to have a normal Gaussian distribution.

4.1 EVALUATION PER TEST

In this section the results are discussed per test. The test methods which were used by the various laboratories were taken into account for explaining the observed differences when possible and applicable. These test methods are also in the table in appendix 1 together with the original data. The abbreviation used in these tables are explained in appendix 4.

Test method EN71-11 does mention precision data, unfortunately only at a low level of 0.03 mg BPA/L aqueous migrate. Therefore, the calculated reproducibility was compared against the estimated reproducibility calculated from the Horwitz equation.

Test method EN 71-10 does not describe whether the sample should be used one-sided or two-sided. Therefore, some test conditions like sample size (width and length) and surface area were prescribed. However, it was also requested to report the sample size (width and length) and the surface area used for the migration. All test results were evaluated as one-sided exposure as the sample is very thin. All participants calculated the test result based on a single sided surface.

In the instructions for this PT iis prescribed to use a piece of 5 x 6 cm (or one-sided surface area of 30 cm^2). iis meant to prescribe 2 x 5 cm^2, one-sided surface area of 10 cm^2 which is prescribed by EN71-10. Next to 10 cm2 this test method also prescribes to use 100 mL simulant (surface to volume ratio of 0.1). As a result some laboratories used the conditions of EN71-10 10 cm^2 and 100 mL and some used iis prescribe condition of 30 cm^2, but also used 300 mL to keep the ratio the same. And some used both iis conditions and EN71-10: 30 cm^2 and 100 mL (resulting in a higher surface to volume ratio of 0.3). In order to compare the results of the latter group the reported results by these laboratories were divided by three. One laboratory treated the sample as two-sided and used 5 cm^2 and 100 mL (resulting in a lower surface to volume ratio of 0.05). This test result was multiplied by two.

Sample #20695

BPA (migratable): This determination may be problematic. No statistical outliers were observed. The calculated reproducibility is not in agreement with the estimated reproducibility calculated from the Horwitz equation.

4.2 PERFORMANCE EVALUATION FOR THE GROUP OF LABORATORIES

A comparison has been made between the reproducibility as declared by the estimated target reproducibility using the Horwitz equation and the reproducibility as found for the group of participating laboratories. The number of significant test results, the average, the calculated reproducibility (2.8 * standard deviation) and the estimated target reproducibility are presented in the next table.

Parameter	unit	n	average	2.8 * sd	R(target)
Bisphenol A (migratable)	mg/L	21	7.1	5.7	2.4

Table 4: reproducibility on sample #20695

Without further statistical calculations it could be concluded that for migration of BPA there is not a good compliance of the group of participating laboratories with the reference method.

4.3 COMPARISON OF THE PROFICIENCY TEST OF OCTOBER 2020 WITH PREVIOUS PTS

The performance of the determinations of the proficiency test was compared, expressed as relative standard deviation (RSD) of the PTs, see table below.

Parameter	October 2020	December 2019	December 2018	December 2017	R(target)
Bisphenol A (migratable)	29%	14%	34-39%	8.3%	12-13%

Table 5: development of uncertainties over the years

The uncertainty observed in this PT is larger than the uncertainty observed in the PT conducted in 2017 and 2019 and is comparable with the uncertainty of the PT conducted in 2018.

4.4 EVALUATION OF THE ANALYTICAL DETAILS

In this PT also some analytical details were asked. See appendix 2 for the reported details. The majority of the participants (thirteen of the twenty particpants reporting analytical details) is ISO/IEC17025 accredited for this test.
Not all participants used the same surface area and simulant volume, resulting in different surface to volume ratios. This is also explained in paragraph 4.1. All participants reported one-sided surface areas based on the dimensions of (part of) the sample used. The temperatures used were between 20 and 24.5°C, the rotation speed per minute used by all participants was 60 and the time used for the migration was for all 60 minutes.
The differences of the surface to volume ratios are the most notable. Where possible the test results have been corrected for these differences (see appendix 1).

5 DISCUSSION

In this proficiency test, the average of the homogeneity test results is not in line with the average (consensus value) from the PT results. There are several reasons for this.
Firstly, the goal of homogeneity testing is very different from the goal of the evaluation of the reported PT results. In order to prove the homogeneity of the PT samples, a test method is selected with a high precision (smallest variation). The accuracy (trueness) of the test method is less relevant.
Secondly, the homogeneity testing is done by one laboratory only. The test results of this (ISO/IEC17025 accredited) laboratory will have a bias (systematic deviation) depending on the test method used. The desire to detect small variations between the PT samples leads to the use of a sensitive test method with high precision, which may be a test method with significant bias.

Finally, each test result reported by the laboratories that participate in the PT will have a bias. However, some will have a positive bias and others a negative bias. These different biases compensate each other in the PT average (consensus value). Therefore, the PT consensus value may deviate from the average of the homogeneity test. At the same time the accuracy of the PT consensus value is more reliable than the accuracy of the average of the results of the homogeneity test.

6 CONCLUSION

All participants, except one, found sample #20695 to be positive on BPA (above the limit of EN71-9 (0.1 mg/L) and directive EU/2017/898 (0.04 mg/L).

It is to be expected that the variation of the migration test results in real life practise will be larger than observed in this PT as the test conditions like sample size, simulant, exposure temperature, exposure time and rotation speed will not be prescribed but will be selected by the individual laboratories.

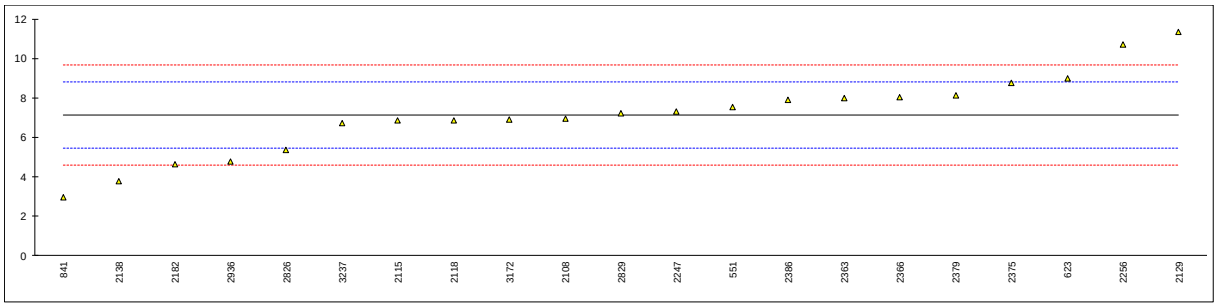
Each laboratory has to evaluate its performance in this study and make decisions about necessary corrective actions. Therefore, participation on a regular basis in this scheme could be helpful to improve the performance and the quality of the analytical results.

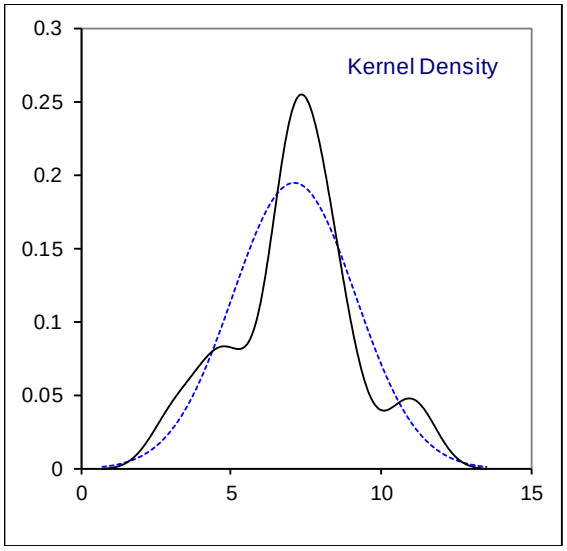
APPENDIX 1

Determination of BPA in aqueous migrate on sample #20695; (1-sided surface) results in mg/L

lab	method	Reported	Corr. In bold	mark	z(targ)	remarks
339		-----	-----		-----	
551	EN71-11	7.538	7.538		0.49	
623	EN71-11	27	**9**		2.21	used surface to volume ratio of 0.3 cm^2/mL
841	EN71-11	8.81	**2.937**		-4.94	used surface to volume ratio of 0.3 cm^2/mL
2108	EN71-11	3.4628	**6.9256**		-0.24	used surface to volume ratio of 0.05 cm^2/mL
2115	EN71-11	6.84	6.84		-0.34	
2118	EN71-11	6.86	6.86		-0.31	
2129	EN71-11	34.0	**11.33**		4.96	used surface to volume ratio of 0.3 cm^2/mL
2138	EN71-11	3.78	3.78		-3.94	
2182	EN71-11	4.611	4.611		-2.96	
2247		21.88	**7.293**		0.20	used surface to volume ratio of 0.3 cm^2/mL
2256	EN71-11	10.72	10.72		4.24	
2363	EN71-11	8.0	8.0		1.03	
2366	EN71-11	8.03	8.03		1.07	
2375	EN71-11	8.75	8.75		1.91	
2379	EN71-11	24.360	**8.120**		1.17	used surface to volume ratio of 0.3 cm^2/mL
2386	EN71-11	7.88	7.88		0.89	
2685	EN71-11	<0,02	<0,02		<-8.37	possibly a false negative test result?
2826	EN71-11	5.378	5.378		-2.06	
2829	EN71-11	21.706	**7.2353**		0.13	used surface to volume ratio of 0.3 cm^2/mL
2936	EN71-11	14.33	**4.777**		-2.77	used surface to volume ratio of 0.3 cm^2/mL
3172	EN71-11	20.75	**6.917**		-0.25	used surface to volume ratio of 0.3 cm^2/mL
3237	EN71-11	20.18	**6.727**		-0.47	used surface to volume ratio of 0.3 cm^2/mL
3238		-----	-----		-----	

normality	OK	
n	21	
outliers	0	
mean (n)	7.126	
st.dev. (n)	2.0513	RSD = 29%
R(calc.)	5.744	
st.dev.(Horwitz)	0.8484	
R(Horwitz)	2.376	





APPENDIX 2 Analytical details

lab	ISO/IEC 17025 accredited?	length test portion (cm)	width test portion (cm)	surface area migration (cm^2)	volume simulant migration (mL)	surface to volume ratio calc. by iis	temp. simulant (°C)	rotation speed (r/min)	time used migration (min)
339	-----	-----	-----	-----	-----	-----	-----	-----	-----
551	Yes	5.00	6.00	30.00	300.00	0.1	20.00	60.00	60.00
623	Yes	6	5	30	100	0.3	24	60	60
841	Yes	5.8	2.7	15.35	51.2	0.3	24.0	60	60
2108	Yes	5	1	5	100	0.05	20	30	60
2115	No	2	5	10	100	0.1	20	60	60
2118	No	6	5	30	300	0.1	20	60	60
2129	Yes	6	5	30	100	0.3	20	60	60
2138	Yes	2.01	5.01	10.07	100.00	0.1	20.50	60.00	60.00
2182	Yes	2	5	10	100	0.1	20	60	60
2247	-----	6	5	30	100	0.3	20	60	60
2256	Yes	3.979	2.518	10.019	100	0.1	21	60	60
2363	Yes	5	2	10	100	0.1	25	60	60
2366	Yes	5	2	10	100	0.1	20	60	60
2375	No	5	2	10	100	0.1	20	60	60
2379	No	6.00	5.00	30.00	100	0.3	20	60	60
2386	Yes	6	5	30	300	0.1	20	60	60
2685	No	3.3	3.3	10	100	0.1	20	60	60
2826	No	6	5	30	300	0.1	24	60	60
2829	No	6	5	30	100	0.3	21	60	60
2936	-----	5.0	6.0	30.0	100.0	0.3	24.5	60.0	60.0
3172	Yes	6	5	30	100	0.3	25	60	60
3237	Yes	6	5	30	100	0.3	20.5	60	60
3238	-----	-----	-----	-----	-----	-----	-----	-----	-----

APPENDIX 3

Number of participants per country

1 lab in BELGIUM
1 lab in BRAZIL
2 labs in FRANCE
3 labs in GERMANY
2 labs in HONG KONG
1 lab in INDIA
1 lab in INDONESIA
3 labs in ITALY
3 labs in P.R. of CHINA
1 lab in SERBIA
1 lab in SOUTH KOREA
1 lab in SPAIN
1 lab in THAILAND
2 labs in TURKEY
1 lab in VIETNAM

APPENDIX 4

Abbreviations

C	= final test result after checking of first reported suspect test result
D(0.01)	= outlier in Dixon's outlier test
D(0.05)	= straggler in Dixon's outlier test
G(0.01)	= outlier in Grubbs' outlier test
G(0.05)	= straggler in Grubbs' outlier test
DG(0.01)	= outlier in Double Grubbs' outlier test
DG(0.05)	= straggler in Double Grubbs' outlier test
R(0.01)	= outlier in Rosner's outlier test
R(0.05)	= straggler in Rosner's outlier test
E	= calculation difference between reported result and result calculated by iis
W	= test result withdrawn on request of participant
ex	= test result excluded from statistical evaluation
n.a.	= not applicable
n.d.	= not detected
n.e.	= not evaluated
fr.	= first reported

Literature

1 iis Interlaboratory Studies, Protocol for the Organisation, Statistics & Evaluation, June 2018
2 ASTM E178:02
3 ASTM E1301:03
4 ISO5725:86
5 ISO5725, parts 1-6, 1994
6 ISO13528:05
7 Directive 2014/81/EU amending Appendix C of Annex II to Directive 2009/48/EC of the European Parliament and of the Council on the safety of toys, as regards Bisphenol A
8 M. Thompson and R. Wood, J. AOAC Int, 76, 926, (1993)
9 W.J. Youden and E.H. Steiner, Statistical Manual of the AOAC, (1975)
10 IP367:84
11 DIN38402 T41/42
12 P.L. Davies, Fr. Z. Anal. Chem, 331, 513, (1988)
13 J.N. Miller, Analyst, 118, 455, (1993)
14 ASTM F963, Standard consumer safety specification on toy safety
15 Analytical Methods Committee, Technical brief, No 4, January 2001.
16 P.J. Lowthian and M. Thompson, The Royal Society of Chemistry, Analyst, 127, 1359-1364 (2002)
17 www.chemicalwatch.com/44942/bpa-poised-for-classification-as-category-1-reprotoxin
18 Annex XVII to REACH Regulation 1907/2006
19 Bernard Rosner, Percentage Points for a Generalized ESD Many-Outlier Procedure, Technometrics, 25(2), 165-172, (1983)