universitätbonn

institut für landtechnik

sektion haushaltstechnik

Report

Round Robin Test ‚Small Cakes'

Empfänger	IEC 59 K WG1+2
Kontaktperson	Fr. Stolz - BSH
Test Institut	Sektion Haushaltstechnik des Instituts f. Landtechnik der Friedrich-Wilhelms Universität Bonn
Prüfer	Diane Lommel
Verantwortlich	Prof. Dr. Rainer Stamminger
Datum	10.06.2005

universitätbonn

Introduction

This round robin test was carried out on demand of IEC sub commission (SC) 59 K WG 1 and 2 in collaboration with the University of Bonn, Section of Household Appliance Technology. The goal was to investigate sub clause 8.4.2. “small cakes” of the IEC standard 60350 Ed 2.0. Especially repeatability and reproducibility were analysed. Small cakes are used to test oven performance, especially evenness of heat distribution in the cavity of an oven. Up to now they are not prepared with standardized ingredients but with ingredients which can be bought everywhere without special defined specifications. These different kind of local ingredients do not lead to reproducible results. This ring test analysed if using standardized ingredients for small cakes ameliorates repeatability and reproducibility.

During the ring test a very big data base was collected and several items were analysed. This report presents only the main goals in order to give an overview on the results. Not all details of the investigation could be mentioned here in this context.

Institut f. Landtechnik
Haushaltstechnik
Nussallee 5, D-531151 Bonn

Tel.: 0228-733117
Fax: 0228-732596
Email: stamminger@uni-bonn.de

universitätbonn

Institut f. Landtechnik
Haushaltstechnik
Nussallee 5, D-531151 Bonn

Tel.: 0228-733117
Fax: 0228-732596
Email: stamminger@uni-bonn.de

universitätbonn

List of abbrevations

Wg: working group
IEC: International Electronical Commission
SC: sub commission
NCS: Natural Colour System
BI: Browning intensity
ave: Average
std. dev.: Standard deviation
S_r: Repeatability standard deviation
S_R: Reproducibility standard deviation
att: attempt

Institut f. Landtechnik
Haushaltstechnik
Nussallee 5, D-531151 Bonn

Tel.: 0228-733117
Fax: 0228-732596
Email: stamminger@uni-bonn.de

universitätbonn

1. Organisation

The round robin test was initiated by IEC SC 59 K working groups 1 and 2. Three big suppliers financed this test and gave technical support, namely Bosch und Siemens Hausgeräte GmbH in Traunreut, Electrolux Major Appliances in Rothenburg and Miele & Cie. KG in Oelde. They all participate in the working groups. It was coordinated by University of Bonn, Section of Household and Appliance Technology. Moreover, support was given by all participating laboratories. Standardized ingredients needed for this test were bought and delivered free of charge to all participating laboratories by StaDe Testmaterialien.

1.1. Objectives

Main goal of the ring test was to investigate repeatability and reproducibility of IEC 60350 Ed. 2.0 sub clause 8.4.2 “small cakes” in consideration of the changes discussed in working group 1 and 2. Additionally modifications were worked out in a pre- study done at the University of Bonn, Section of Household and Appliance Technology focussing on the selection of ingredients which offer the highest possibility for achieving a repeatable and reproducible result. Due to the fact that these improvements are discussed to become a future amendment to IEC 60350 – A2 Ed. 2.0 (now published as 59K-113e-CD) they had to be well qualified.

Another important task of this ring test was to implement a defined way of preparation to help achieving constant results within all participants. As assessment is always the last step in measuring oven performance, different ways of manual evaluation were investigated in this test as well. To fulfil these three most important tasks a very complex organisation for this test was necessary. Therefore, any participating laboratory underwent a testing procedure which was strictly stated in the instructions. To find out the influence of the ingredients and the preparation the same technical equipment was used during the whole test. Ovens, digital calliper gauge, food mixer, thermometer, NCS shade gauge and standardized ingredients were send around to all participating labs. For a general master assessment, preparation of small cakes in one test and documentation of the entire results one student was send around to every participating laboratory. All results were assessed manually by the same student trained on assessing small cakes with the NCS shade gauge.

Institut f. Landtechnik
Haushaltstechnik
Nussallee 5, D-531151 Bonn
Tel.: 0228-733117
Fax: 0228-732596
Email: stamminger@uni-bonn.de

This organisation helped to compare the different situations in the labs and gave an overview on problems.

1.2. Participants

To grant the participants´ anonymity in the following presentation of the results every laboratory is referred by a capital letter. The following list shows all the participants and does not correspond to the given letters:

- Bosch und - Siemens- Hausgeräte GmbH Traunreut, Germany
- Electrolux Major Appliances Rothenburg ob der Tauber, Germany
- Konsumentverket, Stockholm Sweden
- Intertek Milton Keynes, UK
- V-Zug AG Zug, Switzerland
- Miele& Cie. KG Oelde, Germany
- Electrolux Spannymoor, UK

The results of two labs could not be taken into account in the evaluation; the test failed in one lab and time was too short in another. Nevertheless, both labs were important partners in the work of the ring test. In all, the test was carried out in four different countries.

1.3. Equipment and ingredients

Two **ovens**, identical in construction (Bosch HBN 360) were send around and used in every laboratory. One oven (white) was set to 160 °C in forced air convection on level two from bottom and used in this way in test 1 and 3. This oven had a manipulated fan cover to reach bad results. The other oven (stainless) was set to 170 °C in top and bottom heating on level three from bottom and always used in test 2 and 4. Both appliances had an electronic control which guarantees an exact real temperature in the middle of the cavity compared to the setting temperature on the knob. All tests were carried out on only one level. One single **food mixer** Bosch MUM 4600 is used in any laboratory. As the preparation of the mixture for small cakes has a big influence on the quality of the results it is important to guarantee that the ingredients are always treated the same way. Hence, it is necessary to fix a food mixer. **Measuring instruments** like a thermometer, a digital calliper gauge and NCS shade gauge are always used as well. In three of four tests

Institut f. Landtechnik
Haushaltstechnik
Nussallee 5, D-531151 Bonn
Tel.: 0228-733117
Fax: 0228-732596
Email: stamminger@uni-bonn.de

standardized ingredients delivered by StaDe Testmaterialien are used, coming from the same batch. This guarantees comparable starting conditions in all labs.

2. Methodology

2.1. Changes in test procedure of IEC 60350

The current version of the IEC 60350, sub clause 8.4.2. has been under investigation in several test and has not been able to lead to reproducible results. So SC TC 59 worked out a more detailed text with several improvements. Additional refinements were added as felled necessary. They are the basement for the round robin test.

2.2. Test instructions

Every lab is asked to carry out the test in the same way, while the same equipment is used and the student evaluates every test. For the ring test four tests are carried out with each five repetitions. Table one gives an overview.

	Test 1	**Test 2**	**Test 3**	**Test 4**
Used Oven	Oven White 160°C Forced air convection	Oven Stainless 170°C Top and Bottom	Oven White 160°C Forced air convection	Oven Stainless 170°C Top and Bottom
Ingredients	StaDe	StaDe	Local	StaDe
Preparation	Local	Local	Local	Master
Assessment	Master Local staff	Master	Master	Master

Tab.1: structure of small cakes ring test; 4 tests

Based on this test structure the following topics are taken into account for the evaluation:

- Comparison between StaDe ingredients and local ingredients bought in every country
- Comparison between manual master assessment and assessment of local lab staff

Every attempt leads to one baking tray with each 20 small cakes. For the master evaluation every cake is assed manually according to NCS in 13 segments on top. Bottom is assessed as one area. Height is measured with a digital calliper gauge in mm and rounded to the next whole number. As the student was well trained on assessing small cakes the evaluation can

Institut f. Landtechnik
Haushaltstechnik
Nussallee 5, D-531151 Bonn
Tel.: 0228-733117
Fax: 0228-732596
Email: stamminger@uni-bonn.de

be regarded as an objective way of measurement. Hence, all the results of the assessment following below are based on the **manual master evaluation**.

Institut f. Landtechnik
Haushaltstechnik
Nussallee 5, D-531151 Bonn

Tel.: 0228-733117
Fax: 0228-732596
Email: stamminger@uni-bonn.de

3. Results

3.1. Reproducibility of Ingredients

The tables below show the evaluation of the ring test based on the IEC 61923, an international standard for evaluation of ring tests. Test 1 and test 3 are compared in order to contrast testing small cakes made of standardized ingredients and small cakes made of local ingredients which are used in every laboratory.

Test 1: stand. ingredients in oven white		Lab A	Lab B	Lab C	Lab D	Lab E	mean acc IEC 61923	S_r	S_R	S_R/S_r
ave. browning top	NCS	9,83	9,59	9,77	10,26	10,38	9,97		0,9	0,96
ave. browning top- std. dev.	NCS	0,93	0,85	0,87	1,02	0,98		0,93		
Delta ave. browning	NCS	5,5	4,6	4,4	5	4,8	4,86		0,86	1,03
Delta ave. browning- std. dev	NCS	1,29	0,89	0,55	0,71	0,45		0,83		
ave. browning bottom	NCS	9,81	9,61	10,11	10,75	10,67	10,19		0,87	1,1
ave. brown. bottom- std. dev	NCS	1,2	0,6	0,53	0,8	0,65		0,79		
Height	mm	28,8	27,98	28,85	28,57	27,52	28,34		0,99	1,1
height - std. dev.	mm	0,73	0,72	1,02	1,17	0,77		0,9		

Tab.2: Results of test 1

Test 3: local ingredients in oven white		Lab A	Lab B	Lab C	Lab D	Lab E	mean acc IEC 61923	S_r	S_R	S_R/S_r
ave. browning top	NCS	10,21	9,73	9,56	9,84	11,63	10,19		1,14	1,32
ave. browning top- std. dev.	NCS	1,05	0,81	0,77	0,99	0,63		0,86		
Delta ave. browning	NCS	4,4	4,4	4	5	3,4	4,24		0,86	1,23
Delta ave. browning- std. dev	NCS	1,14	0,55	0,71	0	0,55		0,69		
ave. browning bottom	NCS	10,23	9,58	9,7	10,12	11,13	10,15		0,85	1,29
ave. Brown. bottom- std. dev	NCS	0,92	0,59	0,52	0,59	0,6		0,66		
Height	mm	29,55	28,94	30,62	29,46	27,98	29,31		1,23	1,43
height - std. dev.	mm	0,84	0,84	0,82	0,82	0,98		0,86		

Tab.3: Results of test 3

S_r means the repeatability standard deviation and gives information about how repeatable the measurements are within one laboratory.

Institut f. Landtechnik
Haushaltstechnik
Nussallee 5, D-531151 Bonn

Tel.: 0228-733117
Fax: 0228-732596
Email: stamminger@uni-bonn.de

S_R is the reproducibility standard deviation of the performance test small cakes over all laboratories.[1]

As there are no reference tolerances for testing with small cakes defined by now (as needed for an assessment according to IEC 61923), a grading based on experiences with other household appliances tests is used. (e.g. in testing the reproducibility of dish washer cleaning S_R is 1,7; see attached table in annex). Following ranking of S_R is proposed following this experience:

- $S_R / S_r < 1,2$: reproducibility is very good
- $S_R / S_r < 1,5$: reproducibility is still good
- $S_R / S_r > 1,5$: reproducibility not sufficient

Comparing test 1 with StaDe ingredients and test 3 with the different kind of local ingredients, test 1 always shows a very good reproducibility ($S_R / S_r < 1,2$) in all performance criteria. The quotient is always smaller than it is in test 3 which can be interpreted as: StaDe ingredients lead to more reproducible results than the different kind of local ingredients.

S_R / S_r	ave. browning top	delta ave. browning	ave. browning bm	Height
Test 1	0,96	1,03	1,10	1,10
Test 3	1,32	1,23	1,29	1,43

Tab.4: Comparison of results from test 1 and test 3

It is obvious that the average browning of top side with StaDe ingredients leads to the most reproducible results. The average browning of bottom and height has a slightly worse quotient because when removing a cake from its paper case some dough sticks to the paper. That is the reason why it is more difficult to measure these criteria.

If the level of repeatability and reproducibility achieved is sufficient for establishing and controlling a scheme for classification of the baking performance cannot be answered by this ring test.

[1] Formulas for calculation can be found in annex

Institut f. Landtechnik
Haushaltstechnik
Nussallee 5, D-531151 Bonn
Tel.: 0228-733117
Fax: 0228-732596
Email: stamminger@uni-bonn.de

Another, more detailed way of evaluating the database of the ring test is to compare the browning intensity (BI) of every segment. By means of reproducibility small cakes are supposed to show a similar way of browning when baked in one oven with the same ingredients prepared in a nearly identical way and baked for app. the same time. The pattern of browning on a tray has to show an similar result compared within several attempts. Therefore, the results of test 1 assessed by master manually (BI on top) are investigated. As every segment on a certain small cake has a defined "address" (e.g. Cake 2 in line 3, segment 10) it can be compared within all attempts of the ring test. This is done with all of the 260 segments. One attempt of test 1 is chosen as reverence by random. The BI of this reverence attempt is compared to all the other attempts of test 1. The figures below show how many segments in % have the same BI, how high the fit in % is.[2]
Following definitions are used:

"Fit": no difference between BI of reverence segment and compared segment. BI is identical.

"Accepted": BI of compared segment is the fit +/-1 NCS step from reverence segment (e.g. reverence segment has BI 10. Compared segment is allowed to have a BI of 9, 10 or 11). This corresponds to an acceptable level of deviating BI as there are several sources for a difference in BI of 1 NCS step.[3]

	Lab A		Lab B		Lab C		Lab D		Lab E	
att.	Fit	Accepted	Fit	Accepted	Fit	Accepted	Fit	Accepted	Fit	Accepted
1	28,9%	83,5%	24,6%	72,7%	28,1%	75,4%	41,2%	91,5%	-	-
2	48,5%	92,7%	36,9%	83,8%	25,8%	77,3%	47,7%	92,7%	46,5%	92,3%
3	42,7%	93,1%	50%	94,4%	35,4%	87,7%	36,20%	90,4%	50,4%	95,8%
4	47,7%	91,%5	40,8%	91,2%	44,6%	94,2%	40,4%	88,1%	51,2%	96,2%
5	-	-	45%	94,2%	53,1%	93,5%	40,8%	86,5%	52,3%	96,5%

Tab.5: Comparison of every segment in test1; fit and Var+/-1 in %

The same comparison based on manual master evaluation of BI on top of every small cake segment was made for test 3.

[2] One attempt leads to 20 small cakes per tray with each 13 segments. So there are 260 segments per attempt. 10 segments are to be counted as 3,85%.

Institut f. Landtechnik
Haushaltstechnik
Nussallee 5, D-531151 Bonn
Tel.: 0228-733117
Fax: 0228-732596
Email: stamminger@uni-bonn.de

	Lab A		Lab B		Lab C		Lab D		Lab E	
att.	Fit	Accepted	Fit	Accepted	Fit	Accepted	Fit	Accepted	Fit	Accepted
1	40,8%	82,7%	3,1%	22,8%	1,15%	31,9%	5,8%	35,4%	Reference	
2	40,8%	89,2%	3,1%	33,5%	3,5%	40,4%	11,6%	59,6%	70,7%	99,6%
3	18,1%	78,5%	1,5%	54,2%	6,5%	46,5%	14,6%	61,9%	58,5%	96,5%
4	24,2%	86,2%	1,2%	35,8%	15,4%	72,3%	12,3%	53,8%	56,2%	96,2%
5	12,3%	43,8%	8,1%	55,4%	8,5%	52,7%	16,5%	60,0%	55,8%	96,9%

Tab.6: Comparison of every segment in test 3; fit and Var+/-1 in %

Comparing test 1 and test 3 shows obviously that in test 1 the fit of the results is more constant within the labs than it is in test 3. Looking at test 3 shows that ingredients from lab A and ingredients from lab E still lead to quite comparable results (judging by Accepted NCS). But in lab B, C and D ingredients are far from showing a comparable pattern of browning. In lab B for example the maximum level of comparable BI level is 55%. On the other hand in test 1 with StaDe ingredients results have a Accepted NCS level of 72,7%. This is a high level because the analysis is very detailed and many sources for mistakes are taken into account.

Regarding the average over all attempts in all labs from table 6, it shows clearly that in test 3 results are poor. Only 62% of the segments are within Accepted NCS level. Baking small cakes prepared in an nearly identical way and assessed by one person does not show an similar BI within several laboratories in test 3. But in test 1 88% of the segments are in an acceptable range of BI which enables to show a nearly identical pattern of browning within all participating labs.

	Average Fit	Average Var +/- 1
Test 1	42,8%	88,0%
Test 3	20,6%	62,0%

Tab. 7: Comparing the average fits of test 1 and test 3 in %

Summing up the results of comparing StaDe ingredients with different local ingredients, standardized ingredients have to be preferred when maintaining a higher reproducibility.

[3] Light influences, a slightly different baking time or a rotation of the cake when removing from the paper cases e.g. can have an impact on the BI which cannot be eliminated.

Institut f. Landtechnik
Haushaltstechnik
Nussallee 5, D-531151 Bonn
Tel.: 0228-733117
Fax: 0228-732596
Email: stamminger@uni-bonn.de

3.2. Local assessment

Another important task of the ring test was to compare local assessment to master assessment. Therefore, test 1 was assessed by master and by lab internal staff which is experienced with manual assessment. It has to be stated that there are different systems in every laboratory used to assess small cakes. The table below lists up which criteria are used for evaluation.

	Lab A	Lab B	Lab C	Lab D	Lab E
BI top	5 segments; each one BI	1 area; lightest and darkest BI	5 segments; each one BI	1 area; lightest and darkest BI	11 segments; each one BI
BI bottom	1 area	1 area	1 area	1 area	1 area
Height	Ruler, accuracy in m; cluster	digital calliper gauge mm	Mechanical calliper gauge; accuracy in mm	Mechanical calliper gauge; accuracy in 1/10 mm	??

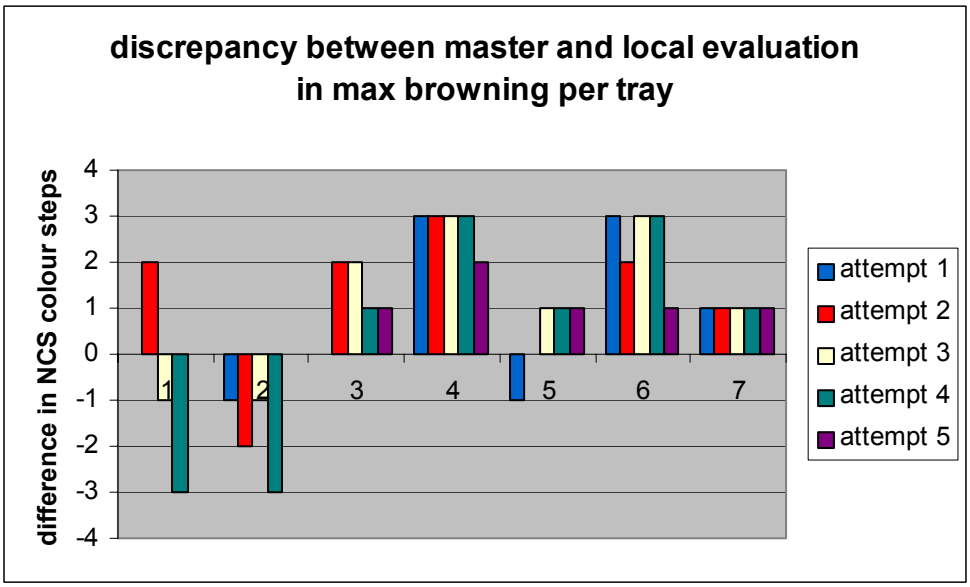
Tab. 8: Different evaluation systems in laboratories

As there are various kinds of evaluation systems used over all participating labs, the results have a different structure as well. Hence, it is not possible to compare them directly in detail. The average BI – usually used in an analysis like this- cannot be taken for a comparison either because of the same reason. An average based on the lightest and the darkest BI does not give the same information as an average of 13 segments. But other important characteristics of local assessment like the darkest BI measured on a cake can be related to master evaluation. The diagram below shows the difference between the darkest BI measured by master and local staff.

Institut f. Landtechnik
Haushaltstechnik
Nussallee 5, D-531151 Bonn

Tel.: 0228-733117
Fax: 0228-732596
Email: stamminger@uni-bonn.de

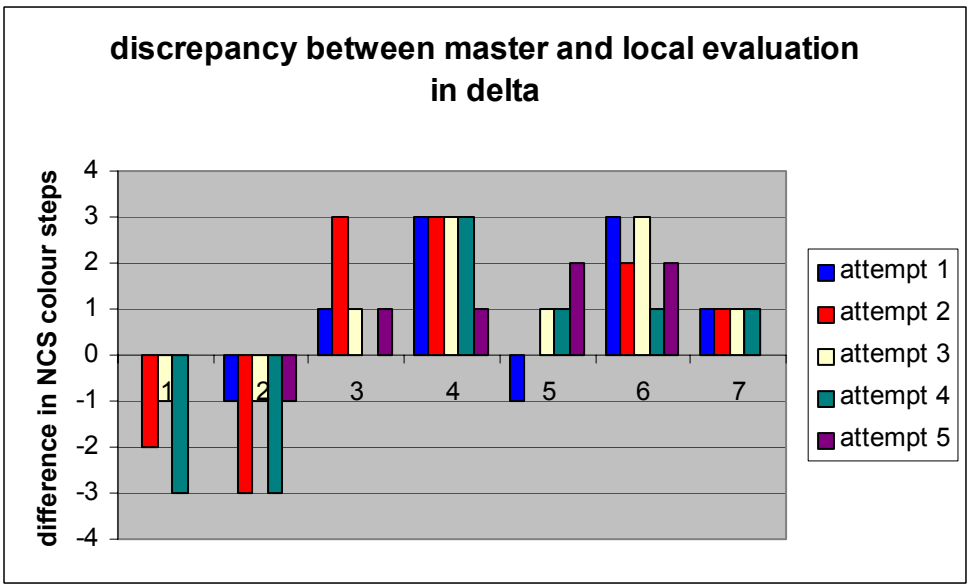
Diagram 1: Discrepancy between master and local evaluation in max BI measured on a tray



The numbers represent evaluating persons
Zero line means no difference between master and local evaluation
e.g. -3 means evaluating 3 NCS steps lighter than master

Although local staff is experienced with assessing small cakes dark BI are difficult to be recognised constantly and correctly. But especially dark BIs play an important role in qualification of oven performance. Therefore, it is a big problem when different people are not able to assess the same maximum BI on a tray of small cakes. More over, this effect touches the results of the delta (difference of darkest and lightest BI measured on a tray).

Diagram 2: discrepancy between master and local evaluation in delta.



The numbers represent evaluating persons
Zero line means no difference between master and local evaluation
e.g. 3 means evaluating a 3 NCS steps bigger Delta than master

Institut f. Landtechnik
Haushaltstechnik
Nussallee 5, D-531151 Bonn

Tel.: 0228-733117
Fax: 0228-732596
Email: stamminger@uni-bonn.de

It can be stated that the important criteria delta BI per tray in assessing small cakes is not assessed in a reproducible way by local staff, although they have many experiences with small cakes.
Comparing only these two criteria leads to the fact that manual assessment of different persons cannot lead to the same results. All the different systems of assessment and human influence have a negative influence on repeatability and reproducibility.
Based on this ring test it is proposed to synchronise the evaluation systems between different laboratories. Additionally, an objective way of measurement is recommended, e.g. a digital system working with a camera and a computer based assessment.

3.3. A defined way of preparation

This round robin test also wants to help implementing a defined way of preparation. Therefore, all participating labs had to follow detailed instructions concerning e.g. ambient conditions, processing of the ingredients or removing small cakes from the paper cases. This was done in all labs without major problems.
As test 1 with StaDe ingredients could prove people all over the participating labs were able to prepare small cakes reproducibly. In order to achieve small cakes of constant quality and high reproducibility it is proposed to retain the instructions of the ring test.

4. Conclusion

Small cakes made of standardized ingredients are definitely able to lead to more reproducible results than different local ingredients. Both ways of analysing reproducibility of small cakes proved this. Especially the more detailed version based on every segment shows that the differences in BI caused by local ingredients from every country are not acceptable. Moreover, this ring test proves that the way of preparation has to be defined as detailed as the instructions for the ring test are. The test shows that all labs are able to prepare small cakes in a good quality. The various assessing systems of small cakes in every lab do not allow a general and reproducible evaluation of oven performance. An improvement based on the system of manual master evaluation has been added to IEC 60350 – A2 Ed. 2.0 (now published as 59K-113e-CD). Further more, manual assessment

Institut f. Landtechnik
Haushaltstechnik
Nussallee 5, D-531151 Bonn
Tel.: 0228-733117
Fax: 0228-732596
Email: stamminger@uni-bonn.de

of different people naturally leads to different results in BI. An objective system is recommended.

A defined way of preparation could be implemented in the test.

Institut f. Landtechnik
Haushaltstechnik
Nussallee 5, D-531151 Bonn

Tel.: 0228-733117
Fax: 0228-732596
Email: stamminger@uni-bonn.de

universitätbonn

Annex

CECED Dishwasher ring test 2003 summary of statistics

	machine	s_r in %	s_R in %	s_R/s_r
Cleaning	H	3,22	5,49	1,70
	L	4,52	5,74	1,27
Drying	H	2,73	3,58	1,31
	L	3,67	8,19	2,23
Energy	H	0,71	2,39	3,37
	L	1,64	3,36	2,05

General accepted conclusion:

Reproducibility is too weak - a lot of action have started within CECED and CENELEC and IEC to improve this situation

$$s_r = \sqrt{\frac{1}{p}\sum_{i=1}^{p} s_{L,i}^2}$$

$$s_{L,i} = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}(x_k - \bar{x}_i)^2}$$

$$s_R = \sqrt{\frac{1}{p-1}\sum_{i=1}^{p}(\bar{x} - x_m)^2 + \frac{n-1}{n}s_r^2}$$

Institut f. Landtechnik
Haushaltstechnik
Nussallee 5, D-531151 Bonn

Tel.: 0228-733117
Fax: 0228-732596
Email: stamminger@uni-bonn.de