PRECISION TRIALS CAN IMPROVE TEST METHODS FOR ALKALI AGGREGATE REACTION (AAR) – PART OF THE PARTNER-PROJECT

Björn Schouenborg*, Urban Åkesson, Lotta Liedberg

SP Technical Research Institute of Sweden, BORAS, Sweden

Abstract

This paper summarizes the results and conclusions that could be drawn from the precision trials carried out within the European research project PARTNER focussing on assessment and development of existing tests for AAR. The overall objective of the project was to provide the basis for unified test procedures for evaluating the alkali reactivity of aggregates. The PARTNER project (www.partner.eu.com) was funded by the European Community.

The methods included in the precision trial were RILEM AAR-1 Petrographic method, AAR-2 Accelerated mortar-bar test, AAR-3 Concrete prism test and AAR-4 Accelerated Concrete prism – Alternative method.

The precision of the test methods was improved from the pre-trials made earlier in the project. The different accelerated tests discriminate between potentially reactive and innocuous aggregates. However, the precision trials also clearly indicated that most laboratories would benefit from more experience with the methods, especially the petrographic analysis. An education package, including a photo atlas, was therefore prepared. By education and practice, the precision of the methods will improve and the methods will also be possible to use for evaluating different grades of reactivity.

Keywords: Alkali-Silica Reaction, concrete aggregates, test-methods, accuracy, precision trials, ISO 5725

1 INTRODUCTION

This paper describes part of the results of PARTNER [1], a project partly funded by the European Community, which had the overall objective of establishing a unified test procedure for evaluating the alkali reactivity of aggregates across the different European economic and geological regions. It is intended that the results of the project will be implemented by CEN in the form of new standard methods of test and, subsequently, specifications.

The main objective of work package (WP) no. 4, dealing with precision data, was to determine the precision of the test methods used in the PARTNER project. This is part of the aim to fill the gap in the Work agenda of "CEN TC 154 Aggregates" in relation to the mandate. There are presently no European standardised test methods for determining the potential alkali reactivity of concrete aggregates. The project aim has also been to primarily use the existing RILEM methods and assess whether they are suitable for the aggregate types generally used in Europe and where necessary, and possible, to improve the test procedures.

The precision is determined by inter-comparison trials. The international standard ISO 5725 [2] describes, in detail, the way an inter-comparison trials should be carried out and evaluated. To satisfy the requirements of ISO 5725, a minimum of eight participating laboratories is needed. All trials were therefore set up to fulfill this basic requirement. In addition, the sample division, coding of samples, preparation of test instructions and result sheets were all prepared in order to avoid the influence by possible errors from other sources than the methods themselves.

Four RILEM methods were included in this WP:

AAR 1 Petrographich method, [3] *AAR 2 Ultra-accelerated mortar-bar test,*[4] *AAR 3 Concrete prism test* [5] and the *AAR 4 Accelerated Concrete prism – Alternative method*[6].

2 ORGANISATION OF THE PRECISION TRIAL

2.1 General

Precision data is used, for example, to assess test results in relation to uncertainty of measurement and in relation to requirements. It can also be used to develop quality classes provided in product standards. In addition, it can be used to assess deficiencies in test methods and the performance of test laboratories. The latter is called proficiency testing.

* Correspondence to: bjorn.schouenborg@sp.se

It is important to note that the real use of inter-comparison trials is as a tool for quality assurance. Inter-comparison-trials should therefore be conducted recurrently, with a certain frequency.

Ideally, all European test methods include a clause describing the precision of the method. In addition, this statement should be based on an inter-comparison trial strictly in accordance with ISO 5725. Further, the precision data should also be the basis for development of quality classes. The reason for this is that if a method has a poor precision, too narrow classes will result in many cases when different laboratories will produce different test result on the same material, they will not classify the material/product in the same way. SP was responsible for the precision trial and is accredited for organizing and evaluating inter-comparison trials according to, for example, ISO 5725.

2.2 Definitions

Repeatability (r_1) conditions. Where test results are obtained with the same test method on different test portions of the same laboratory sample of a material, in the same laboratory, by the same operator, using the same equipment, and within short interval of time.

Reproducibility (R_1) conditions. Where test results are obtained with the same test method on test portions of different laboratory samples of the same bulk sample of a material, in different laboratories, by different operators, using different equipment.

Outlier: A member of a set of values, which is inconsistent with the other members of that set. The consistency can be tested using graphical or numerical techniques [7].

Straggler: Similar to an outlier but with less inconsistency [7].

r = repeatability. This is a measure of the spread in results obtained on the individual samples, tested at the same laboratory, using the same aggregate combinations and same concrete mixture design.

R = Reproducibility. This is a measure of the difference in the mean value between the different laboratories.

COVSR = Coefficient of variation for the Reproducibility. By using the coefficient of variation (COV) one relates the spread to the actual expansion. The COV is the standard deviation divided by the mean value. The COVSR is thus used to compare the difference in the spread between the different laboratories.

*The two types of stragglers and outliers have been used for the precision calculations in this project. Many techniques can be used for testing the consistency of the test results reported from various laboratories. In ISO 5725-2, Cochran's test is recommended for examining the within-laboratory consistency and Grubbs' test for examining the between-laboratory consistency. Grubbs' test can also be used to test the consistency of results measured in one laboratory on the identical material.

ASTM E 691 recommends a graphical technique (Mandel's *k* and *h* statistics. Fig. 5), while ISO 5725 recommends both graphical (Mandel's *k* and *h* statistics) and numerical techniques (Cochran's statistic and Grubbs' statistic). These techniques are also included in SPs soft-ware and shown in the report and have been used for the omission of two outliers.

2.3 The laboratories

Eight laboratories were chosen to participate in each trial (table 1). The criteria for participation were either experience with the method from routine commission work or, as a minimum, participation in the pre-trial tests in PARTNER. ISO 5725 asks for some indication of competency of the participating laboratories. A questionnaire was therefore developed in order to determine the experience of the participants. This, together with the pre-trial tests, was considered sufficient information for the selection of participating laboratories in each trial.

2.4 Selecting the materials

It is preferred to use at least 3 different qualities of materials in a precision trial. Ideally, these materials shall represent the entire working range of the test method. This is a valid principle for all accelerated expansion tests but more difficult to follow for the petrographic analysis (AAR 1).

The materials for the AAR 1 were primarily chosen to represent as large a variation as possible and to include as many different reactive components as possible since the identification of reactive components and classification in different reactivity categories had earlier been identified as one of the main obstacles.

Three materials were chosen for the AAR1 precision trials, consisting of two natural sands and one coarse aggregate. Table 2 list the reactive rocks of the aggregates selected for AAR 1 only. The reason for the selection of these materials was that they are composed of various types of reactive rocks, and most of these are globally known and accepted as reactive.

Low/non-reactive, medium reactive and highly reactive materials were considered for inclusion in the tests AAR 2, 3 and 4-alternative methods. In order to identify three candidates for each test, a questionnaire was sent out asking the partners for their knowledge of the reactivity of the different materials that could serve as suitable candidates for use in the project.
Cataclasite from Norway was chosen for the highest grade of reactivity (Class 3), Belgium siliceous limestone for the medium reactive aggregate (Class 2) and a Norwegian natural gravel for the non-reactive (Class 1).

2.5 Preparing the test specimens

Petrographic analysis (AAR 1)

The materials were sampled at each quarry according to EN-932-1 [8]. The materials were then sent to SP for preparation of test specimens. The two sands, samples S and D were divided using a rotary sample divider, a small one for the sand and gravel and a large for the crushed rocks (fig. 1). The materials were sieved into 0.063/1, 1/2 and 2/4 mm fractions. The crushed material was sieved and the 2/4 mm fraction was used for the analysis. The sieved fractions were divided into eight specimens using a small-scale rotary sample divider. The specimens were sent to the participating laboratories for preparation of thin sections. One thin section was prepared for the 0.063/1 mm fraction and two thin sections for each of the 1/2 and 2/4 mm fractions.

The materials were delivered to SP either directly by the producer or from one of the PARTNER's. They were thus considered to represent a ready product and no further treatment beyond sample splitting was undertaken. This is how they were prepared prior to being shipped to a test laboratory.

Accelerated expansion tests

The sample splitting for the AAR 2 was carried out by use of a large rotary sample divider (fig 1). The samples for the concrete methods AAR 3 and 4 Alternative was carried out by means of fractional shovelling (fig 2), a technique used in the MAT1 - CT 93 – 0040 [9].

2.6 Test instructions and evaluation

In order to simplify the calculations and avoid misunderstanding, detailed instructions were prepared for all methods. Reporting templates were also prepared for all necessary information in addition to test results. Excel files were prepared for the documentation and calculation of all results in a uniform way. All knowledge gained from the pre-trials was used to prepare any necessary additional instruction. In addition, SP developed an Excel template for all calculations in accordance with ISO 5725:1994 (figure 3).

4 TEST RESULTS

4.1 AAR 1 Petrography

It is very difficult to apply a strictly statistical approach to the results of the petrographic analyses, since there is no true value/result. We have to rely on the competence on the laboratories providing the test material(s). The three laboratories providing the test materials are the ones with the most "nuanced" test results. Their results were used as a kind of reference.

The results from the AAR1 petrography show a large spread between the different laboratories. Almost every laboratory has identified the rocks and minerals correctly. However, the classification of the degree of alkali silica reactivity is different. This is probably due to the regional experience and inexperience with the different reactive rocks. Only porous/opaline flint and mylonite/cataclasite were classified as reactive by all laboratories. The coefficient of variation is high and varies between 40 and 140 %.

4.2 AAR 2 Ultra-accelerated mortar-bar method

The reference method has been to use thin and long prisms; 25x25x285 mm and evaluating the expansion after 14 days. In PARTNER, another common dimension was included, the 40x40x160 mm. The reason for this is that several European laboratories already use the latter dimensions for several other tests, for example, the Nordic accelerated mortar bar test [10] and for cement testing according to EN 196-1 [11]. One purpose was to investigate whether the same requirements (quality classes) can be used and if the precision is equivalent for both procedures/dimensions. Two options were therefore used, the traditional with thin and long prisms and one using the thicker and shorter prisms. The precision evaluation of the expansion was made after 14 and 28 days and on 3 and 5 prisms. However, several measurements were made in between.

Using 5 prisms instead of 3 did not improve the precision. Many laboratories found it necessary to use two containers to accommodate 5 prisms, which brings about slightly different conditions. The final precision calculations were therefore made on results obtained for 3 prisms.

The reaction is initially (14 days results) about two times quicker in the thin prisms for low-expanding material and it is only after about 28 days that prisms of both dimensions show comparable expansion (figure 4).

4.3 AAR 3 Concrete prism test

This test takes one year to complete and the possibilities to correct any errors are few. All laboratories did not manage to report fully. Nevertheless, after the recalculation, between 6 and 7 laboratories remained for the final evaluation, which is acceptable.

The standard deviation of all laboratories taken together was very large. The standard deviation of the reproducibility is about half that of the test result itself. It is therefore very difficult to identify any laboratory that has an "outlier" test result.

When applying the uncertainty of the measurement in a very rigid way, it may be concluded that the method may be useful for discriminating between not reactive and likely reactive. It has not proven satisfactory in discriminating between likely reactive and clearly/highly reactive. Another approach for interpreting the results is to say that perhaps test material "Class 3" is not highly reactive as previously thought?

4.4 AAR 4 Concrete prism – Alternative method

The method has a better precision than the AAR 3. The "overlap" between the different materials (for example, categories of reactivity) is not as large: approximately half that of the AAR 3 test.

The coefficient of variation, COV_{SR}, is about one third of the mean test result. The repeatability COVsr is about 10 to 15 % which is acceptable and slightly better compared to the AAR 3 with 14 to 21 %. Figure 5, below, is one of the resulting diagrams from the ISO 5725 based software. It displays a measure of the reproducibility, i.e. the spread between laboratories as the number of standard deviations from the overall average. The zero-line is the average and the columns represent one material, with the height representing the deviation from the average. Laboratory 7 did not complete the test and laboratory 5 had one outlier (extreme result) that was omitted from the final calculation. The diagram gives a clear illustration of the quality of the test/trial. Note that the lab number does not represent the numbering of laboratories listed, Table 1.

5 CONCLUDING DISCUSSION

5.1 Petrographic analysis

It can be concluded that Petrographic analysis is a subjective method and that the results are therefore highly dependent on the petrographer's experience with AAR, the means of carrying out the point counting and the classification of the reactivity according to the method. According to the results from this precision trial, the best way to decrease the influence of subjectivity is to count only the reactive mineral grains/components as they are encountered at the cross hairs and not whole particles. If a particle is composed of both reactive and non-reactive material, the statistical approach to use a virtual grid, of a suitable size, covering the entire test area will ensure a correct quantitative estimate of amount of these two components. The size of the grid is equal to the step size between the points and the rows to count. However, if the petrographer has a certain experience with some local material it may be adequate to count the whole grain.

In addition, the conclusion of the previous STAR-project [12] was confirmed, i.e. the need for education. Petrographers should be familiar with the potentially reactive components in aggregates from different countries in order to use the proper analytical techniques and to make a relevant assessment. Education in combination with proficiency trials is therefore our recommendation for future constructive improvement in this field.

5.2 Expansion tests

All expansion tests rank or categorize the test materials in the same way (figure 6) and from what was initially expected in terms of reactivity. However, the expected difference between the medium and highly expansive materials is very small.

The AAR 2 has seemingly the biggest "difficulties" to discriminate between them. However, in practice, this is not true. The diagram must be evaluated together with the precision of each method. The precision is very good and it is possible to detect small differences in reactivity. Two varieties of the same test have been included, the traditional long and thin prisms (25x25x285 mm) and the "new" short and thick prisms (40x40x160 mm). Both methods seem to work well. The standard 14 days give different results, with the long prisms having the highest/quickest expansion. However, the results equalize after 21 days and more so after 28 days. Thus, it is clear that the same requirement for discriminating between non-reactive and reactive can not be used after 14 days exposure. The difference between laboratories is generally less than 1 standard deviation of the overall mean value. Another important finding is that it is sufficient to use 3 replicates (test specimens) instead of 5.

Concerning the one year long AAR-3 Concrete prism method, the repeatability is good, whereas the reproducibility is rather poor. There is a very good correlation between the spread and the mean result. The overall precision of AAR 3 is not sufficient to be used to discriminate between different types of reactivity. The conclusion of the precision trial is that the method can only be used to discriminate between non-reactive and reactive materials.

The 6 months long AAR 4 Alt. method has a better overall precision than AAR 3. The possibility to ensure a stable climate under testing is most likely better due to the well-defined equipment. Similarly as for the AAR 2 (small prism method), the AAR 4 alternative method can therefore detect smaller differences in reactivity compared to the AAR 3 method. The smaller spread between the three tested materials is therefore not a significant concern but merely to be expected due to the shorter time of exposure.

Table 3 gives a summary of the most important precision data, i.e. the test results as the average and the coefficient of variation for the repeatability and reproducibility standard variation. It is a measure of the spread as a percentage of the mean. The mean is the closest one can get to the true value in this kind of precision trial.

6 REFERENCES

[1] Nixon, P.J, Lingård, J, Borchers, I, Wigum B.J and Schouenborg, B. (2007): The EU "PARTNER" Project- European Standard Tests to Prevent Alkali Reactions in Aggregates, Final Results and Recommendations.

[2] ISO 5725–94: Precision of test methods - Determination of repeatability and reproducibility for a standard test method by inter-laboratory tests.

[3] RILEM (2003): AAR-1 - Detection of potential alkali-reactivity of aggregates – Petrographic method, Materials and Structures (36):480-496.

[4] RILEM (2000): AAR-2 - Detection of potential alkali-reactivity of aggregates - the ultra-accelerated mortar bar test. Materials & Structures (33): 283-289.

[5] RILEM (2000): AAR-3 - Detection of potential alkali-reactivity of aggregates - Method for aggregate combinations using concrete prisms, Materials & Structures, (33): 290-293.

[6] RILEM (to be published): AAR-4 - Detection of potential alkali-reactivity of aggregates: Accelerated (60°C) concrete prism test.

[7] Luping & Schouenborg (2000): Methodology of Inter-comparison Tests and Statistical Analysis of Test Results. NORDTEST Project No. 1483-99. SP REPORT 2000:35.

[8] EN-932-1. Tests for general properties of aggregates – PART 1: Methods for sampling.

[9] MAT – CT93 – 0040 (1993-1997): Testing of industrial products; aggregates for construction. Final report – Cross testing.

[10] NT Build 295: Sand. Alkali silica reactivity – Accelerated test.

[11] EN 196-1: Methods of testing cement – Part 1: Determination of strength.

[12] STAR-Project, SMT4-CT96-2128.

TABLE 1: Table of participating laboratories in the different methods. The laboratory number was later deleted for confidentiality reasons.

Name of the laboratory	Lab number	Country	Petrography AAR 1	Mortar bars AAR 2 (40x40x160 mm) (25x25x285 mm)	Concrete prisms AAR 3	AAR 4 Alt. (incl wrapping) standard Rilem container
BRE	1	GB				
STATS	2	GB				
PC-lab	3	DK				
SINTEF	4	N				
SP	5	S				
NORCEM	6	N				
RAMBOLL	7	DK				
LCPC	8	F				
ISSEP	9	B				
VDZ	10	D				
CRIC	11	B				
CESI	12	I				
LABEIN	13	ES				
IMBiGS	14	PL				
DTI	15	DK				
Holcim	16	B				
Cemex	17	ES				
AIDICO	18	ES				
IBRI	19	IC				
VOZ	20	A				
TITAN	21	EL				
Hönnun	22	IS				

TABLE 2: Tested samples for the AAR1 precision trials
and the constituent of reactive rocks in each sample.

Reactive rocks (class II/III according to AAR 1)	Sample S Swedish sand/gravel (0/8 mm)	Sample D Danish sand (0/4 mm)	Sample N Norwegian gravel/stone (8/16 mm)
Mylonite/cataclasite	X	X	X
Quartzite	X	X	X
Sandstone	X	X	X
Greywacke	X		
Meta rhyolite	X		X
Flint		X	
Porous/opaline flint		X	
Slate	X		
Siltstone	X		

Note: Class II – alkali-reactivity uncertain; Class III – very likely to be alkali-reactive

TABLE 3: Precision statements for all assessed methods. *m* is the mean value.

	repeatability standard		coefficient of correlation	reproducibility standard		coefficient of correlation
Method	**deviation**		**r2**	**deviation.**		**R2**
	Sr			**SR**		
AAR- 1	Not determined			SR =0,143*m*		0.60
AAR- 2, 14 days 25x25x285 mm 5 prisms	Sr = 0.015*m*	0.539	0.50	SR = 0.194*m*	1.812	0.82
AAR- 2, 14 days 40x40x160 mm 5 prisms	Sr = 0.002 +0.039*m*		0.78	SR = 0,922*m*	1.854	0.85
AAR- 2, 28 days 25x25x285 mm 5 prisms	Sr = 0.022*m*		0.35	SR = 0,245*m*	2.593	0.80
AAR- 2, 28 days						
40x40x160 mm	Sr = 0.005 +0.015*m*		0.87	SR = 0,161*m*	0.704	0.63
3 prisms						
AAR- 3	Sr = 0.257*m*	1.232	0.99	SR = 0.001 + 0.512*m*		1.0
AAR- 4 Alternative	Sr = 0.741*m*	1.919	0.93	SR = 0.453*m*	1.32	0.92

Note: See section 2.2 for further definitions. Sr relates to the differences/spread within one laboratory, whereas SR relates to the spread between laboratories. The coefficient of correlation r^2 (and R^2) is 1.0 if the correlation is perfect; a good correlation is about 0.9 or better. In the table r^2 is used to assess the validity of the formula, i.e. if r^2 is about 0.9 or higher the formula can be used, with a reasonable amount of certainty, to calculate the expected difference in test results if the same material is sent to different laboratories.

Figure 1: Rotary sample divider with a 50 liter hopper and an outlet that can take maximum 50 mm big particles.

Figure 2: Fractional shovelling, in accordance with EN 932-2.

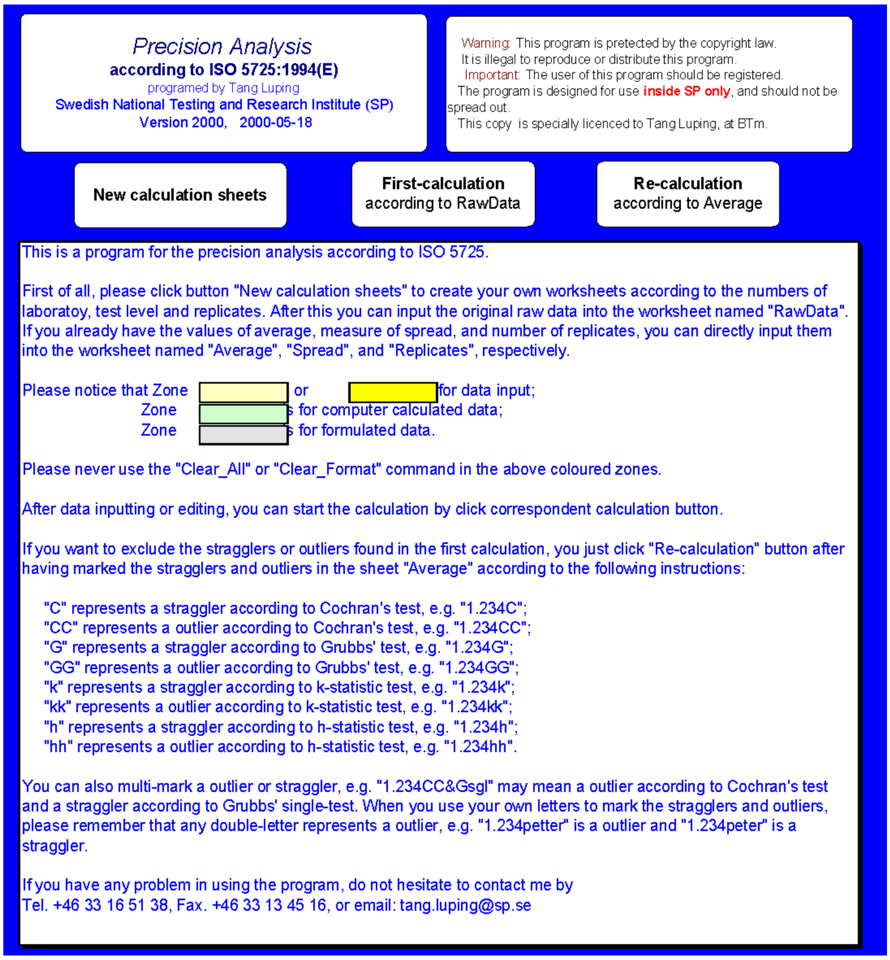

Figure 3: Excel template for Precision analysis according to ISO 5725.

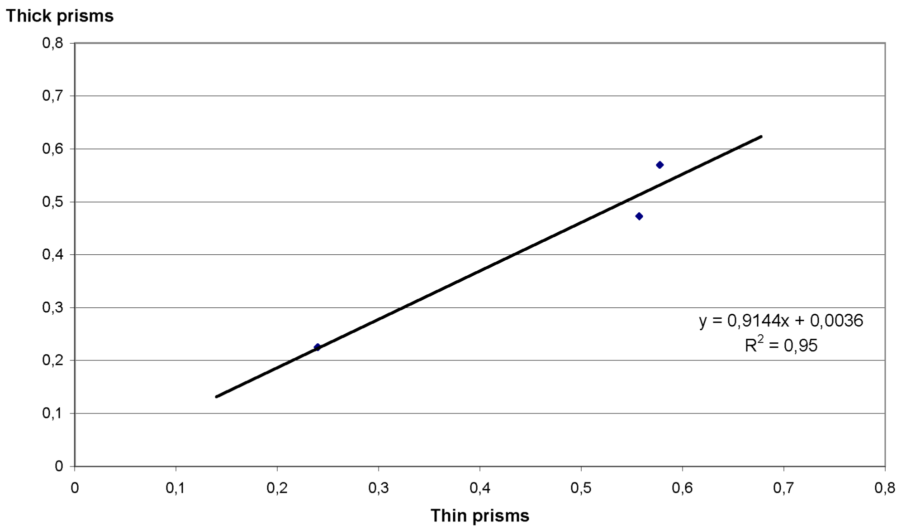

Figure 4: Comparison of average test values for 3 thin and 3 thick prisms after 28 days exposure.

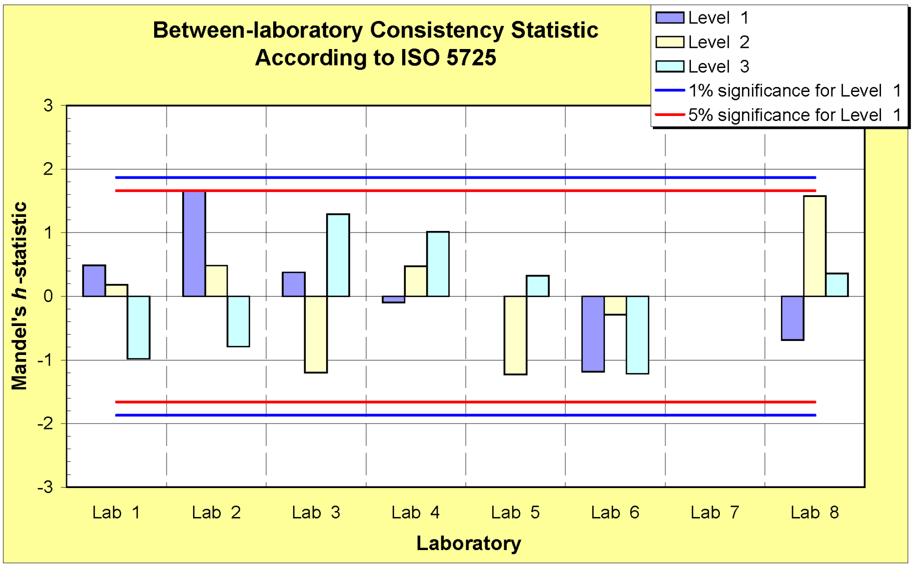

Figure 5: Mandel's h-statistics for the AAR 4 results, showing the spread in test results between all participating laboratories.

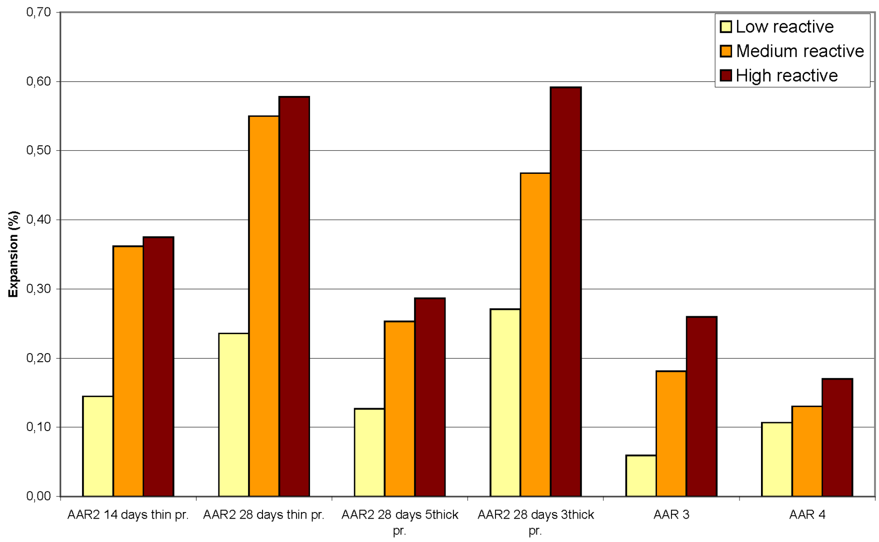

Figure 6: Classification of the three test materials according to the Rilem-methods. Yellow indicates the results of the least expansive material combinations.