MULTI-LABORATORY PRECISION OF MARSHALL DESIGN RELATED TESTS

E. Denneman, H. Marais*, and E. S. Sadzik**

CSIR, P.O. Box 395 Pretoria, 0001, South Africa, Email: edenneman@csir.co.za
*Much Asphalt, P.O. Box 1089, Benoni, 1500, South Africa,
Email: herman.marais@murrob.com
**GDPTRW, Private bag X3, 0039, South Africa, Email: elzbieta.sadzik@gauteng.gov.za

ABSTRACT

The Marshall method is still the method of choice for the design of Hot-Mix Asphalt (HMA) in South Africa. During the validation of a HMA mix design, considerable variability was encountered in Marshall test results for the same mix supplied by different laboratories. The variability was of such a degree that it would influence design decisions or could lead to conflict over product acceptance. A study was undertaken to investigate the extent and consequences of the inter-laboratory variability. The results, presented in this paper, show that the tests included in the Marshall design method are subject to considerable inter-laboratory variability. The variability should be taken into account during the design phase as well as during quality control and product verification. One of the deficiencies in the current standard test method is that it is lacking an effective method of calibrating the compaction effort for different hammers. The results indicate a need to calibrate the Marshall hammers of all parties taking part in a project and for every HMA mix that is used. Also, a need was identified for a national proficiency testing scheme to aptly identify, and subsequently correct, sources of variability in the testing processes.

1. INTRODUCTION

The Marshall design method for Hot-Mix Asphalt (HMA) was originally developed by Bruce Marshall of the Mississippi highway department around 1939. To this day the method remains the back-bone of HMA design in many places in the world. In South Africa it is the dominant method of design for most HMA types. The contemporary Marshall method for determination of the optimum binder content described TMH1:1986 includes the preparation of test briquettes in a mould, using the compaction effort of the Marshall hammer. After compaction the density of the briquettes is determined, along with other volumetric properties. The stability and flow of the compacted samples is measured in the Marshall breaking head apparatus. The results of these test as well as volumetric parameters such as density, air void content, voids in mineral aggregate (VMA) are compared for different binder contents, and from this comparison, an optimum binder content is determined for the mix. Marshall method related testing is performed on a large scale in HMA design laboratories, as well as in site laboratories for quality control and product acceptance testing during the construction phase.

An underlying assumption to the current TMH1:1986 procedures and the use of the Marshall hammer in industry processes is that different Marshall hammers will generate the same standard compaction effort. This assumption is flawed as will be shown in present paper.

In a recent study to validate the optimum binder content for a standard HMA, considerable inter-laboratory variability in Marshall test results was encountered (Denneman, 2006). The inter-laboratory variability was of such a degree that it could have serious implications for the reliability of Marshall HMA designs in South Africa. The Marshall design method is expected to maintain its dominant position in HMA design practice in South Africa for the foreseeable future. Design engineers and client bodies require reliable and comparable Marshall test results to predict and assess the field performance of HMA, during the design phase. Inter-laboratory variability between the contractor's quality control tests and the client's verification tests may lead to conflict over product acceptance. Quality data is one further a requisite to understanding the field performance of HMA when compared to Marshall design parameters.

In a meeting with representatives of several specialized laboratories it was proposed that the origin and the extent of the variability should be investigated in greater detail so that problems with reliability can be addressed effectively. The main findings of this study are presented in present paper, the full results can be found in Denneman and Marais (2007).

2. METHODOLOGY

2.1 Initial study

The initial mix validation study (Denneman, 2006) eventually involved five laboratories. Each laboratory compacted Marshall HMA specimens, the material was mixed from raw materials (binder and aggregate) at a range of binder contents in accordance with TMH1: 1986 method C2. The Marshall stability and flow were determined in accordance with the same method. The bulk relative density (BRD) of the compacted samples were determined in accordance with TMH1: 1986 method C3 and the maximum theoretical relative density (MTRD) of the loose mix was determined in accordance with TMH1: 1986 method C4.

2.2 Detailed investigation

For the detailed follow-up study of inter-laboratory variability, the laboratories of two asphalt producers, one client body, a research organization and three commercial laboratories (seven laboratories in total) were approached and found willing to participate in round-robin testing. Most Marshall method design parameters are determined through a number of consecutive testing processes. To be able to assess the different sources of variability in isolation, the results of each process needed to be compared separately. Therefore, the study was divided into separate tasks, with each task targeting variability in specific aspects of Marshall method related testing.

2.2.1 Task 1: Assessment of variability in determining material density

The objective of the first task was to investigate inter-laboratory variability in the determination of bulk relative density (BRD), and maximum theoretical relative density (MTRD). Each laboratory determined the density of identical samples circulated amongst the participating laboratories.

Three test briquettes of a single standard continuously graded medium HMA mix were compacted in a single laboratory and then circulated amongst the other laboratories. Two samples of loose mix of the same production batch were also supplied for determination of the MTRD by each laboratory. Detailed work instructions, as well as standard report forms were provided to the laboratory.

2.2.2 Task 2: Assessment of variability originating from sample preparation methodology
This task aimed to assess possible differences in test results caused by the methodology of mixing the HMA material from loose components (i.e. aggregate and bitumen). This was done by compacting at the different laboratories a mix that was centrally produced at a single asphalt plant, thus mix eliminating the mix production factor from the comparison. By comparing the results from task 2 to the results from the initial study, which included the mixing process variability related to the mixing process can be identified. Representatives of the different laboratories met at the asphalt plant to collect mixed material in hot boxes. The mix was of a different composition than the mix used for the initial study, but both mixes were of the continuously graded medium variety. Each of the participating laboratories used this material to produce six Marshall briquettes and determine the BRD (6x), Marshall stability and flow (3x), and indirect tensile strength (ITS) (3x) in accordance with ASTM D4123-82. Detailed work instructions, as well as standard report forms were provided to the laboratories.

2.2.3 Task 3: Assessment of equipment and operator dependent variability
In the last task the objective was to isolate the equipment related variability from operator dependent variability. To this end, an experienced technician visited the different laboratories and prepared Marshall briquettes in each laboratory. The technician used the same loose material from a single batch (aggregate as well as bitumen) in every laboratory. The HMA mix design and raw materials were the same as used for the mix of task 1. The technician used the same mixing procedure in every laboratory and brought his own spatula which he used every time. In each laboratory the technician used the most frequently operated Marshall hammer to produce specimens. The technician then determined BRD, Marshall stability and flow for three of the specimens with the equipment available in the laboratory. The same technician then tested the three remaining briquettes from each lab using the equipment available at his place of work (using the same equipment every time).

The results from this task allows a comparison of Marshall hammers used in the different laboratories as well as a comparison of the equipment used for BRD, Marshall stability and Marshall flow. By comparing the results from this task to the results from the initial study, which involved multiple operators, the operator dependent variability can be assessed as well.

3. RESULTS

3.1 Initial study
The Marshall method design parameters determined from the results provided by the laboratories involved in the initial mix validation study are shown in Figure 1. The mix design criteria applied to the design are also shown in the figure. The optimum binder content for the mix as determined by the lab from which the design originated (laboratory 1) was 5.0 per cent. If assessed using the results from the other laboratories that binder content is at the dry side of the Bulk Relative Density (BRD) curve. The lower density at 5 per cent binder content achieved at three of the laboratories is the probable cause of those samples not meeting the minimum requirements for Marshall stability. For some of the laboratories the mix does not meet the requirements for percentage Voids Filled with Binder (VFB) or the percentage air voids in the mix.

The results obtained were cause for concern. If these sets of values were used as an input to the Marshall design method, the conclusions on the optimum binder content for the mix could be different depending on which laboratory supplied the results.

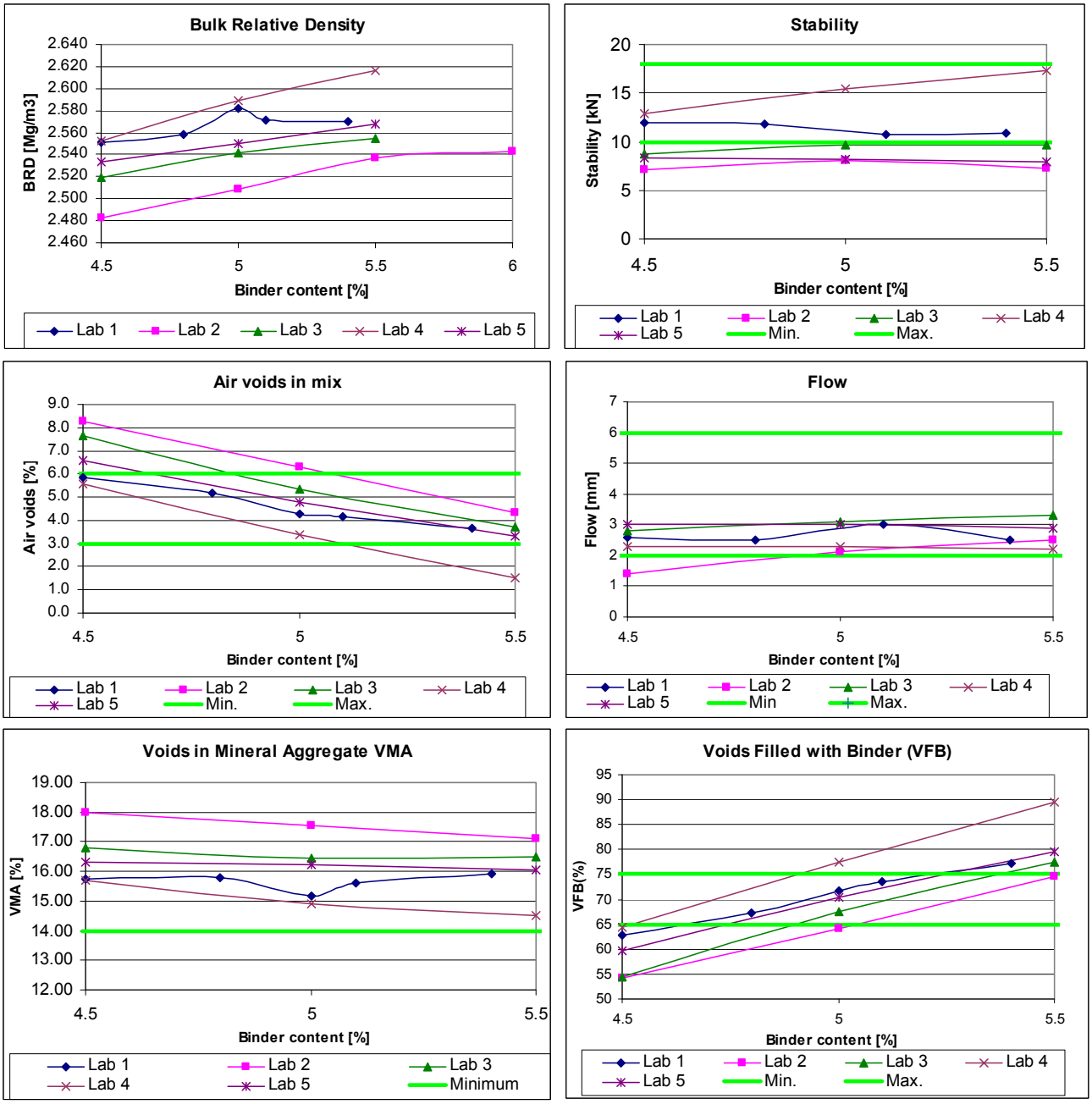

Figure 1 Marshall method design parameters initial study

3.2 Analysis of sources of variability and comparison to international data

The test property statistics for inter-laboratory precision for density results obtained in the different tasks are shown in Table 1. The table also contains the statistics from the American AASHTO Materials Reference Laboratory (AMRL) proficiency testing programme for the past three years (available from www.amrl.net). The scale of the AMRL proficiency scheme, with hundreds of laboratories participating every year, dwarfs the present study, but provides a valuable benchmark for comparison. The table shows the number of laboratories for which valid results were obtained after outliers that could be related back to systematic errors were ignored. Reported are the standard deviation (σ) and the Coefficient of Variation (CV), which is the ratio between σ and μ as per Equation 1.

$$CV = \frac{\sigma}{\mu} \cdot 100\% \qquad (1)$$

The CV value is dimensionless and can be used to compare distributions with different mean values or units. It is applied in this report to compare test results from different mixes and results reported in SI units to different, and sometimes undefined American units.

A comparison between the results from the initial study and the data from the American proficiency testing scheme shows that the coefficients of variation obtained from the local study for BRD, MTRD and air voids are within the range of what is to be expected for the precision of producing samples from loose components using the 75 blow Marshall compaction effort. The precision has been relatively stable over the years, Cassidy et al (1994) reported that the multi-laboratory CV for air voids range from 20 to 35 per cent for nine years of AMRL proficiency testing data. An analysis of the latest three years of AMRL proficiency data by Azari et al (2007) shows that the standard deviation for multi-laboratory precision of the percentage air voids is 1.0 per cent, which is similar to the findings of the present study. The present study shows the multi-laboratory precision for a limited number of local laboratories to be comparable to international standards. However, there are key differences in project specific calibration requirements, to mitigate the high variability, in the American standard (ASTM D6926-04) and TMH1:1986, as will be discussed in Section 4.

The results from task 1 show that multi-laboratory precision of the determination of BRD and MTRD on a set of specimen for a single mix is high. It needs to be mentioned however that the results from one laboratory had to be ignored due to a systematic error in determining the BRD and the results of two laboratories had to be treated as outliers for MTRD results. This in itself is cause for concern. As a consequence the percentage air voids was determined for four of the seven laboratories only.

For task 2 the material mixed centrally at an asphalt plant was compacted at the participating laboratories The out coming data, when compared to the results of the initial study, indicate that the mixing procedure accounts of a major portion of the variability inherent to the test procedure. The mix used for task 2 was not the same as used for the initial investigation and a one on one comparison is therefore not possible.

Table 1 Multi-laboratory precision of density results

Task	BRD			MTRD			Air voids		
	No Labs	σ [Mg/m^3]	CV [%]	No Labs	σ [Mg/m^3]	CV [%]	No Labs	σ [%]	CV [%]
Initial study 4.5% b/c	5	0.028	1.13	4	0.011	0.40	5	1.1	16.9
Initial study 5.0% b/c	5	0.032	1.27	4	0.006	0.21	5	1.1	23.0
Initial study 5.5% b/c	4	0.034	1.33	3	0.002	0.08	4	1.2	38.0
Task 1 specimen 1	6	0.001	0.06	5	0.004	0.16	4	0.002	4.72
Task 1 specimen 2	6	0.003	0.11	N/A	N/A	N/A	4	0.002	3.94
Task 1 specimen 3	7	0.005	0.20	N/A	N/A	N/A	5	0.002	3.39
Task 2	7	0.014	0.59	7	0.011	0.45	7	0.87	26.3
Task 3	7	0.016	0.68	N/A	N/A	N/A	N/A	N/A	N/A
AMRL 2007*	585	0.029	1.22	519	0.007	0.27	566	1.24	17.9
AMRL 2006*	593	0.021	0.86	541	0.008	0.31	578	0.94	28.6
AMRL 2005*	564	0.025	1.01	557	0.015	0.56	552	1.00	18.4

Note: *AMRL results are shown are for uneven sample numbers only.

A single operator visited the different laboratories to produce Marshall briquettes from mix components for task 3. The coefficient of variation of the BRD results is lower than for the initial study, in which the sample was prepared by different operators. The technician used the same mixing procedure at every laboratory. With both mixing procedure variability and operator dependent variability eliminated the remaining (substantial) variability may be

assumed to be equipment related.

Table 2 shows the multi-laboratory precision for Marshall stability and flow, as well as ITS results for the present study and the equivalent AMRL statistics where available. The stability results from the initial study exhibit high variability. The CV results from the AMRL proficiency scheme typically range between 15 and 20 per cent (Cassidy, 1995). The latest AMRL data shown in the table also falls within this range. It appears some improvement can be made in local multi-laboratory precision for stability. The flow results on the other hand are within the 15 to 25 per cent range mentioned in the Cassidy (1995) study. The multi-laboratory precision for stability increases as sources of variability are eliminated. Task 2, which excludes multi-laboratory mixing of the material, resulted in a higher precision for stability, as well as flow. A single operator visiting the various labs to produce Marshall briquettes for Task 3 results in a reduction of variability in stability and flow results, with exception of the flow results for the samples tested by the technician in the visited laboratory.

Table 2 Multi-laboratory precision of stability, flow and ITS results

Task	Stability			Flow			ITS		
	No Labs	σ [kN]	CV [%]	No Labs	σ [mm]	CV [%]	No Labs	σ [kPa]	CV [%]
Initial study 4.5% b/c	5	2.5	25.6	5	0.6	25.9	N/A	N/A	N/A
Initial study 5.0% b/c	4	3.5	34.0	4	0.5	19.0	N/A	N/A	N/A
Initial study 5.5% b/c	4	4.6	44.1	4	0.5	17.6	N/A	N/A	N/A
Task 2	7	2.0	17.9	7	0.4	12.5	7	103	10.7
Task 3 home lab	7	1.1	11.0	7	0.2	8.7	N/A	N/A	N/A
Task 3 visited lab	4	1.2	13.9	4	0.7	21.9	N/A	N/A	N/A
AMRL 2007*	541	2.4	16.2	533	-	19.0	N/A	N/A	N/A
AMRL 2006*	544	2.0	15.1	546	-	16.5	N/A	N/A	N/A
AMRL 2005*	531	2.0	16.7	535	-	22.0	N/A	N/A	N/A

Note: *AMRL results are shown are for uneven sample numbers only.

3.3 Observations with respect to equipment

Figure 2 shows the range of different Marshall hammers that were encountered during the present study. One of the laboratories used a manual compactor (refer Photo 1), the other six laboratories used mechanical compactors. All the hammers complied with the TMH1: 1986 specification as far as the hammer drop height and drop weight are concerned, but the configuration of the mechanical hammers were not the same. Refer to Photo 2 for the type of mechanical hammer used by 5 laboratories and Photo 3 of the mechanical hammer used by one laboratory. The major differences between the hammers were the number of blows per minute, type of release mechanism of hammer, size of counter weight on top of hammer, total weight of hammer as well as the way the hammers were bolted to the floor. The number of blows per minute ranged from 90 (manual compaction), 60 – 74 (mechanical compactors Photo 2) and 52 (mechanical compactor Photo 3). Quicker compaction reduces the amount of heat lost during the compaction process, resulting in higher density of the briquette due to lower viscosity of the binder. At slower speeds, the asphalt has longer to cool down but on the other hand, the hammer can rebound between blows and spends more time on the sample, which results in extra compactive effort being applied.

The type of release mechanism of the mechanical hammers requires different sizes of counter weights on top of the hammer to prevent the hammer from being lifted out of the mould. With a manual compactor the operator also tend to apply a static load on the hammer. The varying static loads imposed by the different sizes of counter weights as well

as the total weight of the hammer can also have an slight influence on the compactive effort being applied. One of the laboratories also had a hammer without a counter weight (Refer Photo 4), this hammer was not used for the present study.

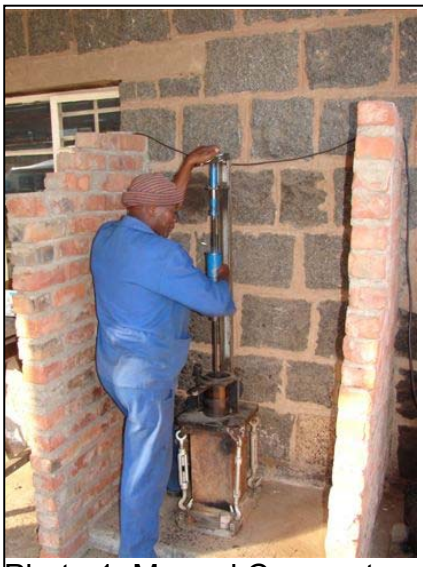

Photo 1: Manual Compactor

Photo 2: Mechanical compactor

Photo 3: Mechanical Compactor

Photo 4: MC without counter weight

Figure 2 Various types of Marshall compactor equipment

4. REDUCING VARIABILITY

Cassidy (1995) identified the following sources of variability in the Marshall compaction method:

- Operator related variability,
- Preparation methodology related variability,
- Equipment related variability, and
- Mix properties.

Present study shed some light on the relative importance of the first three sources of variability. The compactability of the mix is a source of variability that was not investigated and can only be investigated based on large sets of data (e.g. through proficiency testing).

The Marshall compactor related variability can be further broken down into:

- Variability in the weight of hammer,
- Variability in the height of free fall,
- Variability in the friction between rod and hammer,
- Variability in the base type and foundation,
- Variability in the mould restraint (rotating vs fixed), and
- Variability in the alignment of hammer.

During present study, differences in the speed of compaction, were identified as a possible source of variability.

It is not practically feasible to eradicate all sources of variability. In South Africa the “penny test” is used as a tool to reduce the variability of compaction hammers. The penny test however, only evaluates the pedestal reaction and does not address other sources of variability.

The use of different hammers will always result in different densities at a given number of blows. This is recognized in the current American standard ASTM D 6926-04 (p1) which states: *Although the mass and height of mass drop for each apparatus are the same, density achieved in compacted specimens with the same number of blows will be different. It is up to the user to establish the specific required number of blows to be used for compaction of the specimen in relation to the field.* Such calibration of all hammers would be required for every new mix and every compactive effort.

Adjusting the number of blows is the only known effective method of calibrating Marshall compactors. Calibration for a given mix can be achieved using the following procedure: Several samples of the mix are compacted with first compactor at the required compaction effort (e.g. 75 blows Marshall). The average BRD of the specimens is considered the target BRD. Specimens of the same mix are then prepared with the second hammer using a range of compactive efforts. The BRD values of these samples are plotted against the number of blows applied. The required number of blows to get to the target BRD can now be determined from this plot.

An alternative method to calibrate Marshall compaction equipment was developed by Cassidy et al (1994). This method involves the electronic measurement of the impulse of the blow of the hammer, followed by adjustment of the number of blows to get to a standard compaction effort. The published results from the development of devise and field validation (Cassidy, 1995) look promising, but it is unclear whether this method was implemented on a large scale.

5. CONCLUSIONS AND RECOMMENDATIONS

The overriding conclusion is that the Marshall design method is subject to considerable multi-laboratory variability and that this should be taken into account during the design phase as well during quality control and product verification.

The multi-laboratory coefficients of variation for air voids and stability and flow are of such an extent that they can create dispute when results from different laboratories are used for validation, or product acceptance.

Note that replacing the Marshall compaction hammer with the gyratory compactor will not provide a solution. A recent study of precision estimates of different compaction equipment

by Azari et al (2007) shows that the inter-laboratory variability for air voids is 1.08 per cent for the Marshall hammer and 1.01 for the Gyratory compactor.

The requirement in the ASTM D6926-04, that compactors need to be calibrated for every project, is absent from the procedure in TMH1 1986. The assumption underlying the local use of the method, that every hammer yields the same compaction effort, is clearly flawed.

It is therefore recommended that:

- Marshall compaction hammers from different parties involved in a project be calibrated for every mix design using the method described in Section 4,
- The procedures in TMH1 be updated to reflect the calibration method. A precision statement with regards to the method should also be included, to highlight the limitations of the method,
- A national proficiency testing scheme be implemented to allow a better picture of multi-laboratory precision and to allow laboratories to correct their processes. From the results it would appear that especially the multi-laboratory precision of Marshall stability may benefit from proficiency testing.

6. ACKNOWLEGEMENTS

The authors wish to express their sincere appreciation to the Gauteng Department of Public Transport Roads and Works (GDPTRW) for sponsoring present study. Thanks are also extended to the representatives of the laboratories participating in the study for their cooperation and input.

7. REFERENCES

[1] ASTM D4123-82, 1982. Standard method of indirect tension test for resilient modulus of bituminous mixtures. ASTM international, West Conshohocken, USA.

[2] ASTM D 6926-04, 2004. Standard practice for preparation of bituminous specimens using Marshall apparatus. ASTM international, West Conshohocken, USA.

[3] Azari, H., Lutz, B., Spellerberg, B., 2007. Precision estimates for AASHTO test method T 269 determined using AMRL proficiency sample data. NCHRP web-only document 114, AASHTO materials reference laboratory, Gaithersburg, USA.

[4] Cassidy, M.M., Spellenberg, P.A, Savage, D.A., 1994. A system for the calibration of the Marhall compaction hammer. Report No: NISTIR 5338. Prepared for the Federal Highway Administration, NISTIR, Maryland.

[5] Cassidy, M.M., 1995. Field evaluation of the system for calibration of the Marshall compaction hammer. Report No. NISTIR 5553, Prepared for Federal Highway Administration.

[6] Denneman, E., 2006. Validation of reference mix for use in the HMA research project. Contract Report CSIR/BE/IE/ER/2006/0077/B. Prepared for GDPTRW, CSIR, Pretoria, RSA.

[7] Denneman, E., Marais, M., 2007. Investigation of multi-laboratory precision of Marshall design related test results. Technical Memorandum CSIR/BE/IE/ER/2007/023B. Prepared for GDPTRW, CSIR, Pretoria, RSA.

[8] TMH1:1986. Technical methods for highways. Standard methods of testing road construction materials. 2nd ed. National institute for transport and road research, CSIR, Pretoria, RSA.