# NCAT REPORT 22-01

January 2022

## NCAT Performance Testing Round Robin

Adam J. Taylor, Jason Moore, Nathan Moore

NCAT Performance Testing Round Robin

NCAT Report 22-01

By

Adam J. Taylor, PE
Assistant Research Engineer

Jason Moore, PE
Lab Manager

Nathan Moore, PE
Assistant Research Engineer

National Center for Asphalt Technology
Auburn University, Auburn, Alabama
January 2022

**TABLE OF CONTENTS**

**DISCLAIMER**



**ACKNOWLEDGMENTS**

The authors gratefully acknowledge the following members of the NCAT Applications Steering Committee for their review of this technical report: Imad Al-Qadi, Tim Aschenbrener, Erv Dukatz, Cheng Ling, and Robert Rea.

The authors would also like to thank all of the laboratories that participated in this round robin study. Without their funding and participation, this work would not have been possible.

## EXECUTIVE SUMMARY

A round robin study was conducted by the National Center for Asphalt Technology (NCAT) to assist with the implementation of mixture performance testing for balanced mix design (BMD) efforts. The study was conducted to help participating labs benchmark their results against a large body of data, as well as to collect preliminary data on the variability of the different tests. Four mixture performance tests were included in the study and a single plant-produced mixture was tested. The Hamburg Wheel-Track Test (HWTT) and the Asphalt Pavement Analyzer (APA) were included as tests for evaluating rutting resistance. The Illinois Flexibility Index Test (I-FIT) and the Indirect Tensile Asphalt Cracking Test (IDEAL-CT) were included as tests to assess cracking resistance. Forty-one unique labs participated in this round robin study, with some participating in multiple tests. For each test, participating labs were sent a sample of loose mix and detailed instructions to fabricate and test the necessary specimens. An additional phase was added to include cracking tests where each participating lab received specimens for testing that had been fabricated in the NCAT lab. The ASTM E691 procedure was used to develop limited, preliminary variability estimates for the data that were collected in this study.

# 1 OBJECTIVES

In 2018, NCAT initiated a round robin study for mixture performance tests being considered for balanced mix design (BMD) implementation. There were two primary objectives for this study. The first objective was to provide a basis of comparison for new performance test users. Having guidance for performing a test and the opportunity to assess their results compared to results from other labs was intended to help them gain experience and confidence in their testing abilities. The second objective was to collect variability data on mixture performance tests being considered by multiple agencies as part of BMD implementation efforts. This round robin study would generate data to help users with understanding the variability (both within lab and between labs) for various performance tests.

# 2 SCOPE

The NCAT round robin was performed on a single mix with four laboratory performance tests: the Asphalt Pavement Analyzer (APA), Hamburg Wheel-Track Test (HWTT), Illinois Flexibility Index Test (I-FIT), and the Indirect Tensile Asphalt Cracking Test (IDEAL-CT). The NCAT study was conducted in two phases. The first phase (Phase I) was the originally planned study where the participating labs fabricated and tested the specimens for the various tests in their labs using plant produced mix provided by NCAT. Participating labs selected the tests they wanted to evaluate. Phase I included both rutting (APA and Hamburg) and cracking tests (I-FIT and IDEAL-CT). The second phase of the study (Phase II) was added to help assess the effect of specimen fabrication on the variability of the cracking tests only (I-FIT and IDEAL-CT). A flowchart describing the scope of the study is shown in Figure 1.

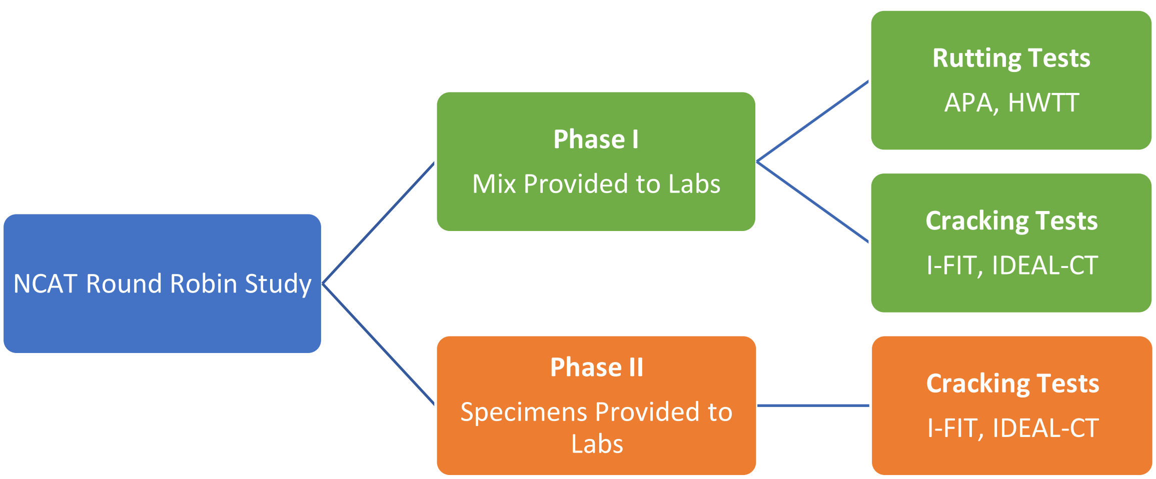



**Figure 1. Study Scope Flowchart**

# 3 METHODOLOGY

## 3.1 Mixture Sampling

The mix selected for this project was designed using a BMD approach with the APA and the IDEAL-CT. A BMD mix was selected because the research team wanted to prevent having a mix that would only give extreme results in the selected rutting and cracking tests. A mixture with extreme results (a very low rut depth or cracking index) would increase the likelihood of labs getting similar results and suppressing the true variability of the tests. The mixture was a 9.5 mm nominal maximum aggregate size (NMAS) blend containing 30% reclaimed asphalt pavement (RAP) and utilizing a PG 64-22 base binder. The mixture did not contain any reclaimed asphalt shingles (RAS) or rejuvenating additives.

A large volume of plant-produced mix was sampled for this round robin study (Figure 2). Two hundred five-gallon buckets of mix were sampled from a stockpile that had been passed through a material transfer vehicle for consistency. The buckets were sealed and stored at NCAT's main laboratory before being shipped to the participating laboratories.

**Figure 2. Mixture Stockpile Sampling**

### 3.2 Specimen Fabrication

During Phase I, each participating lab received enough mixture to fabricate specimens for their respective tests. The participating labs were sent detailed, test-specific instructions for specimen fabrication and testing along with a data file for reporting results back to NCAT. Participating labs were asked to return this summary data file along with the raw data files from testing in the event that odd results required further investigation. The detailed instructions and summary data file for the IDEAL-CT test are attached in Appendix A as an example.

For Phase II, each participating laboratory received specimens of the same mixture used in Phase I. These specimens were all fabricated at NCAT. Eight specimens were provided to participating labs for the I-FIT while five specimens were provided for the IDEAL-CT. The buckets of mixture used for Phase II were all homogenized using a quartermaster after re-heating, and the specimens were all prepared by the same technician using the same equipment (gyratory compactor, oven, wet saw, etc.). Each laboratory received a set of specimens with almost equal average air voids. Participating labs provided a summary of the testing results (referenced in Appendix A) as well as the raw data files back to NCAT for compilation. All testing for this study was performed on plant-produced mix in the re-heated condition with no additional aging.

## 3.3 Study Participation

An advertisement for participation in the NCAT round robin was sent out in late summer 2018. This advertisement offered five tests to participants (IDEAL-CT, I-FIT, Overlay Tester, APA, HWTT). Study participants included state highway agency labs, contractor labs, and research labs. The research team set a threshold of a minimum of six participating labs to move forward with each of the tests in the advertisement. This is the minimum number of labs recommended by ASTM E691-19 *Standard Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method*.

The participation threshold of a minimum six participating labs was met for all tests except the Overlay Tester. Hence, the OT was not included in the study moving forward. Table 1 shows a summary of the number of participating labs for each test for both phases. There was no minimum experience requirement for study participation in this round robin since the study objective was to encourage newer users to get experience running laboratory performance tests. In total, 41 unique labs participated in the study with several labs electing to perform multiple tests.

**Table 1. Number of Participating Labs**

| Test | Participating Labs – Phase I | Participating Labs – Phase II |
|---|---|---|
| HWTT | 32 | N/A |
| I-FIT | 20 | 13 |
| IDEAL-CT | 15 | 14* |
| APA | 10 | N/A |

* = Includes a second set of specimens tested by a single lab on a secondary load frame.

Mix samples and instructions were shipped to the study participants in early 2019. Specimens for Phase II were fabricated at NCAT and shipped to the participating labs in late summer 2019. Upon completion of each test and phase, a summary report was sent to the participating labs. These reports did not disclose laboratory names, but instead assigned each laboratory a unique number ID and revealed that number only to the participating lab submitting the data. That reporting scheme is also used in this report so that the identities of the participating labs are kept blind to the other participants.

## 3.4 Mixture Performance Tests

### *3.4.1 Hamburg Wheel-Track Test*

The Hamburg Wheel Track Test (HWTT or Hamburg) (Figure 3) was conducted per AASHTO T324-17. Specimens were loaded for a maximum of 20,000 passes with a 158-pound wheel load while submerged in a 50°C water bath. Participating labs were asked to fabricate four Hamburg specimens to a height of 62 mm tall with target air voids of 7.0 ± 0.5 percent. In the Hamburg, two specimens are trimmed and loaded together as a single replicate. Each lab was asked to test two replicates (four total specimens).

Several states have developed and implemented HWTT criteria (*1*). The majority of states specify a minimum number of passes (such as 10,000 or 20,000) to reach a defined failure threshold (commonly 12.5 mm) based on factors such as the grade of the virgin binder or traffic level. A few states also require their mixtures to reach a defined number of passes without exhibiting a stripping inflection point (SIP) – a quantity that is defined in AASHTO T324-19 which can be calculated as a measure of moisture resistance in the HWTT. Figure 3 shows the HWTT equipment and an example of the rut depth versus wheel passes data collected by the HWTT, including an example AASHTO SIP.

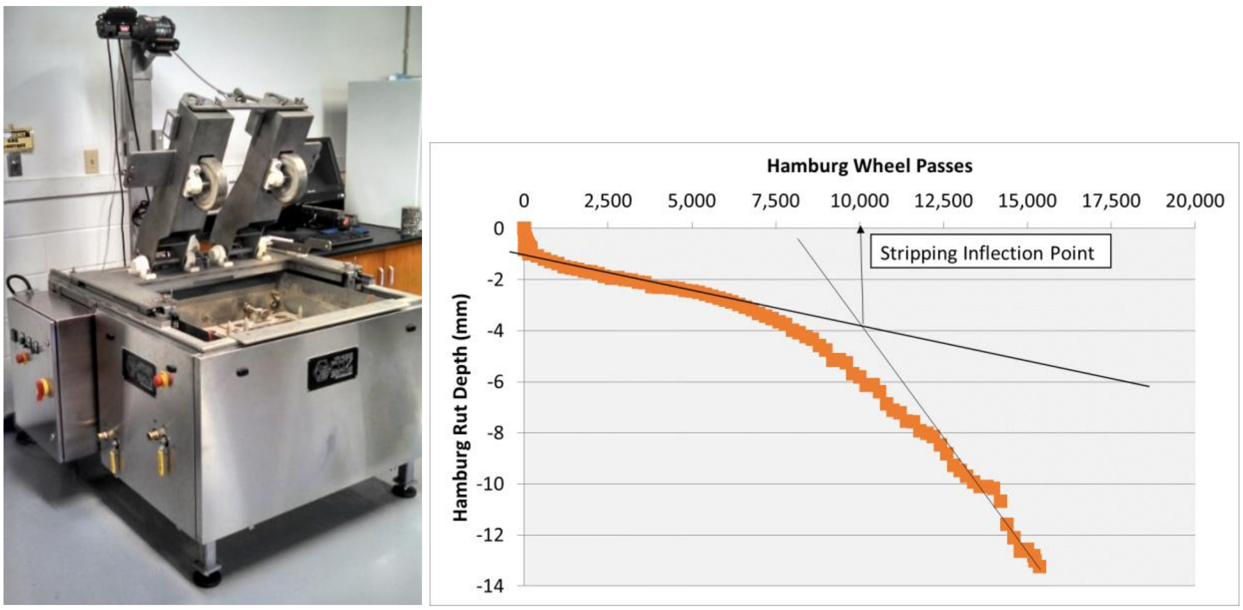
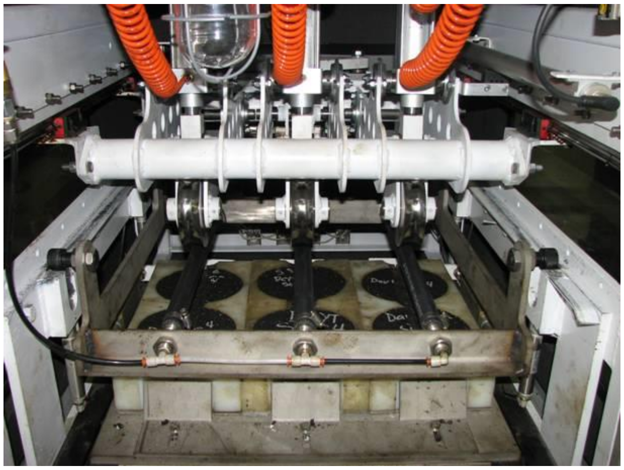



**Figure 3. Hamburg Wheel-Tracking Machine (left) and Example Data (right)**

### *3.4.2 Asphalt Pavement Analyzer (APA)*

The Asphalt Pavement Analyzer (APA) test was conducted per AASHTO T340-10 (2015). For this study, participating labs were asked to fabricate and test one full set of APA specimens prepared to a height of 75 mm and a target air void level of 7.0 ± 0.5 percent air voids. Depending on the age of their machine, this may have been either four or six replicates (the older model machines will accommodate six replicates while the newer model machines will hold four). The APA test applies a repeated load to the specimens via a loaded wheel atop inflated rubber hoses. For this study, the test temperature was set to 64°C and the wheel load and hose pressure were set to 100 lb. and 100 psi, respectively. The test was performed for 8,000 cycles. Participating labs were asked to provide raw data files along with both the manual (hand-measured) and automated (machine instrumentation measured) rut depths. However, several of the labs did not have the equipment required to make the manual measurements, so the study analysis is primarily based on the automated readings.

Several states have available APA criteria (*1*). These criteria vary from state to state but will typically specify a maximum APA rut depth based on a target test temperature and other factors: such as mix type and traffic level. Previous studies at the NCAT Test Track indicate that a manually read rut depth of less than 5 mm in the APA would yield a rut-resistant mix in the field (*2*). Figure 4 shows the APA equipment and an example of APA results.



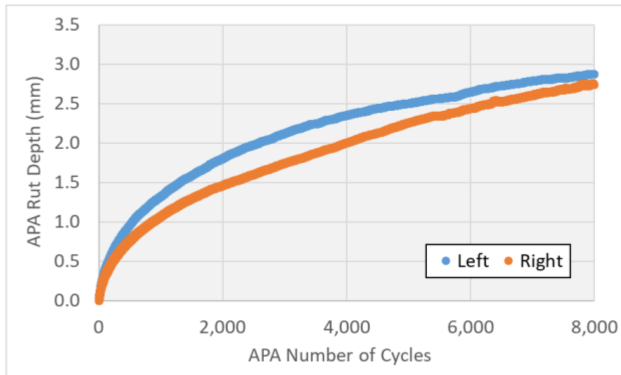



**Figure 4. Asphalt Pavement Analyzer (APA) Equipment (left) and Example Results (right)**

### *3.4.3 Illinois Flexibility Index Test (I-FIT)*

The Illinois Flexibility Index Test (I-FIT) was conducted per AASHTO TP 124-18. For Phase I, participating labs were asked to fabricate and test eight replicates prepared to the air void level of 7.0 ± 0.5 percent after saw trimming. For Phase II, eight replicates meeting the same air void criteria were fabricated at NCAT and shipped to the participating labs. For each semi-circular I-FIT specimen, a notch in the flat side of the specimen is cut at a depth of 15 ± 1.0 mm and width of 1.5 ± 0.5 mm. The specimens meeting the air void tolerance were conditioned either in an environmental chamber or water bath for two hours at 25°C before testing. The specimens are loaded monotonically at a rate of 50 mm/min until fracture while a plot of specimen load versus displacement is generated. The NCAT test setup as well as examples of the raw data generated during the test are shown in Figure 5 below.

Flexibility index (FI) is an index used as a measure of mixture cracking propagation resistance. The FI is essentially the area under the load-displacement curve (fracture energy) divided by the slope at the curve inflection point post-peak. The slope is related to the speed of the crack propagation. Mixtures with a higher FI are considered more cracking resistant than mixtures with a lower FI. The FI calculation is shown as Equation 1. For consistency, participants were instructed to calculate the FI using the software available from the Illinois Center for Transportation (ICT). The Illinois DOT originally recommended a minimum FI criteria of 8 for AC surface mixes without long-term aging (*1, 3*). However, state-specific FI criteria are likely needed to be more representative of mixtures in different climates.

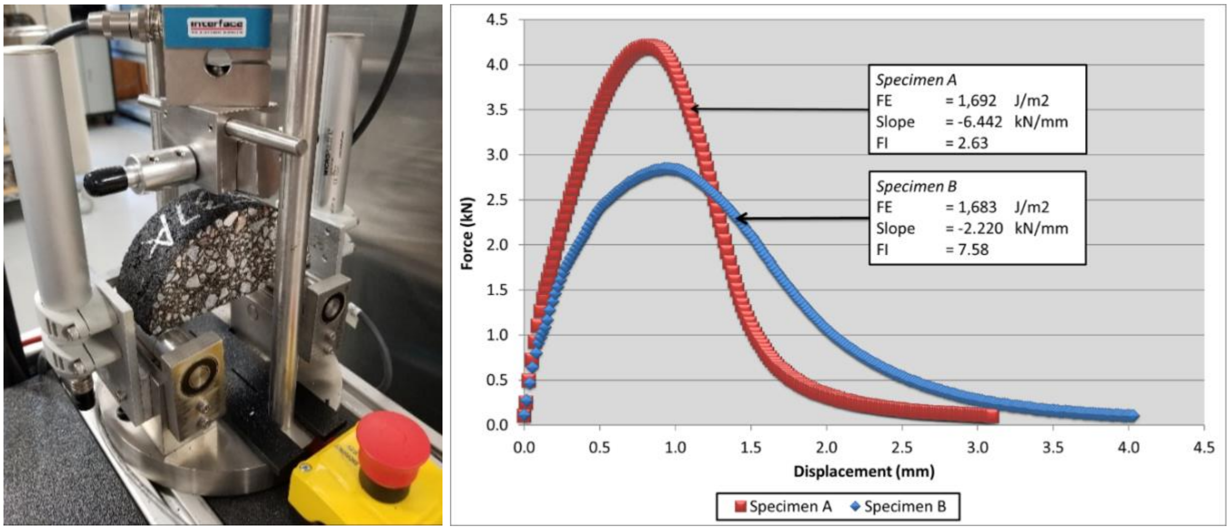



**Figure 5. I-FIT Test Setup (left) and Example Raw Data (right)**

$$FI = \frac{G_f}{|m|} \, x \, A \qquad (1)$$

Where:

- $G_f$ = fracture energy (J/m$^2$);
- $FI$ = flexibility index;
- $m$ = post-peak slope (kN/mm); and
- $A$ = scaling factor (0.01 for gyratory specimens).

### *3.4.4 Indirect Tensile Asphalt Cracking Test (IDEAL-CT)*

The Indirect Tensile Asphalt Cracking Test (IDEAL-CT) was performed per ASTM D8225-19. The NCAT IDEAL-CT test setup is shown in Figure 6. For Phase I, each participating lab was asked to fabricate and test a minimum of five 62 mm tall gyratory specimens prepared to a target air void level of 7.0 ± 0.5 percent. For Phase II, five specimens meeting the desired air void tolerance were fabricated at NCAT and shipped to the participating labs. In the IDEAL-CT, specimens are loaded monotonically in indirect tension at a rate of 50 mm/min until failure while load line displacement (LLD) is recorded. A plot of load versus LLD is generated for each specimen and is then analyzed to determine the $CT_{Index}$ (Figure 6).

The $CT_{Index}$ equation from ASTM D8225-19 is shown as Equation 2 below. Three major parameters factor into the calculation of the $CT_{Index}$. The area under the load-displacement curve ($G_f$) and the post-peak slope $|m_{75}|$ both factor into the results. The slope value for the $CT_{Index}$ is fixed at 75% of the peak load after the peak. Additionally, the $CT_{Index}$ calculation also includes the $l_{75}$ parameter (the displacement of the specimen at 75 percent of the peak load after the peak). A higher $G_f$ and $l_{75}$ would increase the $CT_{Index}$ while a higher $|m_{75}|$ would lower the $CT_{Index}$. A higher $CT_{Index}$ is generally representative of increased mixture cracking resistance. The Virginia Department of Transportation is currently proposing to use a minimum $CT_{Index}$ of 70 for the design of surface mixes (no long-term oven aging) using balanced mix design (*4*).

$$\mathrm{CT_{Index}} = \frac{t}{62} x \frac{l_{75}}{D} x \frac{\mathrm{G_f}}{|m_{75}|} \mathrm{x}\ 10^6 \qquad (2)$$

Where:

$CT_{Index}$ = cracking tolerance index;

$G_f$ = fracture energy (J/m$^2$);

$|m_{75}|$ = absolute value of the post-peak slope $m_{75}$ (N/m);

$L_{75}$ = displacement at 75% of the peak load after the peak (mm);

$D$ = specimen diameter (mm); and

$t$ = specimen thickness (mm).

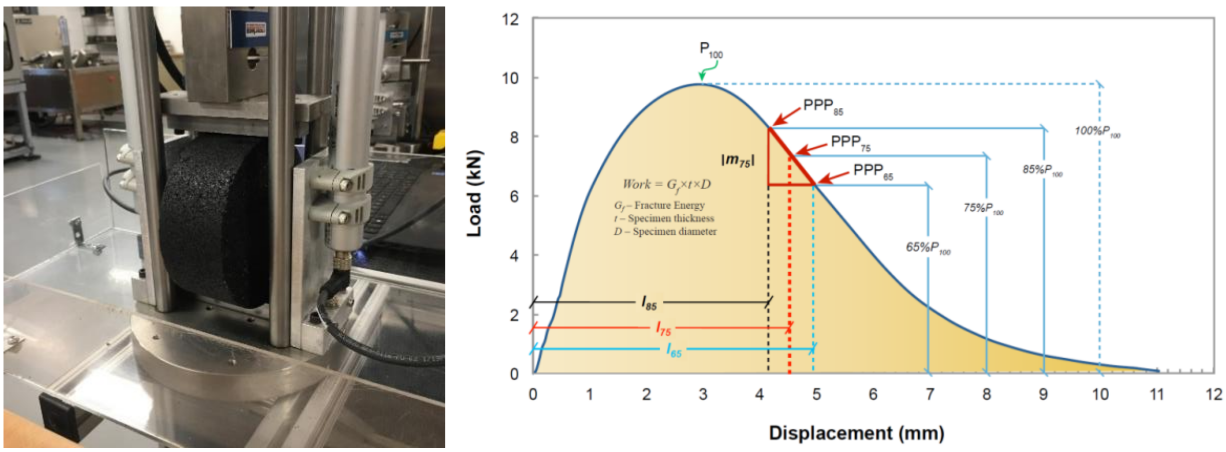



**Figure 6. IDEAL-CT Test Setup (left) and Plot of Load vs. LLD (*5*)**

### 3.5 Variability Analysis

Test results were inspected for reasonableness and uploaded to a database when received from the participating labs. Replicates with data quality issues were evaluated and removed from the database on a case-by-case basis. Additionally, the individual lab results were examined for statistical outliers per the procedure outlined in ASTM E178-16a *Standard Practice for Dealing with Outlying Observations.* Statistical outliers (90% confidence) that were screened using this process were not included in the final database. A limited number of outliers were removed from the cracking test results using this method. For the IDEAL-CT, only one total replicate from each phase failed this criteria and was removed. For the I-FIT, three replicates failed this criteria during Phase I and no replicates failed this criteria for Phase II.

The data collected from the round robin study were used for test variability analysis. ASTM E691-19 offers a methodology for calculating both the within-lab (single-lab) and between-lab (multiple-lab) coefficients of variation (CV). It should be noted that ASTM E691-19 recommends between three and six materials (mixes in this case) be used to develop precision statements and this study includes only one mix. However, the data can still be used to provide preliminary estimates of test variability. Hence, any ASTM E691 within-lab and between-lab variability estimates given in this report should be considered as preliminary and specific to the single mixture that was tested.

# 4 TEST RESULTS AND ANALYSIS

## 4.1 Hamburg Wheel-Tracking Test (HWTT)

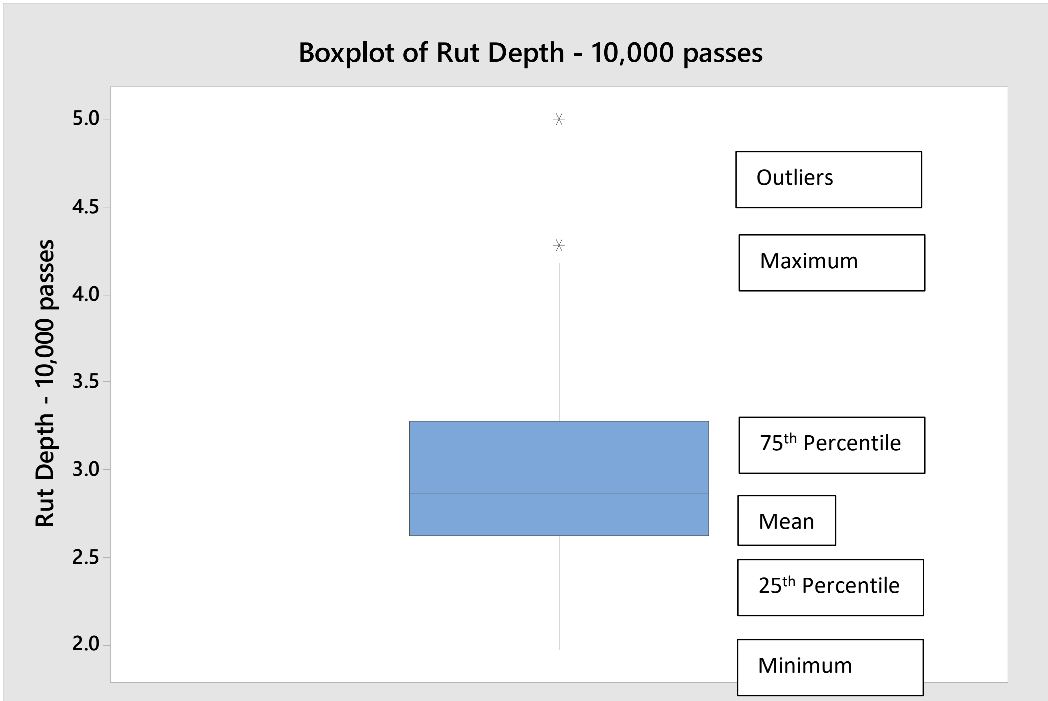
Thirty-two labs participated in the Phase I Round Robin study for the Hamburg Wheel-Tracking Test using equipment from four different machine manufacturers. Each lab reported the maximum rut depth profile from their machine for each replicate tested and the results from those replicates were averaged for data analysis. The average results from all 32 of the individual labs are shown in Appendix B. Summary statistics for all 32 labs are summarized in Table 2 at multiple rut depth benchmarks (2,500, 5,000, 7,500, 10,000, 15,000, and 20,000 passes). The overall average rut depth for the selected mixture was approximately 4 mm at 20,000 passes with values ranging from a low of 3.1 mm to a high of 8.4 mm. Again, this was a mix designed with BMD and high rut depths were not expected. Hence, none of the test results exceeded the common Hamburg failure criteria of 12.5 mm rut depth at 20,000 passes.

**Table 2. Summary Statistics for HWTT Rut Depths (mm) - Phase I - All Participating Labs**

| Variable | N | Mean | SE Mean | St Dev | Min | Q1 | Median | Q3 | Max |
|---|---|---|---|---|---|---|---|---|---|
| Rut Depth - 2,500 passes | 32 | 2.12 | 0.07 | 0.40 | 1.08 | 1.86 | 1.99 | 2.33 | 3.07 |
| Rut Depth - 5,000 passes | 32 | 2.52 | 0.09 | 0.50 | 1.31 | 2.23 | 2.41 | 2.78 | 3.76 |
| Rut Depth - 7,500 passes | 32 | 2.81 | 0.10 | 0.59 | 1.62 | 2.47 | 2.66 | 3.07 | 4.41 |
| Rut Depth - 10,000 passes | 32 | 3.05 | 0.12 | 0.67 | 1.97 | 2.62 | 2.87 | 3.28 | 5.01 |
| Rut Depth - 15,000 passes | 32 | 3.45 | 0.16 | 0.92 | 2.34 | 2.86 | 3.13 | 3.64 | 6.22 |
| Rut Depth - 20,000 passes | 32 | 3.91 | 0.25 | 1.42 | 2.53 | 3.10 | 3.37 | 4.01 | 8.42 |

Where: N = Number of Labs
SE Mean = Standard Error of the Mean
St. Dev = Standard Deviation
Min = Minimum Value
Q1 = First Quartile
Q3 = Third Quartile
Max = Maximum Value

Figures 7 and 8 below graphically show the variability and distribution of the HWTT rut depths at 10,000 and 20,000 passes, respectively. For each measurement, a boxplot was utilized to show the distribution of results and a histogram was used to show the frequency at which results were obtained (i.e. which results were most common and least common). The critical components of the boxplots (minimum and maximum values, mean, quartiles, and outliers) are labeled within the plots shown in Figure 7. At 10,000 passes, 2 of the 32 labs (6.3 percent) were shown as statistical outliers by the boxplot analysis - values more than 1.5 times the interquartile range (IQR) from the mean. At 20,000 passes, 4 of the 32 labs (12.5 percent) were shown as outliers. In each case, these outlier values were on the high side of the range. For the rut depth at 20,000 passes, the IQR was 0.9 mm. This indicates that half of the participants reported final rut depths in a very tight range of less than 1 mm.



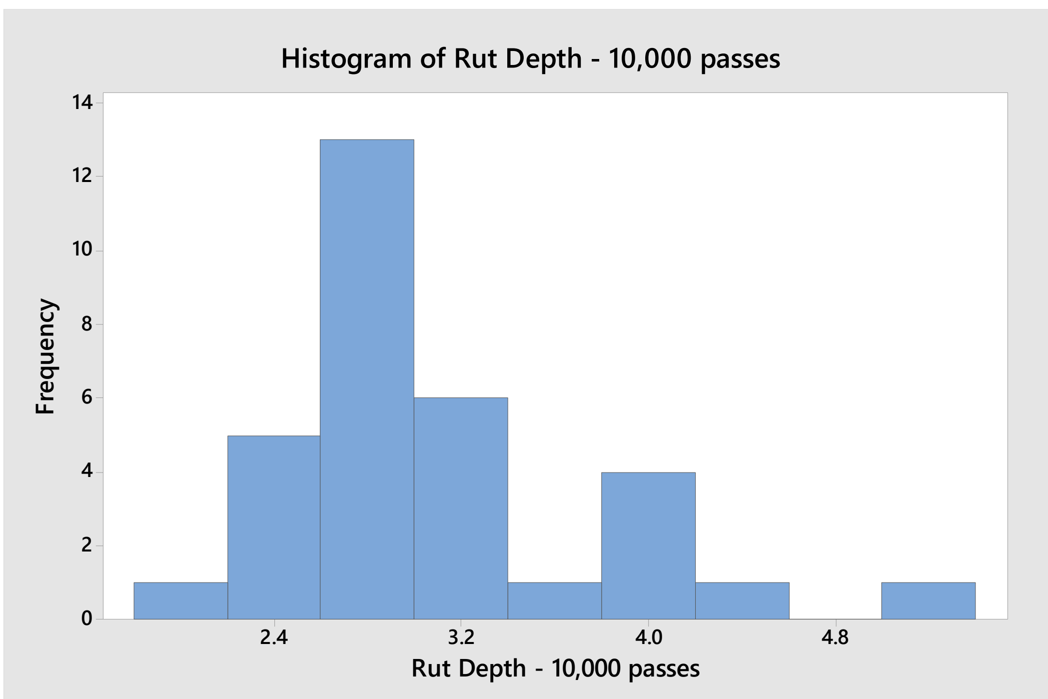
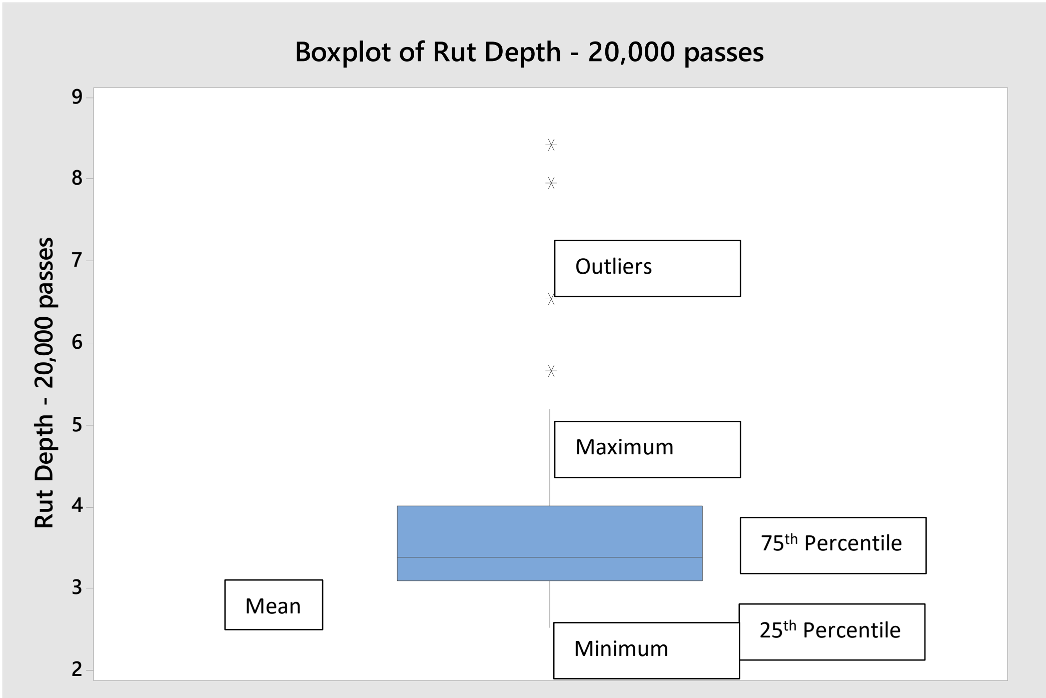




**Figure 7. Boxplot (top) and Histogram (bottom) of HWTT Rut Depths (mm) at 10,000 Passes**



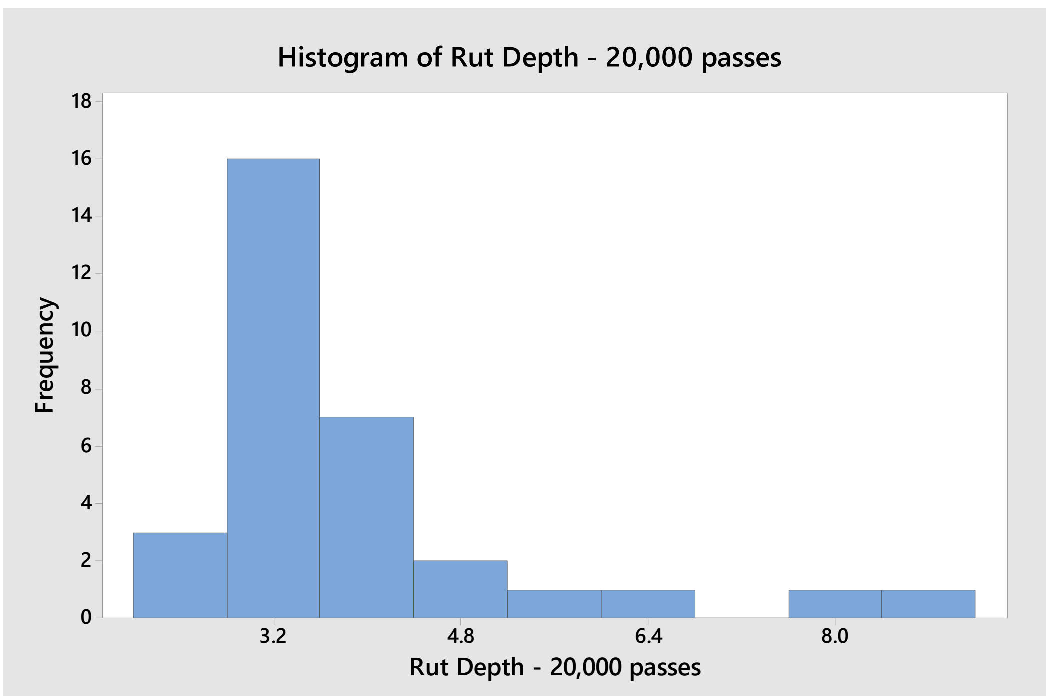




**Figure 8. Boxplot (top) and Histogram (bottom) of HWTT Rut Depths (mm) at 20,000 Passes**

ASTM E691-19 was utilized to estimate the within-lab and between-lab coefficients of variation (CV) at both 10,000 and 20,000 wheel passes for the HWTT (Table 3). The within-lab variability was estimated using the individual wheel-tracks or replicates reported by each lab and the between-lab variability was estimated by looking at the overall variability in the rut depth database of 32 labs. The within-lab CV was very similar at both 10,000 and 20,000 wheel passes– approximately 9 percent for both. The between-lab CV did increase from approximately 21 percent at 10,000 wheel passes to approximately 26 percent at 20,000 wheel passes. Note that the ASTM estimate included 29 labs. One lab was excluded due to noting an LVDT issue and the other two failed the ‘h’ statistic (the between-lab consistency statistic in the ASTM E691 analysis). The underlying cause of this was determined to be late-test stripping (SIP between 15,000 and 20,000 passes) for those two labs. In general, the Hamburg test showed good within-lab repeatability and reasonable between-lab repeatability in this round robin study.

**Table 3: ASTM E691-19 Within-Lab and Between-Lab Estimates for Coefficient of Variation (CV) - HWTT Rut Depth (29 labs)**

| Hamburg Wheel Passes | Mean Rut Depth (mm) | Within-Lab CV (%) | Between-Lab CV (%) |
|---|---|---|---|
| 10,000 | 2.91 | 9.0 | 21.1 |
| 20,000 | 3.53 | 9.4 | 25.9 |

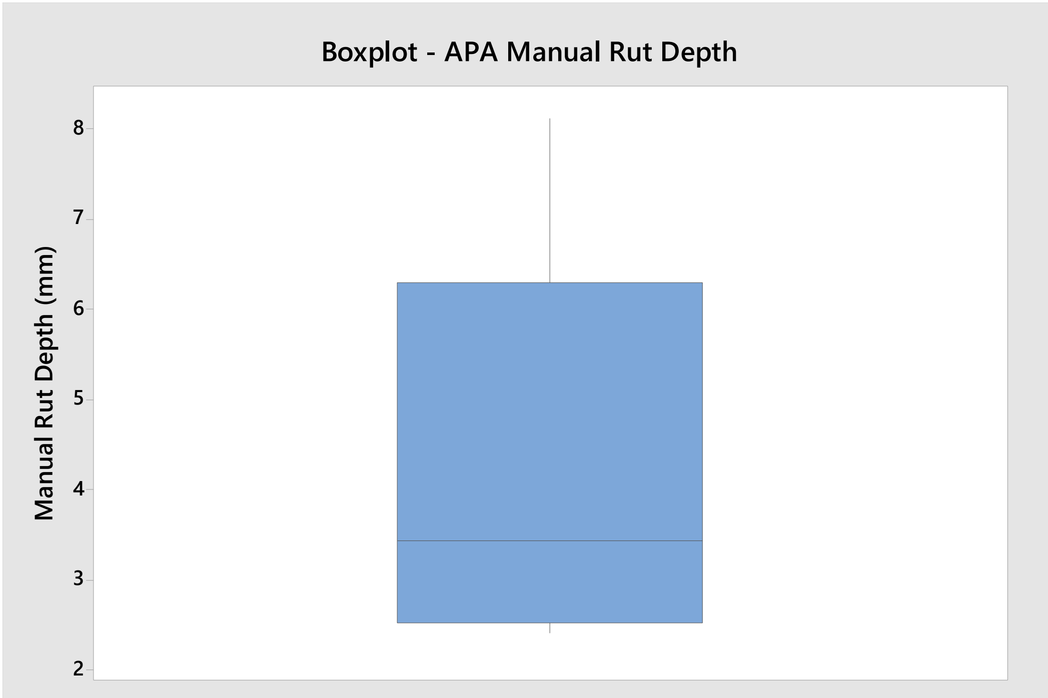
### 4.2 Asphalt Pavement Analyzer (APA)

Ten labs participated in the Round Robin study for the asphalt pavement analyzer (APA). Labs were asked to provide both manual (caliper) and automated (machine) rut depths for the specimens tested. Several labs reported that they did not have the equipment to test the manual rut depth measurements for their particular model of APA. Additionally, one lab reported an equipment issue with their automated rut depth measurements and was only able to provide the manual rut depths. In total, nine labs were able to provide automated rut depth measurements and five labs were able to provide manual rut depth measurements for their set of specimens.

The summary statistics for both the automated and manual rut depth measurements are shown in Table 4. The average results from all 10 of the individual labs are shown in Appendix B. Figures 9 and 10 show the boxplots and histograms for the manual and automated rut depths, respectively. The majority of the results for both the manual and automated rut depths fell between 2 and 4 millimeters. It should be noted that the spread of the data in the boxplots (Figures 9 and 10) appears to show the manual rut depths to have a much wider range than the automated rut depths. This is being driven by data from a single lab that reported a higher manual rut depth result but was unable to report an automated rut depth result. The manual rut depth data from that lab were repeatable within that dataset, but their results were significantly higher than the rut depths from the other participating labs.

**Table 4. Summary Statistics for APA Rut Depths (mm) - Phase I - All Participating Labs**

| Rut Depth Type (mm) | N | Mean | SE Mean | St Dev | Min | Q1 | Median | Q3 | Max |
|---|---|---|---|---|---|---|---|---|---|
| Automated | 9 | 2.91 | 0.25 | 0.75 | 2.20 | 2.30 | 2.81 | 3.23 | 4.55 |
| Manual | 5 | 4.21 | 1.04 | 2.33 | 2.40 | 2.52 | 3.43 | 6.29 | 8.13 |



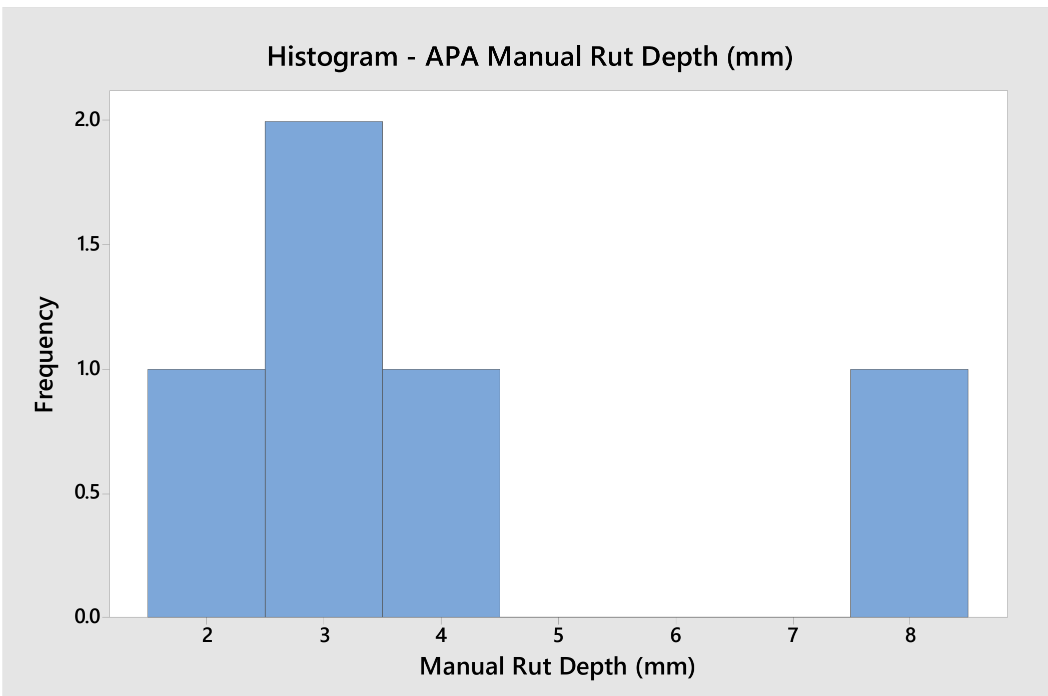
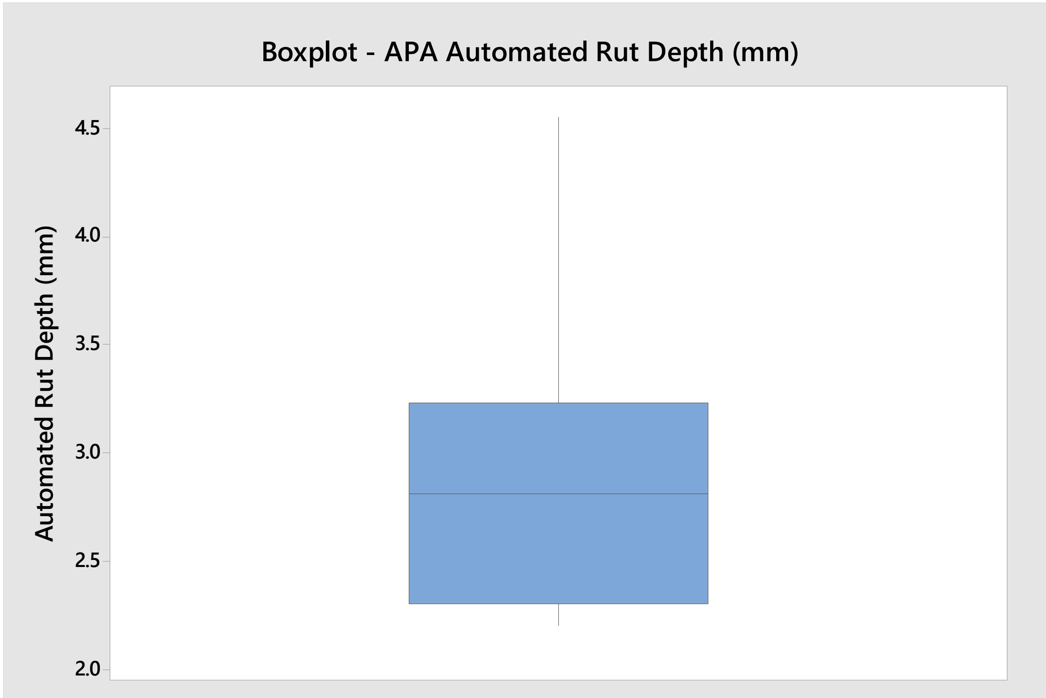




**Figure 9. Boxplot (top) and Histogram (bottom) of APA Manual Rut Depths (mm)**



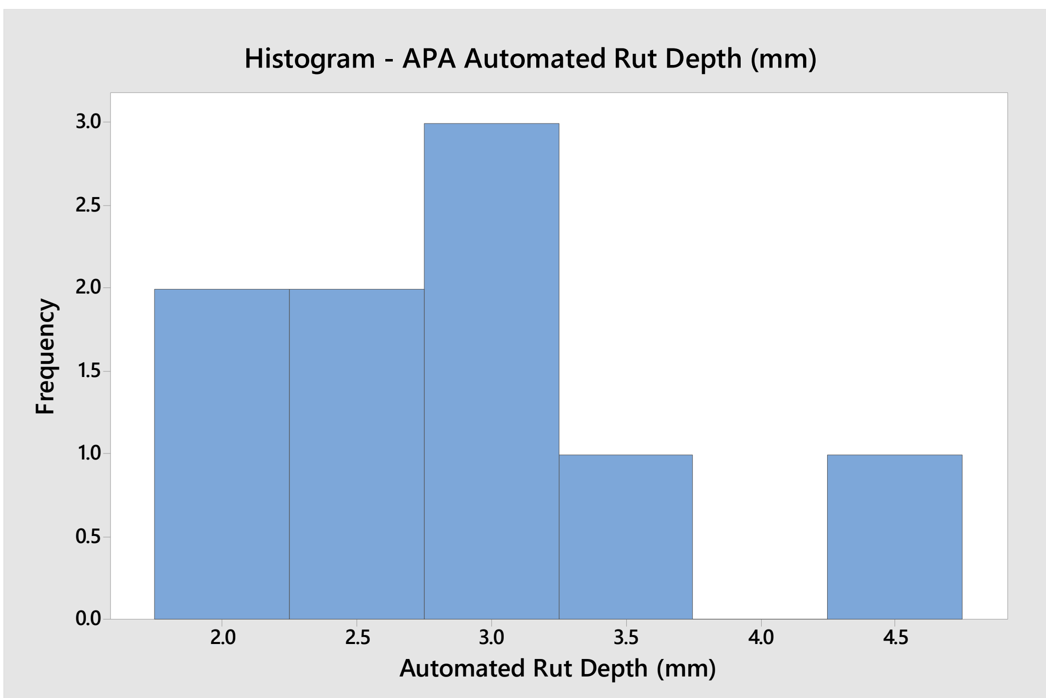




**Figure 10. Boxplot (top) and Histogram (bottom) of APA Automated Rut Depths (mm)**

Table 5 shows the ASTM E691-19 within-lab and between-lab variability estimates for the APA automated rut depth measurements. This calculation was not performed for the manual rut depths since less than the requisite six labs' worth of data were available. The APA had a within-lab CV of less than 20 percent (18.3%) and a between-lab CV of just under 30 percent (29.6%) for the measurement. These values were above what was obtained for the Hamburg, albeit with only about a third of the participating labs (29 for Hamburg versus 9 for APA).

**Table 5: ASTM E691-19 Within-Lab and Between-Lab Estimates for Coefficient of Variation (CV) – APA Automated Rut Depth (9 labs)**

| APA Measurement | Mean APA Rut Depth (mm) | Within-Lab CV (%) | Between-Lab CV (%) |
|---|---|---|---|
| Automated Rut Depth (mm) | 2.9 | 18.3 | 29.6 |

### 4.3 Illinois Flexibility Index Test (I-FIT)

Twenty labs participated in Phase I (specimens fabricated by participating labs) while twelve labs participated in Phase II (specimens provided to participating labs) for the I-FIT. The FI summary statistics for both Phase I and Phase II are shown in Table 6. The average results of the individual labs are shown in Appendix B.

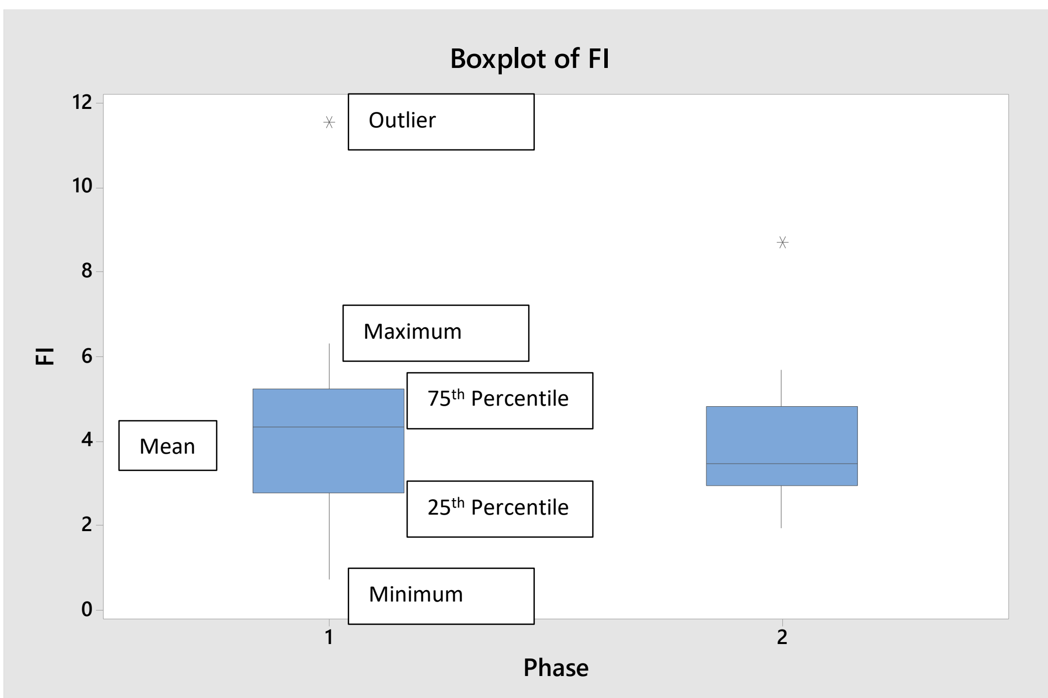
**Table 6: Summary Statistics – I-FIT Flexibility Index – Phase I and II**

| Variable | Phase | N | Mean | SE Mean | StDev | Min | Q1 | Median | Q3 | Max |
|---|---|---|---|---|---|---|---|---|---|---|
| FI | I | 20 | 4.33 | 0.51 | 2.28 | 0.75 | 2.79 | 4.35 | 5.24 | 11.52 |
| FI | II | 12 | 4.04 | 0.51 | 1.77 | 1.95 | 2.96 | 3.46 | 4.81 | 8.67 |

A boxplot and histogram are shown in Figure 11 to illustrate the spread of the data and the frequency of the different results received. These figures show a reduction in the spread of the data between Phase I and Phase II. This can be attributed to the effect of specimen fabrication being performed in a single laboratory (Phase II) relative to the variation inherent to specimens being fabricated by multiple operators in multiple laboratories (Phase I). Table 6 shows the standard deviation of the overall dataset was reduced from 2.28 to 1.77 – a 22% reduction – from Phase I to Phase II. Figure 11 does show one lab in each Phase I and Phase II to be an outlier relative to the main data set.

For this study, it was notable that the within lab coefficient of variation (CV) for flexibility index (FI) tended to be significantly higher than typical. NCAT experience has been that the within lab CV averages between 20 and 30 percent when an untrimmed (no replicates removed) data set is utilized. For both study phases, several labs had within lab CV values significantly above this threshold. This was driven by having multiple specimens in each set of eight have an FI fall within the expected range for this mixture (FI between 2 and 6) and having at least two specimens have an FI below 1. This would prevent the low values from failing the ASTM E178 outlier screening. An example of this would be the FI raw data from Lab #3 in Phase I. The raw FI values from this set were as follows: 0.45, 1.11, 0.58, 0.83, 2.47, 3.65, 2.42, and 4.60. This data yielded a high CV of 76.1 percent for FI with no clear outliers. Similar data was seen among several of the participating labs across both phases of the study. After compiling the Phase I data, the research team envisioned that a Phase II study with tightly controlled sample fabrication would mitigate the high variability. While Phase II did reduce the overall spread of the FI data between the labs, high CV values for FI within the labs were still prevalent in the Phase II data. The root cause of this is unknown and could be attributable to several factors, but the behavior occurring in several labs across both study phases suggests it may be behavior specific to this mix in the I-FIT test.

The research team conducted a small investigation to determine if the higher variability in some labs was equipment-related. For this study, the participating labs in the Phase I I-FIT testing used both servo-hydraulic and mechanical (screw-driven) load frames. A review of the database showed that sets with high variability were present in datasets from both the servo-hydraulic and mechanical (screw drive) I-FIT equipment used by the individual participating labs. The available raw data from the participating labs were evaluated to ensure the data sampling rate was appropriate (greater than 20 Hz) and the loading rate was within the recommended range of 50 +/- 1 mm per minute. The machines in this study largely met these criteria, though three labs conducted the testing on older load frames and their raw data did not provide the needed time stamp for rate verification. The same type of device was used by 10 of the participating labs in Phase I, and this device was verified to have a data sampling rate and loading rate per AASHTO TP 124-18. This subset of FI data from a single device (dubbed 'Device A') was then compared with all of the data from Phase I. The boxplots in Figure 12 below show a similar spread and average of the dataset between the two groups (All devices versus just 'Device A'). Additionally, the CVs for the individual datasets from Device A were compared to those from the other devices used in the study. The average CV of FI within each lab from Device A was 37.5 percent while the average CV of FI within lab from 10 labs with other devices was 34.8 percent. Hence, the data sets with higher variability did not seem to be equipment-related or associated with the use of any particular device. This suggests that the high FI variability seen in some labs is more related to the mixture used for this study than the testing equipment that was being used by the participating labs.



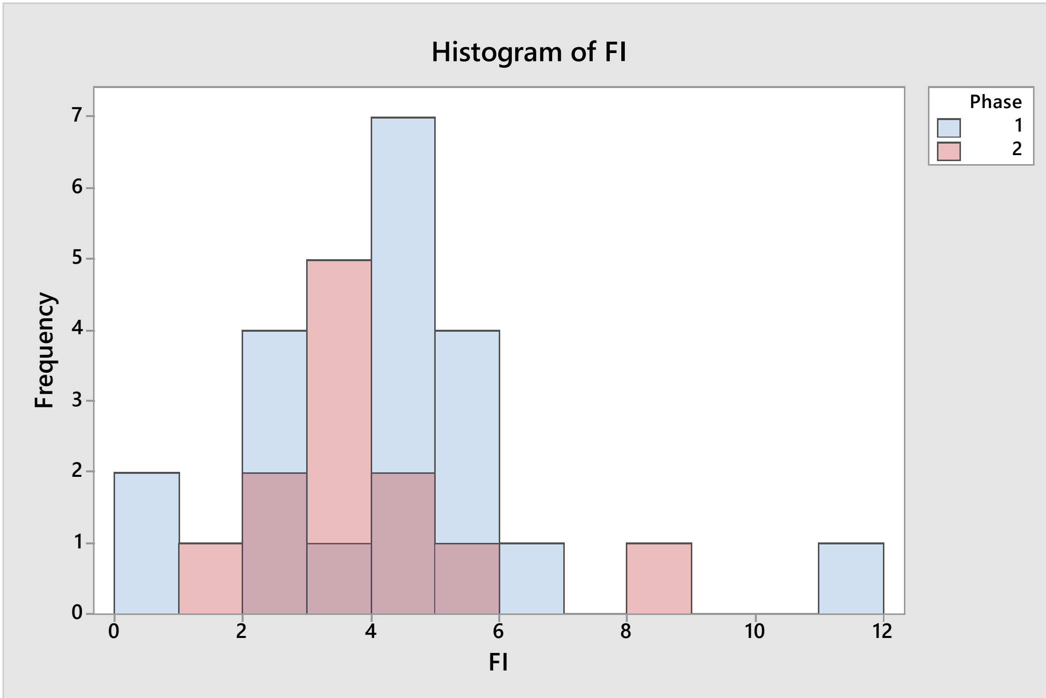



**Figure 11. Boxplot (top) and Histogram (bottom) - I-FIT Flexibility Index (FI)**

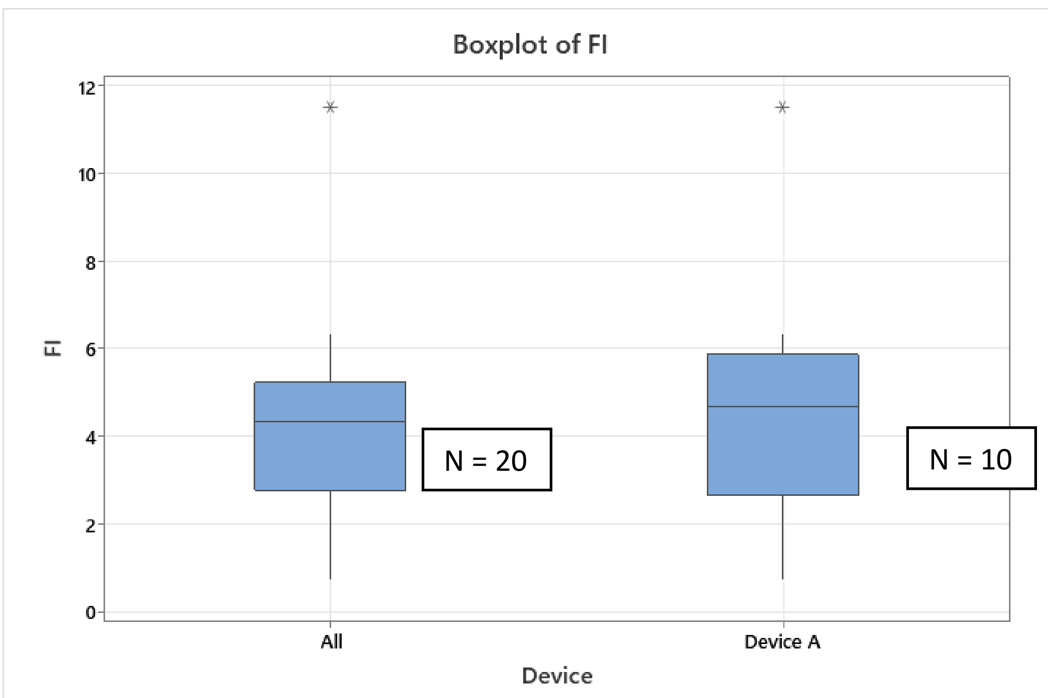



**Figure 12. Boxplot Comparison – I-FIT Phase I – Flexibility Index – Data from All Devices versus Data from a Single Device (Device A)**

The ASTM E691-19 within-lab and between-lab estimates for CV of FI are shown in Table 7, below. For both study phases, one outlier lab (shown in the Figure 11 boxplots) was removed for these calculations following ASTM E691-19 guidelines. Table 7 shows the repeatability (within-lab) CV was significantly higher for Phase II (47 percent) than Phase I (30 percent). This result was counterintuitive given the tight control of sample fabrication and may have been influenced by having a smaller number of labs participate in Phase II than Phase I. The reproducibility (between lab) CV values were similar for Phase I (48 percent) and Phase II (53 percent) with Phase II being slightly higher. This was unexpected given that the data in Figure 11 show Phase II to have a tighter data spread than Phase I. However, repeatability is a significant component of the reproducibility calculation. Hence, for Phase II, the reproducibility CV could not be less than the repeatability CV. The effect of multiple versus single lab specimen fabrication is shown when looking at the difference between the within lab and between lab CV for each phase. The between lab CV is 17 percent higher than the within lab CV for Phase I while this difference drops to around 6 percent for Phase II.

AASHTO TP 124-20 recently published precision estimates based on I-FIT data from three separate round robin studies conducted by IDOT. These precision estimates have a within-lab CV of 27.1 percent (which agrees with NCAT's experience with the I-FIT) and a between-lab CV of 34.1 percent for FI. The variability estimates from the NCAT round robin were significantly above these precision estimates. Further information and discussion regarding how the variability of the I-FIT test is related to fracture mechanics can be found in Al-Qadi et. al (*6*).

**Table 7: ASTM E691-19 Within-Lab and Between-Lab Estimates for Coefficient of Variation (CV) – I-FIT FI – Phase I and II**

| Phase | Number of Labs | Mean FI | Within-Lab CV (%) | Between-Lab CV (%) |
|---|---|---|---|---|
| I | 19 | 3.95 | 30.3 | 47.6 |
| II | 12 | 3.53 | 46.9 | 53.4 |

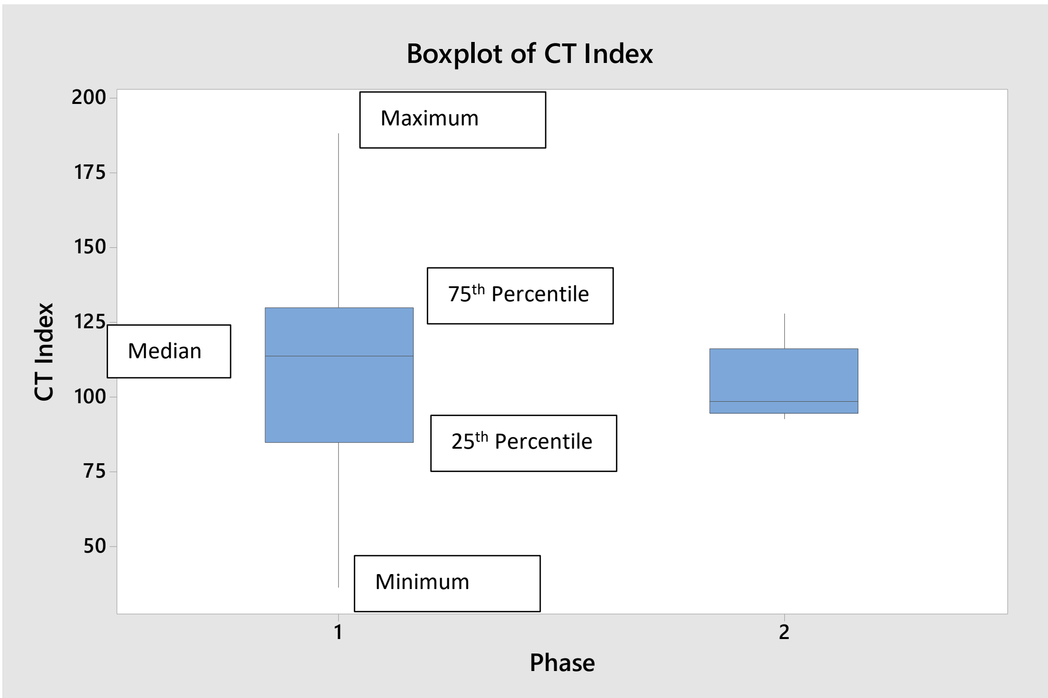
### 4.4 Indirect Tensile Asphalt Cracking Test (IDEAL-CT)

Fifteen labs participated in the Phase I study for IDEAL-CT while fourteen unique labs were able to participate in Phase II. It should be noted that one of the labs in Phase II is a second set of specimens tested on a second load frame by one of the participating labs (two sets of specimens, two separate load frames, one participating lab). This set was treated as a separate lab for analysis since it was a unique set of specimens tested on a separate load frame. The average $CT_{Index}$ results from the individual labs are provided in Appendix B.

The summary statistics for $CT_{Index}$ for both Phase I and Phase II are summarized in Table 8, while a boxplot and histogram of the data are shown in Figure 13. These figures clearly show a large reduction in the spread of the data between Phase I and Phase II. The data from Table 8 shows the standard deviation of $CT_{Index}$ was reduced by two-thirds going from Phase I to Phase II. This reduction in variability can be attributed to the effect of specimen fabrication being performed in a single laboratory (Phase II) relative to the variation inherent to specimens being fabricated by multiple operators in multiple laboratories (Phase I).

**Table 8: Summary Statistics – IDEAL-CT $CT_{Index}$ – Phase I and II**

| Variable | Phase | N | Mean | SE Mean | StDev | Min | Q1 | Median | Q3 | Max |
|---|---|---|---|---|---|---|---|---|---|---|
| $CT_{Index}$ | I | 20 | 111.1 | 9.2 | 35.6 | 36.5 | 84.9 | 113.7 | 130.0 | 188.0 |
| $CT_{Index}$ | II | 12 | 103.7 | 3.1 | 11.5 | 92.6 | 94.6 | 98.8 | 116.4 | 127.8 |



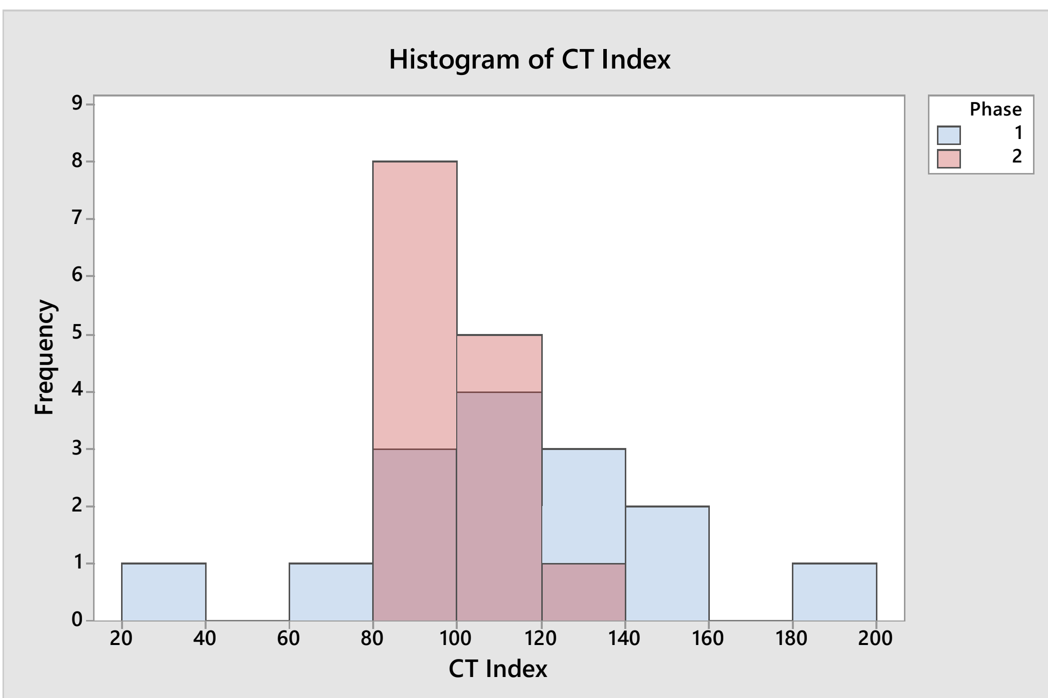




**Figure 13. Boxplot (top) and Histogram (bottom) – IDEAL-CT $CT_{Index}$**

The ASTM E691-19 within-lab and between-lab variability estimates for CV of $CT_{Index}$ are shown in Table 8. For the IDEAL-CT $CT_{Index}$, a similar within-lab CV of just under 20 percent was noted for both Phase I (samples prepared in participating labs) and Phase II (samples provided to participating labs). The within-lab variability also agrees with NCAT's experience with the IDEAL-CT. However, the between-lab CV for $CT_{Index}$ dropped from 35.3 percent to 20.2 percent when all of the specimens were fabricated in a single lab for Phase II. This highlights how impactful consistent sample fabrication practices (such as mix heating times and appropriate mix splitting techniques) can be on the variation of $CT_{Index}$ results.

**Table 8: ASTM E691-19 Within-Lab and Between-Lab Estimates for Coefficient of Variation (CV) – IDEAL-CT $CT_{Index}$ – Phase I and II**

| Phase | Number of Labs | Mean $CT_{Index}$ | Within-Lab CV (%) | Between-Lab CV (%) |
|---|---|---|---|---|
| I | 15 | 111.1 | 19.5 | 35.3 |
| II | 14 | 103.7 | 18.8 | 20.2 |

## 5 SUMMARY AND FUTURE WORK

Over 40 unique labs participated in the first NCAT round robin study regarding performance tests for balanced mix design. These labs prepared and tested samples on a single mix using a minimum of one of four laboratory performance tests (APA, HWTT, I-FIT, and IDEAL-CT). For each test, specimens were prepared and tested by the participating labs with a sample of loose mix provided by NCAT. For the cracking tests (I-FIT and IDEAL-CT), a second phase of the study was included so that the participating labs would also test a set of specimens prepared at NCAT.

For each test and phase, the within-lab and between-lab variability of each test under evaluation was estimated using the procedure recommended in ASTM E691-19. Those variability estimates are summarized in Table 9 below. It should be re-iterated, however, that these are estimates from a single study where a single mix was used for investigation. A summary of the test-specific results are as follows.

- The Hamburg rut depths (both at 10,000 and 20,000 passes) had a within-lab CV of less than 10 percent. The Hamburg rut depths at 10,000 and 20,000 passes had a between-lab CV of 21.1 and 25.9 percent, respectively.
    - It should also be noted that stripping was not observed in the labs included in the HWTT ASTM E691 variability estimates. A mixture that exhibited stripping would have likely increased the variability or necessitated the use of an alternate method of analysis.
- The APA had a within-lab CV of 18.3% and a between-lab CV of 29.6% for the automated rut depth measurement. These values were above what was obtained for the Hamburg, albeit with only about a third of the participating labs (29 for Hamburg versus 9 for APA).
- The variability for the I-FIT Flexibility Index (FI) was significantly higher than expected for both phases of the study. AASHTO TP 124-20 recently published precision estimates based on I-FIT data from three separate round robin studies. These precision estimates have a within-lab CV of 27.1 percent (which agrees with NCAT's experience with the I-

FIT) and a between-lab CV of 34.1 percent for FI. The variability estimates from the NCAT round robin study were above these precision estimates.

  - For the mixture used in this study, several of the labs reported replicates in the expected range (FI between 2 and 6) along with multiple replicates with an FI below 1. This caused several of the labs to report very high CV values for both phases, driving the high variability. The root cause of this is unclear. The research team hypothesizes that it may be a mixture-specific issue in the I-FIT test based on the consistency of the issue across multiple labs and different testing devices plus the lack of experience with such high variability in previous I-FIT testing at NCAT.

- For the IDEAL-CT $CT_{Index}$, a similar within-lab CV of just under 20 percent was noted for both phases. The within-lab variability also agrees with NCAT's experience with the IDEAL-CT. However, the between-lab CV for $CT_{Index}$ dropped from 35.3 percent to 20.2 percent when all of the specimens were fabricated in a single lab for Phase II. This highlights the importance of consistent sample fabrication on $CT_{Index}$ results.

**Table 9: ASTM E691-19 Within-Lab and Between-Lab Variability Estimates – NCAT Round Robin (Single Mixture)**

| Test ID | Phase | Participating Labs | Test Parameter | Average Result | Within-Lab CV (%) | Between-Lab CV (%) |
|---|---|---|---|---|---|---|
| Hamburg | I | 29 | Rut Depth – 10,000 passes (mm) | 2.91 | 9.0 | 21.1 |
| Hamburg | I | 29 | Rut Depth – 20,000 passes (mm) | 3.53 | 9.4 | 25.9 |
| APA | I | 9 | Auto. Rut Depth (mm) | 2.9 | 18.3 | 29.6 |
| IDEAL-CT | I | 15 | $CT_{Index}$ | 111.1 | 19.5 | 35.3 |
| IDEAL-CT | II | 14 | $CT_{Index}$ | 103.7 | 18.8 | 20.2 |
| I-FIT | I | 19 | Flexibility Index (FI) | 3.95 | 30.3 | 47.6 |
| I-FIT | II | 12 | Flexibility Index (FI) | 3.53 | 46.9 | 53.4 |

NCAT will be conducting another round robin study in the future to assist with the implementation of performance testing for BMD, which would be useful to continue to provide proficiency testing for existing users along with benchmarking for newer users. Future round robin studies may also be expanded to evaluate newer tests being developed for BMD, such as quick high temperature rutting resistance tests.

## REFERENCES


1. West, R., C. Rodezno, F. Leiva, and F. Yin. *Development of a Framework for Balanced Mix Design.* National Cooperative Highway Research Program (NCHRP). Washington, D.C. 2018.
2. Tran, N. H., R. West, B. Powell, and A. Kvasnak. Evaluation of AASHTO Rut Test Procedure Using the Asphalt Pavement Analyzer. *Journal of the Association of Asphalt Paving Technologists, Volume 78*, 2009, pp. 1-24.
3. Al-Qadi, I. L., D. L. Lippert, S. Wu, H. Ozer, G. Renshaw, I. M. Said, and J. W. Vespa. *Utilizing Lab Tests to Predict Asphalt Concrete Overlay Performance.* Illinois Center for Transportation. Urbana, IL, 2017.
4. *Special Provision for Balanced Mix Design (BMD) Surface Mixtures Designed Using Performance Criteria*. Virginia Department of Transportation. 2019.
5. Zhou, F., S. Im, L. Sun, and T. Scullion. Development of an IDEAL Cracking Test for Asphalt Mix Design and QC/QA. *Journal of the Association of Asphalt Paving Technologists*. 2017. pp 549-577.
6. Al-Qadi, I. L., Said, I. M., Ali, U. M., & Kaddo, J. R. (2021). Cracking Prediction of Asphalt Concrete Using Fracture and Strength Tests. *International Journal of Pavement Engineering*.

## APPENDIX A: EXAMPLE INSTRUCTIONS TO PARTICIPANTS

### Blank Results Form Provided to Participating Labs – IDEAL-CT

| NCAT IDEAL-CT Round Robin - Data Summary Sheet | | | | | |
|---|---|---|---|---|---|
| | | | | | |
| *General Equipment Information* | | | | | |
| Gyratory Compactor Make and Model: | | | | | |
| IDEAL-CT Make and Model: | | | | | |
| Screw-Drive (mechanical) or Servo-Hydraulic IDEAL-CT machine? | | | | | |
| | | | | | |
| *General Specimen Info* | | | | | |
| Bucket ID (Number on Side of Bucket) | | | | | |
| Sample ID (only the 5 used for testing) | | | | | |
| Date Compacted | | | | | |
| Number of Gyrations to achieve 62 mm | | | | | |
| Gmb - Dry Mass in Air, g | | | | | |
| Gmb - Mass of Specimen underwater, g | | | | | |
| Gmb - SSD Mass of Specimen, g | | | | | |
| Gmb | | | | | |
| Air Voids (%) - Use Provided Gmm of 2.691 | | | | | |
| *IDEAL-CT Results Summary**** | | | | | |
| Conditioning Time (hr) | | | | | |
| Conditioning Chamber* | | | | | |
| Peak Load (kN)** | | | | | |
| Fracture Energy (J/m$^2$)** | | | | | |
| Tensile Strength (kPa)** | | | | | |
| CT Index | | | | | |
| | | | | | |
| * Environmental Chamber in Air or Inside Bags in a Water Bath | | | | | |
| ** = If your machine or template outputs the information in English Units, please report those and make a note | | | | | |
| *** = In addition to the results summary below, please send the raw rut depth versus cycles to failure | | | | | |

### Instructions to Participating Labs – IDEAL- CT

A. Mix re-heating procedure
   1. Remove rubber lining from inside of mix bucket lid, and cut the plastic handle off the bucket handle. Put bucket lid back on bucket.
   2. Place mix bucket with lid in an oven set to the compaction temperature.
      - Compaction Temperature for this mix is 300°F.
   3. Remove bucket after 2 to 3 hours once mix is workable enough to split into individual samples.
   4. Pour entire bucket of mix into splitting device and split out samples to desired mass according to AASHTO R47.
      - The “Quartering Method” is preferred in the NCAT lab. However, a mechanical splitter in accordance with AASHTO R47 may also be used.
      - Recommended weight for each pan would be around 2,700 grams. This should give enough material for the testing specimens in the event the air voids are too high on the initial trial specimen.
      - One bucket of mix may yield between 8 and 12 specimens worth of material for IDEAL-CT specimens.

5. Place samples in aging pans. Make sure to flatten the mix out into a uniform thickness so that it is aged evenly.
6. Split entire bucket out and place in aging pans even though all of the specimens will not be compacted at this time. This will keep from having to reheat the mix in the bucket again, which reduces overall mix aging time.
    - The samples that will not be compacted that day may be left in aging pans, labeled, and stored away until they are needed.
    - Note: If you have already performed this process for the 62 mm tall Hamburg specimens in the NCAT round robin, you should perform one single verification trial using the Hamburg specimen weight prior to proceeding to compacting IDEAL-CT testing specimens.
7. Once each sample is placed into an aging pan, place the pan containing the mix for the trial weight specimen back into the oven. The oven should be set to compaction temperature + 10°F (310°F for this mix).
8. Place dial thermometer in the mix and compact as soon as it reaches desired compaction temperature ± 5°F (between 295 and 305°F for this mix). This usually takes 0.5 to 1.5 hours.
    - Be sure to check the operation and accuracy of the oven and dial thermometer before starting.
    - If mix is not up to compaction temperature after 1 hour, stir with a spatula.
9. Once each sample reaches the desired compaction temperature, compact according to the desired test method.
    - For the IDEAL-CT Specimens in this study, specimens shall be compacted to a 62 mm height.
    - Use 2.691 as the mix Gmm in all air void calculations.
    - A good trial mass for the first IDEAL-CT specimen would be 2,645 grams. This should get you close to the target of 7.0 ± 0.5 percent air voids.

B. Compacting Testing Specimens
1. Determine bulk specific gravity of the trial specimen in accordance with AASHTO T166.
2. Calculate the air voids of the trial specimen using the $G_{mm}$ provided (2.691).
3. Use the provided spreadsheet 'Trial Weight Calculation – NCAT Round Robin' to calculate the adjusted mass in the mold for the testing specimens.
    - Input data into the green cells (Mass of Trial Specimen in grams, percent air voids of the trial sample, and the target air voids – which should be 7 percent for IDEAL-CT specimens).
    - The calibrated mass will appear in the orange cell after this data is input into the spreadsheet.
4. Place six (6) additional pans (split out previously) of material into an oven at the compaction temperature +10°F (310°F for this mix).
5. When the mix reaches the compaction temperature ±5°F (between 295 and 305°F for this mix), compact to 62 mm using the new calibrated mass.
6. Allow specimens to cool completely and determine bulk specific gravity using the provided $G_{mm}$. Calculate air voids.

7. If you have five (5) specimens that have 7.0 ± 0.5 percent air voids, proceed to IDEAL-CT testing.
    - If not, use the bulk data from the compacted specimens to recalibrate the trial mix weight. Compact remaining specimens at calibrated weight.
    - One approach would be to take the average specimen mass and the average air voids of the compacted specimens and input those values into the Trial Mass Spreadsheet.

C. IDEAL-CT Testing
1. Allow specimens to dry (either under a fan or using a vacuum drier) prior to testing.
    - No saw cutting is required prior to testing IDEAL-CT specimens.
2. Specimens shall be conditioned at 25°C for a minimum of 2 hours prior to testing.
    - Please note the conditioning method (air environmental chamber or water bath in bags) on the attached data form.
3. Conduct the IDEAL-CT testing in accordance with your manufacturer's instructions and the parameters listed below.
    - There is currently not a finalized national standard on the IDEAL-CT test method.
    - If you have any questions about the method, please contact Adam Taylor (tayloa3@auburn.edu).
4. IDEAL-CT testing shall be performed at a 50 mm/minute load rate.
    - Please note the equipment manufacturer and model and whether your machine is screw-drive or servo-hydraulic on the attached data form.
5. Target seating load is 0.1 kN.
6. Termination load at the end of the test is 0.1 kN.
7. A full load versus displacement curve should be collected by the testing machine for each specimen.
8. Calculate the IDEAL-CT value for each specimen using the NCAT template or software provided by your equipment manufacturer.
    - If you are in need of a calculation template, please contact Adam Taylor at NCAT and he can provide you with one.
9. Complete the attached summary sheet and return it, in addition to the raw data files from the IDEAL-CT machine, to Adam Taylor.
    - Summary sheet is titled 'NCAT Round Robin – IDEAL-CT Results.'
    - Please 'Save As' and add your company name to the end of the file to help with organization.

D. Mix re-heating don'ts (Practices to Avoid)
1. Do not re-heat the bucket of mix multiple times. Re-heat the bucket once and split out the samples you will need. Multiple re-heats will age the mix excessively and may severely skew the results.
2. Do not re-heat the bucket for more than 4 hours prior to splitting out the mix.
3. Do not scoop material directly out of the bucket. This will lead to segregated samples and will skew the results.

## APPENDIX B: INDIVIDUAL TEST RESULTS

**Table B1: Summary of Hamburg Results – Avg. Rut Depth (mm) from Individual Labs – Phase I**

| Lab ID | Replicates | Rut Depth (mm) | | | | | |
|---|---|---|---|---|---|---|---|
| | | 2,500 passes | 5,000 passes | 7,500 passes | 10,000 passes | 15,000 passes | 20,000 passes |
| 1 | 2 | 2.17 | 2.53 | 2.80 | 3.09 | 3.65 | 4.29 |
| 2 | 2 | 1.91 | 2.24 | 2.48 | 2.69 | 3.11 | 3.59 |
| 3 | 2 | 2.45 | 3.03 | 3.50 | 4.06 | 5.42 | 8.42 |
| 4 | 2 | 1.86 | 2.23 | 2.47 | 2.62 | 2.87 | 3.06 |
| 5 | 2 | 2.31 | 2.72 | 2.99 | 3.23 | 3.57 | 3.86 |
| 6 | 2 | 1.91 | 2.20 | 2.47 | 2.58 | 2.79 | 3.09 |
| 7 | 2 | 1.79 | 2.18 | 2.31 | 2.64 | 2.86 | 3.04 |
| 8 | 2 | 2.34 | 2.89 | 3.19 | 3.42 | 3.76 | 4.04 |
| 9 | 2 | 1.93 | 2.25 | 2.44 | 2.63 | 2.89 | 3.12 |
| 10 | 3 | 1.61 | 1.92 | 2.04 | 2.22 | 2.41 | 2.63 |
| 11 | 2 | 2.16 | 2.47 | 2.67 | 2.87 | 3.14 | 3.35 |
| 12 | 2 | 2.10 | 2.56 | 2.85 | 3.04 | 3.35 | 3.58 |
| 13 | 2 | 1.96 | 2.42 | 2.71 | 2.99 | 3.40 | 3.73 |
| 14 | 2 | 2.01 | 2.40 | 2.65 | 2.86 | 3.13 | 3.36 |
| 15 | 2 | 1.78 | 2.29 | 2.52 | 2.43 | 2.81 | 2.91 |
| 16 | 4 | 2.82 | 3.38 | 3.83 | 4.29 | 5.35 | 6.53 |
| 17 | 2 | 2.33 | 2.44 | 2.57 | 2.60 | 2.79 | 2.94 |
| 18 | 2 | 1.92 | 2.28 | 2.50 | 2.76 | 3.00 | 3.18 |
| 19 | 6 | 2.74 | 3.39 | 3.82 | 4.19 | 4.88 | 5.65 |
| 20 | 2 | 1.88 | 2.39 | 2.68 | 2.86 | 3.15 | 3.35 |
| 21 | 2 | 1.96 | 2.30 | 2.52 | 2.69 | 2.90 | 3.18 |
| 22 | 2 | 3.07 | 3.41 | 3.73 | 3.94 | 4.19 | 4.62 |
| 23 | 2 | 1.08 | 1.31 | 1.62 | 1.97 | 2.34 | 2.53 |
| 24 | 2 | 2.24 | 2.68 | 3.02 | 3.26 | 3.62 | 3.94 |
| 25 | 2 | 2.82 | 3.76 | 4.41 | 5.01 | 6.22 | 7.95 |
| 26 | 2 | 1.94 | 2.23 | 2.57 | 2.84 | 3.01 | 3.38 |
| 27 | 2 | 2.53 | 3.08 | 3.54 | 3.86 | 4.50 | 5.19 |
| 28 | 2 | 1.75 | 2.21 | 2.61 | 2.89 | 3.00 | 3.37 |
| 29 | 2 | 2.33 | 2.81 | 3.09 | 3.29 | 3.61 | 3.88 |
| 30 | 2 | 1.78 | 2.02 | 2.19 | 2.34 | 2.47 | 2.63 |
| 31 | 2 | 2.06 | 2.51 | 2.80 | 3.03 | 3.43 | 3.78 |
| 32 | 2 | 1.76 | 2.08 | 2.32 | 2.51 | 2.85 | 3.15 |

Table B2: Summary of APA Results – Averages from Individual Labs – Phase I

| Lab ID | N | Air Voids (%) | Manual Rut Depth (mm) | | | Automated Rut Depth (mm) | | |
|---|---|---|---|---|---|---|---|---|
| | | Average | Avg. | St. Dev. | CV (%) | Avg. | St. Dev. | CV (%) |
| 1 | 4 | 7.0 | n/a | n/a | n/a | 3.28 | 0.20 | 6.2 |
| 2 | 6 | 6.8 | n/a | n/a | n/a | 3.18 | 1.03 | 32.4 |
| 3 | 4 | 6.6 | 2.64 | 0.50 | 18.9 | 2.40 | 0.10 | 4.0 |
| 4 | 4 | 6.9 | n/a | n/a | n/a | 2.40 | 0.68 | 28.5 |
| 5 | 4 | 7.3 | 4.46 | 0.51 | 11.4 | 4.55 | 0.18 | 4.0 |
| 6 | 6 | 6.9 | 8.13 | 0.71 | 8.7 | n/a | n/a | n/a |
| 7 | 4 | 6.9 | 2.40 | 0.48 | 19.9 | 2.20 | 0.28 | 12.6 |
| 8 | 4 | 7.1 | n/a | n/a | n/a | 2.81 | 0.35 | 12.5 |
| 9 | 6 | 7.1 | 3.43 | 0.55 | 15.9 | 3.14 | 0.52 | 16.5 |
| 10 | 6 | 7.3 | n/a | n/a | n/a | 2.20 | 0.30 | 13.6 |

Table B3: Summary of I-FIT Results – Individual Labs – Phase I

| Lab ID | Replicates | Air Voids (%) | FE (J/m²) | ITS (psi) | Slope (kN/mm) | Flexibility Index | | |
|---|---|---|---|---|---|---|---|---|
| | | Average | Average | Average | Average | Average | St Dev. | CV (%) |
| 1 | 7 | 7.0 | 2,524 | 112.1 | -31.18 | 0.97 | 0.56 | 57.8 |
| 2 | 7 | 6.8 | 1,538 | 127.8 | -38.70 | 0.75 | 0.73 | 98.2 |
| 3 | 8 | 6.8 | 2,373 | 110.7 | -20.27 | 2.01 | 1.53 | 76.1 |
| 4 | 8 | 7.0 | 2,911 | 95.7 | -5.62 | 5.24 | 0.96 | 18.3 |
| 5 | 16 | 6.9 | 2,533 | 115.9 | -12.59 | 2.76 | 1.07 | 38.6 |
| 6 | 8 | 7.0 | 2,829 | 100.3 | -5.12 | 5.71 | 1.58 | 27.6 |
| 7 | 8 | 7.1 | 2,133 | 93.0 | -5.14 | 4.29 | 1.27 | 29.6 |
| 8 | 8 | 7.1 | 2,750 | 106.7 | -5.75 | 4.89 | 0.95 | 19.4 |
| 9 | 7 | 7.3 | 2,565 | 106.7 | -17.39 | 2.86 | 2.09 | 73.2 |
| 10 | 30 | 7.1 | 3,355 | 85.0 | -2.91 | 11.52 | 3.21 | 27.8 |
| 11 | 8 | 6.9 | 1,933 | 82.3 | -4.44 | 4.41 | 0.60 | 13.7 |
| 12 | 8 | 7.2 | 2,631 | 97.7 | -5.02 | 5.45 | 1.32 | 24.2 |
| 13 | 8 | 6.9 | 2,589 | 106.3 | -8.49 | 4.19 | 1.87 | 44.7 |
| 14 | 8 | 6.9 | 2,669 | 104.7 | -5.91 | 4.66 | 1.06 | 22.9 |
| 15 | 10 | 7.2 | 2,554 | 113.7 | -13.62 | 2.39 | 1.25 | 52.4 |
| 16 | 8 | 6.8 | 2,730 | 106.3 | -5.94 | 4.73 | 1.07 | 22.7 |
| 17 | 8 | 6.9 | 2,757 | 106.7 | -6.93 | 4.29 | 0.91 | 21.3 |
| 18 | 5 | 7.3 | 3,017 | 99.7 | -4.90 | 6.33 | 1.31 | 20.7 |
| 19 | 8 | 7.0 | 3,297 | 113.9 | -6.41 | 5.23 | 0.66 | 12.7 |
| 20 | 8 | 7.0 | 2,480 | 95.1 | 6.52 | 3.90 | 0.83 | 21.4 |
| All | | 7.0 | 2,609 | 104.0 | -10.0 | 4.33 | | |

**Table B4: Summary of I-FIT Results – Individual Labs – Phase II**

| Lab ID | Replicates | Air Voids (%) | FE (J/m$^2$) | ITS (psi) | Slope (kN/mm) | Flexibility Index | | |
|---|---|---|---|---|---|---|---|---|
| | | Average | Average | Average | Average | Average | St Dev. | CV (%) |
| 1 | 5 | 7.1 | 2,676 | 102.1 | -8.23 | 3.41 | 0.78 | 22.9 |
| 2 | 8 | 7.1 | 2,491 | 73.6 | -2.90 | 8.67 | 1.02 | 11.8 |
| 3 | | | | | | | | |
| 4 | 7 | 7.1 | 2,360 | 95.5 | -7.76 | 3.76 | 1.52 | 40.5 |
| 5 | | | | | | | | |
| 6 | 8 | 7.1 | 2,510 | 107.1 | -10.19 | 3.51 | 1.38 | 39.4 |
| 7 | 8 | 7.1 | 2,144 | 101.2 | -12.29 | 2.95 | 1.78 | 60.6 |
| 8 | | | | | | | | |
| 9 | 8 | 7.1 | 2,458 | 102.4 | -18.03 | 1.95 | 1.22 | 62.3 |
| 10 | 6 | 7.1 | 2,660 | 85.7 | -10.33 | 4.54 | 2.44 | 53.7 |
| 11 | 7 | 7.1 | 1,754 | 89.3 | -6.83 | 2.85 | 1.29 | 45.2 |
| 12 | | | | | | | | |
| 13 | | | | | | | | |
| 14 | 8 | 7.1 | 2,455 | 108.4 | -14.22 | 3.02 | 2.22 | 73.7 |
| 15 | 8 | 7.1 | 2,510 | 108.9 | -8.32 | 3.23 | 0.91 | 28.2 |
| 16 | | | | | | | | |
| 17 | | | | | | | | |
| 18 | 8 | 7.1 | 2,642 | 95.7 | -5.14 | 5.70 | 2.31 | 40.5 |
| 19 | 8 | 7.1 | 2,967 | 107.5 | -6.29 | 4.90 | 1.14 | 23.3 |
| 20 | | | | | | | | |
| All | | 7.1 | 2,469 | 98.1 | -9.2 | 4.0 | | |

**Table B5: Summary of IDEAL-CT Results – Individual Labs – Phase I**

| Lab ID | Replicates | Air Voids (%) | Peak Load (kN) | FE (J/m$^2$) | ITS (kPa) | CT Index | | |
|---|---|---|---|---|---|---|---|---|
| | | Average | Average | Average | Average | Average | St Dev. | CV (%) |
| 1 | 5 | 7.1 | 18.6 | 12,273 | 1,271.5 | 117.5 | 22.5 | 19.1 |
| 2 | 5 | 7.0 | 20.0 | 11,954 | 1,375.1 | 82.5 | 13.9 | 16.9 |
| 3 | 5 | 7.0 | 20.6 | 13,370 | 1,406.9 | 113.7 | 29.5 | 26.0 |
| 4 | 5 | 6.9 | 15.0 | 6,176 | 990.3 | 36.5 | 13.0 | 35.6 |
| 5 | 12 | 7.0 | 18.5 | 11,960 | 1,266.2 | 100.6 | 15.1 | 15.0 |
| 6 | 5 | 7.0 | 19.4 | 12,683 | 1,328.4 | 97.4 | 13.9 | 14.3 |
| 7 | 5 | 6.9 | 17.2 | 12,496 | 1,180.3 | 144.1 | 22.4 | 15.5 |
| 8 | 5 | 7.1 | 21.1 | 12,412 | 1,452.6 | 74.7 | 12.9 | 17.2 |
| 9 | 5 | 7.0 | 18.0 | 12,452 | 1,240.5 | 126.0 | 23.0 | 18.2 |
| 10 | 5 | 7.0 | 20.6 | 12,265 | 1,407.7 | 84.9 | 14.7 | 17.3 |
| 11 | 5 | 6.9 | 17.4 | 11,471 | 1,191.1 | 102.2 | 15.3 | 14.9 |
| 12 | 5 | 7.0 | 18.8 | 14,937 | 1,285.1 | 188.0 | 25.2 | 13.4 |
| 13 | 5 | 6.8 | 15.0 | 10,539 | 1,027.4 | 122.1 | 13.7 | 11.2 |
| 14 | 5 | 7.1 | 18.7 | 13,475 | 1,286.3 | 146.6 | 21.0 | 14.3 |
| 15 | 8 | 7.2 | 18.0 | 12,347 | 1,233.6 | 130.0 | 39.8 | 30.6 |
| All | | 7.0 | 18.5 | 12,054 | 1,263 | 111.1 | | |

**Table B6: Summary of IDEAL-CT Results – Individual Labs – Phase II**

| Lab ID | Replicates | Air Voids (%) | Peak Load (kN) | FE (J/m²) | ITS (kPa) | CT Index | | |
|---|---|---|---|---|---|---|---|---|
| | | Average | Average | Average | Average | Average | St Dev. | CV (%) |
| 1 | 5 | 7.1 | 19.6 | 12,196 | 1,340 | 97.2 | 26.6 | 27.3 |
| 2 | 5 | 7.1 | 19.3 | 12,216 | 1,320 | 97.9 | 17.8 | 18.2 |
| 3 | 4 | 7.1 | 20.5 | 13,171 | 1,404 | 94.9 | 16.2 | 17.1 |
| 4 | | | | | | | | |
| 5 | 4 | 7.0 | 18.6 | 12,050 | 1,272 | 101.8 | 8.3 | 8.2 |
| 6 | 5 | 7.1 | 17.8 | 11,864 | 1,220 | 118.7 | 29.7 | 25.0 |
| 7 | 5 | 7.1 | 19.1 | 12,218 | 1,308 | 99.1 | 14.7 | 14.8 |
| 8 | 5 | 7.1 | 18.7 | 12,442 | 1,278 | 115.7 | 12.0 | 10.3 |
| 9 | 5 | 7.1 | 19.0 | 12,553 | 1,302 | 118.9 | 23.4 | 19.7 |
| 10 | 5 | 7.1 | 20.3 | 12,542 | 1,387 | 92.6 | 22.9 | 24.8 |
| 11 | 5 | 7.1 | 17.8 | 11,300 | 1,217 | 93.4 | 11.1 | 11.9 |
| 12 | 5 | 7.1 | 17.6 | 12,251 | 1,186 | 127.8 | 22.5 | 17.6 |
| 13 | 5 | 7.1 | 18.6 | 11,734 | 1,276 | 93.3 | 19.0 | 20.4 |
| 14 | 5 | 7.1 | 18.7 | 12,242 | 1,281 | 101.9 | 19.5 | 19.2 |
| 15 | | | | | | | | |
| 16* | 5 | 7.1 | 19.6 | 12,542 | 1,340 | 98.6 | 13.9 | 14.1 |
| All | | 7.1 | 18.9 | 12,214 | 1,292 | 103.7 | | |

* = Second set of specimens for Lab 8 tested using an additional load frame