VERMONT
AGENCY OF TRANSPORTATION

INTER-LABORATORY STUDY (ILS) OF BITUMINOUS CONCRETE BALANCED MIX DESIGN (BMD) TESTS FOR USE ON VTRANS PROJECTS

Alex Bernier, PE, Program Director
James McLaughlin, PE, Research Engineer
James Mahoney, Executive Program Director
Iliya Yut, PhD, Research Associate
Scott Zinke, Research Engineer
The Connecticut Transportation Institute
University of Connecticut
Connecticut Advanced Pavement Laboratory

November 2023

Research Project
Reporting on Project # GR-1675

Final Report 2023-02

This material is based upon work supported by the Federal Highway Administration under SPR VTRC 22-2. Any opinions, findings and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the Federal Highway Administration.

TECHNICAL DOCUMENTATION PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
2023-02		

4. Title and Subtitle Inter-Laboratory Study (ILS) of Bituminous Concrete Balanced Mix Design (BMD) Tests for Use on VTrans Projects	**5. Report Date** January 3, 2024 **6. Performing Organization Code**
7. Author(s) Mahoney, James Bernier, Alex McLaughlin, James Yut, Iliya Zinke, Scott	**8. Performing Organization Report No.**
9. Performing Organization Name and Address The Connecticut Transportation Institute University of Connecticut Connecticut Advanced Pavement Laboratory 270 Middle Turnpike, Unit 5202 Storrs, CT 06269-5202	**10. Work Unit No.** VTRC 22-2 **11. Contract or Grant No.** GR1675
12. Sponsoring Agency Name and Address Vermont Agency of Transportation (SPR) Research Section One National Life Drive Montpelier, VT 05633	**13. Type of Report and Period Covered** Final Report 2022-2023 **14. Sponsoring Agency Code**

15. Supplementary Notes
Conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration.

16. Abstract
Many state departments of transportation across the US are in various stages and levels of implementation of performance testing for Balanced Mix Design (BMD). The tests are intended to provide data to the designer with the goal of developing a mix design that balances adequate resistance to rutting and stripping with resistance to thermal and stress cracking. The objective of this Inter-Lab Study (ILS) was to determine the variability between eleven (11) different laboratories in the region using typical testing methods included as part of Balanced Mix Design (BMD) procedures including AASHTO T 324 Hamburg Wheel-Track Test (HWTT), the AASHTO T 393-21 Illinois Flexibility Index Test (I-FIT), and the ASTM D8225-19 Indirect Tension Asphalt Cracking Test (IDEAL-CT). To ensure the consistency of samples used in this study, all specimen sampling, compaction and sample preparation were undertaken by a single lab (the UConn CT Advanced Pavement Lab). The aforementioned BMD testing methods were then conducted across multiple machines and operators from various laboratories/agencies to establish minimum variabilities expected across a single sample for these tests.

17. Key Words	18. Distribution Statement
Inter-Laboratory Study, Balanced Mix Design, Bituminous Concrete, Hamburg Wheel-Track Test, (HWTT), Illinois Flexibility Index Test (I-FIT), Indirect Tension Asphalt Cracking Test (IDEAL-CT)	No restrictions. This document is available through the National Technical Information Service, Springfield, VA 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	15	

DISCLAIMER

The information contained in this report was compiled for the use of the Vermont Agency of Transportation. Conclusions and recommendations contained herein are based upon the research data obtained and the expertise of the researchers and are not necessarily to be construed as Agency policy. This report does not constitute a standard, specification, or regulation. The Vermont Agency of Transportation assumes no liability for its contents or the use thereof.

This material is based upon work supported by the Federal Highway Administration under SPR [insert work project]. Any opinions, findings and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the Federal Highway Administration.

ACKNOWLEDGEMENTS

The author(s) wish to acknowledge the support of personnel from the Technical Advisory Committee for this project from the Vermont Agency of Transportation: Aaron Schwartz, Emily Parkany, Ian Anderson, Brandon Kipp, Tanya Miller, and Ashlie Mercado.

METRIC CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS				
SYMBOL	**WHEN YOU KNOW**	**MULTIPLY BY**	**TO FIND**	**SYMBOL**
LENGTH				
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
AREA				
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
VOLUME				
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
NOTE: volumes greater than 1000 L shall be shown in m^3				
MASS				
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
TEMPERATURE (exact degrees)				
°F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	°C
ILLUMINATION				
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
FORCE and PRESSURE or STRESS				
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

Table of Contents

Figures

Tables

CHAPTER 1 Introduction and Background

Unlike other construction materials, asphalt concrete pavement has yet to be characterized by a single physical test. This is due to many factors such as the geographic variability of aggregate, the viscosity of binder, traffic levels, and environmental damage. These factors change the material over time and also manifest different material responses. For decades, researchers have sought to develop asphalt performance tests that best characterize an asphalt concrete pavement as *good* or *bad*. Early examples include the Marshall Flow and Stability test (AASHTO T 245)(*1)* and the Superpave Shear Tester (SST)(*2*). While the Marshall test was widely used for many years, it has significant limitations to considering the holistic performance of an asphalt pavement. The SST on the other hand was a high-cost machine which required time consuming specimen fabrication and advanced analytical capabilities to interpret the test results. While the SST can adequately characterize pavements it could not be widely adopted for project-level quality testing.

For many years, the global asphalt pavement research community has proposed alternative methods for testing and characterizing asphalt concrete pavement. The long-term vision has been to find *one* test which could produce assurance of the quality and longevity of a bituminous mixture. To date, a singular testing method has not been achieved. Over time, a collection of tests provide a comprehensive characterization of pavements and have gained wide acceptance as performance tests. These tests will collectively provide an indication of a pavement's quality. While these are known individually as asphalt pavement performance tests, a movement to incorporate these tests into construction specifications has come to be known as Balanced Mix Design (NCHRP 2018)(*3*). Per AASHTO PP 105-20, Balanced Mix Design (BMD) is defined as "asphalt mix design using performance tests on appropriately conditioned specimens that address multiple modes of distress taking into consideration mix aging, traffic, climate and location within the pavement structure..". The idea behind this methodology is that an asphalt pavement producer would no longer be constrained to individual mix design targets for gradations, asphalt content, and other volumetric constraints, but that any design that passes a series of these performance tests would satisfy the construction specification requirements.

The challenge for Agencies as they look to implement Balanced Mix Design is multi-fold. First, any threshold values that performance tests may incorporate into the specifications must be established. Second, the inherent variability of the tests need to be determined, and third the repeatability of the test for a given material sample or material stream must be established to determine when action limits or engineering limits would be violated for a given performance test.

To address the first challenge, the Vermont Agency of Transportation (VTrans) has developed initial threshold values and material producers have begun submitting test data to the state (4). VTrans is now working towards resolving the second and third challenges discussed above. To this end, an inter-laboratory study was undertaken whereby one large plant-produced sample was compacted into gyratory samples at a single laboratory and shipped to nine other labs (a total of ten labs participated including the lead lab) where the specimens were tested in three different asphalt pavement performance tests: Hamburg Wheel Tracking Test (AASHTO T 324)(*5*), Illinois Flexibility Index Test [i-FIT] (AASHTO T 393)(*6*), and the Indirect Tension Asphalt Cracking Test [IDEAL-CT] (ASTM 8225)(*7*). This report summarizes the efforts and findings of this inter-laboratory study.

CHAPTER 2 Research Approach

2.1 Selection of Test Procedures to be Studied

In recent years, VTrans updated their mix design submittal policy to include HMA performance testing including the follow three tests:

- AASHTO T 324 Hamburg Wheel-Track Test (HWTT)
- AASHTO T 393-21 Illinois Flexibility Index Test (I-FIT)
- ASTM D8225-19 Indirect Tension Asphalt Cracking Test (IDEAL-CT)

While submission during a mix design is one component of a balanced design, it is crucial to understand the tests' repeatability across labs and during production prior to full implementation. These three tests have gained significant use across the paving industry over the past 10 years but the push for implementation of BMD has given rise to the need to understand the constraints of the equipment used to perform these tests.

2.2 Collection and Initial Testing of Samples

Prior to collecting samples, an estimate of the amount HMA mix sample material required was generated to ensure that a sufficient quantity would be collected to allow for an adequate number of Superpave Gyratory-Compacted (SGC) specimens to be prepared. Enough material would be required for each test method to be performed by each laboratory two separate times. For example the Hamburg Wheel-Tracking test requires four specimens for each test, If 10 laboratories were to perform that test two times, it would require a total of 80 specimens to be prepared. Based on the number of different tests and the number of agencies/laboratories that would be involved, it was estimated that approximately 1000 kg or 2200 lbs of HMA mix would be required.

The required material was collected from the Pike Industries, Waterford, Vermont asphalt plant on September 7, 2022, Material was shoveled into 5-gallon steel buckets with covers which were then transported to the Connecticut Advanced Pavement (CAP) Laboratory at the University of Connecticut. Samples of the mix from four separate randomly selected buckets were then tested to determine Maximum Theoretical Specific Gravity (G_{mm}) using AASHTO T 209 Method "A", resulting in G_{mm} values of between 2.617 and 2.627. AASHTO T 166 Bulk Specific Gravity testing was also performed on 150-mm diameter x 62.5 mm high SGC specimens as well as 150 mm diameter x 115 mm high SGC specimens which are both required for the various tests in the study to dial-in specimen fabrication and then to confirm air voids of every manufactured specimen. Once target batch weights were established for production of all the SGC specimens, a G_{mm} was performed on nearly every bucket of material to ensure consistency of the mix as sample production progressed.

The material sampled was an 80-gyration 9.5-mm polymer-modified Superpave mix with 20% RAP. The binder grade was a PG70-28 and the mix had a 4.8% virgin binder content with a 5.6% total binder content when accounting for the RAP. The material utilized 4 virgin aggregate sources including 6.8% natural sand and 3.9% dust content. The manufacturer's approved mix design reported 2.3 mm rutting on the HWTT and a 12.5 Flexibility Index (AASHTO T 393 – Formerly TP 124).

2.3 Selection of Participating Laboratories and Distribution of Initial Round of Samples

VTrans coordinated with HMA producers, State Agency materials laboratories, and University laboratories within the Northeast region who had the necessary performance test equipment available and who were willing to participate in the ILS study. Participants were advised that they would be testing two rounds of specimens, a set to be provided to them in Fall/winter of 2022/2023 and then a set to be provided in early Spring of 2023. The results from the second round of specimens and testing would serve as a means of identifying sources of errors or outlier results from specific laboratories during the first round of testing. Each participant was allowed to select a preference for the time frame when the test rounds were to be completed. Table 1 below summarizes the participating laboratories and the number of tests to be performed in a single round of testing. It is of note that not every lab was able to perform all 3 tests.

TABLE 1 NUMBER OF LABORATORIES AND NUMBER OF DIFFERENT MANUFACTURER'S DEVICES USED IN THE STUDY

State DOTs	University Laboratories	Contractor Labs
4	2	5

Hamburg Machine (AASHTO T324)	I-FIT Machine (AASHTO T393)	IDEAL-CT Machine (ASTM 8225)
4	3	3

2.4 *Preparation and Distribution of Specimens for Testing*

The first round of Hamburg, IDEAL-CT and I-FIT specimens were prepared in the CAP Lab in accordance with the AASHTO specifications for each test. 150-mm diameter x 62mm high SGC specimens were prepared for the Hamburg and IDEAL-CT tests. 150 mm diameter x 115-mm high I-Fit specimens were cut into notched semi-circular samples by CAP Lab personnel. Triplicate testing of CAP Lab prepared specimens was performed to ensure consistency of the specimen materials and preparation process prior to shipping the SGC specimens to other labs. Additionally, air voids were checked on every sample and were determined to be stable at the 7.0 ± 1.0% target value throughout the preparation process. Specimens were then labeled, carefully packaged to minimize potential shipping damage (Figure 1), and shipped by US Postal Service Priority Mail to participating laboratories on December 13, 2023, along with detailed written instructions for handling, testing, and report preparation. The second round of samples was sent on May 18, 2023

FIGURE 1 PREPARATION AND PACKAGING OF SPECIMENS FOR MAILING

An example of the form and instructions labs were provided is included in this report as Appendix A. CAP Lab personnel worked with all participants to ensure testing went smoothly and if needed, spare samples were provided to labs in the case of problems with test equipment.

CHAPTER 3 Findings and Applications

AASHTO T 324 Hamburg Wheel Tracking

Samples were tested for 20,000 cycles in each Hamburg device, tested at 45° C. Because the material sampled was polymer modified, at the test temperature a *Stripping Inflection Point* was not readily identified by most machines. The devices that run this test identify the maximum rut depth by averaging the middle 5 sensors across each set of specimens and report the overall average rut depth (across 11 sensors) at this point in accordance with the AASHTO standard. Both runs are averaged together for the final reported rut depth. The overall results are summarized below in Figure 2 and Table 2

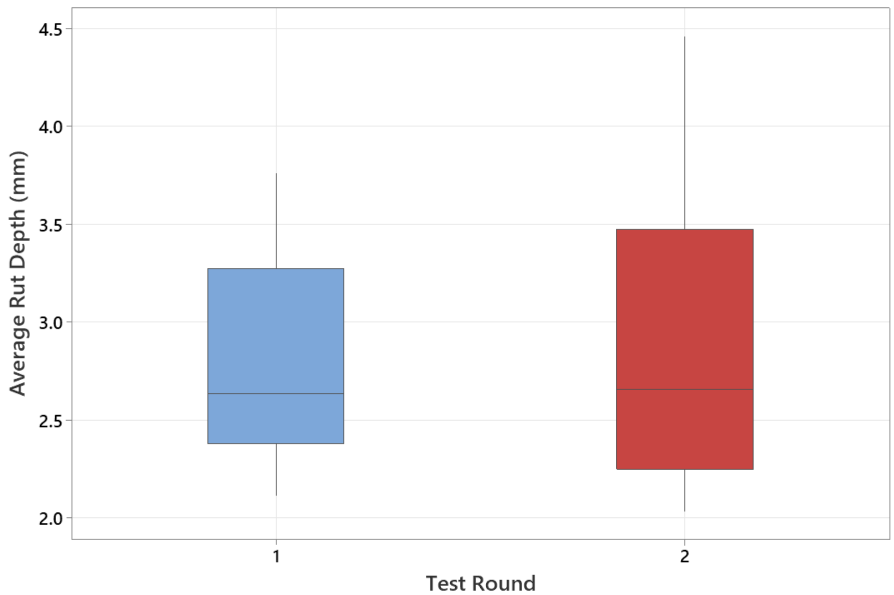

FIGURE 2 BOX PLOT OF HAMBURG WHEEL TRACKING RESULTS

The average of all test data is 2.82 and a coefficient of variation of 23.4%. This is in-line with test data reported by a joint study of VTrans and FHWA's Mobile Asphalt Technology Center (*8*).

TABLE 2 SUMMARY OF HAMBURG WHEEL TRACKING TEST RESULTS

Test Round	N	Mean (mm)	Standard Deviation (mm)	Minimum (mm)	Maximum (mm)	Median (mm)
1	11	2.769	0.497	2.115	3.761	2.635
2	11	2.872	0.861	1.871	4.458	2.482

Of the 22 Hamburg Wheel Tracking tests reported, the results demonstrate exceptional repeatability. While a precision and bias statement for AASHTO T 324 does not presently exist, the results from this study have a standard deviation of less than 1.0 mm, which the research team considers highly repeatable given the variety of devices and operators. In the same 2022 VTrans and FHWA study (*8*), the standard deviation found within a production lot of HMA was found to be 1.6 mm for HWT testing which is approximately double what was found in this study. A two-sample t-test failed to reject the hypothesis (p-value = 0.947 at 95% confidence) that Round 1 and Round 2 are equal (meaning the pool of each set of data are statistically equal). A one-way Analysis of Variance (ANOVA) did indicate the machine type as a significant factor for test results. In none of the test results were the difference between the two runs/wheels greater than 1.0 mm (well below the VTrans limit of 6.0 mm)

What does this mean for actionable outcomes from this testing? The researchers suggest these results strengthen the use of a threshold value to hold a mix design to, however, they caution against comparing two machines' data to each other, realizing there could be approximately 2 mm of variability across different testing machines.

Another way one could interpret the HWT data is using the Normalized Rutting Resistance Index (NRRI) as developed by Luiza Helena Barros at the University of Texas at El Paso in 2018 *(9*). The NRRI advances the concept of the RRI initially developed by Wu et al, 2017 (*10*). Essentially, the NRRI considers the presence of polymer and the rate as well as the total rut depth of the test. Since this was all the same material and it all went to 20,000 cycles, the NRRI was exceptionally stable with an average of 1.76.

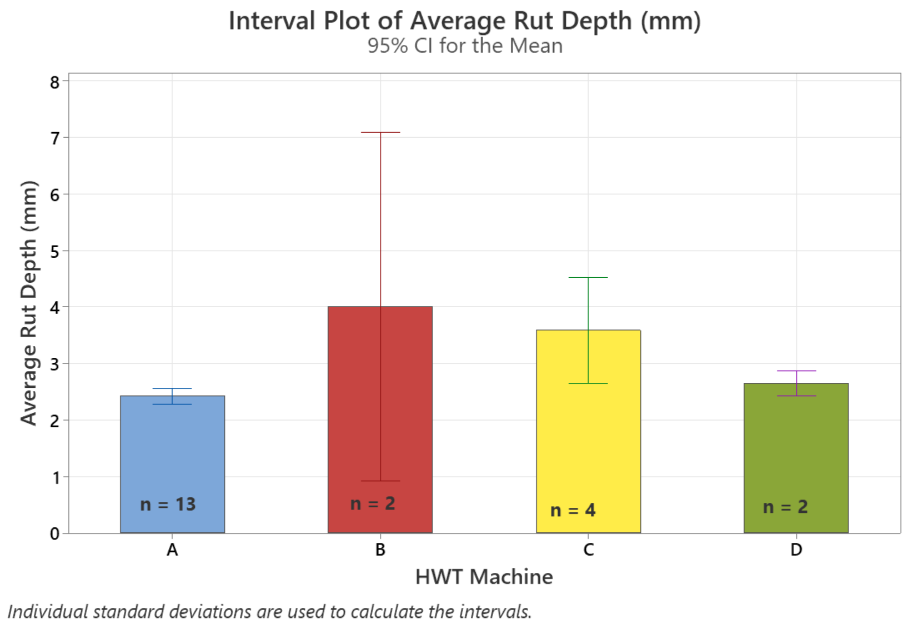

FIGURE 3 RUT DEPTH BY MACHINE TYPE

AASHTO T393 Illinois Flexibility Index Test

For the Illinois Flexibility Index Test, 3 semi-circular specimens were sent to each laboratory. In preparing the specimens, the research team found that cutting 115 mm gyratory specimens by trimming 32.5 mm from the top and the bottom of each SGC was a simpler laboratory procedure resulting in a higher number of conforming specimens. The resulting 50 mm (+/- 1 mm) thick disk was then cut into a hemispherical shape and a 15 mm deep notch was cut into the tangent portion of the sample at the UConn lab prior to distribution. The test is performed by applying a constant displacement from a load frame at 50 millimeters per minute. Load and displacement data is collected and the results are reported by interpreting the area under the fracture energy curve as well as its slope characteristics. In the case of the summary data presented in Figure 4, each participating laboratory interpreted the fracture curves by their own means (typically a readout from the instrument). While the reported FI value from AASHTO T 393 is the average of all three specimens tested by a given laboratory, the preliminary data presented below explores the difference between individual samples as well as the triplicate average for reasons discussed later in this chapter.

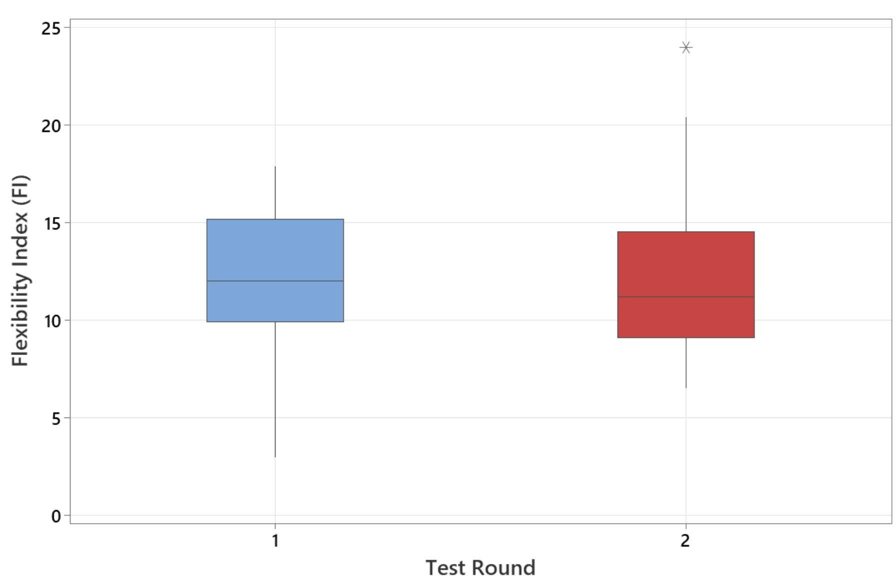

FIGURE 4 BOX PLOT OF ILLINOIS FLEXIBILITY INDEX (AASHTO T393) INDIVIDUAL SAMPLE RESULTS

While the overall spread of all specimen tests between Round 1 and Round 2 (Fig. 4). are very similar, a breakdown by test device tells a more detailed story (Figure 5). A visual analysis of the I-FIT data can be seen in Figure 5 through Figure 7 below where the post-peak slope |m|, fracture energy G_f and Flexibility Index *FI* are displayed by device and test round. It can be observed that device B had the best repeatability across the entire study. Devices A and C both had wider variability in the second round of testing, which was the opposite of the case for CT Index shown later in this section, where all test metrics seemed to improve in the second round. There are many factors to consider when attempting to extract meaning from this data. For example, while the air voids and dimensions all met the AASHTO T 393 standard, in the CAP LAB a different technician fabricated the samples between the two test rounds and a new masonry saw was used (identical, but new) for the initial cuts of the Semi-Circular Beam samples down to the 50 +/- 1 mm sample thickness.

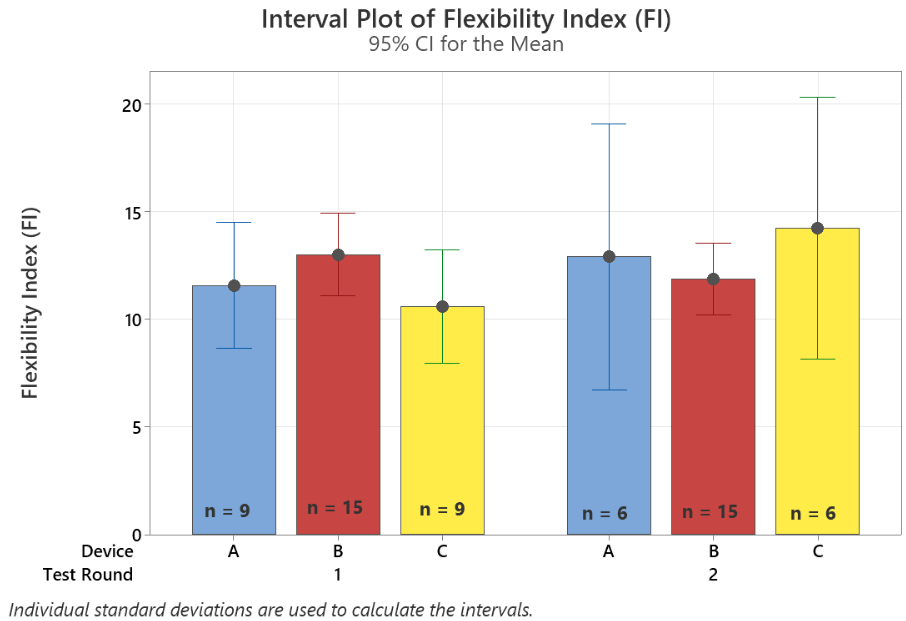

FIGURE 5 INTERVAL PLOT OF FLEXIBILITY INDEX VALUES WITHIN DEVICES IN ROUND 1 AND ROUND 2 (ERROR BARS REPRESENT 95% CONFIDENCE INTERVAL)

The fracture energy across all tests however shows very consistent test results, where the confidence interval is uniform across all devices in both rounds and the results are statistically the same across all data as well (Figure 6). The influence of the post-peak slope (Figure 7) can be seen on the two prior figures, where in Round 2 the |m| was relatively the same across all devices, the trend seen in fracture energy is reflected in the FI, whereas in Round 1, the |m| values were higher for device C, it lowered the FI relatively to the other devices, giving it ultimately the lowest average FI and the wider |m| for device A pulled the average FI further from the average of device B which had the closest average post-peak slope values across the two rounds of testing. Researchers compared calculated values from raw data from devices B and C using several different post-processing techniques including RAATPack analysis software as developed by Rutgers University. However, without delving into some of the device manufacturers' software, it is unclear what type of modeling is being used to determine post-peak slope and fracture energy with 'off the shelf' software packages.

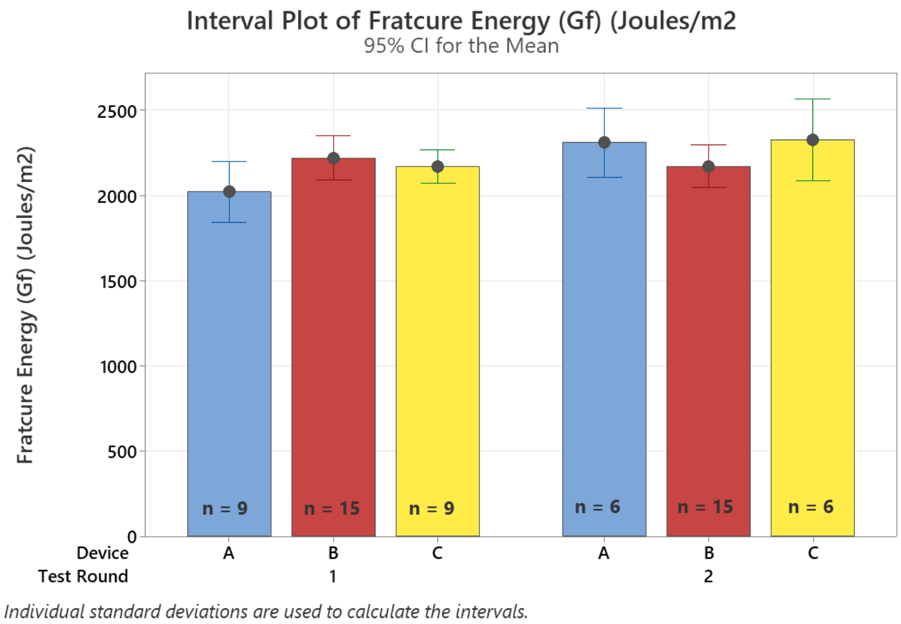

FIGURE 6 INTERVAL PLOT OF G_F (FRACTURE ENERGY) VALUES WITHIN DEVICES IN ROUND 1 AND ROUND 2 (ERROR BARS REPRESENT 95% CONFIDENCE INTERVAL)

It is not until inspection of the |m| or post-peak slope that the differences in FI seen in Figure 5 may be fleshed out. As seen in Figure 7, the variability of the post-peak slope is widest for device C and smallest for device B. While the variability for device A is consistent between Rounds 1 and 2, the average value is lower in Round 1 resulting in a higher average FI for that subset of test data.

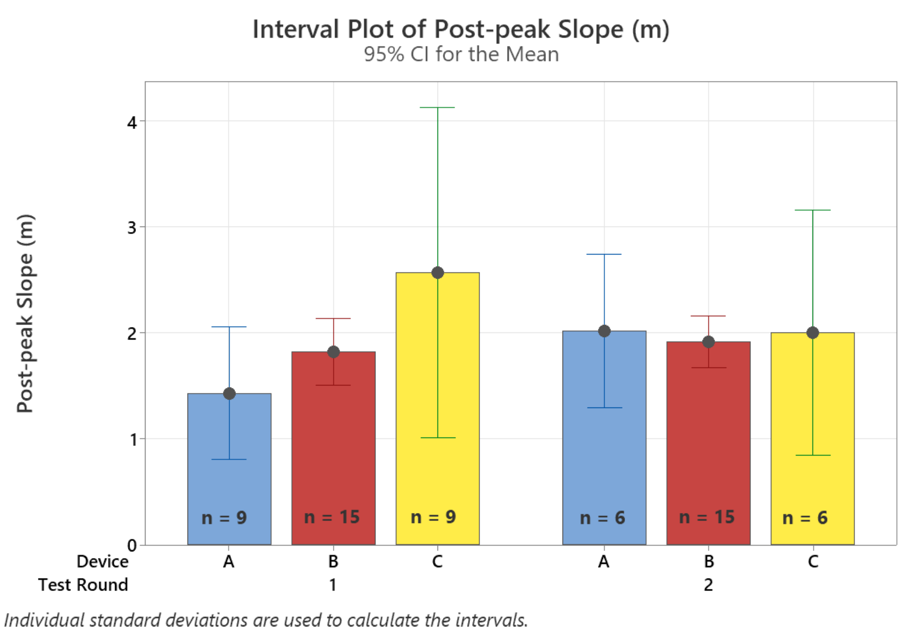

FIGURE 7 INTERVAL PLOT OF |M| (POST-PEAK SLOPE) VALUES WITHIN DEVICES IN ROUND 1 AND ROUND 2 (ERROR BARS REPRESENT 95% CONFIDENCE INTERVAL)

TABLE 3 OVERALL SUMMARY OF FLEXIBILITY INDEX (INDIVIDUAL SAMPLES)

Test Round	N	Mean	Standard Deviation	Minimum	Maximum	Median

1	33	11.9	3.6	2.99	17.88	12.0
2	27	12.6	4.4	6.5	23.98	11.2

TABLE 4 OVERALL SUMMARY OF FLEXIBILITY INDEX (AVERAGED TRIPLICATE SAMPLES)

Test Round	N	Mean	Standard Deviation	Minimum	Maximum	Median
1	11	11.9	2.775	8.2	16.1	11.9
2	9	12.6	2.076	10.0	15.7	13.3

The overall difference in AASHTO T 393 test results was evaluated by t-test for Round 1 testing vs. Round 2 for both individual specimens as well as the triplicate averages and in both cases the difference in the means were found to be not significant (α = 0.05).

For Round 2 results, a deeper dive into the data was taken whereby multiple t-tests were performed to test for significance of difference in mean Flexibility Index values between the devices (α = 0.05). As shown in Table 5, there is no evidence of statistical differences between devices. This analysis was performed on individual specimen results.

TABLE 5 SUMMARY OF T-TESTS FOR SIGNIFICANCE OF DIFFERENCE IN FI BETWEEN DEVICES FOR ROUND 2

t-Test Parameter	*C*	*A*	*C*	*B*	*A*	*B*
Mean	14.23	12.91	14.23	11.87	12.91	11.87
Variance	33.60	34.68	33.60	9.02	34.68	9.02
Observations	6	6	6	15	6	15
Hypothesized Mean Difference	0.00		0.00		0.00	
df	10		6		6	
t Stat	0.39		0.95		0.41	
P(T<=t) one-tail	0.351		0.189		0.347	
t Critical one-tail	1.81		1.94		1.94	
P(T<=t) two-tail	0.702		0.378		0.695	
t Critical two-tail	2.23		2.45		2.45	
Difference	Not Significant		Not Significant		Not Significant	

To verify the effect of Lab and Device variables on FI values, separate one-way ANOVA tests were performed as summarized in Table 6. While the effect of Lab on the variability in FI values was evident in Round 1, it was found to be statistically insignificant for Round 2. Ultimately no statistical difference between devices was found in either Round 1 or in Round 2.

TABLE 6 SUMMARY OF ONE-WAY ANOVA TESTS FOR ROUNDS 1 AND 2

Source of Variance	**Round 1**		**Round2**	
	F	Prob (F>Fcrit.)	F	Prob (F>Fcrit.)
Device	0.77	0.471	0.63	0.540
Lab	3.73	0.0062	0.7	0.670

This data presents a key decision that needs to be addressed regarding the treatment of outliers. A wide error bar is seen on device 'A,' test Round 1. When all the test data from the study is checked, one outlier can be found from device 'A' (Figure 8). This outlier was not removed from the dataset since it is not an outlier when analyzed in its triplicate average (Figure 9), and in fact Lab 2 (Device B) has one outlier analyzed in triplicate due to the fact that two of the three results were nearly identical from their test results. Additionally, a sample threshold line of 10.0 for the FI is projected on both Figure 8 and Figure 9 to demonstrate the number of individual samples that would be considered below the acceptance threshold.

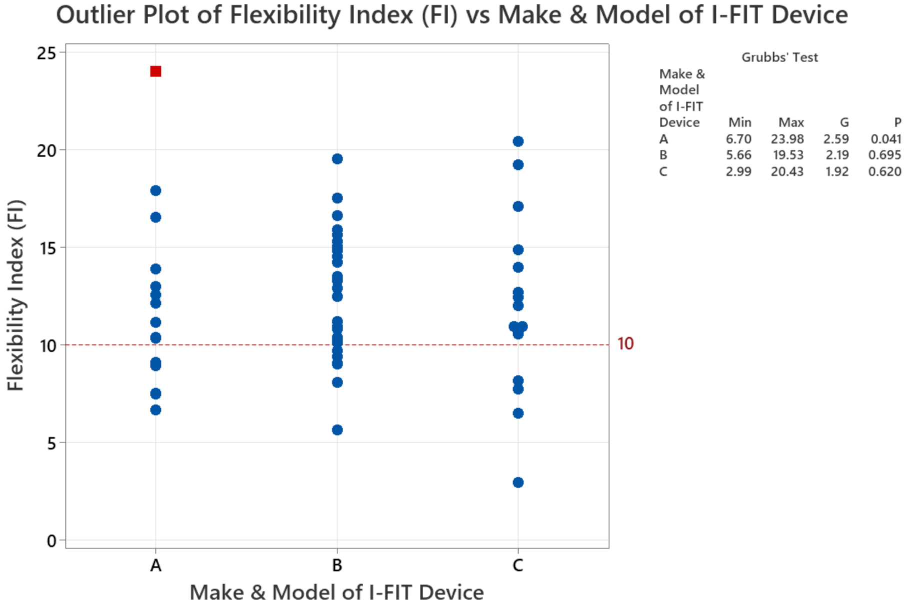

FIGURE 8 OUTLIER TEST ON FLEXIBILITY INDEX FOR INDIVIDUAL SPECIMEN RESULTS BY DEVICE

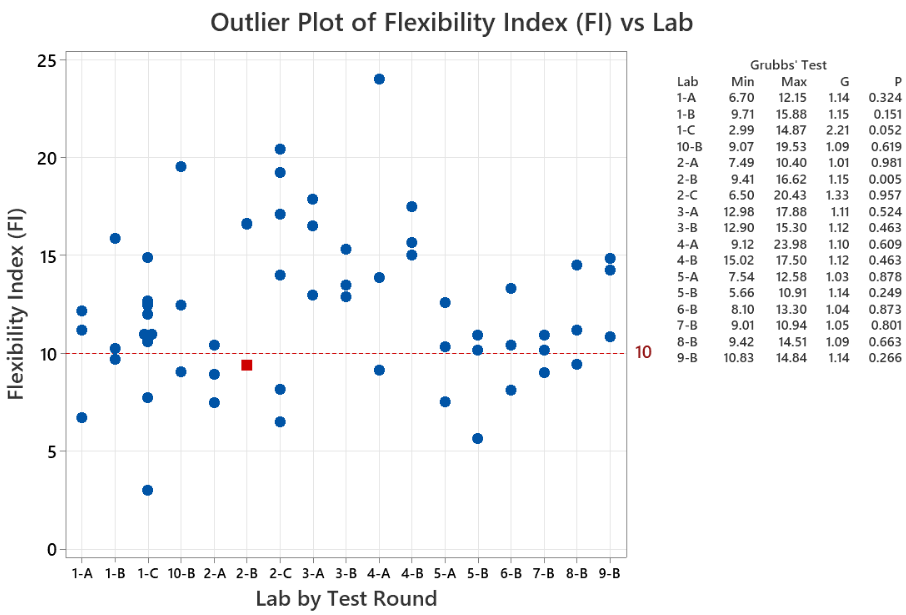

FIGURE 9 OUTLIER TEST ON FLEXIBILITY INDEX FOR INDIVIDUAL SPECIMEN RESULTS BY LAB & TEST ROUND

ASTM D 8225 Indirect Tension Asphalt Cracking Test
Comparison of Test Designs

TABLE 7 SUMMARY OF TEST DESIGNS IN ROUND 1 AND 2 FOR ASTM8225 CT INDEX

Parameter	Round 1	Round 2
Number of Labs	10	8
Number of Devices	3	3
Number of Tests	36	31
Air Voids	7% ± 0.2%	6.9% ± 0.2%
Mean Lab CTIndex	148 ± 16	139 ± 29
Within-Lab Variation	Mean: 15% Minimum: 4% Maximum: 28%	Mean: 13% Minimum: 5% Maximum: 22%
Between-Labs Variation	11%	21%

Differences in ASTM D 8225 Results Within Labs
As shown in Figure 10, there was no significant changes between the two rounds of testing for CT Index values when looking at all test results. However, when the data is divided more, the three devices used did produce notably different results in Round 2. For example, Devices B and C produced lower values with less variability compared to Device A in Round 2 (Figure 11).

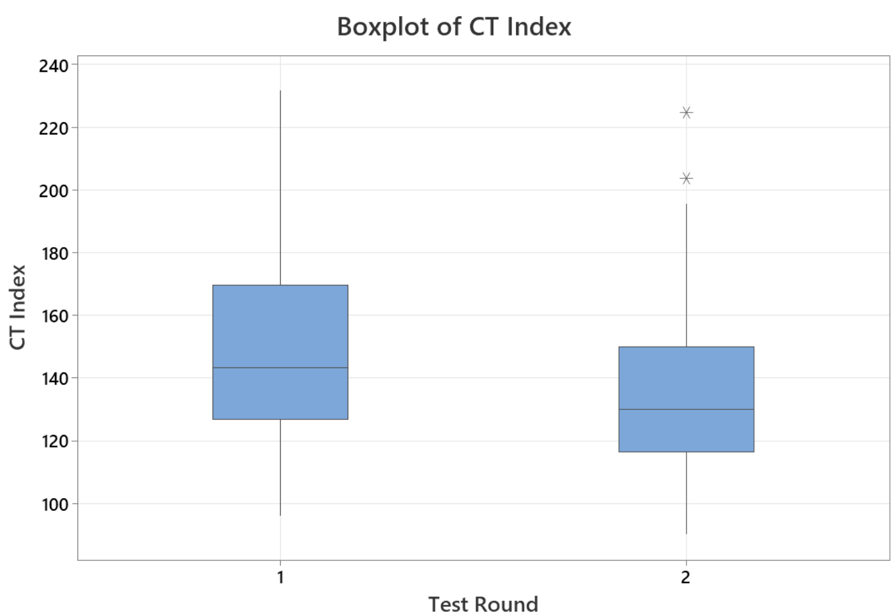

FIGURE 10 DISTRIBUTIONS OF CT INDEX VALUES

While the average CT Index across the devices were closer during the first round (Figure 11), the post-peak slope values in the second round were slightly higher, but also more consistent (Figure 13).

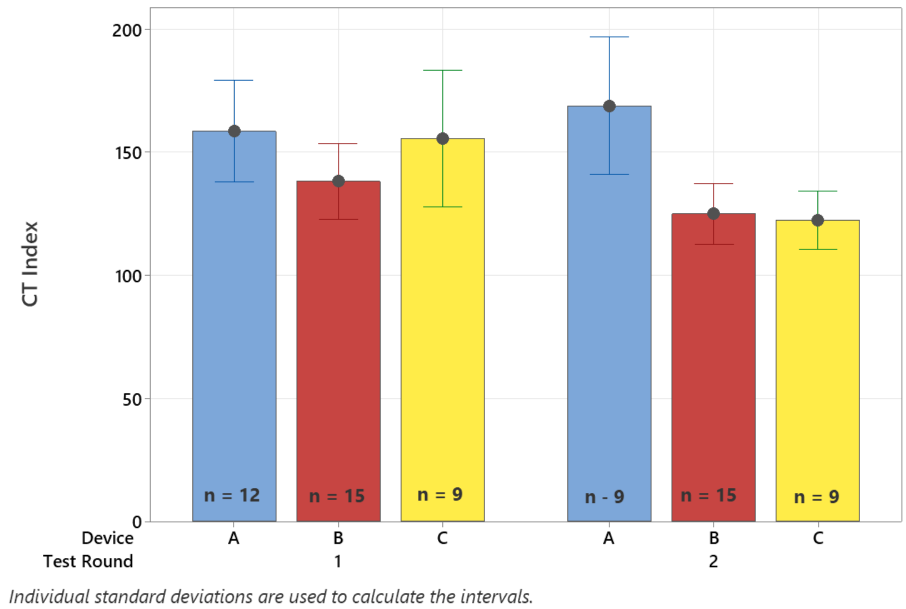

FIGURE 11 INTERVAL PLOT OF CT INDEX VALUES WITHIN DEVICES IN ROUND 1 AND ROUND 2 (ERROR BARS REPRESENT 95% CONFIDENCE INTERVAL)

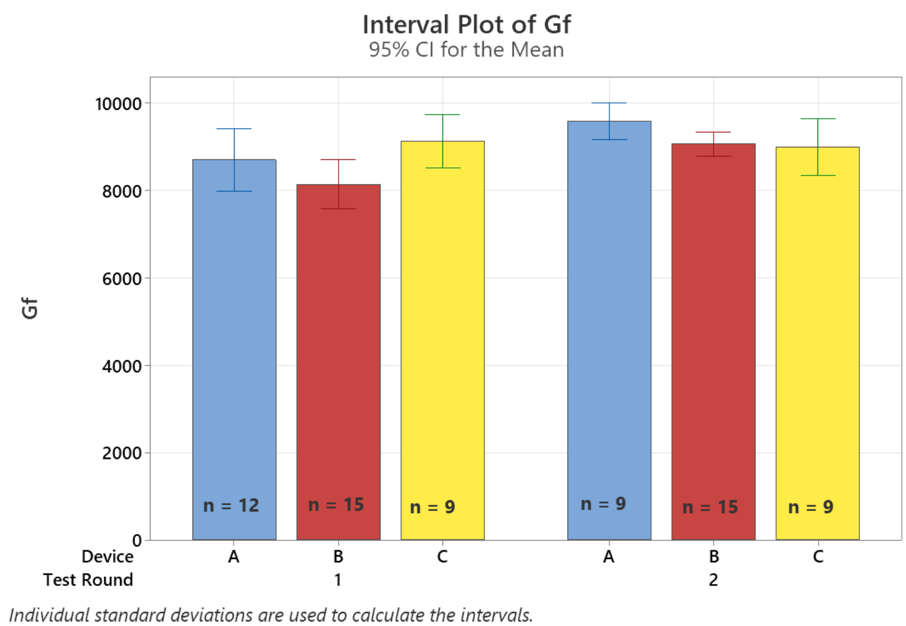

FIGURE 12 INTERVAL PLOT OF G_F (FRACTURE ENERGY) VALUES WITHIN DEVICES IN ROUND 1 AND ROUND 2 (ERROR BARS REPRESENT 95% CONFIDENCE INTERVAL)

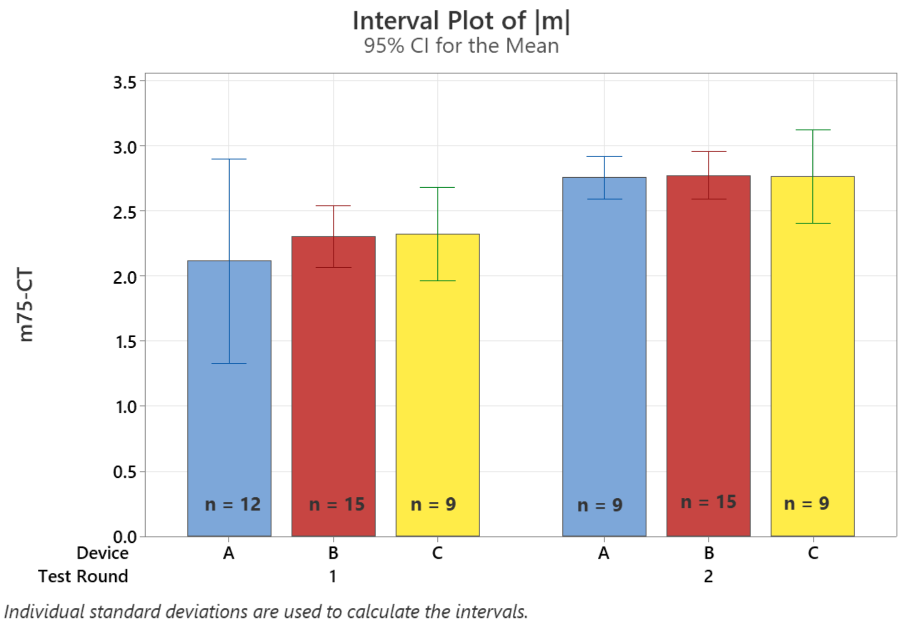

FIGURE 13 INTERVAL PLOT OF POST-PEAK SLOPE VALUES $|M|_{75}$ WITHIN DEVICES IN ROUND 1 AND ROUND 2 (ERROR BARS REPRESENT 95% CONFIDENCE INTERVAL)

Statistical Analysis of Difference in CT Index due to Influencing Factors

The multiple t-tests for significance of difference in mean CT Index values between the devices manufactured by different companies were performed at level of significance α=0.05. As shown in Table 8 a statistically strong variance was found between device A as compared with the other two devices.

TABLE 8 SUMMARY OF T-TESTS FOR SIGNIFICANCE OF DIFFERENCE IN CT INDEX BETWEEN DEVICES FOR ROUND 2

t-test parameter	***C***	***A***	***B***
Mean	120	208	125
Variance	250	221	498
Observations	10	3	15
Hypothesized Mean Difference		0	0
df		3	23
t Stat		-8.82	-0.59
P(T<=t) one-tail		0.002	0.281
t Critical one-tail		2.35	1.71
P(T<=t) two-tail		0.003	0.561
t Critical two-tail		3.18	2.07
Difference with C		**Significant**	**Not Significant**

t-test parameter	***A***	***B***
Mean	208	125
Variance	221	498
Observations	3	15
Hypothesized Mean Difference	0	
df	4	
t Stat	8.04	
P(T<=t) one-tail	0.001	
t Critical one-tail	2.13	
P(T<=t) two-tail	0.001	
t Critical two-tail	2.78	
Difference	**Significant**	

To verify the effect of Lab and Device variables on the CT Index values, separate one-way ANOVA tests was done as summarized in Table 9. The highly significant F-statistics values indicate highly important effect of both Lab and Device. Note that in Round 1, both factors were found to be insignificant. The only other independent variable in the CT Index dataset was Air Voids. As shown in Table 9, the Air Voids affected the CT Index significantly in Round 1 but not in Round 2.

TABLE 9 SUMMARY OF ONE-WAY ANOVA TESTS FOR ROUND 2

Source of Variance	Round 1		Round 2	
	F	Prob ($F>F_{crit.}$)	F	Prob ($F>F_{crit.}$)
Device	1.62	0.213	17.5	1.64E-06
Lab	0.86	0.568	7.1	1.42E-04
Air Voids	2.29	0.043	0.52	0.831

Statistical Analysis of Difference in CT Index between Round 1 and Round 2

The final phase of analysis investigated the difference in CT Index between Round 1 and Round 2 for "suspect" pair of variables using the t-test for two means with unequal variance. When all the factors were analyzed together, no significant difference between Round 1 and Round 2 was found at α=0.05 (Table 10). However, Device C results differed between rounds due to higher variance in Round 1 (Table 11).

TABLE 10 SIGNIFICANCE OF DIFFERENCE BETWEEN CT INDEX FOR OVERALL ROUND 1 AND ROUND 2

t-test parameter	*Round 1*	*Round 2*
Mean	149	135
Variance	1032	1110
Observations	36	31
Hypothesized Mean Difference	0	
df	63	
t Stat	1.761	
P(T<=t) one-tail	0.042	
t Critical one-tail	1.669	
P(T<=t) two-tail	0.083	
t Critical two-tail	1.998	
Difference	**Not significant**	

TABLE 11 SIGNIFICANCE OF DIFFERENCE BETWEEN CT INDEX FOR DEVICE C ROUND 1 AND ROUND 2

t-test parameter	*Round 1*	*Round 2*
Mean	156	120
Variance	1312	250
Observations	9	10
Hypothesized Mean Difference	0	
df	11	
t Stat	2.691	
P(T<=t) one-tail	0.010	
t Critical one-tail	1.796	
P(T<=t) two-tail	0.021	
t Critical two-tail	2.201	
Difference	**Significant**	

While these overall results are promising, there are some 'caution flags' to be found – in that certain devices have better repeatability and even how the specimens are fabricated can play a significant role in the test results. Due to the wider variability of the CT Index testing, there were no outliers identified by Round (1 or 2) or but when split by device or laboratory, individual outliers could be found (Figure 14 and Figure 15) which lie within the realm of all results, similar to what was found with the i-FIT data (Figure 8 and Figure 9).

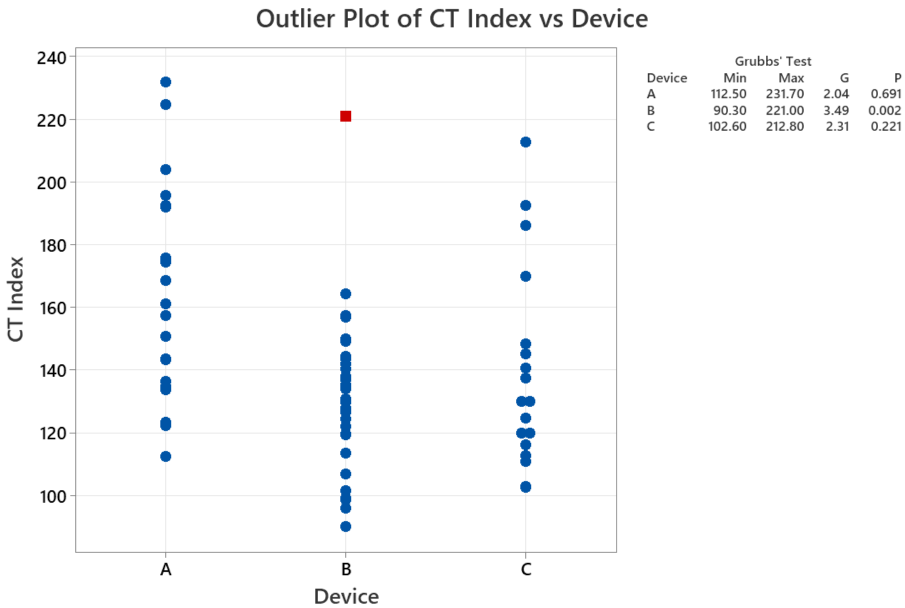

FIGURE 14 OUTLIER TEST ON CT INDEX FOR INDIVIDUAL SPECIMEN RESULTS BY DEVICE

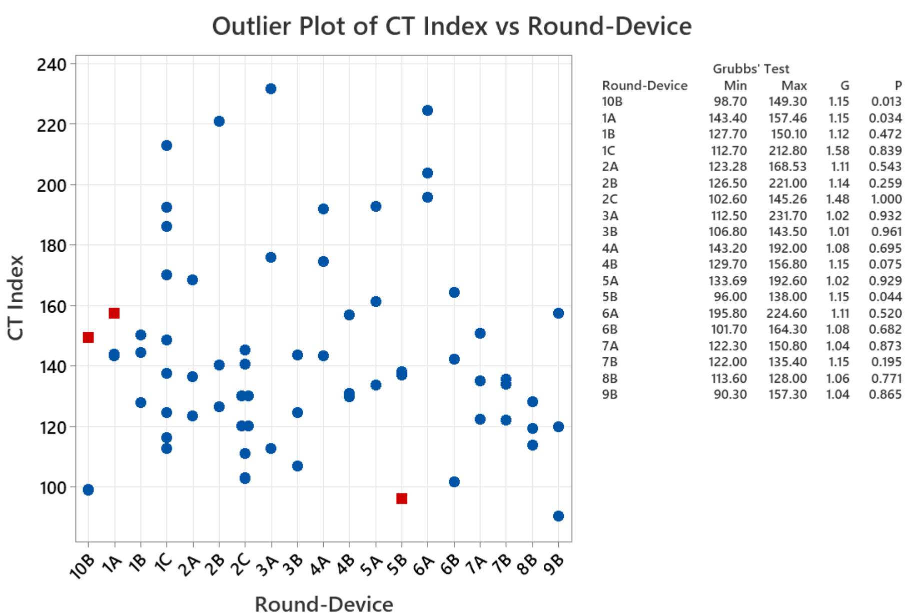

FIGURE 15 OUTLIER TEST ON FLEXIBILITY INDEX FOR INDIVIDUAL SPECIMEN RESULTS BY LAB & TEST ROUND

CHAPTER 4 Conclusions, Recommendations and Suggested Research

While the results in Chapter 3 demonstrate reasonable levels of repeatability for a research study, it can be said that this study successfully established variability of this test protocols to help Vermont AoT to appropriately administer these performance tests in specifications for asphalt pavements. The research performed as an inter-laboratory study of performance tests for asphalt pavement mixtures to assist VTrans in establishing benchmarking for these tests with Vermont-specific material. The research team sampled sufficient plant-produced material in a single event to produce specimens for Hamburg Wheel Tracking (AASHTO T 324), Flexibility Index (AASHTO T 393), and IDEAL-CT (ASTM 8225) testing to be performed on 2 separate occasions by 11 different laboratories across the region ranging from DOTs to Contractors to Universities.

The research team kept all production factors the same to minimize variability in specimen preparation and ultimately, the results showed some of the lowest variability to date of inter-laboratory testing for mixes using the above-mentioned tests.

- In all cases, the Round 1 and Round 2 test data were statistically equal.
- For Hamburg Wheel Tracking, all specimens met the 20,000 cycle test length without a Stripping Inflection Point
- Variability of results across all HWT tests was minimal (<3 mm) between the minimum and maximum results from all testing
- iFIT variability was smaller than that of IDEAL-CT testing, however both had several samples that would have been considered 'low' for acceptance.
- For both IDEAL-CT and iFIT testing, the fracture energy (G_f) and post-peak slope |m| values were quite uniform across the dataset. In both tests, the post-peak slope was more uniform across the test devices in the 2nd round of testing.

CHAPTER 5 Implementation of Research Results

Implementation of the findings of this report will primarily be left to the bituminous engineers at VTrans. The research team disseminated test results to all participating laboratories. Additionally, a presentation was given at the 2023 NEAUPG meeting in Providence, Rhode Island. The researchers plan to prepare a journal article based on the results of this paper in conjunction with members of the VT AOT TAC to submit to a relevant journal.

Effects of this research will be implemented as updates to the *Bituminous Concrete Mix Design Submittal Policy* at VTrans.

References

1. AASHTO, T 245 (2022) Resistance to Plastic Flow of Asphalt Mixtures Using Marshall Apparatus. Digital Publication. American Association of State Highway and Transportation Officials, Washington DC.

2. AASHTO, T 320 (2022) Determining the Permanent Shear Strain and Stiffness of Asphalt Mixtures Using the Superpave Shear Tester (SST). Digital Publication. American Association of State Highway and Transportation Officials, Washington DC.

3. West, R., C., Rodezno, F., Leiva, and F., Yin. NCHRP Project 20-07/Task 406 Development of a Framework for Balanced Mix Design. Transportation Research Board, Washington, D.C., 2018.

4. *Bituminous Concrete Mix Design Submittal Policy,* Vermont Agency of Transportation Highway Division, February 8, 2022; https://outside.vermont.gov/agency/VTRANS/external/docs/construction/04MatTestCert/BitConcMat/Bituminous%20Concrete%20Mix%20Design%20Submittal%20Policy.pdf

5. AASHTO, T 324 (2022) Hamburg Wheel-Track Testing of Compacted Asphalt Mixtures. Digital Publication. American Association of State Highway and Transportation Officials, Washington DC.

6. AASHTO, T 393 (2022) Determining the Fracture Potential of Asphalt Mixtures Using the Illinois Flexibility Index Test (I-FIT). Digital Publication. American Association of State Highway and Transportation Officials, Washington DC.

7. ASTM D8225, Standard Test Method for Determination of Cracking Tolerance Index of Asphalt Mixture Using the Indirect Tensile Cracking Test at Intermediate Temperature. Digital Publication. Book 04.03 American Society for Testing and Materials. DOI: 10.1520/D8225-19

8. Nener-Plant, D. and A. Schwartz, "Balanced Mix Design Benchmarking of Field-Produced Mixtures in Vermont." Presented at the Northeast Asphalt User and Producer's Group Annual Meeting, Albany, NY, November, 2022. https://neaupg.engr.uconn.edu/wp-content/uploads/sites/2932/2022/11/2022_NEAUPG_BMD-Benchmarking-of-Field-Produced-Asphalt-Mixtures-in-VT_20221014.pdf

9. Barros, Luiza Helena, "Influence of Mix Design Parameters on Performance of Balanced Asphalt Mixtures" (2018). Open Access Theses & Dissertations. 1402.; https://digitalcommons.utep.edu/open_etd/1402

10. Wu, S., Zhang, W., Shen, S., Muhunthan, B., & Mohammad, L. N. (2017). Short-Term Performance and Evolution of Material Properties of Warm- and Hot-Mix Asphalt Pavements: Case Studies. Transportation Research Record, 2631(1), 39-54. https://doi.org/10.3141/2631-05

Appendix A – Laboratory Instructions

2023 VTrans Asphalt Mixture Performance Testing Instructions

General:

1. If you have not received proficiency samples by December 19th, 2022 please contact:

   Connecticut Advanced Pavement Laboratory (CAP Lab) at UConn
   (860) 486 5956

2. Be sure to complete the Organization contact information at the top of the Report Form spreadsheet.

3. Complete each of the procedures in accordance with the instructions provided below. Please return the completed form to Scott Zinke (scott.zinke@uconn.edu) **no later than Friday, February 3, 2023.**

4. **The following instructions will reference the VTrans 2022 Bituminous Mix Design Submittal Policy available here and relevant AASHTO and ASTM standards identified below.**

   Note: Please be sure to submit the Excel Workbook as opposed to a .pdf version so macros may used for data extraction.

T 324:

1. Prior to testing, ensure the HWT device is set for data acquisition such that the deformation at each of the 11 sensor locations (AASHTO T 324, Section 5.2.1) is recorded at every 2 passes or 1 cycle.
2. Use the four (4) sample specimens labeled "T 324".
3. Cut specimens along equal secant lines in accordance with AASHTO T 324, Section 6.4.2. The gap between specimens joined together shall not exceed 7.5 mm.
4. Mount the prepared specimens in accordance with AASHTO T 324, Section 8.2.
5. Follow the instructions for placement of the mounted specimens into the device as well as machine initiation in accordance with AASHTO T 324, Sections 8.3, 8.4, 8.5 and 8.6.
6. The temperature (AASHTO T 324, Section 8.6.1) shall be set to 45°C.
7. The maximum allowable rut depth (AASHTO T 324. Section 8.6.2) shall be set to 12.5 mm or 0.5 in.
8. The maximum number of passes (AASHTO T 324, Section 8.6.3) shall be set to 20,000. If the machine uses the number of cycles as opposed to the number of passes then the maximum number of cycles shall be set to 10,000.

9. Submerge the samples at the test temperature of 45°C. Precondition the submerged samples at test temperature for a minimum of 45 minutes and a maximum of 60 minutes.
10. Proceed with testing in accordance with AASHTO T 324, Section 8.8 or Section 8.9
11. Report the number of passes/cycles to failure in the appropriate entry section on the report form for each set.
12. Report the average rut depth in the appropriate entry section on the report form for each set.
13. Report the Creep Slope, Strip Slope and Stripping Inflection Point if reported by the machine software in the appropriate entry sections on the report form for each set.

D8225:

1. Use the three (3) sample specimens labeled "D8225".
2. Precondition the specimens at 25°C ± 1.0 for 2 h C ± 10 min in accordance with Section 9.1 of ASTM D8225. If the water bag method is used to condition the sample, ensure that the samples are not saturated by placing them in a plastic bag. If the bag leaks and the specimen becomes wet, dry the specimen to a constant mass and begin conditioning again.
3. Inspect the test head fixture to ensure all parts are freely moving and contact surfaces are clean.
4. Load and test each specimen in accordance with ASTM D8225. Sections 9.3, 9.4 and 9.5.
5. Testing of each specimen shall be completed within 4 minutes of removal from conditioning.
6. The CT_{index} shall be calculated using the software provided by the equipment manufacturer. As an alternative, the calculation may be completed using the Rutgers Asphalt Analysis Tool-pack (RAAT-Pack) software which is available through the Rutgers Asphalt Laboratory website.
7. Report the specimen thickness and diameter in the appropriate entry section on the report form for each specimen.
8. Report the specimen displacement at 75% of peak load in the appropriate entry section on the report form for each specimen.
9. Report the absolute value of the post-peak slope in the appropriate entry section on the report form for each specimen.
10. Report the work of Failure Energy in the appropriate entry section on the report form for each specimen.
11. Report the Work of Failure in the appropriate entry section on the report form for each specimen.
12. Report the CT_{index} in the appropriate entry section on the report form for each specimen.

T 393:

1. Use three (3) of the four (4) semi-circular (a.k.a half-moon) sample specimens labeled T 393.
2. Precondition the specimens at 25°C ± 0.5 for 2 h C ± 10 min in accordance with Section 11.1 of AASHTO T 393. If the water bag method is used to condition the sample, ensure that the samples are not saturated by placing them in a plastic bag.

If the bag leaks and the specimen becomes wet, dry the specimen to a constant mass and begin conditioning again.

3. Load and test each specimen in accordance with AASHTO T 393, Sections 11.2, 11.3, 11.3.1, 11.3.2, 11.3.3 to 11.3.2.
4. Testing of each specimen shall be completed within 4 - 6 minutes of removal from conditioning.
5. The Flexibility Index (FI) shall be calculated using the software provided by the equipment manufacturer. As an alternative, the calculation may be completed using the Rutgers Asphalt Analysis Tool-pack (RAAT-Pack) software which is available through the Rutgers Asphalt Laboratory website. Another option for calculation is the software provided by the Illinois Center for Transportation available on their website.
6. Report the specimen thickness, notch length and ligament length in the appropriate entry sections on the report form for each specimen.
7. Report the post-peak slope in the appropriate entry section on the report form for each specimen.
8. Report the Fracture Energy in the appropriate entry section on the report form for each specimen.
9. Report the FI in the appropriate entry section on the report form for each specimen.