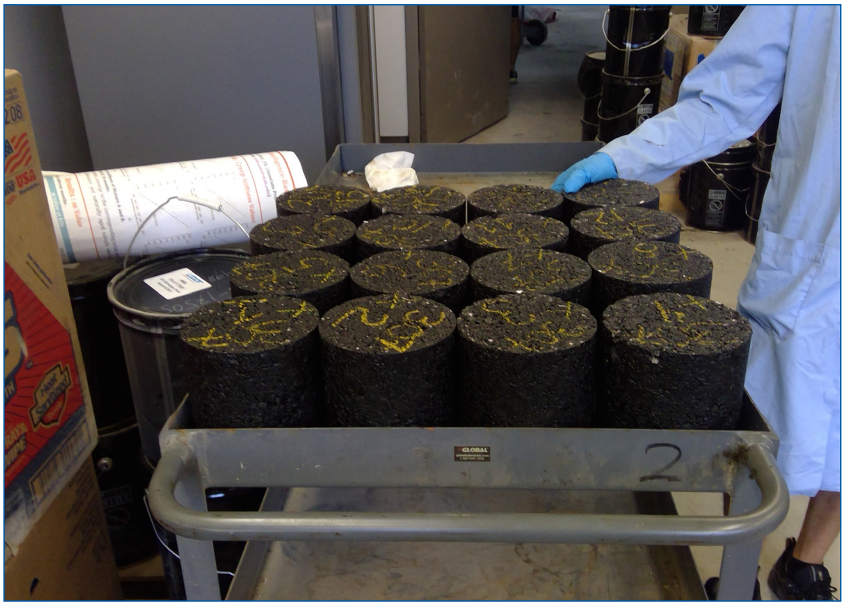
MOUNTAIN-PLAINS CONSORTIUM

MPC 23-492 | P. Romero-Zambrana

VARIABILITY OF THE IDEAL-CT TEST FOR PAVEMENT CRACKING TO ACHIEVE A BALANCED ASPHALT MIX DESIGN



MOUNTAIN-PLAINS CONSORTIUM REGION 8

A University Transportation Center sponsored by the U.S. Department of Transportation serving the Mountain-Plains Region. Consortium members:

Colorado State University
North Dakota State University
South Dakota State University
University of Colorado Denver
University of Denver
University of Utah
Utah State University
University of Wyoming

Technical Report Documentation Page

1. Report No. MPC-589	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Variability of the IDEAL-CT Test for Pavement Cracking to Achieve a Balanced Asphalt Mix Design		5. Report Date February 2023
		6. Performing Organization Code
7. Author(s) Pedro Romero-Zambrana, Ph.D., P.E.		8. Performing Organization Report No. MPC 23-492
9. Performing Organization Name and Address The University of Utah 110 S Central Campus Drive Suite 2000 Salt Lake City, UT 84112		10. Work Unit No. (TRAIS)
		11. Contract or Grant No.
12. Sponsoring Agency Name and Address Mountain-Plains Consortium North Dakota State University PO Box 6050, Fargo, ND 58108		13. Type of Report and Period Covered Final Report
		14. Sponsoring Agency Code
15. Supplementary Notes Supported by a grant from the US DOT, University Transportation Centers Program		

16. Abstract

This report describes the research efforts aimed at determining the repeatability of the IDEAL-CT test for use as part of the balanced asphalt mix design process. Three mixes in common production in the Northern part of Utah were investigated by three different labs experienced in performance testing. Within lab and between lab variability were studied as well as a number of factors likely to cause variability. Recommendations were developed for specimen preparation, minimum number of replicates, and how to discard outliers. It was found that the test could be repeated within laboratory within a coefficient of variation, CV, of 15% and between laboratories within a CV value of 20%. using properly trimmed results. Four replicates are recommended with the result farthest from the mean discarded. No correlation between higher CT index and greater durability was identified in this study.

17. Key Word asphalt mixtures, cracking, mix design, pavement performance, test procedures		18. Distribution Statement Public distribution	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 34	22. Price n/a

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Variability of the IDEAL-CT Test for Pavement Cracking to Achieve a Balanced Asphalt Mix Design

Pedro Romero, Ph.D., P.E.
Department of Civil and Environmental Engineering
The University of Utah
Salt Lake City, Utah

February 2023

Acknowledgments

The authors acknowledge the Utah Department of Transportation (UDOT) for funding this research, and the following individuals from UDOT on the Technical Advisory Committee for helping to guide the research:

- Scott Nussbaum
- Howard Anderson
- David Stevens

The authors also acknowledge the following individuals and their work. Most of the information contained in this report is the result of their hard work. This includes the following:

- Abu Sufian Mohammad Asib, Ph.D. University of Utah
- Abdullahal Mamun, Ph.D. Candidate, Civil Engineering, University of Utah
- Tim Biel, PEPG Consulting
- Faramarz Safazadeh, Ph.D., PEPG Consulting
- Michael VanMilligen, PEPG Consulting
- Clark Allen, Utah Department of Transportation
- Mike White, Utah Department of Transportation

The authors also acknowledge the financial contribution from the Mountains-Plains Consortium.

Disclaimer

ABSTRACT

This report describes the research efforts aimed at determining the repeatability of the IDEAL-CT test for use as part of the balanced asphalt mix design process. Three mixes in common production in northern Utah were investigated by three different labs experienced in performance testing. Within-lab and between-lab variability were studied as well as a number of factors likely to cause variability. Recommendations were developed for specimen preparation, minimum number of replicates, and how to discard outliers. Using properly trimmed results, it was found that the test could be repeated within a laboratory with a coefficient of variation (CV) of 15% and between laboratories within a CV value of 20%. Four replicates are recommended with the result farthest from the mean discarded. No correlation between a higher CT index and greater durability was identified in this study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

This report describes the research efforts aimed at using the IDEAL-CT test as part of a balanced mix design process. Three dense graded asphalt mixtures in common production in northern Utah were investigated by three different labs experienced in performance testing.

The within-lab and between-lab variability were studied as well as a number of factors likely to cause variability. Using properly trimmed results, it was found that the test could be repeated within laboratory within a coefficient of variation (CV) of 15% and between laboratories within a CV value of 20%. Preparation and testing of four replicate samples is recommended with the result farthest from the mean discarded if the variability exceeds 15%.

Regarding the test procedures, it was found that if the number of gyrations exceeded the design number of gyrations (Ndes) of the mixtures, some aggregates could be broken and the variability increase. In those cases, preparation of thicker samples is recommended. This is possible since specimen thickness is already considered as part of the calculation of the cracking tolerance index.

Finally, it was found that the IDEAL-CT tests meet many requirements for a quality control test; however, no correlation between a higher CT index and greater durability was identified in this study.

1. INTRODUCTION

1.1 Problem Statement

Many transportation agencies across the region instituted their first performance-based test into specification in the early 2000s when the Hamburg Wheel Tracking Device (HWT) was required for mix design verification. Although this action might have solved a rutting issue in pavement systems, it tended to favor harder mixes, leading to cracking issues. Furthermore, many materials engineers feel that the adoption of the SuperPave system alone, notwithstanding HWT effects, caused mixes to be over dry and raveling prone. It was surmised that increasing binder content and binder modification would reduce raveling and cracking tendencies, so a number of specification modifications were undertaken. Some improvement resulted from increasing voids in the mineral aggregate (VMA), reducing maximum nominal aggregate size, reducing voids at compaction, and limiting recycled material. These goals needed to be balanced against rutting for a good performing mix. Since specification modification continued to produce spotty results and because a number of cracking tests were being developed, there has been a nationwide effort to find an appropriate test that could discriminate between mixes that crack and mixes that do not crack. The following criteria was desired to be met:

- The test must have meaning. It must be able to rank mixes in order from less cracking to more cracking susceptible.
- The test must be rapid. It must produce results within 24 hours of sampling.
- The test must be precise. Tests with high variation cannot be used for specification. A low number of samples and a low number of replicates must be required.
- The test must be repeatable between labs. QC and QA must be equivalent or processes cannot be controlled.
- The test must be simple. It must lend itself to field lab conditions. Sample fabrication must involve the least amount of manipulation. Conditioning must not require sophisticated procedures. Existing training and equipment must be exploited. Analysis must not be complex.
- The test must be inexpensive. A $50k setup limit is desired.

It must be noted here that most of the cracking tests being developed assume a non-uniform material in which a crack is always present (Fujie Zhou, et al. 2022) and the tests focus on the crack propagation properties of the mix. However, this interpretation may or may not always be valid given that some of the tests demonstrate a crack initiation energy and therefore some resistance to crack development. In some mixtures, crack resistance may prove to be as important as resistance to crack propagation.

Many highway agencies have rejected tests such as Semi-Circular Bend, Jc, Direct Compact Tension, Illinois Flexibility Index and others due to complexity, timeliness, cost, and sample preparation issues. The most promising candidate to date is the IDEAL CT test developed at the Texas Transportation Institute and described in ASTM D8225-19.

Development and adoption of a laboratory test that can evaluate asphalt mixtures for intermediate temperature properties has been an important goal during the last few years. The high-temperature properties of asphalt mixtures are being addressed with the HWT device. The low-temperature properties are being addressed with the bending beam rheometer (BBR) test on asphalt mixtures. After some testing and evaluation, the IDEAL-CT test (ASTM D8225) was identified as a promising test to determine intermediate temperature properties. Once the tests for both temperature extremes as well as intermediate temperature properties of asphalt mixtures have been adopted, development of a performance-based specification will be possible, resulting in mix designs that can be tailored to specific environments. The result can be a balanced performance-based mixture optimized on all distresses. This concept is referred to as balanced mix design.

Adoption of an intermediate-temperature cracking test is the final component of the performance-related specifications being adopted. However, as part of adopting the test, it is necessary to understand the within- and between-laboratory variability of the test. Even though the IDEAL-CT test is becoming popular with state agencies, there is no information regarding the variability and reproducibility of the test; therefore, before the test is adopted, it is necessary to understand its variability. Furthermore, some asphalt producers are starting to use the test for mix design, thus it is important that a proper variability study be conducted.

1.2 Objectives

The objectives of this research are:

- Develop a within- and between-laboratory variability statement for the IDEAL-CT test in support of its adoption as a specification for intermediate temperature testing.
- Recommend a preferred sample height.
- Recommend a number of replicate samples and analyses to achieve a reasonable variability.

1.3 Scope

As part of this work, testing of three, plant-produced, "standard" asphalt mixtures was performed at three labs following the procedures outlined in ASTM D8225: *Standard Test Method for Determination of Cracking Tolerance Index of Asphalt Mixture Using the Indirect Tensile Cracking Test at Intermediate Temperature*. All three labs have significant experience in testing and evaluation of asphalt mixtures using this test and all of them used equipment manufactured by the same supplier. The variables for this research are shown in Table 1.1.

Table 1.1 Variables Used in this Study

Variable	Number	Comments
Testing laboratories	3	Lab 1, Lab 2, Lab 3
Asphalt mixtures	3	Plant produced
Aggregate size	1	12.5-mm nominal maximum size
Binder grade	1	All mixtures used PG 64-34 from different suppliers
Sample thickness	2	62-mm and 75-mm
Target air voids	1	7% ± 0.5% - Values outside target were still included
Loading head speed	1	50 mm/min, displacement controlled
Test temperature	1	25ºC
Samples per condition	9	

2. RESEARCH METHODS

2.1 Overview

Evaluating the variability of the IDEAL CT tests was done using three different laboratories with extensive testing experience. One was a state materials lab, one was a university lab, and the other was a private consulting lab. Hot-mix asphalt was collected at the plant from three different suppliers and distributed among the different labs. Each lab proceeded to test the samples in a consistent manner. The results from each lab were compared with the purpose of evaluating the variability of the testing procedure and not the expected performance of the mix itself. Since the performance of the mix is not being considered, the source of the mix is given as an identifier only; mix properties are not relevant and are not included in this report.

2.2 Sample Preparation

Hot-mix asphalt was delivered in metal buckets to each of the participant laboratories. While it is known that slight differences between materials can exist due to segregation and sampling error, it was assumed that all of the buckets contained material with identical compositions.

When the labs were ready for testing, the buckets were heated overnight to a temperature of 120ºC while keeping the lid on to prevent further aging. Once the mix was pliable, the material for each sample replicate was separated by quartering or using a riffle splitter, as appropriate. The hot-mix was weighted based on the given maximum specific gravity (Gmm) so that specimens could be compacted to the target air voids at the two specified heights. The mix was heated to the appropriate compacting temperature and compacted to height using the Superpave Gyratory Compactor (SGC) following the procedures described in AASHTO T312: *Standard Method of Test for Preparing and Determining the Density of Asphalt Mixtures by Means of the Superpave Gyratory Compactor*. Once compacted, the air voids of each sample were determined following the procedures described in AASHTO T269: *Percent Air Voids in Compacted Dense and Open Asphalt Mixtures*. The number of gyrations to reach compaction and the air voids for each sample were recorded. Samples whose air voids fell outside the specified range of 7% ± 0.5% were still tested.

2.3 IDEAL CT Testing

Testing was done based on ASTM D8225: *Standard Test Method for Determination of Cracking Tolerance Index of Asphalt Mixture Using the Indirect Tensile Cracking Test at Intermediate Temperature*. To ensure consistency, the samples were tested within eight to 20 hours after compaction.

All three laboratories used the same brand of servo-hydraulic equipment, thus the effects of equipment type (i.e., servo-hydraulic vs. electromechanics) are not evaluated as part of this study. To ensure proper temperature during testing, a “dummy sample” fitted with a calibrated thermocouple accompanied the sample during conditioning and then testing. The speed of the loading head was constant at 50 mm/min. A picture of the testing equipment at the University of Utah is shown in Figure 2.1.

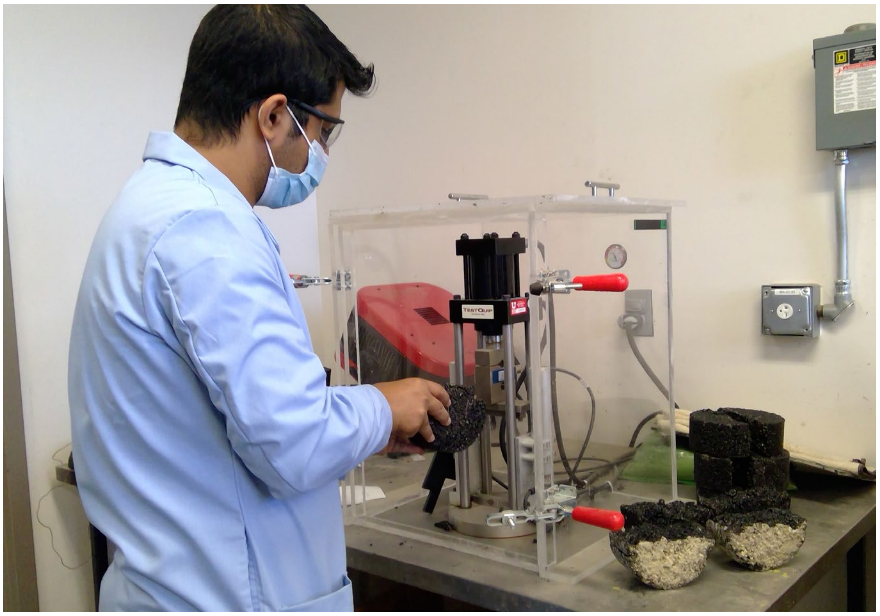


Figure 2.1 IDEAL CT Testing at the University of Utah

2.4 Data Analysis

The testing equipment digitally collected load-displacement data and calculated the cracking tolerance (CT) index for each sample. This process resulted in 18 values per lab for each mix (9 reps x 2 heights) for a total of 162 values. The CT index was collected and analyzed using different statistical techniques as described in Chapter 3.

2.5 Summary

This chapter describes how asphalt mix was collected from three different plants and distributed to the three different labs. Standard protocols were used to handle the material as well as the sample compaction. Samples were compacted to target air voids of 7% ± 0.5% and heights of 62 mm and 75 mm. Testing was done within 20 hours of compaction and all three laboratories used the same brand of servo-hydraulic equipment. The CT index of each sample was determined and used for analysis.

3. DATA ANALYSIS

3.1 Overview

This chapter presents the data obtained from the IDEAL CT tests on three different mixtures as reported by the three participant laboratories.

3.2 Data Collected

A total of 167 values were collected for this study. A summary of the data is shown in Table 3.1 with the complete set of data shown in Appendix A. All mixtures are production mixes used in actual paving projects. However, it should be noted that the labels on the mixtures are used as identifiers only; the results are not meant to compare the mixtures to each other since they were not designed based on their CT index. However, the data show that the mixtures selected are different in terms of their CT index and thus their potential performance.

Table 3.1 Summary of Data Collected

	All Data	Mix 1	Mix 2	Mix 3
Mean	124.1	91.9	106.5	178.7
Standard Error	3.831	2.967	4.177	5.681
Median	112.6	88.0	102.6	177.2
Mode	106.1	87.5	102.1	158.3
Standard Deviation	49.512	23.172	30.407	41.355
As % of mean	--	25.2	28.5	23.1
Sample Variance	2451.42	536.92	924.60	1710.21
Kurtosis	0.3974	-0.3081	0.2456	0.1804
Skewness	0.9019	0.2652	0.6510	0.4546
Range	235.5	106.7	129.0	182.1
Minimum	43.5	43.5	55.4	96.9
Maximum	279.0	150.2	184.4	279.0
Sum	20722.8	5604.3	5647.1	9471.4
Count	167	61	53	53
Average Gyrations*	--	55	122	50

* the average number of gyrations required to compact all samples by all labs

3.2.1 Variance

Table 3.1 shows that the variances (standard deviation) of the results are not the same for each mix (heteroscedasticity). To determine if the difference in variance is statistically significant, an F-test for equal variance was performed between the largest and the smallest variance. The results, shown in Table 3.2, indicate that the hypothesis, which states that variances for these two mixtures are the same, is rejected ; therefore, we conclude that the two variances are different. While it is possible that if the variances for all three mixtures were analyzed together using a

Bartlett's test, the conclusion might be different; in reality, the mixtures are not meant to be compared to each other. Based on these results, each mix will be treated separately in this report.

Table 3.2 F-Test for Equal Variance

	Mix 3	*Mix 1*	*Conclusion*
Mean	178.7	91.9	
Variance	1710.21	536.92	
Observations	53	61	
df	52	60	
F	3.1852		
P(F≤f) one-tail	1.0492E-05		$p < 0.05$ reject
F Critical one-tail	1.5534		

Table 3.1 also shows that the variance increases as the mean increases. When the square root of the variance (i.e., the standard deviation) is expressed as a percent of the mean, the values range from 23% to 28%. This gives an indication that the coefficient of variation of the population is around 25% before any outlying observations are evaluated and possibly eliminated. The implications of this value will be discussed in Chapter 4.

3.2.2 Normality of Data

Table 3.1 shows that there is some positive skewness on the data. High levels of skewness indicate the data are not normally distributed, a situation generally not desirable since it can undermine some statistical analysis. While there are no official criteria about cut-off values to decide how large the skew must be to indicate non-normality of the data, values between -2 and +2 are considered as acceptable (George and Malley, 2010). As seen on the table, all values are less than 1, thus it is reasonable to consider the data as normally distributed.

An additional method to determine the normality of the data is to graphically plot the data on a quantile-quantile (Q-Q) plot, which shows the distribution of the data against the expected normal distribution. For normally distributed data, observations should lie approximately on a straight line. Possible outliers are points at the ends of the line, distant from the bulk of the observations. This is shown in Figure 3.1.

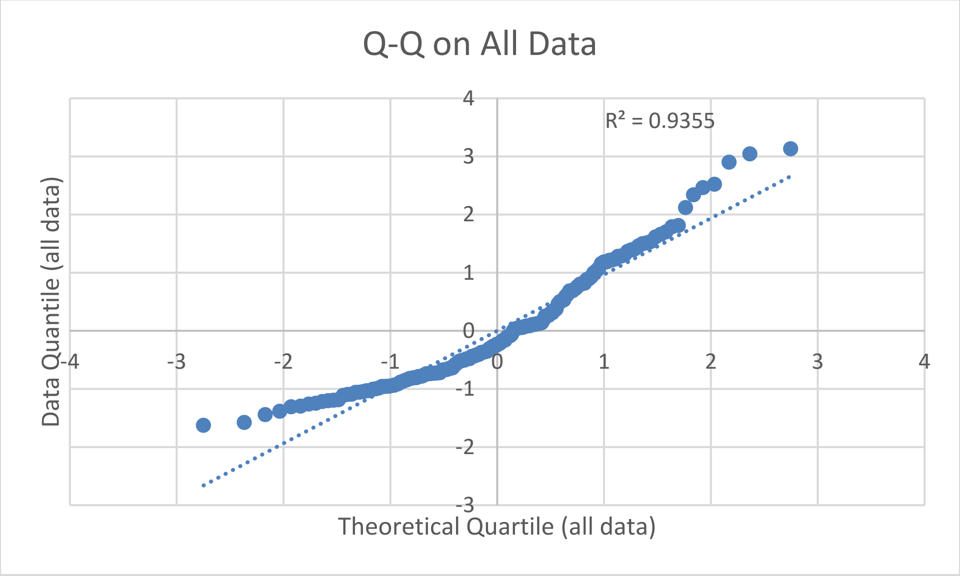


Figure 3.1 Q-Q Plot of All Data

Figure 3.1 shows that the data could be considered normally distributed even though there are some deviations from a straight line. The deviations are likely caused by outliers that require further investigation. While it is impossible to know the cause of all outliers, some of them may be derived from differences in material composition, either from inherent inhomogeneities or from issues during the compaction process.

3.3 Number of Gyrations

An important parameter recorded during compaction was the number of gyrations. Even though all mixtures were designed for 75 gyrations, one of the mixtures seemed to require, on average, a high number of gyrations in relation to the other mixtures to reach the desired height of the specimen (122 vs ~50 gyrations). It was hypothesized that once the number of gyrations reaches values above the Ndes (75 gyrations for these mixtures), aggregates are broken, resulting in uncoated surfaces within the mix, which affects the CT index and increases the variability of the results. To verify this hypothesis, the data for each mixture were separated into those values that required more than 75 gyrations and those that did not. A student's t-test assuming unequal variance was performed for each mix with the results shown in Table 3.3.

Table 3.3 Analysis of Number of Gyrations

		Gyrations > 75	***Gyrations < 75***	**Conclusion**
Mix 1	Mean	74.8	94.8	
	Variance	771.26	450.66	
	Observations	9	52	
	t-Stat	-2.0565		
	t-critical (2-tail)	2.2281		
	$p(T \leq t)$ two tail	0.0667		$p > 0.05$ fail to reject
Mix 2	Mean	99.8	122.3	
	Variance	800.08	907.52	
	Observations	37	16	
	t-Stat	-2.5440		
	t-critical (2-tail)	2.0518		
	$p(T \leq t)$ two tail	0.0170		$p < 0.05$ reject
Mix 3	Mean	159.5	181.2	
	Variance	828.03	1789.21	
	Observations	6	47	
	t-Stat	-1.6293		
	t-critical (2-tail)	2.3060		
	$p(T \leq t)$ two tail	0.1419		$p > 0.05$ fail to reject

The data in Table 3.3 show that, regardless of the mixture analyzed, those samples that required more than 75 gyrations have lower CT. In other words, too many gyrations decrease the CT index. The student's t-test indicates that for those mixtures that, on average, required less than 75 gyrations to compact, the hypothesis that the means are equal is not rejected, meaning there is no statistical difference in their average CT index; however, for the mix that, on average, required 122 gyrations, the hypothesis that the means are equal is rejected. The variance of the results, which is an indicator of the variability, does not show a consistent trend with number of gyrations. In two cases it increased and in one case it decreased.

The data presented in Table 3.3 indicate that data from samples compacted using a high number of gyrations can result in different CT index values. As was previously mentioned, these different values can be considered outliers and not representative of the actual mixture properties. To further verify this hypothesis, the data from samples compacted to more than 75 gyrations were excluded from the analysis and a new Q-Q plot was generated. This is shown in Figure 3.2.

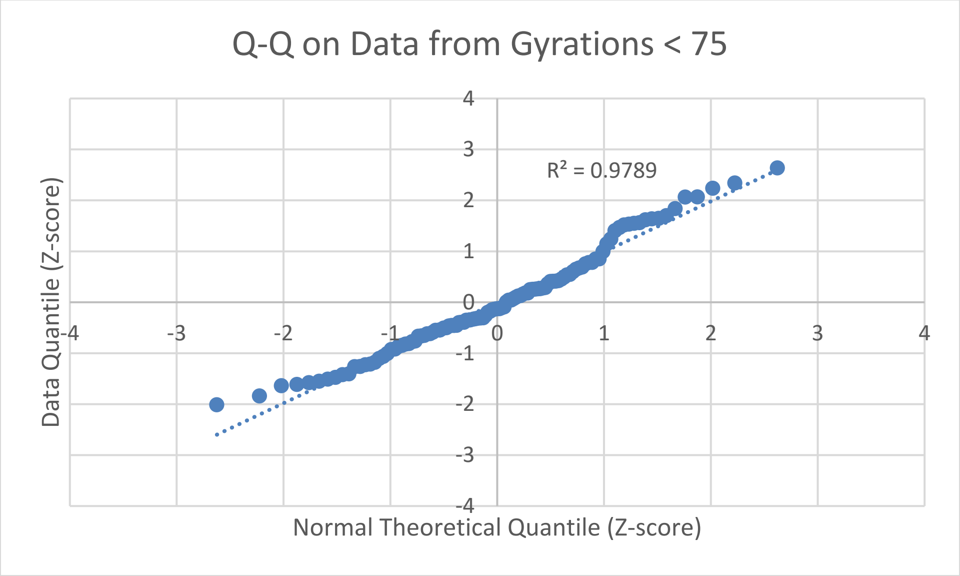


Figure 3.2 Q-Q Plot of Data from Samples Compacted Using Less than 75 gyrations

Comparing Figure 3.1 and Figure 3.2, it is evident that once the data from samples compacted to more than 75 gyrations are removed, the data more closely follow a straight line in the Q-Q plot. In other words, when samples can be compacted using less than 75 gyrations, the resulting CT index values more closely follow a normal distribution. The implications of these results on routine testing will be discussed in Chapter 4; however, based on these observations, it was decided that further data analysis would be done only on specimens that required less than 75 gyrations to compact.

The data shown in Table 3.3 and Figure 3.2 combine samples compacted to heights of 62 mm and 75 mm. It can be argued that the taller samples allow for better aggregate distribution during compaction, thus resulting in a lower number of gyrations (and a more representative value). Further analysis of the data shows that to be the case; for two of the mixtures, the only samples that required more than 75 gyrations were compacted to a height of 62 mm. In other words, samples with a height of 75 mm are easier to compact. For the mix that, on average, required a larger number of gyrations, the majority (22 out of 37) of the 62-mm samples required more than 75 gyrations. This indicates that, in some cases, a height of 75 mm might be preferred and might result in more representative samples.

3.4 Specimen Height

The discussion in Chapter 3.3 indicates that compacting laboratory samples to heights greater than 62 mm might be desirable in some mixtures to prevent the need to use gyrations in excess of the Ndes that decrease the CT index and potentially decrease the number of outlying observations. Conceptually, as long as the volume tested is representative of the material and the plain stress condition is maintained, the results of any indirect-tension-based tests, such as the IDEAL CT, should not depend on the specimen thickness as both the thickness and the diameter are accounted for in the calculations. According to ASTM D8225, for laboratory compacted samples using mixtures with a nominal maximum aggregate size of 19 mm or smaller, the specimen shall be 150 ± 2 mm in diameter and 62 ± 1 mm in thickness. However, the same standard provides the following equation to calculate the CT index:

$$CT_{index} = \frac{t}{62} x \frac{l_{75}}{D} x \frac{G_f}{|m_{75}|} x10^6$$

Where:

G_f = failure energy in Joules/m^2
$|m_{75}|$ = absolute value of post-peak slope
l_{75} = displacement at 75% post peak load in mm
D = specimen diameter in mm
t = specimen thickness in mm

From the CT index equation, it is clear that specimen thickness is already part of the calculations and that a height of 62 mm was chosen for factors other than the mechanics of materials. Nonetheless, to verify that the CT index of specimens with a height of 62 mm is the same as the CT index of specimens with a height of 75 mm (i.e., specimen height does not matter in the results), a student's t-test assuming equal variances was performed on the data with the results shown on Table 3.4. As mentioned in Chapter 3.3, those specimens that required more than 75 gyrations were considered questionable under some circumstances and were excluded from the analysis.

Table 3.4 Analysis of Specimen Height

		62 mm	***75 mm***	**Conclusion**
Mix 1	Mean	90.4	96.5	
	Variance	408.66	456.62	
	Observations	19	32	
	t-Stat	-0.9970		
	t-critical (2-tail)	2.0095		
	p(T≤t) two tail	0.3237		$p > 0.05$ fail to reject
Mix 2	Mean	114.8	125.7	
	Variance	774.80	1010.58	
	Observations	5	11	
	t-Stat	0.6576		
	t-critical (2-tail)	2.1447		
	p(T≤t) two tail	0.5215		$p > 0.05$ fail to reject
Mix 3	Mean	194.8	171.9	
	Variance	2512.93	1153.69	
	Observations	19	28	
	t-Stat	1.876		
	t-critical (2-tail)	2.014		
	p(T≤t) two tail	0.0683		$p > 0.05$ fail to reject

The results presented in Table 3.4 confirm the theory and show there is no statistical difference in the CT index for specimens compacted to a thickness of 62 mm or 75 mm. Furthermore, there is no clear pattern indicating that one or the other would result in a higher CT index or higher variance. In fact, based on the material mechanics, any thickness between those values, or perhaps even thicker, is acceptable. This is clearly reflected in the equation shown in ASTM D8225. It is recognized that samples cored from asphalt pavements might be thinner. In those cases, ASTM D8225 states that the minimum thickness should be 38 mm.

3.5 Effect of Air Voids

It is known that the air void content affects asphalt mixture properties. While ASTM D8225 does not specifically require a target air void content, it states that a typical target value is 7.0%. It also states that while other target air voids can be used, specimens with a difference in air voids greater than ± 0.5% are not comparable.

While Table 1.1 does not identify air voids as a variable in this study, the practicalities of sample compaction resulted in some samples with air voids outside the 7.0% ± 0.5% target. To evaluate the effect of air voids on the CT index, the values were plotted and a trend was determined. Figures 3.3, 3.4, and 3.5 show the effect of air voids for each of the three mixtures tested for samples compacted to less than 75 gyrations.

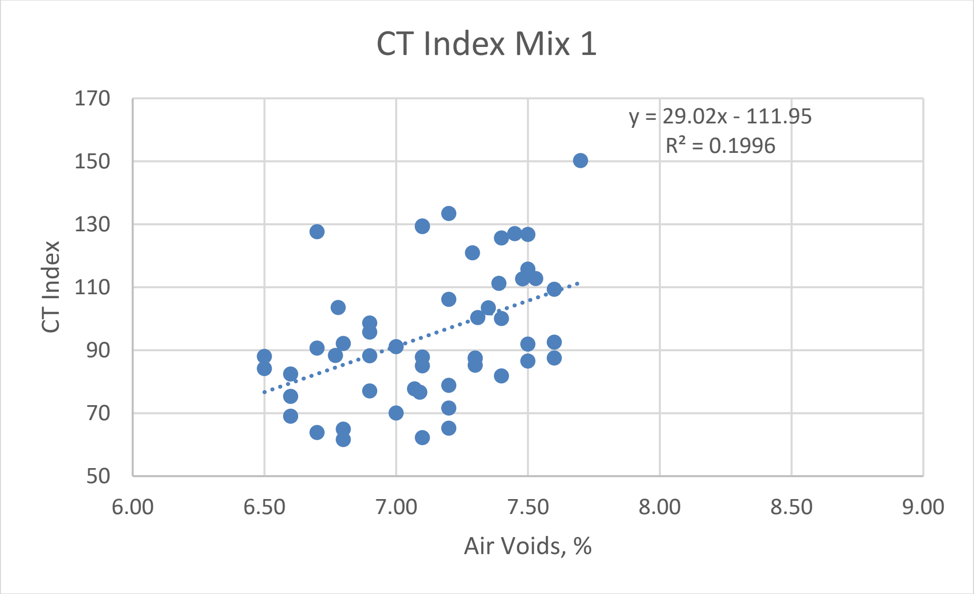


Figure 3.3 Effect of Air Voids on CT Index for Mix 1

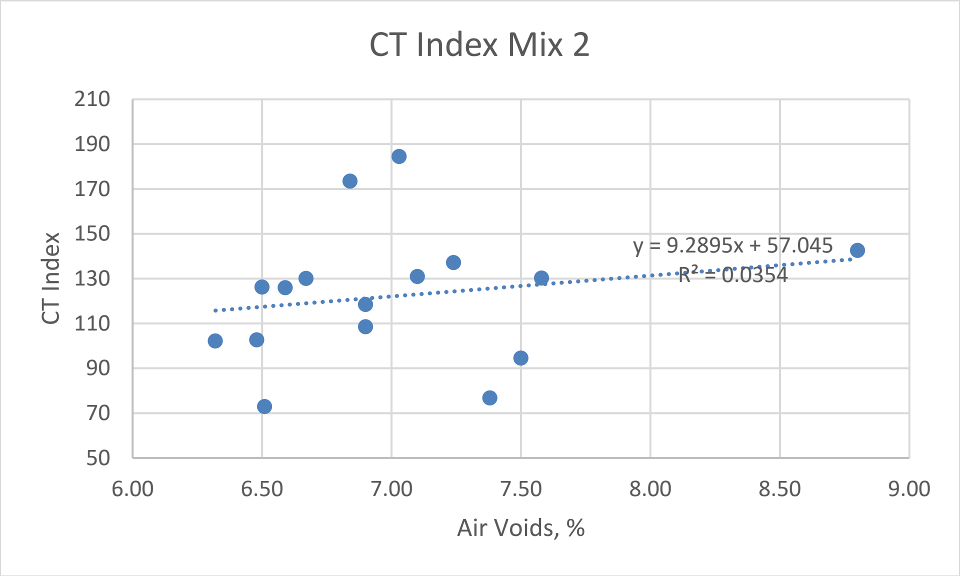


Figure 3.4 Effect of Air Voids on CT Index for Mix 2

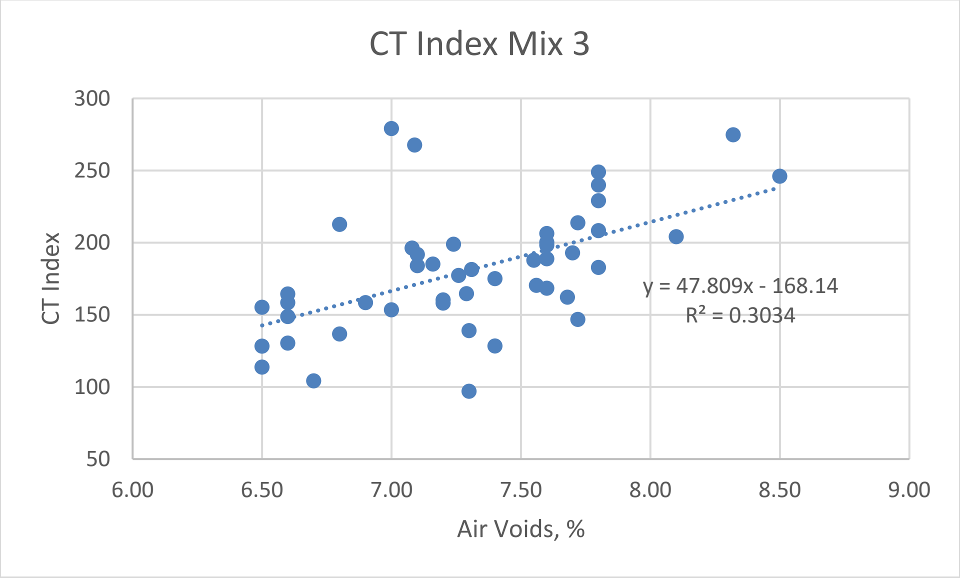


Figure 3.5 Effect of Air Voids on CT Index for Mix 3

Figures 3.3 through 3.5 show that the vast majority of the data fell within the target air void values of 7.0% ± 0.5%. Only two data points fell below the 6.5% lower limit, 21 fell between 7.5% and 7.8% and four had air voids above 8.0%. This limited number of data points outside the range might not be enough to draw definite conclusions regarding a limit for testing (e.g., should samples with 7.8% air voids be accepted?). The correlation coefficient between air voids and CT index ranged from a low of 0.19 to a high of 0.55, indicating a weak correlation. Nonetheless, the correlation is positive, indicating that higher air voids result in higher CT index values. As seen in Figures 3.3 through 3.5, the magnitude of this change is mixture dependent. For example, using the trend developed in Figure 3.5, samples with 7.5% air voids have CT index values of 190 while samples with 7.8% air voids have CT index values of 205. Using the trend developed in Figure 3.4, the same change in air voids result in a CT index increase from 126 to 129. The significance of these changes will depend on how the data are being used. If the objective is to simply rank mixtures, then air voids outside 7.5% would not change the results. On the other hand, if a specific limit, or threshold, in the CT index is ever created, then it is possible that having air voids outside the 7.5% upper limits might "push" the CT index over the edge. However, even in the most sensitive mix, the increase in CT index caused by air voids is less than the overall variability of the test.

3.6 Summary

Based on the analysis of the data, the following was found:

- The data can be reasonably assumed to be normally distributed with some outlier affecting their normality. The coefficient of variation for all results was found to be around 25% before any outliers were identified.

- Different mixtures resulted in different populations of the CT index, indicating that the test is able to distinguish between mixtures from different suppliers. Since no performance data are available and no threshold value of CT index has been determined, no statement is made regarding how any of these mixtures would perform if placed on a road.

- When samples are compacted using a number of gyrations that exceeds their N_{design} (75 gyrations), the data quality can be compromised. Greater reliability in the results was found if samples that required more than 75 gyrations to reach the target height are excluded from the analysis.

- The specimen height is accounted for in the calculations and has no effect on the CT index; therefore, the height of the sample (62 mm or 75 mm) is not considered a variable. However, most of the samples that required more than 75 gyrations to reach compaction had a height of 62 mm. This implies that, for some mixtures, samples should be compacted to a height of 75 mm.

- The sensitivity of air voids on the CT index is mixture dependent but higher air voids result in a higher CT index. Air voids as high as 7.8% were included in this analysis.

4. LABORATORY EVALUATION

4.1 Overview

In Chapter 3, the variables that affect the quality of the data were discussed. It was found that the mixtures should be treated separately. It was also found that sample height was already accounted for in the calculations for the CT index. Based on these findings, the results for each mix were presented separately and values obtained from different height samples were combined. This chapter evaluates the within- and between-lab variability of the data.

4.2 Individual Laboratory Data

The data were separated by the different participant labs. After eliminating the data from samples requiring more than 75 gyrations to compact, each lab had a different number of samples. The results are presented in Table 4.1.

Table 4.1 CT Index for Each Lab and Mix

		Lab 1	**Lab 2**	**Lab 3**	**Pooled***
Mix 1	Mean	101.5	79.9	108.5	94.8
	Standard Deviation	16.63	11.89	21.87	17.00
	As a % of Mean	16.4	14.9	20.2	17.9
	n	12	22	18	52
	95% CI Max†	111.9	85.2	119.3	
	95% CI Min†	91.0	74.6	97.7	
Mix 3	Mean	135.4	191.0	202.3	181.2
	Standard Deviation	22.41	28.08	41.54	32.78
	As a % of Mean	16.6	14.7	20.5	18.1
	N	12	17	18	47
	95% CI Max†	149.5	205.4	222.6	
	95% CI Min†	121.3	176.6	181.7	
Mix 2	Mean	119.26	--	122.9	122.3
	Standard Deviation	11.23	--	33.32	30.43
	As a % of Mean	9.4	--	27.7	24.9
	n	3	0	13	16
	95% CI Max†	139.9		142.9	
	95% CI Min†	98.6		102.9	

*Pooled values based on a weighted variance for standard deviation and a weighted mean
†Maximum and minimum values expected with 95% confidence
Values in red are data significantly different from the other labs

For each of the three mixtures tested and for each lab, Table 4.1 shows the average CT index obtained, the variability in the results (standard deviation), the number of valid samples, and the CT index range expected based on a 95% confidence interval.

Interestingly, for each mixture, there is one lab that shows results different than the other two labs (values shown in red). It is not always the same lab so it is not a consistent bias and is independent of the variability. For Mix 1, Lab 2 seems to have different CT index values when compared with the other two labs; but for Mix 3, it is Lab 1 with different CT index values. Only two labs have valid data for Mix 2 but their results are fairly close to each other. This behavior seems random, and no explanation was found.

An easy way to display the results is through a box and whisker plot. In this type of plot, the mean and median of the data are shown inside the box; the edges of the box indicate the upper and lower quartile of the data. The whiskers represent the extremes. The plots for each mix are shown in Figures 4.1 through 4.3.

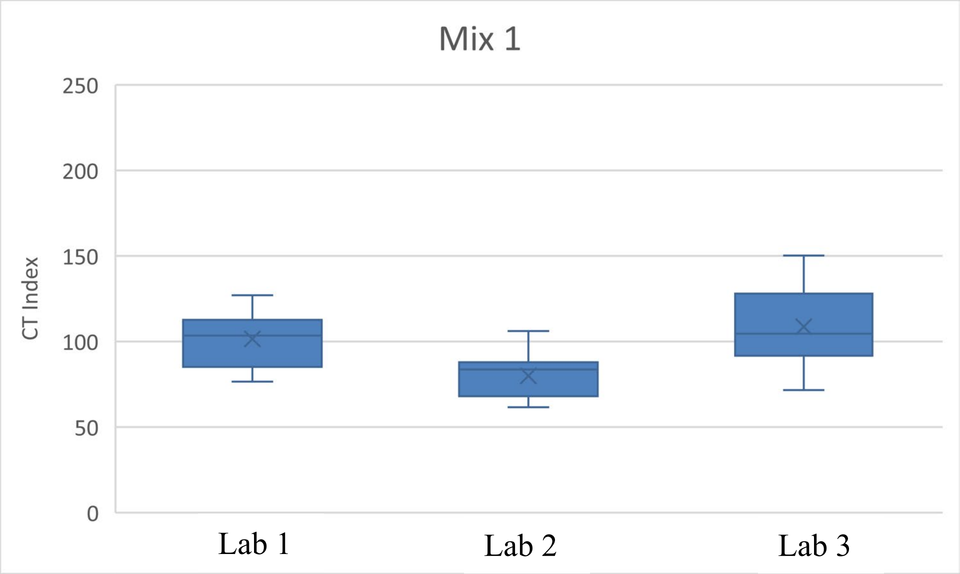


Figure 4.1 Box and Whisker Plot for Mix 1

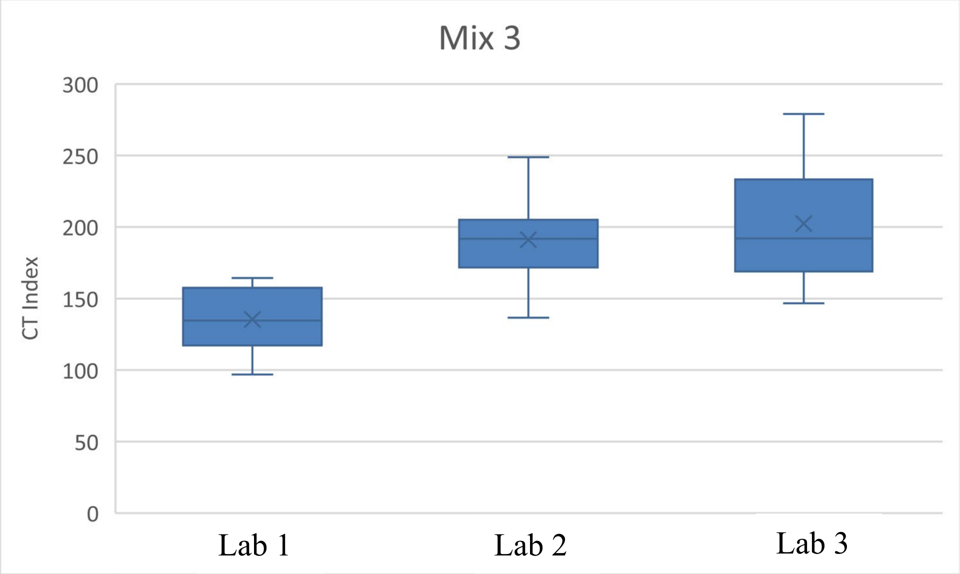


Figure 4.2 Box and Whisker Plot for Mix 3

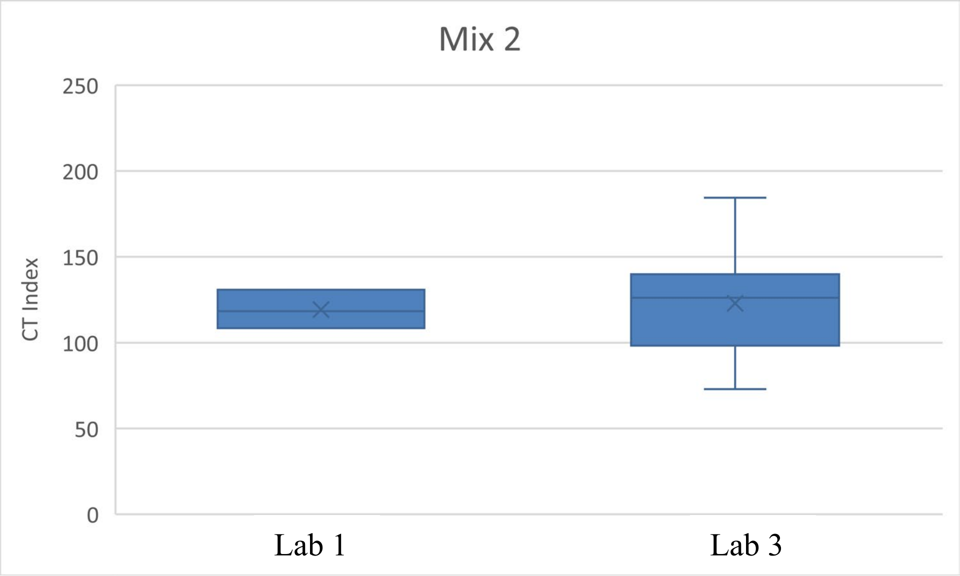


Figure 4.3 Box and Whisker Plot for Mix 2

4.2.1 Within-Lab Variability

According to ASTM D670, the within lab (single operator) standard deviation is defined as: "The standard deviation (or Coefficient of Variation) of test determinations obtained on identical test specimens by a single operator using the same apparatus in the same laboratory over a relatively short period of time. This value is equal to the pooled standard deviation of test determinations obtained by each operator. The coefficient of variation (ratio of standard deviation to the average expressed as a percentage) is used if the standard deviation is proportional to the level of the characteristic being measured. The within lab (single-operator) value, usually considered a property of the test method, will generally be lower than the multi-laboratory value."

To determine the within-lab variability, the pooled coefficient of variation from each lab was obtained by calculating the weighted coefficient of variation shown in Table 4.1 across the three different mixtures. The results are shown in Table 4.2.

Table 4.2 Within-Lab Variability

Laboratory	Coefficient of Variation* Percent
Lab 1	15.7
Lab 2	14.8
Lab 3	22.3

* Pooled across all mixtures

Two of the labs show that the coefficient of variation within lab is about 15%. One lab shows values of 22%. The reason for this one lab showing a larger coefficient of variation is unknown.

4.2.2 Between-Lab Variability

According to ASTM C670, the between-lab (multi-laboratory) standard deviation is defined as: “The standard deviation or coefficient of variation of test results obtained with the same test method on identical test specimens in different laboratories with different operators using different equipment. The multi-laboratory standard deviation, or coefficient of variation, is the fundamental statistic underlying the indexes of precision under multi-laboratory conditions. The multi-laboratory standard deviation is an indication of the variability of a group of test results obtained by different laboratories for identical test specimens.”

The between-lab variability is the coefficient of variation for each mix across the different labs. In other words, it is the variance (standard deviation) of the means between labs from Table 4.1. These values for each mix are shown in Table 4.3.

Table 4.3 Between Lab Variability

Mixture	Standard Deviation	Coeff. of Variation* Percent
Mix 1	14.91	15.7
Mix 3	35.81	19.8
Mix 2[†]	--	--

* Percent based on pooled mean shown in Table 4.1

[†] Only two labs had valid data

4.2.3 Testing Implications

Knowing the within- and between-lab variability provides an estimate of the reliability of the CT index and helps to select a testing protocol for mix evaluation. From a practical point of view, it is unrealistic for a routine test to require a large number of replicate samples; in most cases this number is three to five samples (with the lower number always preferred).

Considering the practical implications and following other material specifications, it is reasonable to require testing of three samples as long as the coefficient of variation from those specimens is below 15% (the within-lab variability described in Chapter 4.2.1). If the variability in the test outcome results in a coefficient of variation above 15%, then a review of the process should be done. If it is found that the high variability derives from one sample, then a fourth sample should be compacted and tested and the outlier should be eliminated. It is acknowledged that going through the mixing, compaction, and testing of just one sample would create logistical difficulties. Therefore, for practical purposes, it is recommended that, as part of routine procedures, four samples be prepared, compacted, and tested from a given mix. The average of the closest three can be used to determine the CT index and the corresponding variability. If the coefficient of variation exceeds 15% and a single outlier cannot be identified, a new set of samples should be prepared and tested.

4.2.4 Quality Control and Acceptance

The process of quality control and acceptance might require the comparison of results between two different laboratories. Table 4.3 shows that the between-lab coefficient of variation is between 16% and 20%. If two labs cannot obtain results within reasonable variability, a referee lab should be used.

4.3 Summary

This chapter presented the results of the CT index separated by the different laboratories. From the data presented, the within- and between-laboratory variability was determined. Based on the results, the following was found.

- Out of three participant labs, one had results that were significantly lower than the other two. It was not always the same laboratory; thus, a bias could not be established. The reason for this difference is not known.

- The within-lab variability, expressed as a coefficient of variation, was determined to be 15%. One lab had higher variability, resulting in a coefficient of variation of 22%.

- Considering the variability of the results and the practical implications of laboratory testing, it was recommended that, for mixture evaluation, four samples be prepared, compacted, and tested. For the value to be valid, the coefficient of variation of the closest three values should be below 15%.

- The between-lab variability, expressed as a coefficient of variation, was determined to be between 16% and 20%.

- The variability in the results seems to be mixture dependent with some mixtures showing a higher level of variability in the results.

5. CONCLUSIONS AND RECOMMENDATIONS

5.1 Summary of Results

The goal of this study was to determine whether the IDEAL-CT test was repeatable and under what conditions repeatability was optimized. This goal was accomplished where within-lab repeatability was determined to have a coefficient of variation less than 15% and a between-laboratory coefficient of variation less than 20%. These measures of variation can be achieved by eliminating outliers and paying attention to voids and gyration counts.

This success fills some of the goals for a test, which include:

- Fairly quick to perform: meets goal
- Precision: meets goal (within-lab CV < 15%)
- Repeatable: meets goals (between-lab CV <20%)
- Simple: meets goals
- Inexpensive: meets goals

The goal of the test having meaning has yet to be demonstrated.

At this point, it is known that the test is sensitive to binder quantity, RAP quantity, binder grade, binder modification, and aging. (Zhou et al., 2017).

5.2 Recommendations

The goal of this research was to determine the repeatability of the IDEAL-CT test and to find the issues that resulted in variability. Three mixes in common use in Utah were tested in three experienced laboratories and the results were investigated to determine factors affecting repeatability.

To apply the statistics of central tendency, the test results must be normally distributed. This was found to be true; however, normalcy was improved with some trimming of identified outliers. The following is a discussion of the factors leading to possible non-normal behavior.

5.2.1 Number of Gyrations

When the number of gyrations needed to compact the sample to the height required exceeds the Ndes (75 for the mixtures tested) it appears from observation of the broken faces and from the test result that the aggregates in the sample are damaged. This damage leaves uncoated faces in the matrix, which increases variability. It is recommended that when the number of gyrations for the given height exceeds the Ndes, the sample should be rejected and the height increased.

5.2.2 Specimen Height

The derivation of the CT index includes parameters that adjust for the size of the sample. There was no difference in the calculated index with either 62-mm or 75-mm sample heights. It was found that when a 62-mm target height exceeded 75 gyrations, the sample target height could be increased and the number of gyrations could be reduced.

It is recommended that if the number of gyrations to reach this height exceeds 75, the target sample height should be increased as needed up to 80 mm. It is also advised that a sample cut from the road be at least 38 mm in height.

5.2.3 Repeatability

It was found that after trimming the data for outliers caused by high air void and by excess gyrations, the data behaved in a more normal fashion. Results settled down such that with three replicates, the laboratories could obtain within-lab repeatability CV of 15% and a between-lab CV of 20%. Given the practicality of making another sample after finding an anomaly, it is recommended that four samples be produced and the result farthest from the mean be discarded. The remaining three samples should be averaged to obtain the final result.

5.3 Limitations and Challenges

It is clear that the IDEAL-CT test meets a majority of the criteria required for adoption. The test has been shown to be sensitive to RAP content, binder content, binder grade, binder modification, and mix aging. The remaining issue is to tie the index to pavement performance. Since it is unknown what the target CT index should be for any environmental or loading condition, caution should be used in applying any standard. There is some disparity in the data previously collected between higher CT index and better durability. It is recommended that short- and long-term aged mixes should be evaluated using the CT index, and roadways using these mixes be observed for performance over a period of time before including the CT index in a specification. The rate of aging should play a role in decisions of mix acceptability.

REFERENCES

American Society of Testing and Materials (2015). *ASTM C670: Standard Practice for Preparing Precision and Bias Statements for Test Methods for Construction Materials*. **DOI:** 10.1520/C0670-15

American Society of Testing and Materials (2019). *ASTM D8225: Standard Test Method for Determination of Cracking Tolerance of Asphalt Mixtures Using the Indirect Tensile Cracking Test at Intermediate Temperature*. **DOI:** 10.1520/D8225-19

George, D., & Mallery, M. (2010). *SPSS for Windows Step by Step: A Simple Guide and Reference, 17.0 update* (10a ed.) Boston: Pearson.

Romero, P. (2021). "Balanced Asphalt Concrete Mix Performance in Utah. Phase V: Field Evaluation for Intermediate and Low-Temperature Cracking." Report UT-21.24. Utah Department of Transportation.

Zhou, F., et al. (2022). "Ruggedness of Laboratory Tests for Asphalt Mixture Cracking Resistance," NCHRP Research Report 987, National Academy of Science

Zhou, F., et al. (2017) "Development of an IDEAL Cracking Test for Asphalt Mix Design and QC/QA" *Road Materials and Pavement Design*, 18:sup4 405-427 DOI: 10.1080/14680629.2017.1389082

APPENDIX A: DATA

Mix	Lab	Height, mm	Sample ID	Gyrations	Voids, %	CT Index
1	2	62	Puck 1	41	7.50	91.90
1	2	62	Puck 2	43	6.50	88.00
1	2	62	Puck 3	49	6.60	75.30
1	2	62	Puck 4	44	6.60	69.00
1	2	62	Puck 5	41	6.80	61.60
1	2	62	Puck 6	39	6.90	88.20
1	2	62	Puck 7	31	7.20	106.10
1	2	62	Puck 8	39	6.60	82.40
1	2	62	Puck 9	28	7.40	81.80
1	2	62	Puck 10	34	7.00	70.00
1	2	62	Puck 11	45	6.80	64.90
1	2	75	Puck 1	25	7.20	65.20
1	2	75	Puck 2	24	7.10	85.00
1	2	75	Puck 3	24	7.50	86.50
1	2	75	Puck 4	33	7.00	91.10
1	2	75	Puck 5	35	6.90	77.00
1	2	75	Puck 6	30	7.30	85.20
1	2	75	Puck 7	30	7.10	62.20
1	2	75	Puck 8	24	7.30	87.40
1	2	75	Puck 9	25	7.10	87.80
1	2	75	Puck 10	31	6.70	63.80
1	2	75	Puck 11	28	7.30	87.50
1	3	62	Puck 1	53	7.10	129.40
1	3	62	Puck 2	57	6.80	92.10
1	3	62	Puck 3	61	6.70	90.60
1	3	62	Puck 4	55	6.70	127.60
1	3	62	Puck 5	64	6.90	98.60
1	3	62	Puck 6	73	7.50	126.70
1	3	62	Puck 7	70	6.90	95.70
1	3	62	Puck 8	63	7.20	78.80
1	3	62	Puck 9	53	7.40	125.60
1	3	75	Puck 1	36	7.20	71.60
1	3	75	Puck 2	35	7.60	87.50
1	3	75	Puck 3	34	7.40	100.00
1	3	75	Puck 4	34	7.60	109.30
1	3	75	Puck 5	35	7.70	150.20
1	3	75	Puck 6	38	7.60	92.50
1	3	75	Puck 7	41	7.20	133.40
1	3	75	Puck 8	36	7.10	129.20
1	3	75	Puck 9	30	7.50	115.70
1	1	62	sp1 62	108	7.49	123.90

1	1	62	sp2 62	161	6.91	52.60
1	1	62	sp3 62	133	6.91	110.30
1	1	62	sp4 62	129	6.52	84.00
1	1	62	sp562	125	6.66	64.50
1	1	62	sp6 62	140	6.54	45.90
1	1	62	sp7 62	103	6.52	43.50
1	1	62	sp8 62	125	6.90	74.60
1	1	62	sp9 62	161	6.53	74.30
1	1	75	sp1 75	53	7.07	77.70
1	1	75	sp2 75	48	7.48	112.60
1	1	75	sp3 75	68	7.29	120.90
1	1	75	sp4 75	44	7.53	112.70
1	1	75	sp5 75	48	7.45	127.00
1	1	75	sp6 75	31	7.35	103.40
1	1	75	sp7 75	40	7.09	76.60
1	1	75	sp8 75	35	7.31	100.30
1	1	75	sp9 75	43	6.78	103.50
1	1	75	sp10 75	53	7.39	111.20
1	1	75	sp12 75	48	6.77	88.30
1	1	75	sp13 75	67	6.50	84.10
2	2	62	Puck 1	271	6.90	55.40
2	2	62	Puck 2	196	7.20	87.70
2	2	62	Puck 3	133	7.10	129.20
2	2	62	Puck 4	299	7.50	83.70
2	2	62	Puck 5	206	7.20	106.60
2	2	62	Puck 6	189	7.40	90.60
2	2	62	Puck 7	299	7.10	85.60
2	2	62	Puck 8	201	7.20	136.50
2	2	62	Puck 9	205	7.50	102.10
2	2	75	Puck 1	109	6.50	71.50
2	2	75	Puck 2	146	6.70	80.90
2	2	75	Puck 3	99	7.00	105.50
2	2	75	Puck 4	112	7.10	80.30
2	2	75	Puck 5	131	6.60	99.00
2	2	75	Puck 6	89	7.00	72.20
2	2	75	Puck 7	190	6.90	98.10
2	2	75	Puck 8	134	7.30	128.00
2	2	75	Puck 9	106	7.50	126.80
2	3	62	Puck 1	61	8.80	142.50
2	3	62	Puck 2	64	7.50	94.50
2	3	62	Puck 3	89	7.54	100.20
2	3	62	Puck 4	84	7.42	105.70
2	3	62	Puck 5	76	7.70	141.20
2	3	62	Puck 6	72	7.38	76.70
2	3	62	Puck 7	80	7.26	116.10

2	3	62	Puck 8	79	7.81	111.10
2	3	62	Puck 9	72	7.58	130.20
2	3	75	Puck 1	68	6.67	130.00
2	3	75	Puck 2	56	7.03	184.40
2	3	75	Puck 3	38	6.51	72.90
2	3	75	Puck 4	58	6.48	102.60
2	3	75	Puck 5	48	6.59	125.90
2	3	75	Puck 6	52	7.24	137.10
2	3	75	Puck 7	46	6.84	173.40
2	3	75	Puck 8	56	6.50	126.20
2	3	75	Puck 9	45	6.32	102.10
2	1	62	m1 62	174	6.50	76.50
2	1	62	m2 62	238	6.20	59.80
2	1	62	m3 62	184	6.60	73.00
2	1	62	m4 62	175	6.80	77.70
2	1	62	m5 62	186	6.50	115.70
2	1	62	m6 62	171	7.10	69.70
2	1	62	m7 62	183	6.80	83.40
2	1	62	m8 62	186	6.50	59.30
2	1	75	m1 75	92	7.40	182.60
2	1	75	m2 75	67	7.10	130.90
2	1	75	m3 75	72	6.90	108.50
2	1	75	m4 75	74	6.90	118.40
2	1	75	m5 75	80	6.80	106.10
2	1	75	m6 75	87	7.00	119.70
2	1	75	m7 75	77	7.10	163.70
2	1	75	m8 75	77	7.00	90.80
2	1	75	m9 75	81	7.00	98.80
3	2	62	Puck 1	89	6.60	167.70
3	2	62	Puck 2	37	7.80	248.80
3	2	62	Puck 3	58	7.60	168.30
3	2	62	Puck 4	69	7.80	239.80
3	2	62	Puck 5	50	7.80	182.70
3	2	62	Puck 6	45	8.10	204.00
3	2	62	Puck 7	107	6.90	139.40
3	2	62	Puck 8	58	7.60	198.20
3	2	62	Puck 9	60	7.40	175.00
3	2	75	Puck 1	52	6.80	136.60
3	2	75	Puck 2	47	7.20	160.20
3	2	75	Puck 3	41	7.10	184.00
3	2	75	Puck 4	50	7.10	191.70
3	2	75	Puck 5	54	6.80	212.60
3	2	75	Puck 6	38	7.60	206.20
3	2	75	Puck 7	39	7.60	200.10
3	2	75	Puck 8	49	7.20	158.00

3	2	75	Puck 9	37	7.70	192.80
3	2	75	Puck 10	36	7.60	188.70
3	3	62	Puck 1	39	7.80	229.00
3	3	62	Puck 2	31	8.32	274.70
3	3	62	Puck 3	35	7.08	196.20
3	3	62	Puck 4	32	7.72	146.70
3	3	62	Puck 5	27	7.72	213.70
3	3	62	Puck 6	34	7.00	153.40
3	3	62	Puck 7	21	8.50	245.90
3	3	62	Puck 8	36	7.00	279.00
3	3	62	Puck 9	33	7.31	181.30
3	3	75	Puck 1	32	7.16	185.10
3	3	75	Puck 2	31	7.24	198.80
3	3	75	Puck 3	33	7.09	267.60
3	3	75	Puck 4	30	7.55	187.70
3	3	75	Puck 5	34	7.29	164.60
3	3	75	Puck 6	31	7.26	177.20
3	3	75	Puck 7	29	7.68	162.10
3	3	75	Puck 8	32	7.80	208.20
3	3	75	Puck 9	30	7.56	170.30
3	1	62	k1 62	85	7.10	187.40
3	1	62	k2 62	76	7.20	119.30
3	1	62	k3 62	95	7.10	193.80
3	1	62	k4 62	82	7.00	149.60
3	1	62	k562	75	7.30	96.90
3	1	62	k6 62	58	7.40	128.30
3	1	62	k7 62	58	7.30	138.90
3	1	75	k1 75	60	6.60	164.30
3	1	75	k2 75	68	6.60	130.30
3	1	75	k3 75	60	6.50	155.10
3	1	75	k4 75	61	6.70	104.10
3	1	75	k5 75	64	6.50	113.60
3	1	75	k6 75	63	6.60	148.70
3	1	75	k7 75	56	6.50	128.20
3	1	75	k8 75	65	6.60	158.30
3	1	75	k9 75	63	6.90	158.30