# Round Robin Testing Program for the Indirect Tensile Cracking Test at Intermediate Temperature: Phase I

http://www.virginiadot.org/vtrc/main/online_reports/pdf/22-r3.pdf

**JHONY HABBOUCHE, Ph.D., P.E.**
**Research Scientist**
**Virginia Transportation Research Council**

**ILKER BOZ, Ph.D.**
**Research Scientist**
**Virginia Transportation Research Council**

**STACEY D. DIEFENDERFER, Ph.D., P.E.**
**Senior Research Scientist**
**Virginia Transportation Research Council**

**YUSUF K. BILGIC, Ph.D.**
**Adjunct Professor**
**University of Rochester**

**Final Report VTRC 22-R3**

**Standard Title Page - Report on Federally Funded Project**

| 1. Report No.: FHWA/VTRC 22-R3 | 2. Government Accession No.: | 3. Recipient's Catalog No.: |
|---|---|---|
| 4. Title and Subtitle: Round Robin Testing Program for the Indirect Tensile Cracking Test at Intermediate Temperature: Phase I | | 5. Report Date: August 2021<br>6. Performing Organization Code: |
| 7. Author(s): Jhony Habbouche, Ph.D., P.E., Ilker Boz, Ph.D., Stacey D. Diefenderfer, Ph.D., P.E., and Yusuf K. Bilgic, Ph.D. | | 8. Performing Organization Report No.: VTRC 22-R3 |
| 9. Performing Organization and Address:<br>Virginia Transportation Research Council<br>530 Edgemont Road<br>Charlottesville, VA 22903 | | 10. Work Unit No. (TRAIS):<br>11. Contract or Grant No.: 118733 |
| 12. Sponsoring Agencies' Name and Address:<br>Virginia Department of Transportation<br>1401 E. Broad Street<br>Richmond, VA 23219 | Federal Highway Administration<br>400 North 8th Street, Room 750<br>Richmond, VA 23219-4825 | 13. Type of Report and Period Covered: Final<br>14. Sponsoring Agency Code: |

15. Supplementary Notes:
This is an SPR-B report.


16. Abstract:

The indirect tensile cracking test (referred to herein as "the IDT-CT") was recommended for use in balanced mix design (BMD) specifications to address cracking susceptibility of dense-graded surface mixtures with A and D designations in Virginia. The test method for the IDT-CT, ASTM D8225-19, does not currently contain a precision statement for the test. This creates potential issues if different test results are measured by individual laboratories conducting testing on the same asphalt mixture.

The purpose of this study was to determine and develop precision estimates and statements for the cracking tolerance index (CT index) of asphalt mixtures determined by performing the IDT-CT at intermediate temperature in accordance with ASTM D8225-19. In addition, precision estimates and statements were developed for the fracture strain tolerance index (FST index), strength (St), and cracking resistance index (CRI) from the same test data. The effects of device and loading rate on the selected IDT-CT indices were also investigated. Moreover, a preliminary assessment of the impact of the shelf life of compacted specimens on the selected IDT-CT indices was conducted. These objectives were achieved by conducting a two-stage round robin study. Stage I focused on non-VDOT (Virginia Department of Transportation) laboratories, and Stage II, conducted 1 year later, focused on VDOT laboratories.

In Stage I, only 14 of 41 participating laboratories submitted results (16 data sets) for both mixtures in full accordance with ASTM D8225-19. The initial data quality resulted in performing the analysis on two groups of data to calculate precision estimates. The precision estimates for the CT index, FST index, St, and CRI were calculated, and the corresponding statements were developed. A significant drop in the precision parameter (i.e., coefficient of variation or standard deviation) of the IDT-CT indices for both single-operator and multi-laboratory conditions were observed when data trimming was performed.

The relatively higher variability observed for data collected in Stage II when compared to the variability for data collected in Stage I could be attributable to a relative lack of operator experience; the need for training; and potential changes in material properties during the storing, handling, shipping, or testing process. However, the analyses and comparisons of data collected in Stage I and Stage II indicated that there was no significant impact of 1 year of climate-controlled storage of compacted specimens on the calculated IDT-CT index.

The study recommends that a second phase of the round robin for the IDT-CT be conducted to assess the impact of variability induced because of specimen preparation to better reproduce the actual state of the practice during design and production. Further, the study recommends that a more comprehensive effort to assess the impact of loading rate on the IDT-CT results be initiated. Further, the study recommends that the impact of testing IDT-CT specimens under saturated surface dry conditions as compared with dry conditions be assessed. Finally, hands-on training and demonstration of the laboratory tests (e.g., the IDT-CT) being considered by VDOT as part of the BMD initiative are recommended.


| 17 Key Words: | 18. Distribution Statement: |
|---|---|
| Balanced mix design, cracking, durability, performance testing, indirect tensile cracking test, cracking tolerance index, fracture strain tolerance index, indirect tensile strength, cracking resistance index, statistical analysis, device, quality, variability, reproducibility, repeatability | No restrictions. This document is available to the public through NTIS, Springfield, VA 22161. |

| 19. Security Classif. (of this report): | 20. Security Classif. (of this page): | 21. No. of Pages: | 22. Price: |
|---|---|---|---|
| Unclassified | Unclassified | 47 | |

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

**FINAL REPORT**

# ROUND ROBIN TESTING PROGRAM FOR THE INDIRECT TENSILE CRACKING TEST AT INTERMEDIATE TEMPERATURE: PHASE I

**Jhony Habbouche, Ph.D., P.E.**
**Research Scientist**
**Virginia Transportation Research Council**

**Ilker Boz, Ph.D.**
**Research Scientist**
**Virginia Transportation Research Council**

**Stacey D. Diefenderfer, Ph.D., P.E.**
**Senior Research Scientist**
**Virginia Transportation Research Council**

**Yusuf K. Bilgic, Ph.D.**
**Adjunct Professor**
**University of Rochester**

In Cooperation with the U.S. Department of Transportation
Federal Highway Administration

Virginia Transportation Research Council
(A partnership of the Virginia Department of Transportation
and the University of Virginia since 1948)

Charlottesville, Virginia

August 2021
VTRC 22-R3

## DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Virginia Department of Transportation, the Commonwealth Transportation Board, or the Federal Highway Administration. This report does not constitute a standard, specification, or regulation. Any inclusion of manufacturer names, trade names, or trademarks is for identification purposes only and is not to be considered an endorsement.

## ABSTRACT

The indirect tensile cracking test (referred to herein as "the IDT-CT") was recommended for use in balanced mix design (BMD) specifications to address cracking susceptibility of dense-graded surface mixtures with A and D designations in Virginia. The test method for the IDT-CT, ASTM D8225-19, does not currently contain a precision statement for the test. This creates potential issues if different test results are measured by individual laboratories conducting testing on the same asphalt mixture.

The purpose of this study was to determine and develop precision estimates and statements for the cracking tolerance index (CT index) of asphalt mixtures determined by performing the IDT-CT at intermediate temperature in accordance with ASTM D8225-19. In addition, precision estimates and statements were developed for the fracture strain tolerance index (FST index), strength ($S_t$), and cracking resistance index (CRI) from the same test data. The effects of device and loading rate on the selected IDT-CT indices were also investigated. Moreover, a preliminary assessment of the impact of the shelf life of compacted specimens on the selected IDT-CT indices was conducted. These objectives were achieved by conducting a two-stage round robin study. Stage I focused on non-VDOT (Virginia Department of Transportation) laboratories, and Stage II, conducted 1 year later, focused on VDOT laboratories.

In Stage I, only 14 of 41 participating laboratories submitted results (16 data sets) for both mixtures in full accordance with ASTM D8225-19. The initial data quality resulted in performing the analysis on two groups of data to calculate precision estimates. The precision estimates for the CT index, FST index, $S_t$, and CRI were calculated, and the corresponding statements were developed. A significant drop in the precision parameter (i.e., coefficient of variation or standard deviation) of the IDT-CT indices for both single-operator and multi-laboratory conditions were observed when data trimming was performed.

The relatively higher variability observed for data collected in Stage II when compared to the variability for data collected in Stage I could be attributable to a relative lack of operator experience; the need for training; and potential changes in material properties during the storing, handling, shipping, or testing process. However, the analyses and comparisons of data collected in Stage I and Stage II indicated that there was no significant impact of 1 year of climate-controlled storage of compacted specimens on the calculated IDT-CT index.

The study recommends that a second phase of the round robin for the IDT-CT be conducted to assess the impact of variability induced because of specimen preparation to better reproduce the actual state of the practice during design and production. Further, the study recommends that a more comprehensive effort to assess the impact of loading rate on the IDT-CT results be initiated. Further, the study recommends that the impact of testing IDT-CT specimens under saturated surface dry conditions as compared with dry conditions be assessed. Finally, hands-on training and demonstration of the laboratory tests (e.g., the IDT-CT) being considered by VDOT as part of the BMD initiative are recommended.

**FINAL REPORT**

# ROUND ROBIN TESTING PROGRAM FOR THE INDIRECT TENSILE CRACKING TEST AT INTERMEDIATE TEMPERATURE: PHASE I

**Jhony Habbouche, Ph.D., P.E.**
**Research Scientist**
**Virginia Transportation Research Council**

**Ilker Boz, Ph.D.**
**Research Scientist**
**Virginia Transportation Research Council**

**Stacey D. Diefenderfer, Ph.D., P.E.**
**Senior Research Scientist**
**Virginia Transportation Research Council**

**Yusuf K. Bilgic, Ph.D.**
**Adjunct Professor**
**University of Rochester**

## INTRODUCTION

The Virginia Department of Transportation (VDOT), like many owner agencies, is interested in ways to improve the durability of asphalt mixtures in an effort to make its roadway network more sustainable, longer lasting, and more economical. The balanced mix design (BMD) method was proposed to address this by incorporating performance criteria into mix design and acceptance. Instead of providing only recipe-type specifications for design and acceptance, the BMD method applies performance test criteria to assess and accept mixtures.

Beginning in 2017, a major initial effort was undertaken at the Virginia Transportation Research Council (VTRC) to provide benchmark indications of performance for a number of asphalt mixtures produced and sampled in 2015 (Bowers and Diefenderfer, 2018; Diefenderfer and Bowers, 2019). The mixtures were extensively characterized, and numerous laboratory performance tests were conducted to determine baseline performance measures. As one outcome of this effort, the indirect tensile cracking test (referred to herein as "the IDT-CT"), developed by Zhou et al. (2017) and specified in ASTM D8225-19, Determination of Cracking Tolerance Index of Asphalt Mixture Using the Indirect Tensile Cracking Test at Intermediate Temperatures (ASTM, 2019a), was recommended for use in BMD specifications to address cracking susceptibility of dense-graded surface mixtures (SMs) with A and D designations in Virginia. From this test, the cracking tolerance (CT) index is calculated using the load-displacement data along with specimen dimensions to indicate the cracking susceptibility of asphalt mixtures. A minimum CT index threshold value of 70 was recommended to mitigate the cracking susceptibility of short-term–aged asphalt SMs.

Another effort was undertaken at VTRC to evaluate the validity of this test and its associated threshold alongside other performance tests such as durability and rutting using the Cantabro and Asphalt Pavement Analyzer (APA) tests, respectively (Diefenderfer et al., 2021). This study recommended that the precision estimates and statements for these test methods be determined and developed so that the tests could be used for quality control, quality assurance, and acceptance of materials.

Currently, ASTM D8225-19 does not contain precision and bias statements for the IDT-CT method. This could create potential disputes if different test results are measured by individual laboratories conducting testing on the same asphalt mixture. Therefore, it is necessary to establish the acceptable variability of the test method to determine if individual test results from the same evaluated asphalt mixture can be considered statistically similar.

In addition to the CT index, numerous performance indices can be determined from the load-displacement data obtained from the IDT-CT including the fracture strain tolerance (FST) index, strength ($S_t$), and cracking resistance index (CRI). Previous studies have shown that these indices are highly correlated, with some showing more promising performance discrimination potential among asphalt mixtures and a better repeatability characteristic, which could be advantageous for quality measurement practices (Diefenderfer et al., 2019; Diefenderfer et al., 2021; Habbouche et al., 2021; Seitllari et al., 2020).

## PURPOSE AND SCOPE

The purpose of this study was to determine and develop precision estimates and statements for the CT index of asphalt mixtures determined by performing the IDT-CT at intermediate temperature in accordance with ASTM D8225-19. In addition, the precision estimates and statements were developed for the other indices calculated from the IDT-CT data: the FST index, $S_t$, and CRI. The effects of equipment type and loading rate on the selected IDT-CT indices were also investigated. Moreover, a preliminary assessment of the impact of the shelf life of compacted specimens on the selected IDT-CT indices was conducted.

To carry out these tasks, a two-stage round robin study was conducted. Stage I focused on non-VDOT laboratories, and Stage II, conducted 1 year later, focused on VDOT laboratories. Both efforts involved the evaluation of specimens fabricated and compacted by a third party laboratory and sent to participant laboratories along with detailed instructions for testing only.

## METHODS

### Literature Review

Literature on previous IDT-CT round robin efforts was identified by a search of various databases related to transportation engineering such as the Transport Research International Documentation (TRID) database. The identified literature was then reviewed to summarize findings and provide a background on relevant work.

## Asphalt Mixtures

Two mixtures, herein referred to as "Mixture A" and "Mixture B," were designed and evaluated in this effort. The mixtures were designed, produced, and compacted by an independent testing laboratory, and volumetric and gradation properties were reported. In addition, the Cantabro mass loss and the APA rut depth were determined for the mixtures.

## Mixture Testing

### Indirect Tensile Cracking Test

The IDT-CT was conducted at 25°C on specimens fabricated and compacted by a third party laboratory in accordance with ASTM D8225-19 (ASTM, 2019a). Tests were performed at a loading rate of 50 ± 2 mm/min on specimens 150 mm in diameter by 62 mm in height compacted with a Superpave gyratory compactor to 7 ± 0.5% air-void content. The CT index, FST index, $S_t$, and CRI were then calculated from the test load-displacement curve shown in Figure 1 using Equation 1, Equation 3, Equation 4, and Equation 5, respectively. Previous studies have shown that these indices are highly correlated, with some showing a better repeatability of characteristics and a more promising performance discrimination potential among asphalt mixtures (Diefenderfer et al., 2019; Diefenderfer et al., 2021; Habbouche et al., 2021; Seitllari et al., 2020).

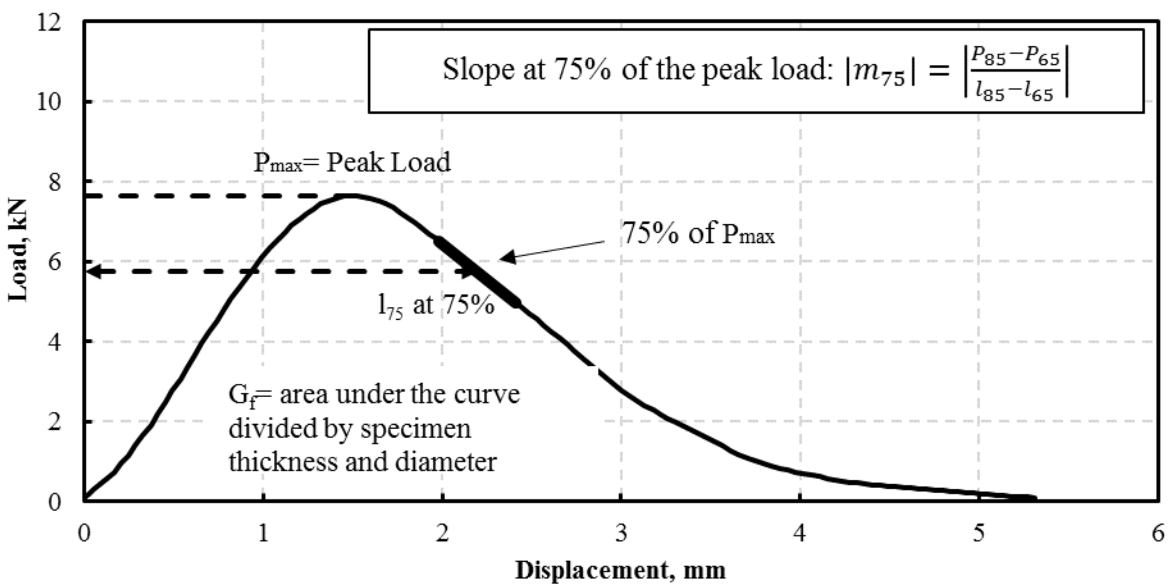



**Figure 1. Indirect Tensile Cracking Test (IDT-CT) Load-Displacement Curve**

$$CT\ index = \frac{G_f}{|m_{75}|} * \left(\frac{l_{75}}{D}\right) * \left(\frac{t}{62}\right)$$ [Eq. 1]

$$m_{75} = \left|\frac{p_{85}-p_{65}}{l_{85}-l_{65}}\right|$$ [Eq. 2]

$$FST = \frac{G_f}{S_t} * 10^6$$ [Eq. 3]

$$S_t = \frac{2000P_{max}}{\pi t D} * 10^3$$ [Eq. 4]

$$CRI = \frac{G_f}{P_{max}}$$ [Eq. 5]

where

CT index = cracking tolerance index expressed in Equation 1
$G_f$ = total area under the load-displacement curve divided by the product of the specimen thickness [t] and diameter [D], kN/mm
$m_{75}$ = slope of interest expressed in Equation 2
$p_{85}$= 85% of the peak load ($P_{max}$) at the post-peak stage, kN
$p_{75}$= 75% of $P_{max}$ at the post-peak stage, kN
$p_{65}$= 65% of $P_{max}$ at the post-peak stage, kN
$l_{85}$ = displacement corresponding to $p_{85}$, mm
$l_{75}$ = displacement corresponding to $p_{75}$, mm
$l_{65}$ = displacement corresponding to $p_{65}$, mm
FST = fracture strain tolerance expressed in Equation 3
$S_t$ = indirect tensile strength expressed in Equation 4, kPa
CRI = cracking resistance index expressed in Equation 5
D = specimen diameter, mm
t = specimen thickness, mm.

## Interlaboratory Study: Phase I, Stages I and II

Phase I of the IDT-CT round robin study included two stages. Stage I focused on non-VDOT laboratories, and Stage II focused on VDOT district laboratories. The interlaboratory study (ILS) was conducted in accordance with ASTM E691-19, Standard Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method (ASTM, 2019b). The procedure includes three basic steps: planning the ILS, guiding the testing phase of the study, and analyzing the test results. The requirements for each step as detailed in ASTM E691-19 were fulfilled in this study with the exception of including at least three materials representing different test levels for developing precision statements. This exception was made because the intent was to involve more laboratories rather than more materials. It was expected that the precision would be relatively constant when compared to the average level over the range of values of interest; thus, a smaller number of materials (in this case two) was included.

ASTM C670-15, Standard Practice for Preparing Precision and Bias Statements for Test Methods for Construction Materials, was used in developing the precision statements. Since there is no accepted reference material suitable for determining the bias in this test method, no statement on bias is made.

### Impact of Shelf Life of Compacted Specimens on IDT-CT Indices

The time (i.e., 1 year) that passed between Stages I and II allowed for a preliminary assessment of the impact of shelf life of compacted specimens stored under climate-controlled conditions (i.e., temperature and humidity) on the IDT-CT data and various indices.

## RESULTS AND DISCUSSION

### Literature Review

In 2018, the National Center for Asphalt Technology (NCAT) initiated a two-phase study to establish a precision statement for the CT index (Taylor, 2019; Taylor et al., 2019) through the evaluation of a single asphalt mixture. Phase I of this effort involved 15 participating laboratories fabricating and testing a minimum of five test specimens using loose mixture sent to the laboratories along with detailed instructions. Phase II involved 14 laboratories testing a set of five specimens fabricated and compacted at the NCAT laboratory using the same mixture evaluated in Phase I.

The data from Phase I showed CT index results between 20 and 200, and the data from Phase II had values with a much smaller range, between 80 and 140, for the same mixture. The within-laboratory coefficients of variation (COVs) for Phase I and Phase II were similar, with values of 19.5% and 18.8%, respectively. However, the multi-laboratory COV values for Phase I and Phase II were 35.3% and 20.2%, respectively, indicating a significant decrease from Phase I to Phase II, which was attributed to the variability induced by individual laboratory specimen preparation.

A similar effort was conducted by researchers at Rutgers University using New Jersey asphalt mixtures (Bennert et al., 2020). A round robin study was conducted to determine the repeatability and reproducibility of the CT index. Five different asphalt mixtures were produced and compacted by a single laboratory to achieve varying levels of performance. The optimum asphalt binder content and CT index of the asphalt mixtures at the design stage had a range of 5.5% to 6.5% and 81 to 456, respectively, to provide widespread applicability of the outcomes. Sets of three test specimens for each mixture were compacted by the Rutgers laboratory to a target air-void level of 5.5% ± 0.5% prior to delivery to nine participant laboratories. The five evaluated mixtures had an average single-operator and multiple-operator COV for the CT index of 15.2% and 23.0%, respectively.

### Design Properties of Evaluated Mixtures

Two asphalt mixtures, Mixture A and Mixture B, were evaluated in this study. Mixture A was a 65-gyration, 9.5 mm nominal maximum aggregate size Superpave mixture containing 30% reclaimed asphalt pavement and produced with performance grade (PG) 76-22 asphalt binder. Mixture B was a 50-gyration 12.5 mm nominal maximum aggregate size Superpave mixture produced with PG 64-22 binder. Mixture B did not contain reclaimed asphalt pavement.

Table 1 summarizes the volumetric and performance properties for the mixtures. Figure 2 shows the aggregate job mix formula for both mixtures. The mixtures were designed such that the CT index values determined for each mixture were significantly different; that the difference was greater than 50; and that the values were greater than and less than 100 to ensure a wider applicability of the study to various ranges of CT index values. Both mixtures were further evaluated at the design stage in terms of durability by the Cantabro test and resistance to rutting by the APA rut test. Both mixtures met the BMD requirements in terms of durability and resistance to rutting for VDOT asphalt SMs with A and D designations.

**Table 1. Volumetric and Performance Properties for Mixture A and Mixture B**

| Mixture ID | Mixture A | Mixture B |
|---|---|---|
| **Composition** | | |
| RAP Content, % | 30 | 0 |
| Asphalt Binder | PG 76-22 | PG 64-22 |
| **Volumetric Property** | | |
| $N_{design}$, gyrations | 65 | 50 |
| NMAS, mm | 9.5 | 12.5 |
| Asphalt Binder Content, % | 5.3 | 5.80 |
| Rice SG ($G_{mm}$) | 2.511 | 2.723 |
| Aggregate Bulk SG ($G_{sb}$) | 2.678 | 2.941 |
| VTM, % | 4.5 | 4.0 |
| VMA, % | 15.2 | 16.4 |
| VFA, % | 70.0 | 75.6 |
| FA Ratio | 1.57 | 1.28 |
| **Performance Property** | | |
| Cantabro Mass Loss at 25°C, % | 6.1 | 3.8 |
| APA Rut Depth at 64°C, mm | 1.350 | 4.160 |
| CT Index at 25°C | Target < 100 | Target > 100 |

$N_{design}$ = number of Superpave design gyrations; NMAS = nominal maximum aggregate size; RAP = reclaimed asphalt pavement; SG = specific gravity; VTM = voids in total mixture; VMA = voids in mineral aggregate; VFA = voids filled with asphalt; FA = fines to asphalt ratio; APA = asphalt pavement analyzer, CT = cracking tolerance.

## Interlaboratory Study: Phase I, Stage I

### Participant Laboratories

A total of 41 laboratories participated in Stage I of the round robin. These laboratories consisted of one VDOT district laboratory; the VTRC laboratory; and numerous contractor, other DOT, and independent testing laboratories. Several laboratories received more than one set of test specimens per mixture to perform testing using machines or load frames from different manufacturers. In total, 46 sets of test specimens for each mixture were distributed and seven devices were evaluated (referred to herein as “Devices I through VII”) (Boz et al., 2021).

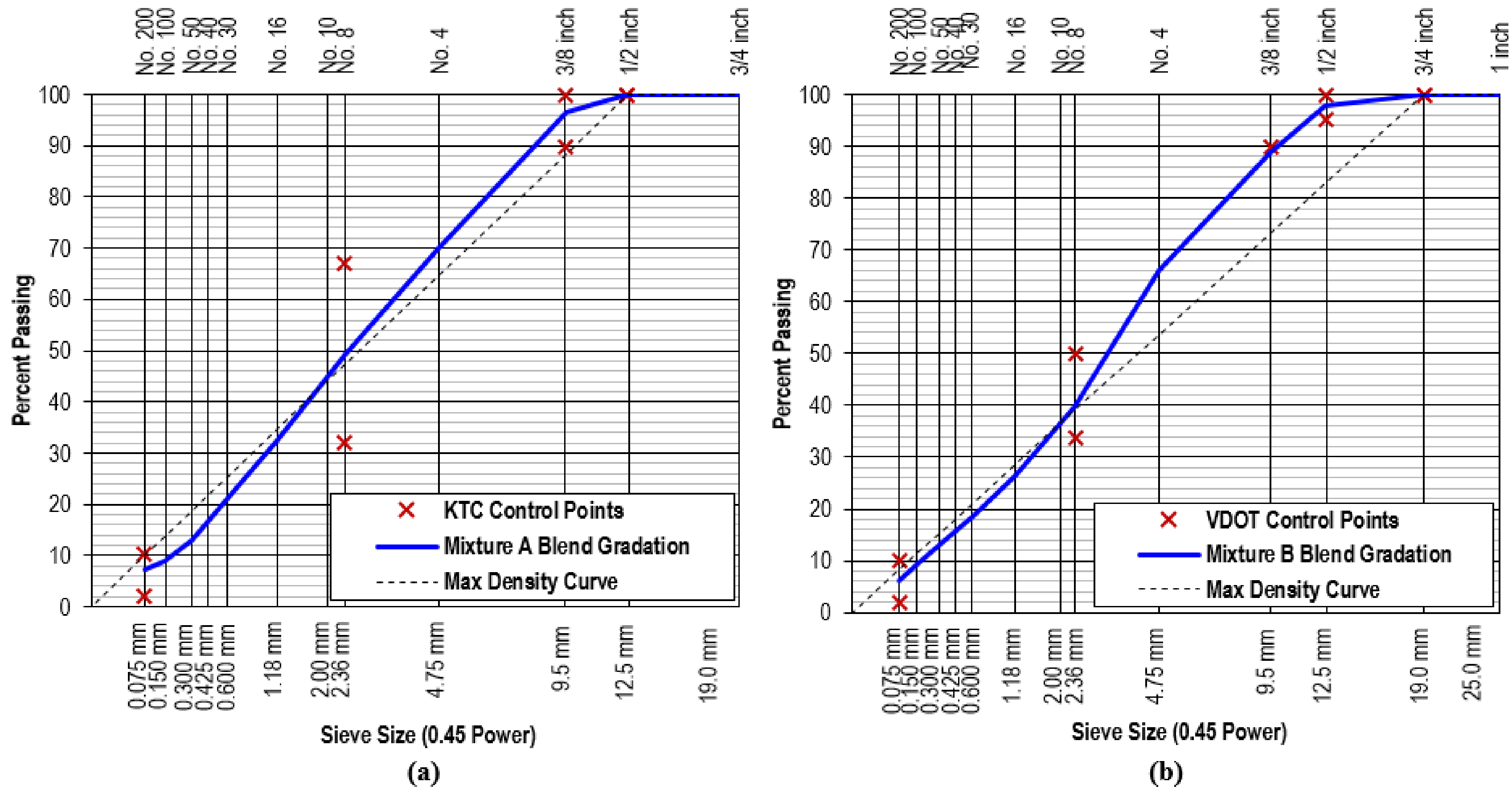



**Figure 2. Job Mix Formula: (a) Mixture A; (b) Mixture B. Max = maximum; KTC = Kentucky Transportation Cabinet; VDOT = Virginia Department of Transportation.**

## Testing Instructions

All participating laboratories were provided with testing instructions and guidelines summarized as follows:

- Inspect all specimens for any visual damage (i.e., cracks, deformation, etc.) and replace in case of evident damage.
- Determine each specimen's diameter by measuring to 0.1 mm at two locations along the specimen. Determine each specimen's thickness by measuring to 0.1 mm at four locations around the specimen.
- Determine the bulk specific gravity of the specimens in accordance with AASHTO T 166, Method A.
- Dry all specimens prior to testing and condition at a temperature of 25°C for at least 2 hours while maintaining the dry condition. Report the method of drying.
- Perform the IDT-CT in accordance with ASTM D8225-19. Report the equipment manufacturer, model, and type (i.e., screw-drive or servo-hydraulic).
- Report additional information for each IDT-CT specimen including testing date and time, data file name, CT index value, and raw data file.

For consistency, all IDT-CTs were performed at 25°C on dry specimens. However, some participating laboratories from both stages reported challenges in keeping the IDT-CT specimens dry when conditioning using a water bath at a temperature of 25°C. Frequent water leaks attributable to tearing of the plastic bags were reported, resulting in wet specimens. These laboratories were asked to dry the IDT-CT specimens again in front of a fan or using a vacuum drier. They were also asked to use higher quality, heavy-duty leak-proof plastic bags or to double-bag the specimens prior to repeating the conditioning process to keep the specimens dry.

## Test Data Quality Evaluation

Data quality checks are important in performance testing. Non-compliance in test data can lead to incorrect IDT-CT index values that do not describe the actual material performance—resulting in unnecessary re-designs or rejected materials. Participating laboratories were asked to submit the raw data files collected by their equipment in addition to reporting the equipment-calculated CT index values for each specimen. These data files were used to perform quality checks on the data before further analysis. Each raw data file was required to include the time, load, and displacement measurements recorded by and obtained through the testing software. In some cases, the operator manual or equipment manufacturer had to be consulted to determine the best way to extract the raw data file. The time, load, and displacement measurements were then plotted in a spreadsheet, and data quality was assessed.

The submitted raw data from each participating laboratory were first evaluated to determine if the tests performed were in accordance with ASTM D8225-19. This was done by evaluating the load versus displacement curves and the displacement versus time curves for each tested specimen. Some key details were assessed to evaluate the data quality:

- verification that no seating load was applied
- verification that the load was applied at the specified constant load rate of 50 ± 2 mm/min
- verification that the test ended only when the applied load dropped to 0.1 kN or less after the peak load was reached.

An example of compliant load-displacement and displacement-time curves are shown in Figures 3a and 3b, respectively. These curves should be checked for each specimen as part of test quality control.

There are many reasons that the load-displacement and displacement-time curves may not meet the test specification. Some load frames have a safety function limiting the ram travel that does not provide enough travel for the post-peak load to drop to 0.1 kN or less during testing. Machine compliance may cause the rate of loading to change with increasing specimen resistance to loading. An improper linear variable differential transformer (LVDT) setup including misalignment, improper zeroing, or incorrect calibration will affect displacement measurements. Data quality checks permit the identification of issues and allow for them to be addressed. If non-compliant data are found, the causes should be determined so that they can be addressed and resolved. A single instance of non-compliance may indicate an isolated or random event; however, recurring events may be a sign of equipment issues, repetitive operator error, or other testing problems. Some examples of non-compliant data submitted during the round robin are shown in Figures 4 through 8 along with some explanations.

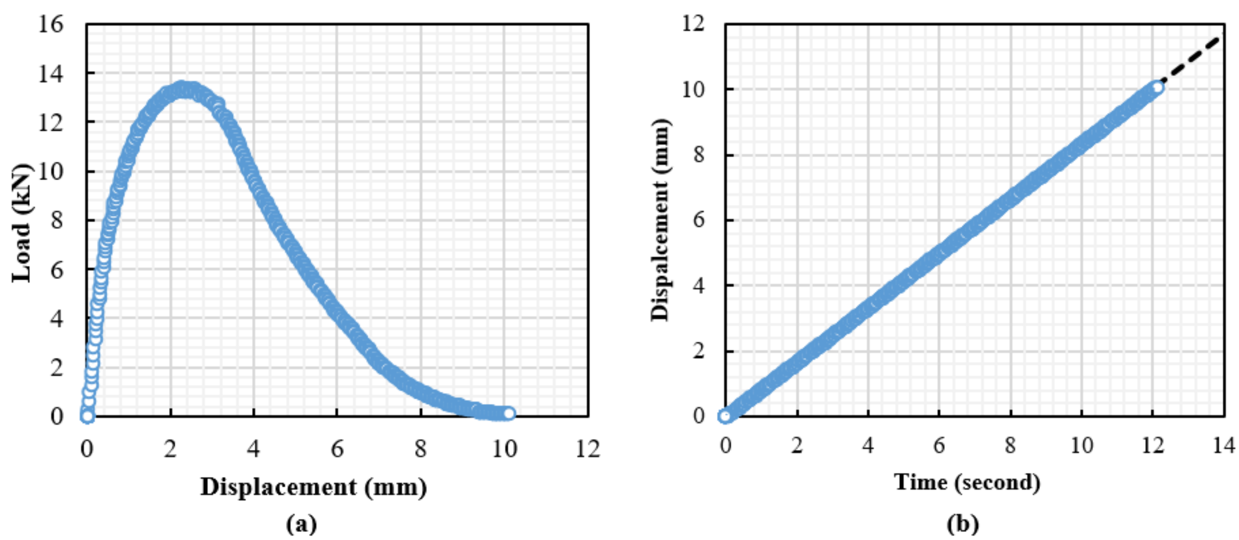



**Figure 3. Examples of Compliant Curves: (a) load-displacement curve; (b) displacement-time curve.**

Figure 4 shows data from a test wherein the data acquisition was incorrectly set up, as the measured load should reach approximately 10 kN at its peak. The difference suggests that the load data shown are in U.S. units (lbf) instead of kN as labelled, whereas displacements are shown correctly in SI units (mm); the mixed units resulted in incorrect calculations of IDT-CT indices. In addition, the test terminated before the load dropped to 0.1 kN.

Figure 5 shows a test with an error in displacement measurements; the LVDT may have slipped out of position or the range may need to be checked. Figure 6 displays a seating load applied at the beginning of the test and an LVDT error toward the end of the test. The test software should be configured to remove the seating load, and the LVDT installation and range should be checked.

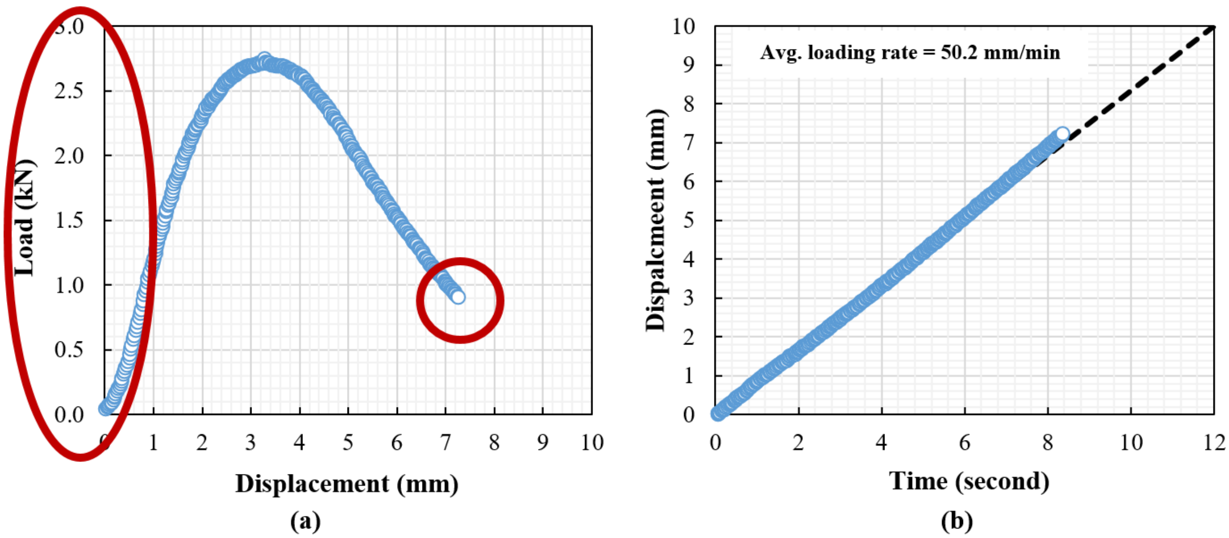



**Figure 4. Example of Non-Compliant Test With Unexpected Load Magnitude, Incorrect Units, and Premature Termination of the Test: (a) load-displacement curve; (b) displacement-time curve.**

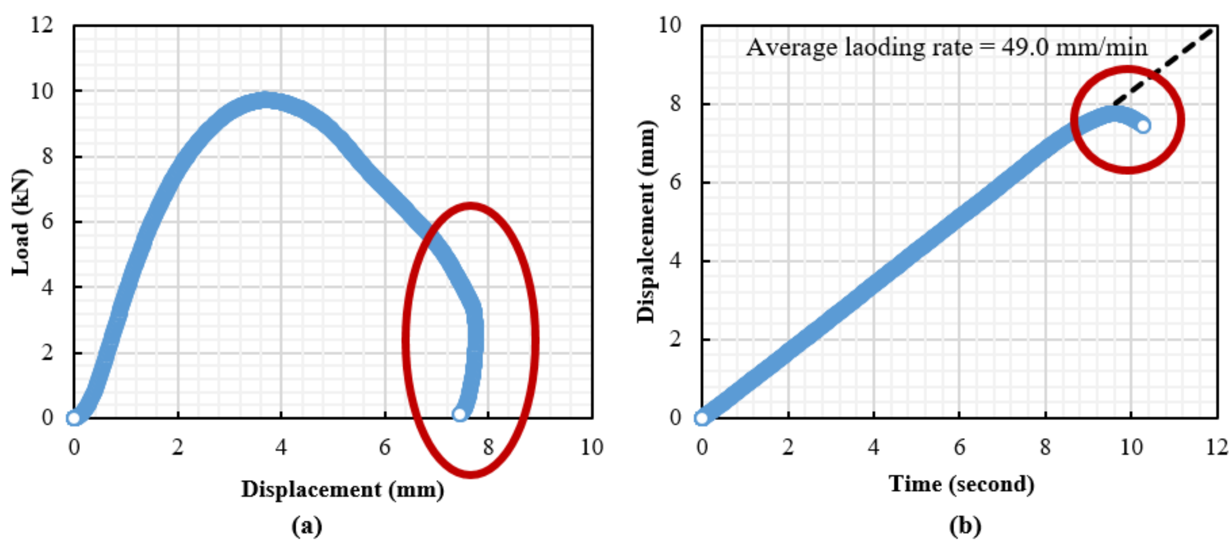
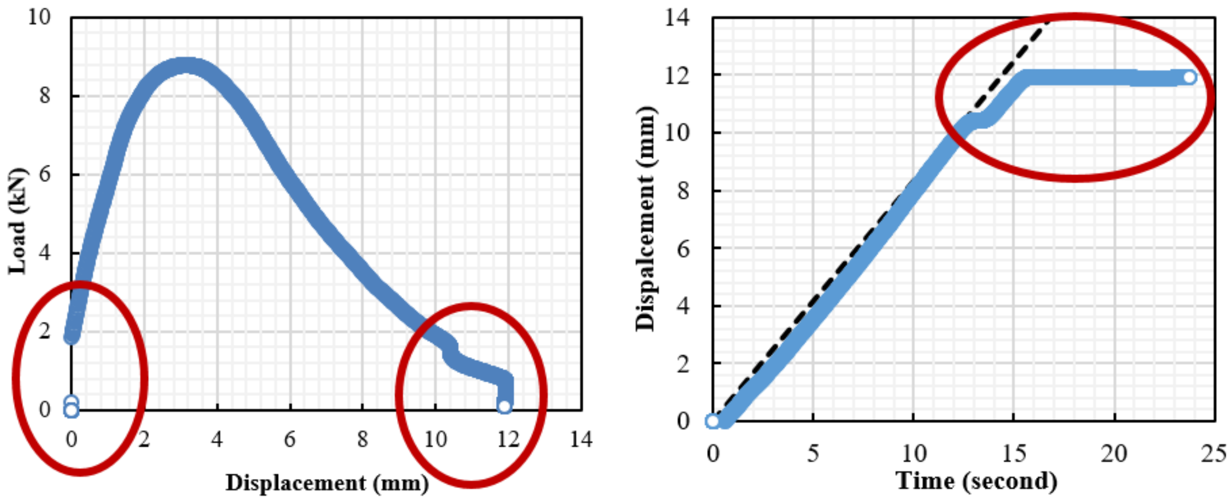


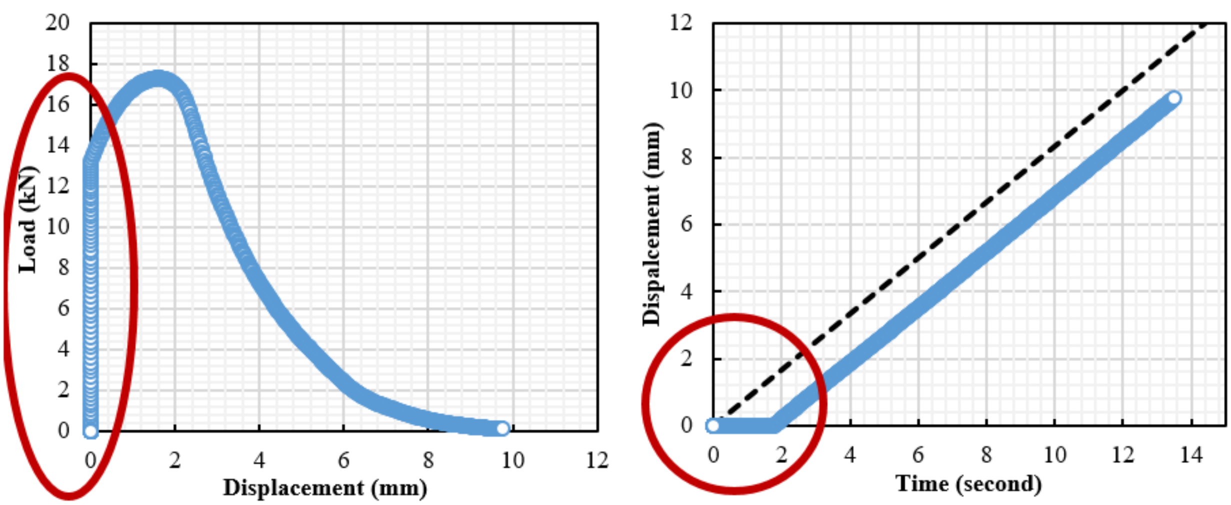

**Figure 5. Example of Non-Compliant Data Because Linear Variable Differential Transformer (LVDT) Was Out of Position: (a) load-displacement curve; (b) displacement-time curve.**




(a) (b)

**Figure 6. Example of Non-Compliant Data Because of Seating Load Applied at Beginning of Test and a Linear Variable Differential Transformer (LVDT) Error Toward the End of the Test: (a) load-displacement curve; (b) displacement-time curve.**

Another issue with the displacement measurement is shown in Figure 7. It appears that the LVDT may not have been installed or zeroed/initialized properly, so that measurements were not collected as the test started. Figure 8 presents a load-displacement curve that is in compliance with test requirements (except for early test termination) although the test was performed using a non-linear loading rate. This demonstrates why evaluating both the load-displacement and displacement-time data is important. If the loading rate is non-compliant, the equipment is not meeting the test requirements and may need troubleshooting or maintenance.




(a) (b)

**Figure 7. Example of Non-Compliant Data Because of Improper Initial Installation of Linear Variable Differential Transformer (LVDT): (a) load-displacement curve; (b) displacement-time curve.**

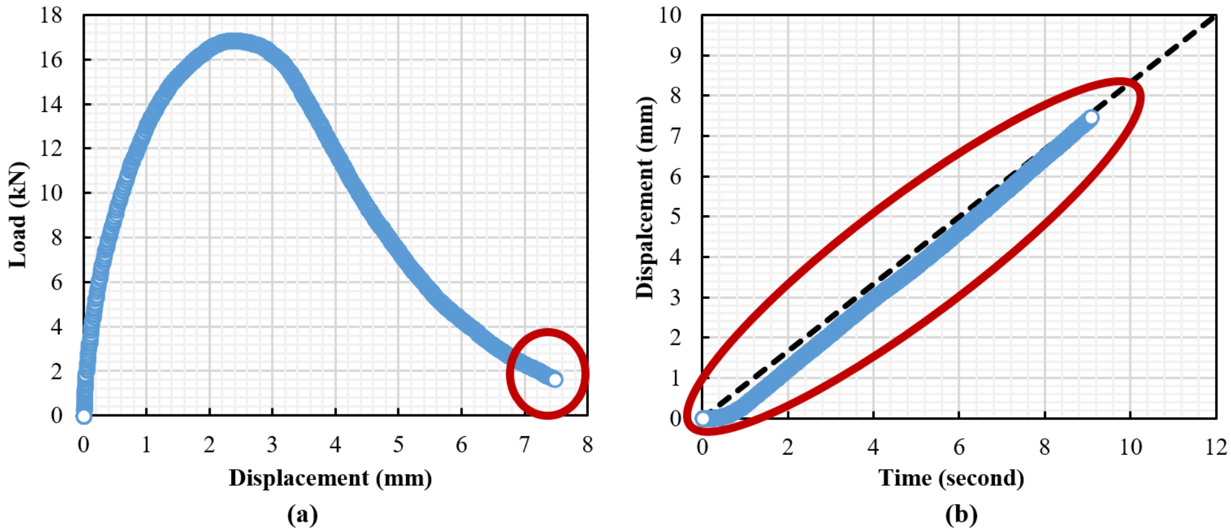



**Figure 8. Example of Non-Compliant Data: (a) load-displacement curve with load not reaching 0.1 kN at end of test; (b) displacement-time curve with rate less than specified and changes occurring during the test.**

Several observations regarding data quality were made regarding the 46 data sets submitted by the 41 participating laboratories for each mixture:

- Three laboratories were unable to perform the testing because of machine-related issues.

- Three laboratories were unable to provide the raw data from their tested specimens.

- Ten laboratories could not perform the test in accordance with ASTM D8225-19 (i.e., 10 data sets). The common issues found with these laboratories were mainly errors with the test setup and data acquisition system such as application of a seating load at the beginning of the test and displacement measurements stopping before the end load was reached; a displacement-measuring device (i.e., LVDT) slipping out of the position; and synchronization issues with a load cell and a displacement measuring device, i.e., the load was recorded but the displacement lagged.

- Fourteen laboratories had issues satisfying the loading rate requirement of 50 ± 2 mm/min for Mixture A (i.e., 14 data sets) (3 of these 14 laboratories received more than one set of test specimens per mixture to perform testing using machines or load frames from different manufacturers), and 10 laboratories had the same issues for Mixture B (i.e., 10 data sets) (1 of these 10 laboratories received more than one set of test specimens per mixture to perform testing using machines or load frames from different manufacturers). This issue was specifically observed for two of seven different devices used in this study. The overall slope (i.e., loading rate) of the displacement-time curves from these laboratories was either 47 ± 1 mm/min or 53 ± 1 mm/min, failing the requirement of 50 ± 2 mm/min.

This resulted in a total of only 14 participating laboratories having submitted results (i.e., 16 data sets) in full accordance with ASTM D8225-19 for both mixtures.

## Analyses and Observations

The obtained data were evaluated for all four indices considered in this study: CT index, FST index, $S_t$, and CRI. Each laboratory reported five replicate measurements for each mixture. Two different approaches were used in performing data analysis: (1) original data, an untrimmed approach using all five replicates, and (2) filtered data, a trimmed approach removing the minimum and maximum values of the index considered and using the remaining three measurements per each mixture. Figure 9 presents the distribution of the individual CT index values (original untrimmed data) reported by the participating laboratories for Mixture A and Mixture B.

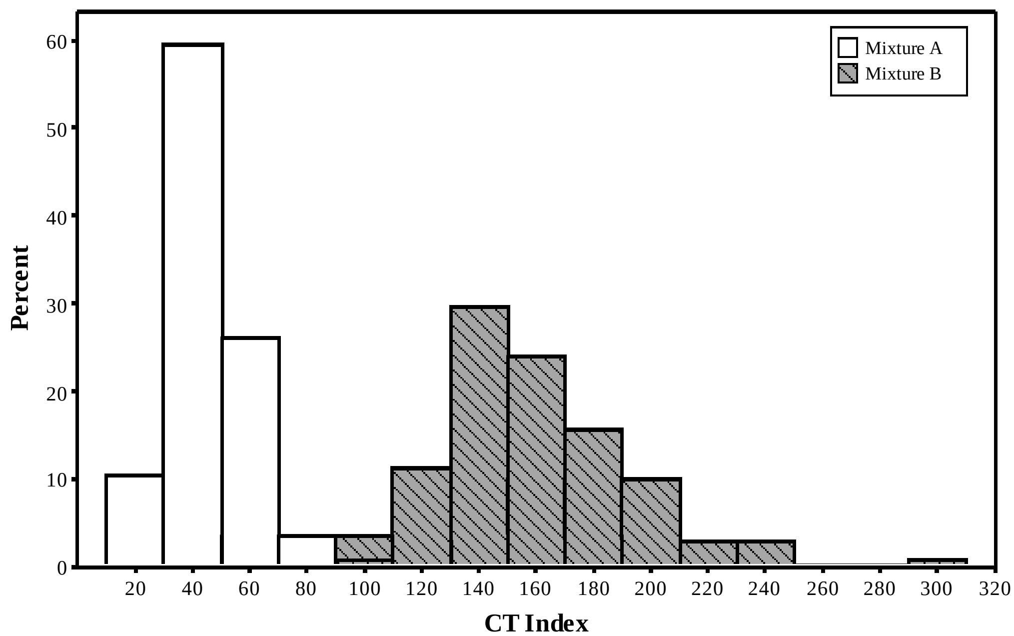



**Figure 9. Individual Reported CT Index Values for Mixture A and Mixture B. CT = cracking tolerance.**

*Effect of Devices and Loading Rate on IDT-CT Indices*

A total of seven devices, labeled Devices I through VII, were identified in this study. Of the participating laboratories, three had multiple devices. This presented an opportunity to investigate from a single-laboratory and single-operator analysis standpoint whether or not the particular device used had a significant effect on the indices considered in this study. To evaluate the effect of a device, two participants (referred to herein as Lab X and Lab Z) were provided three sets of specimens per mixture type to be tested using different devices. An additional participant (referred to herein as Lab Y) was provided two sets of specimens per mixture type for the same reason. Among the equipment used by the three laboratories, four devices were identified and designated Device I (servo-hydraulic machine, hereinafter "SH"); Device II (screw-drive machine, hereinafter "SD"); Device III (SD); and Device IV (SD).

**CT Index.** For the sake of brevity, Figure 10 shows the average CT index of data reported by Lab X (as an example) using different devices. Both data analysis approaches, original and trimmed, were considered. The average CT index values from the original data were similar to the average CT index values from the trimmed data. This observation was statistically confirmed by conducting the paired t-test at a 5% significance level on each pair of data points for each laboratory. The analyses indicated no statistically significant differences between the CT index average values before and after trimming. It is important to note that the COV for the original data ranged from 5.0% to 39.5%, with an average COV of 20.9%. The COV for the trimmed data ranged from 2.1% to 20.2%, with an average COV of 9.5%, thus showing a significant drop in the variability of the test results when trimming was applied.

An analysis of variance (ANOVA) at a 95% confidence interval was performed to determine if there was a statistically significant difference in the CT index results when different devices were used for testing. For the response variable (i.e., CT index), the parameters used as factors in the analysis model were "mixture type" and "device." An interaction term "mixture type*device" was also added into the model as a factor. Initial runs of the model indicated that the assumption of normality was not satisfied for the CT index data obtained from any of the three laboratories, although deviations from the normality assumption were not heavily skewed toward one side (i.e., left skewed or right skewed). In addition, since the data collected were based on subgroups (mixture type and device) and there was independency between the subgroups, not satisfying the normality assumption may not affect the test results significantly. With these considerations in mind, the research team performed an ANOVA on the data as collected and on the data transformed through a Box-Cox transformation (Johnson and Wichern, 2007), which was done to satisfy the normality assumption. Minitab software was used in the analysis, and the Box-Cox transformations were optimized by the software (Minitab, 2018).

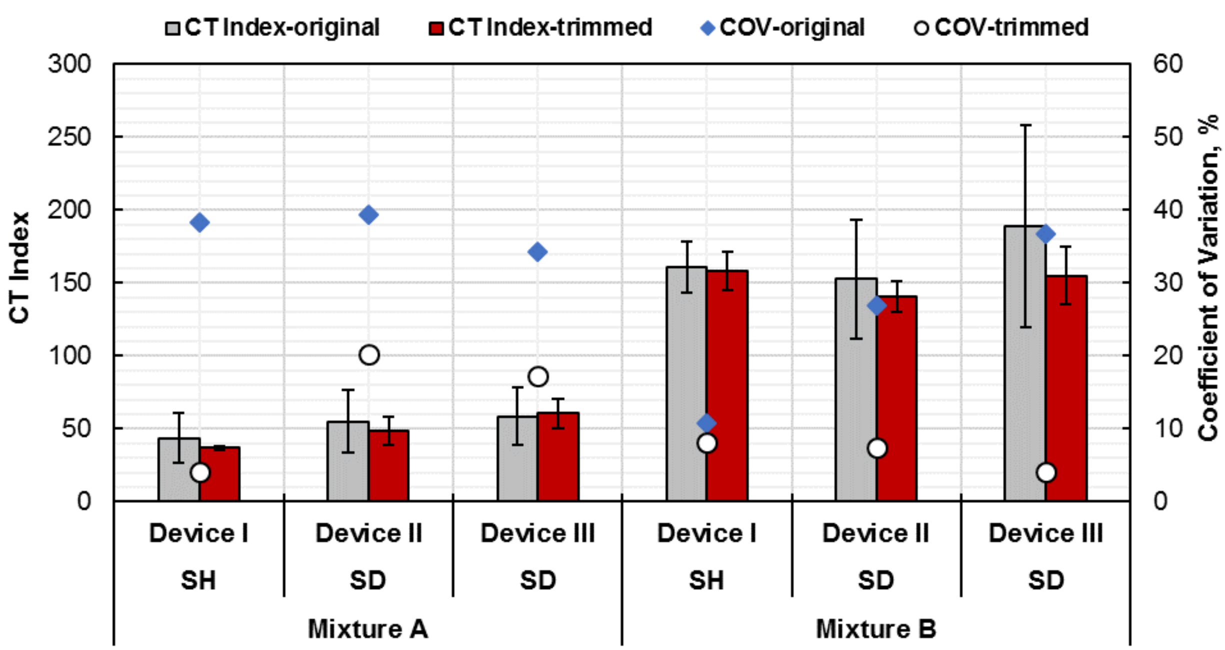



**Figure 10. CT Index Values Reported by Lab X Determined by Testing Using Different Devices / Machine Types. I-bars show ±1 standard deviation. CT = cracking tolerance; COV = coefficient of variation; SH = servo-hydraulic; SD = screw-drive.**

Tables 2 and 3 present the ANOVA statistics for the response variable (i.e., CT index) without transformation (as-collected data) using the original and trimmed data, respectively. The specified loading rate of 50 ± 2 mm/min as per ASTM D8225-19 for the IDT-CT was not achieved for the data obtained from both mixtures using Device III for Lab X. This was also the case for the data obtained for Mixture A using Device IV for Lab Z. For both laboratories, the loading rates varied between 52 and 53 mm/min. In addition, the time data for both mixtures from Lab Y could not be retrieved from Device III. It is assumed that the loading rate for this device was likely out of the specified range, given that it was a common observation from other participating laboratories in this study that tested using Device III. These data were included in the statistical analyses regardless of the fact that the tests were not within the loading rate tolerance limits as per ASTM D8225-19.

As shown in Tables 2 and 3, the mixture type was a statistically significant factor across the three laboratories for both original and trimmed data. The device used was a statistically significant factor for only Lab Z for both the original and trimmed data. A pairwise comparison using the Tukey method indicated that the CT index results obtained using Device II were the source of the statistical difference. It is speculated that this statistical difference was potentially due to the mixture and/or operator variability and not to the device used. After the CT index results from all other laboratories participating in this study were checked, it was found that the results from Lab Z using Device II were on the higher end of the spectrum of the average CT index results of all the laboratories. In addition, the test specimens for the mixtures tested using Device II were shipped to Lab Z after the device was procured. These specimens were fabricated as extra sets as part of the ILS requirement as outlined in ASTM E691-19 and stored at the VTRC laboratory in a climate-controlled environment. There might have been changes in the material properties during storing, handling, shipping, and/or the testing process, thus potentially affecting the test results obtained.

Tables 4 and 5 present the ANOVA statistics for the transformed response variable (i.e., CT index) using the original and trimmed data, respectively. For the original data as shown in Table 4, the mixture type was identified as a statistically significant factor ($p < 0.05$) across the three laboratories.

**Table 2. Summary of ANOVA Results for CT Index Using Original Data**

| | Lab X | | Lab Y | | Lab Z | |
|---|---|---|---|---|---|---|
| **Factor** | **DF** | **p-value** | **DF** | **p-value** | **DF** | **p-value** |
| Mixture Type | 1 | ***0.000*** | 1 | ***0.000*** | 1 | ***0.000*** |
| Device | 2 | 0.344 | 1 | 0.697 | 2 | ***0.004*** |
| Mixture Type*Device | 2 | 0.586 | 1 | 0.461 | 2 | 0.551 |

Bold italic text indicates that the p-values were lower than 0.05. ANOVA = analysis of variance; CT = cracking tolerance; DF = degrees of freedom.

**Table 3. Summary of ANOVA Results for CT Index Using Trimmed Data**

| | Lab X | | Lab Y | | Lab Z | |
|---|---|---|---|---|---|---|
| **Factor** | **DF** | **p-value** | **DF** | **p-value** | **DF** | **p-value** |
| Mixture Type | 1 | ***0.000*** | 1 | ***0.000*** | 1 | ***0.000*** |
| Device | 2 | 0.188 | 1 | 0.429 | 2 | ***0.002*** |
| Mixture Type*Device | 2 | 0.107 | 1 | 0.340 | 2 | 0.561 |

Bold italic text indicates that the p-values were lower than 0.05. ANOVA = analysis of variance; CT = cracking tolerance; DF = degrees of freedom.

**Table 4. Summary of ANOVA Results for Transformed CT Index Using Original Data**

| Factor | Lab X | | Lab Y | | Lab Z | |
|---|---|---|---|---|---|---|
| | DF | p-value | DF | p-value | DF | p-value |
| Mixture Type | 1 | ***0.000*** | 1 | ***0.000*** | 1 | ***0.000*** |
| Device | 2 | 0.371 | 1 | 0.884 | 2 | ***0.000*** |
| Mixture*Device | 2 | 0.563 | 1 | 0.272 | 2 | 0.257 |

Bold italic text indicates that the p-values were lower than 0.05. ANOVA = analysis of variance; CT = cracking tolerance; DF = degrees of freedom.

**Table 5. Summary of ANOVA Results for Transformed CT Index Using Trimmed Data**

| Factor | Lab X | | Lab Y | | Lab Z | |
|---|---|---|---|---|---|---|
| | DF | p-value | DF | p-value | DF | p-value |
| Mixture Type | 1 | 0.000 | 1 | ***0.000*** | 1 | ***0.000*** |
| Device | 2 | 0.079 | 1 | 0.891 | 2 | ***0.000*** |
| Mixture*Device | 2 | ***0.038*** | 1 | 0.262 | 2 | 0.066 |

Bold italic text indicates that the p-values were lower than 0.05. ANOVA = analysis of variance; CT = cracking tolerance; DF = degrees of freedom.

As also observed in the analyses of the as-collected data, the device used was found to be a statistically significant factor for Lab Z only, and this was due to the results from Device II. As shown in Table 5, the ANOVA results for the trimmed data were similar to those observed with the original transformed data except that the device for Lab X turned out to be a statistically significant parameter for Mixture A. Based on the pairwise analysis, it was seen that the CT index results obtained from testing Mixture A using Device III were statistically the same as the CT index results obtained from the same mixture type using Device II but differed from the CT index results of Mixture A determined using Device I.

A regression analysis was performed to quantify the degree of contribution of the factors to the percent variation observed in the CT index. As shown in Figure 11, the mixture type had the most significant effect on the variation observed in the CT index when the original data were used. It was also seen that the experiment factor (combination of operator, mixture, and device factors) contributed more than the device in the variation observed in the CT index. The average adjusted coefficient of determination ($R^2$) for this analysis was 90%. Similar observations were made with the trimmed data.

The data evaluated from Lab X, Lab Y, and Lab Z showed that the CT index results tended to be independent of the device used, regardless of whether data trimming was performed or not. The statistical analyses also indicated that the loading rate limit as specified in ASTM D8225-19 for the IDT-CT may need revising, given that there were devices applying a loading rate of 52 mm/min to 53 mm/min (for these three laboratories) to the test specimens with no statistically significant difference from those devices that applied a loading rate within the predefined allowable tolerance.

The effect of the device on the CT index results was further evaluated using the data from all participating laboratories (a total of 25 laboratories and 30 datasets). Again, 14 of 25 laboratories (i.e., 16 data sets) were able to test in full accordance with ASTM D8225-19, and the remaining laboratories either could not apply the specified loading rate with their device within the tolerance limits or were not able to retrieve the time data from the tests.

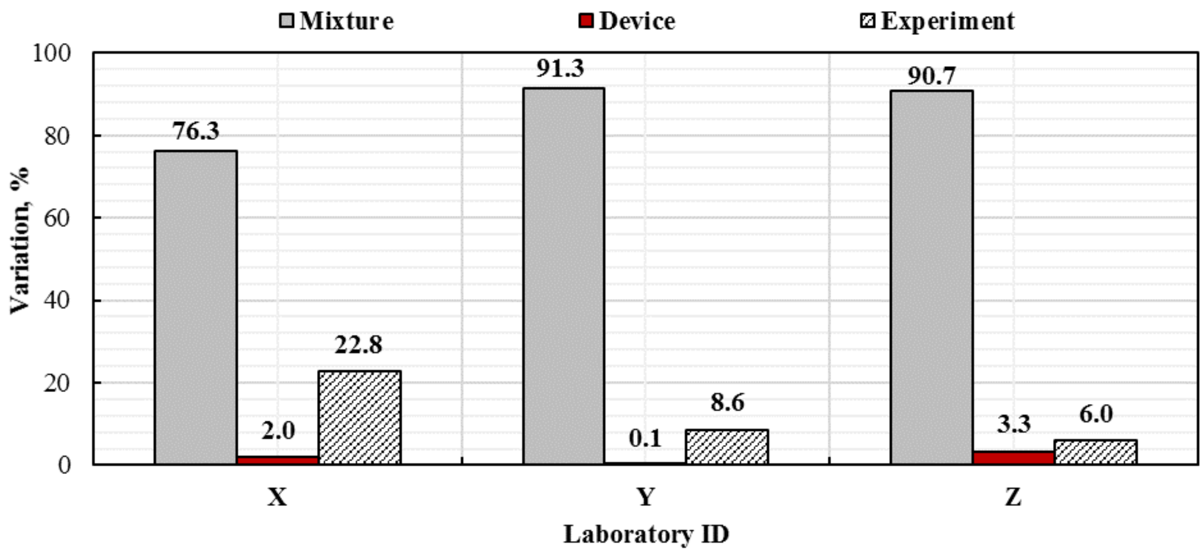



**Figure 11. Factors Contributing to Percent Variation in CT Index Using Original Data. CT = cracking tolerance.**

Table 6 presents the ANOVA results at a 95% confidence interval for the CT index using the original and trimmed data. Several factors were considered in this analysis including mixture type, device, laboratory, and some interaction parameters thereof. In addition, the device was nested as a factor under the laboratory factor. As shown in Table 6, for the original data, the CT index was significantly varied as a function of interaction between the mixture type and laboratory, and hence the device, because it was nested as a factor under the laboratory factor. In order to identify statistically significant CT index results, pairwise comparisons using the Bonferroni test at a 95% confidence level were conducted for the original data. The analysis showed that CT index results were, as expected, significantly varied with differences in the mixture types. It was also observed that the CT index results of two laboratories were significantly different than those of other laboratories for Mixture B only. Of the two laboratories, one laboratory's device was not able to apply the specified loading rate.

Table 6 also presents the ANOVA statistics for the CT index using the trimmed data from all laboratories. The analysis resulted in the same conclusions as with the original data. The pairwise comparisons using the Bonferroni test at a 95% confidence level were performed to identify statistically significant CT index results. For the trimmed data, the CT index did not vary significantly among all laboratories for Mixture A. However, there were six laboratories for which the data resulted in statistically significant different CT index values for Mixture B, one of which did not meet the specified loading rate in the ASTM standard. In addition, four different devices were used across the six laboratories.

The CT index results from all laboratories indicated that the statistical differences were random and not particularly influenced by the devices, including those that failed to maintain the specified loading rate, confirming the findings from the laboratories with multiple equipment types.

**Table 6. Summary of ANOVA Results for CT Index Using Original and Trimmed Data for All Laboratories**

| Factor | Original | | Trimmed | |
|---|---|---|---|---|
| | DF | p-value | DF | p-value |
| Mixture Type | 1 | 0.000 | 1 | 0.000 |
| Device | 6 | 0.654 | 6 | 0.370 |
| Laboratory (Device) | 23 | 0.023 | 23 | 0.000 |
| Mixture Type*Device | 6 | 0.498 | 6 | 0.021 |
| Laboratory*Mixture Type (Device) | 23 | ***0.021*** | 23 | ***0.000*** |

Bold italic text indicates that the p-values were lower than 0.05. ANOVA = analysis of variance; CT = cracking tolerance; DF = degrees of freedom.

**FST Index.** Figure 12 presents the average FST index values of data reported by Lab X using different devices with and without trimming. Similar to the CT index, the average FST index results from the original data were similar to the average FST index results from the trimmed data for all three laboratories (i.e., Labs X, Y, and Z). The results of the paired t-test at a 5% significance level on each pair of data points for each laboratory indicated no statistically significant differences between the FST index results before and after trimming. The COV for the original data with all data points of the three laboratories considered ranged from 1.6% to 11.5%, with an average COV of 6.6%. The COV for the trimmed data ranged from 0.3% to 8.6%, with an average COV of 3%. The results indicated that the FST index has a higher repeatability characteristic compared to the CT index. Given the very high degree of correlation between the two indices (Diefenderfer et al., 2019; Diefenderfer et al., 2021; Habbouche et al., 2021; Seitllari et al., 2020), such a high repeatability characteristic of the FST index can be more advantageous during the quality measurement process because statistical similarities or dissimilarities between two sets of a given mixture (i.e., multi-laboratory evaluation) can be confidently identified without the masking effect from a high-variability test method/index.

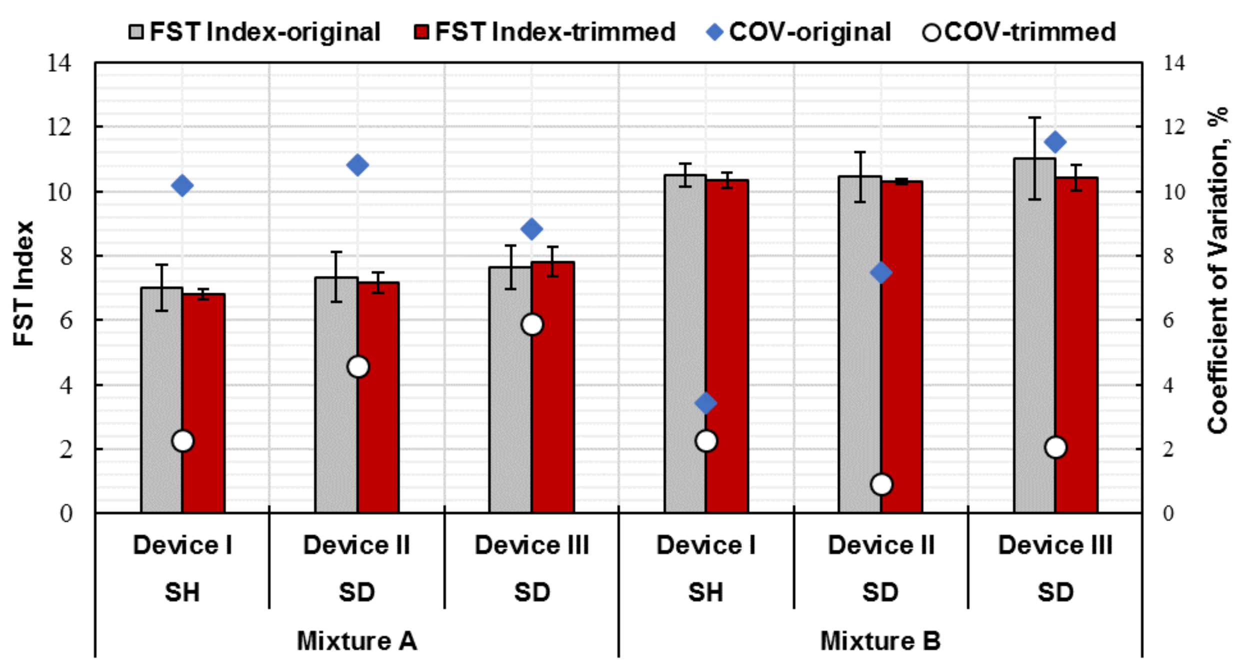



**Figure 12. FST Index Values Reported by Lab X Determined by Testing Using Different Devices. I-bars show ±1 standard deviation. FST = fracture strain tolerance; COV = coefficient of variation; SH = servo-hydraulic; SD = screw-drive.**

Similar to the CT index, a regression analysis was conducted to determine the degree of contribution of the factors to the percent variation observed in the FST index. As shown in Figure 13, when the original data were used, the mixture type had the most significant effect on variation in the FST index followed by the experiment factor. The device had a minimal impact on variation in the FST index among the considered factors. The average adjusted $R^2$ for this analysis was 89.6%. Similar observations were made with the trimmed data.

As also observed for the CT index, the data from the three laboratories indicated that the FST index results were not generally dependent on the device used, regardless of whether data trimming was pursued or not. Further, the statistical analyses also showed that the loading rate tolerance limit of ±2 mm/min for the loading rate of 50 mm/min may not be tightly applicable to the FST index, as there were devices applying a loading rate of 53 ± 1 mm/min (for these three laboratories) to the test specimens that did not result in a statistically significant difference from devices that applied a loading rate within 50 ± 2 mm/min.

The effect of the device on the FST index results was further evaluated using the data from all participating laboratories (a total of 25 laboratories and 30 datasets) using the original and trimmed data. As was the case for the CT index, the results from all laboratories showed that the statistical differences appeared to be random and did not particularly come from the devices, including those that failed to maintain the specified loading rate, thus confirming the findings from the laboratories with multiple devices.

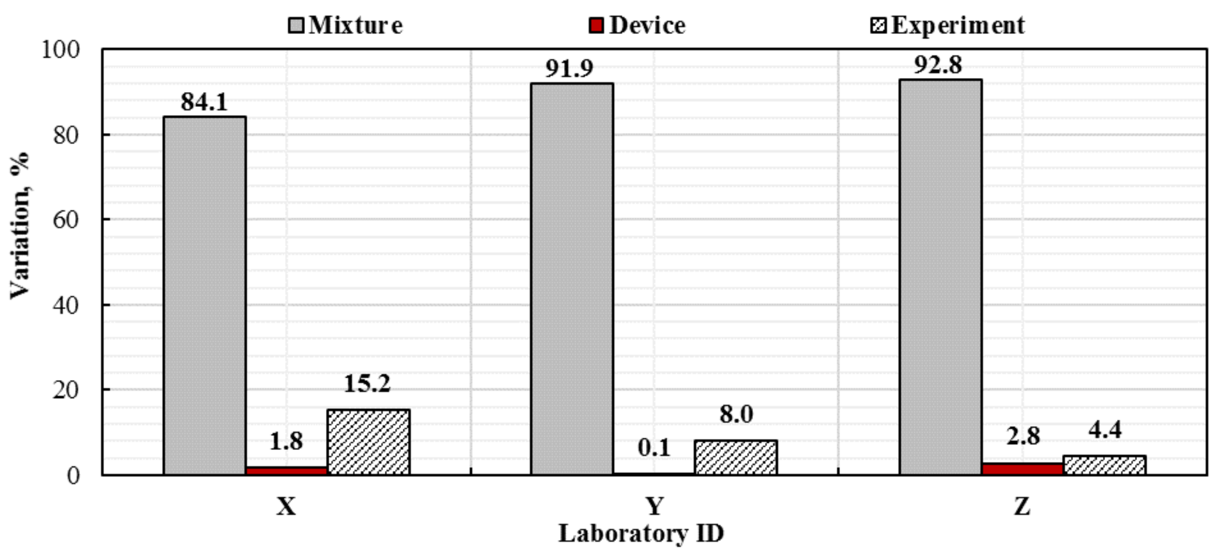



**Figure 13. Factors Contributing to Percent Variation in FST Index Using Original Data. FST = fracture strain tolerance.**

**Indirect Tensile Strength, $S_t$.** Figure 14 provides the average $S_t$ values of data reported by Lab X using different devices with and without trimming. The average $S_t$ values from the original data were similar to the average $S_t$ values from the trimmed data for all three laboratories, the same observation as for the two other indices (i.e., CT index and FST index).

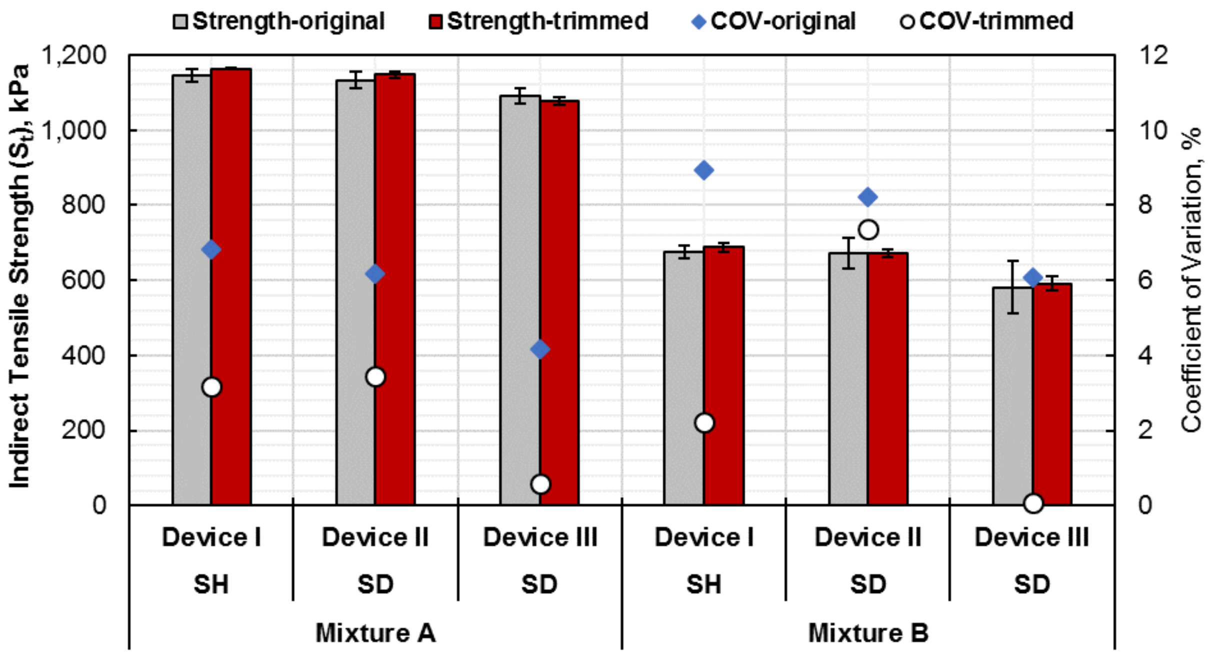



**Figure 14. $S_t$ Values Reported by Lab X Determined by Testing Using Different Devices. I-bars show ±1 standard deviation. FST = fracture strain tolerance; COV = coefficient of variation; SH = servo-hydraulic; SD = screw-drive.**

The results between the original and trimmed data did not indicate any statistically significant differences. The COV for the original data, which reflected all data points of the three laboratories, ranged from 1.8% to 10.7%, with an average COV of 5.9%. The COV for the trimmed data ranged from 0.1% to 8.3%, with an average COV of 3.8%. The results indicated that $S_t$ was a repeatable parameter.

A regression analysis was also performed to determine the degree of contribution of the factors to the percent variation observed in the $S_t$ results for both original and trimmed data. The results for the other two indices were similar for this parameter when both data sets were evaluated (i.e., original and trimmed). For example, as shown in Figure 15 for the case of using original data, the mixture type was the most significant factor in the variation observed in the indirect tensile strength followed by the experiment factor and device. The average adjusted $R^2$ for this analysis was 95.3%. Since such analyses of the indices were based on the limited data, further studies should be performed to evaluate the findings in this regard. In addition, as was the case for the other two indices, the data from the three laboratories indicated that the $S_t$ results were not generally dependent on the device used regardless of whether data trimming was pursued or not.

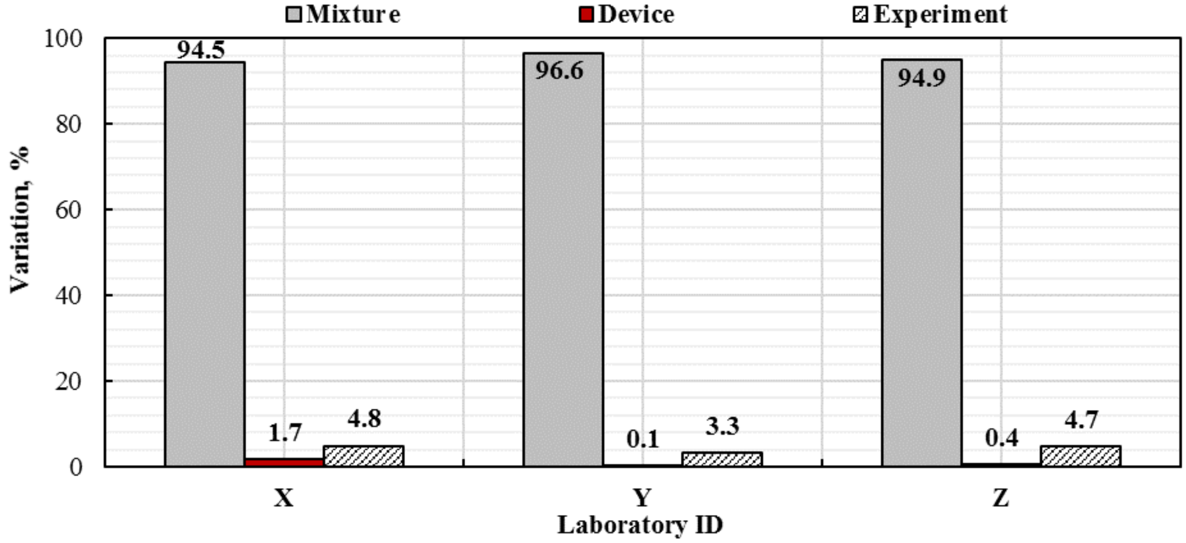



**Figure 15. Factors Contributing to Percent Variation in $S_t$ Using Original Data**

**CRI.** Figure 16 provides the average CRI values of data reported by Lab X using different devices with and without trimming. The average CRI values from the original data were similar to the average CRI values from the trimmed data for all three laboratories, the same observation as for the three other indices and parameters (i.e., CT index, FST index, and $S_t$). The results for the original and trimmed data did not indicate any statistically significant differences. The COV for the original data, with all data points of the three laboratories considered, ranged from 3.5% to 11.5%, with an average COV of 8.7%. The COV for the trimmed data ranged from 1.0% to 5.3%, with an average COV of 2.7%. The results indicated that the CRI was a repeatable index.

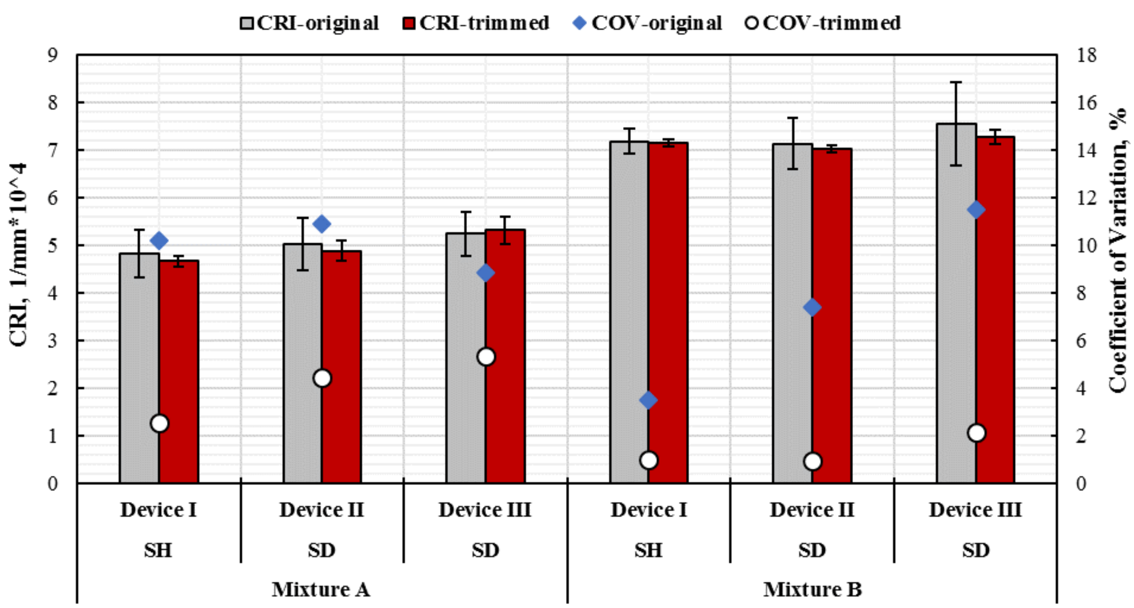



**Figure 16. CRI Values Reported by Lab X Determined by Testing Using Different Devices. I-bars show ±1 standard deviation. CRI = cracking resistance index; COV = coefficient of variation; SH = servo-hydraulic; SD = screw-drive.**

A regression analysis was also performed to determine the degree of contribution of the factors to the percent variation observed in the CRI results for both original and trimmed data. The results for the other three indices and parameters were similar for this parameter when both data sets were evaluated (i.e., original and trimmed). For example, as shown in Figure 17 for the case of using original data, the mixture type was the most significant factor in variation observed in the CRI followed by the experiment factor and device. The average adjusted $R^2$ for this analysis was 89.4%. Since such analyses performed on the indices were based on the limited data, further studies should be performed to evaluate the findings in this regard. In addition, as was the case for the other three indices and parameters, the data from the three laboratories indicated that the CRI results were not generally dependent on the device used regardless of whether data trimming was pursued or not.

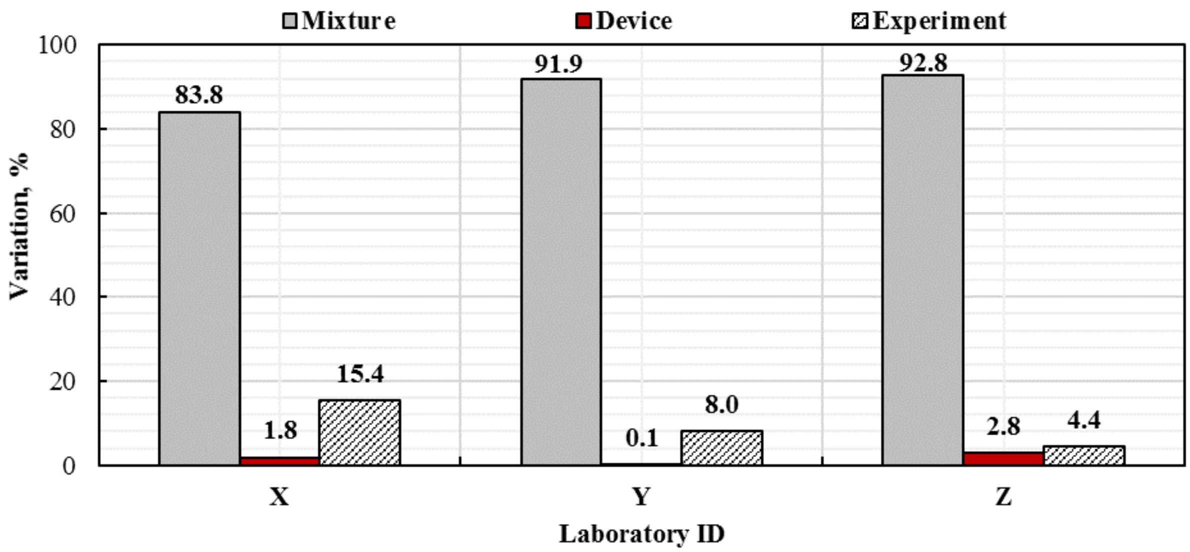



**Figure 17. Factors Contributing to Percent Variation in CRI Using Original Data. CRI = cracking resistance index.**

*Precision Estimates and Statements*

The precision estimates were determined in accordance with ASTM E691-19 for the four considered indices, and the precision statements were developed in accordance with ASTM C670-15. The two approaches (i.e., original and trimmed) were considered when the data for the precision estimates were analyzed. In addition, the data obtained as part of this study were grouped and analyzed in two categories: Category I, 16 data sets from laboratories that were able to perform the test in full accordance with ASTM D8225-19, and Category II, 30 data sets that included laboratories that were able to perform the test in accordance with ASTM D8225-19 with the exception of satisfying the specified loading rate. The second approach was considered because small deviations from 50 ± 2 mm/min have not appeared to significantly affect the values of all calculated IDT-CT indices (i.e., CT index, FST index, $S_t$, and CRI) for data collected to date. This resulted in the generation of four different precision estimates and associated statements for each of the considered indices.

**CT Index.** In order to determine the form of the precision statements, the relationships between the average CT index values and the standard deviation and between the average CT index values and the COVs for single-operator and multi-laboratory were investigated for both data groups (16 and 30 laboratories with both the original and trimmed data) in accordance with ASTM E691-19. The results indicated that the COV had a relatively lower rate of change across data from Mixtures A and B when compared with the standard deviation across the same data sets. Therefore, the COV was the appropriate parameter to be considered for developing the precision statements for the CT index. Table 7 summarizes the precision estimates resulting from this study for the CT index. A substantial drop in the COV for both single-operator and multi-laboratory conditions was observed when the data were trimmed.

**Table 7. Summary of Precision Estimates for the CT Index**

| Data Sets and Conditions | | Precision Estimates, COV, % | |
|---|---|---|---|
| | | **Single-operator** | **Multi-laboratory** |
| 16 data sets per mix | Original data (5 replicates) | 18.3% | 21.3% |
| | Trimmed data (3 replicates) | 11.2% | 15.9% |
| 30 data sets per mix | Original data (5 replicates) | 20.7% | 21.9% |
| | Trimmed data (3 replicates) | 12.8% | 16.9% |

CT = cracking tolerance; COV = coefficient of variation.

The following precision statements for the CT index using the original data of 16 data sets per mixture type were developed in accordance with ASTM C670-15 and are provided as examples. The precision estimates can be similarly developed for the other conditions (i.e., trimmed data of 16 data sets, original data of 30 data sets, and trimmed data of 30 data sets) in accordance with ASTM C670-15.

> Single-Operator Precision—The single-operator coefficient of variation was found to be 18.3%. Therefore, results of two properly conducted tests by the same operator on the same material are not expected to differ from each other by more than 51.1%[A] of their average.
>
> Multi-Laboratory Precision—The multi-laboratory coefficient of variation was found to be 21.3%. Therefore, results of two properly conducted tests by two different laboratories on specimens of the same material are not expected to differ from each other by more than 59.7%[A] of their average.
>
> [A]These numbers represent the difference limits in percent (d2s%) as described in Practice ASTM C670.
>
> Note X—These precision statements are based on an interlaboratory study that involved 14 laboratories (16 data sets), two materials with CT index values ranging from 44 to 162, and five replicate tests per operator.

Table 8 compares the precision estimates for the CT index determined in this study and those determined in other studies (Bennert et al., 2020; Taylor, 2019; Taylor et al., 2019). Overall, the COVs for both single-operator and multi-laboratory conditions were similar for the original data among all studies regardless of the number of data sets considered as part of the analysis (i.e., 16 or 30).

**Table 8. Comparison of Precision Estimates for CT Index Among Various Studies**

| Study | Study Description | | | Precision Estimates, COV, % Single-operator | Multi-laboratory |
|---|---|---|---|---|---|
| NCAT (Taylor, 2019; Taylor et al., 2019) | Phase I:<br>• One mixture<br>• Fifteen participants<br>• Specimens compacted by participating laboratories<br>• Target air voids 7.0% ± 0.5%<br>• Minimum of 5 replicates | | | 19.5% | 35.3% |
| | Phase II:<br>• One mixture (the same mixture as in Phase I)<br>• Fourteen participants<br>• Single lab specimen compaction<br>• Target air voids 7.0% ± 0.5%<br>• Minimum of 5 replicates | | | 18.8% | 20.2% |
| Rutgers University (Bennert et al., 2020) | • Five mixtures<br>• Nine participants<br>• Single lab specimen compaction<br>• Target air voids 5.5% ± 0.5%<br>• Three replicates per mixture | | | 15.2% | 23.0% |
| VTRC | • Two mixtures<br>• 41 participants<br>• 46 data sets<br>• Single lab specimen compaction<br>• Target air voids 7.0% ± 0.5%<br>• Five replicates per mixture | Analysis: 16 data sets per mix | Original data | 18.3% | 21.3% |
| | | | Trimmed data | 11.2% | 15.9% |
| | | Analysis: 30 data sets per mix | Original data | 20.7% | 21.9% |
| | | | Trimmed data | 12.8% | 16.9% |

CT = cracking tolerance; COV = coefficient of variation; NCAT = National Center for Asphalt Technology; VTRC = Virginia Transportation Research Council.

**FST Index.** The analyses and results showed that the standard deviation is the appropriate parameter to be considered for developing the precision statements for the FST index. Table 9 summarizes the precision estimates for the FST index. Similar to the CT index, a significant drop in the standard deviation for both single-operator and multi-laboratory conditions was observed when the data were trimmed.

**Table 9. Summary of Precision Estimates for the FST Index**

| Data Sets and Conditions | | Precision Estimates, Standard Deviation Single-operator | Multi-laboratory |
|---|---|---|---|
| 16 data sets per mix | Original data (5 replicates) | 0.56 | 0.58 |
| | Trimmed data (3 replicates) | 0.31 | 0.43 |
| 30 data sets per mix | Original data (5 replicates) | 0.58 | 0.61 |
| | Trimmed data (3 replicates) | 0.34 | 0.44 |

FST = fracture strain tolerance.

The following precision statements for the FST index using the original data of 16 data sets per mixture type were developed in accordance with ASTM C670-15 and are provided as examples. The precision estimates can be similarly developed for the other conditions in accordance with ASTM C670-15.

Single-Operator Precision—The single-operator standard deviation was found to be 0.56. Therefore, results of two properly conducted tests by the same operator on the same material are not expected to differ by more than 1.57.[A]

Multi-Laboratory Precision—The multi-laboratory standard deviation was found to be 0.58. Therefore, results of two properly conducted tests by two different laboratories on specimens of the same material are not expected to differ by more than 1.62.[A]

[A]These numbers represent the difference limits (d2s%) as described in Practice ASTM C670.

Note X—These precision statements are based on an interlaboratory study that involved 14 laboratories (16 data sets), two materials with FST index values ranging from 7.1 to 10.6, and five replicate tests per operator.

**Indirect Tensile Strength, $S_t$.** The results indicated that the standard deviation is the appropriate basis for developing the precision statements for the indirect tensile strength. Table 10 summarizes the precision estimates for the indirect tensile strength. As with the other two indices, a drop in the standard deviation for both single-operator and multi-laboratory conditions was observed when the trimming approach was applied.

**Table 10. Summary of Precision Estimates for $S_t$**

| Data Sets and Conditions | | Precision Estimates, Standard Deviation, kPa | |
|---|---|---|---|
| | | **Single-operator** | **Multi-laboratory** |
| 16 data sets per mix | Original data (5 replicates) | 49.6 | 106.0 |
| | Trimmed data (3 replicates) | 32.7 | 103.0 |
| 30 data sets per mix | Original data (5 replicates) | 51.5 | 99.6 |
| | Trimmed data (3 replicates) | 33.1 | 94.3 |

The following precision statements for the indirect tensile strength using the original data of 16 data sets per mixture type were developed in accordance with ASTM C670-15 and are provided as examples. The precision estimates can be similarly developed for the other conditions in accordance with ASTM C670-15.

Single-Operator Precision—The single-operator standard deviation was found to be 49.6 kPa. Therefore, results of two properly conducted tests by the same operator on the same material are not expected to differ by more than 138.9 kPa.[A]

Multi-Laboratory Precision—The multi-laboratory standard deviation was found to be 106.0 kPa. Therefore, results of two properly conducted tests by two different laboratories on specimens of the same material are not expected to differ by more than 296.8 kPa.[A]

[A]These numbers represent the difference limits (d2s%) as described in Practice ASTM C670.

Note X—These precision statements are based on an interlaboratory study that involved 14 laboratories (16 data sets), two materials with the indirect tensile strength values ranging from 675.5 to 1159.3 kPa, and five replicate tests per operator.

**CRI.** The results indicated that the standard deviation is the appropriate basis for developing the precision statements for the CRI. Table 11 summarizes the precision estimates for the CRI. As with the other three indices and parameters, a drop in the standard deviation for both single-operator and multi-laboratory conditions was observed when the trimming approach was applied.

**Table 11. Summary of Precision Estimates for CRI**

| Data Sets and Conditions | | Precision Estimates, Standard Deviation, $1/mm*10^4$ | |
|---|---|---|---|
| | | **Single-operator** | **Multi-laboratory** |
| 16 data sets per mix | Original data (5 replicates) | 0.44 | 0.47 |
| | Trimmed data (3 replicates) | 0.21 | 0.29 |
| 30 data sets per mix | Original data (5 replicates) | 0.43 | 0.46 |
| | Trimmed data (3 replicates) | 0.22 | 0.30 |

CRI = cracking resistance index.

The following precision statements for the CRI using the original data of 16 data sets per mixture type were developed in accordance with ASTM C670-15 and are provided as examples. The precision estimates can be similarly developed for the other conditions in accordance with ASTM C670-15.

Single-Operator Precision—The single-operator standard deviation was found to be 0.44 $1/mm*10^4$. Therefore, results of two properly conducted tests by the same operator on the same material are not expected to differ by more than 1.23 $1/mm*10^4$.[A]

Multi-Laboratory Precision—The multi-laboratory standard deviation was found to be 0.47 $1/mm*10^4$. Therefore, results of two properly conducted tests by two different laboratories on specimens of the same material are not expected to differ by more than 1.31 $1/mm*10^4$.[A]

[A]These numbers represent the difference limits (d2s%) as described in Practice ASTM C670.

Note X—These precision statements are based on an interlaboratory study that involved 14 laboratories (16 data sets), two materials with the indirect tensile strength values ranging from 3.8 to 8.4 $1/mm*10^4$, and five replicate tests per operator.

## Interlaboratory Study: Phase I, Stage II

Phase I, Stage I, of the IDT-CT round robin was primarily completed during spring 2020. At that time, most VDOT laboratories were not included because they did not have the needed equipment to perform the IDT-CT. Fortunately, additional sets of compacted specimens for Mixture A and Mixture B were procured, stored at VTRC in a climate-controlled environment, and available for additional testing.

As part of the initial implementation efforts for BMD use in Virginia, servo-hydraulic–type machines (Device I) were procured for all VDOT laboratories during fall 2020. After all machines were installed and VDOT personnel were trained, the additional sets of specimens stored at VTRC were sent to the eight VDOT laboratories to perform testing during spring 2021. Similar to Stage I, VDOT participants were asked to submit the raw data files from their equipment in addition to reporting the CT index values for each specimen. These data files were used to perform quality checks on the data before further analysis. All participating VDOT district laboratories submitted results in full accordance with ASTM D8225-19 for both mixtures. It should be noted that the same device was procured for all VDOT laboratories; therefore, no assessment of the impact of devices could be conducted using this data set.

The precision estimates were determined in accordance with ASTM E691-19 for the four considered indices, as shown in Table 12. The two approaches, original and trimmed, were considered when the data for the precision estimates were analyzed. Similar to Stage I, a decrease in the precision parameter (COV or standard deviation) for both single-operator and multi-laboratory conditions was observed when the trimming approach was applied. Moreover, a relatively higher variability was observed in Stage II when compared to the variability evaluated in Stage I. This could be attributable to several factors including the relative lack of operator experience, a need for additional training, and/or potential changes in the material properties during storing, handling, and the shipping and/or testing process.

**Table 12. Summary of Precision Estimates for IDT-CT Indices Based on Stage II ILS Data**

| IDT-CT Index | Parameter Considered for Precision Estimates | Approach | Precision Estimates | |
|---|---|---|---|---|
| | | | Single-operator | Multi-laboratory |
| CT index | COV | Original data (5 replicates) | 23.4% | 23.8% |
| | | Trimmed data (3 replicates) | 14.8% | 15.8% |
| FST index | Stdv. | Original data (5 replicates) | 0.64 | 0.66 |
| | | Trimmed data (3 replicates) | 0.37 | 0.40 |
| $S_t$ | Stdv. | Original data (5 replicates) | 73.6 kPa | 162.9 kPa |
| | | Trimmed data (3 replicates) | 48.8 kPa | 150.7 kPa |
| CRI | Stdv. | Original data (5 replicates) | 0.44 1/mm*$10^4$ | 0.45 1/mm*$10^4$ |
| | | Trimmed data (3 replicates) | 0.25 1/mm*$10^4$ | 0.27 1/mm*$10^4$ |

IDT-CT = indirect tensile cracking test; ILS = interlaboratory study; CT = cracking tolerance; FST = fracture strain tolerance; $S_t$ = indirect tensile strength; CRI = cracking resistance index; COV = coefficient of variation; Stdv. = standard deviation.

## Impact of Shelf Life of Compacted Specimens on IDT-CT Indices

The data from Stage II were further compared to the data from Stage I for a preliminary assessment of the impact of 1 year of shelf life / storage of compacted specimens on the various IDT-CT indices. For the data from Stage I, several factors were considered and three major analyses were carried out: (i) all data were considered together; (ii) after the data were split into two separate groups based on machine type (screw-drive [SD] vs. servo-hydraulic [SH]), the groups were considered; and (iii) after the data were divided into seven separate groups based on device (Devices I through VII), the groups were considered. A single analysis was considered for the data of Stage II because all testing was performed using one device.

Figures 18 and 19 show box plots of CT index values for Mixtures A and B, respectively. The box plot represents the spread of CT index values for Stages I and II. The line in the box indicates the median, and the interquartile range (IQR) box represents the middle 50%. In addition, the whisker bars extending from either side of the box represent the ranges for the bottom 25% and the top 25% of the CT index values, not including outliers, which are represented by asterisks (*). The average (mean) of the CT index values is identified by the circle in the box. Tables 13 and 14 provide the descriptive statistics of the CT index values for Mixtures A and B, respectively, including the IQR that was used to evaluate the spread of the CT index values. The box plots and descriptive statistics for the other indices (i.e., FTS, $S_t$, and CRI) are shown in the Appendix.

The mean CT index ranged from 33 to 52 and 149 to 166 for Mixtures A and B, respectively (all data sets). For Mixture A, the mean CT index of data collected in Stage II, i.e., 33, was statistically similar to the mean CT index of all data collected in Stage I, i.e., 45 (analysis [i]); the data of group SH, 45 (analysis [ii]); and the data of Device I, 44 (analysis [iii]), indicating no clear impact of 1 year of storage on the CT index. Similarly, for Mixture B, the mean CT index for data collected in Stage II, i.e., 163, was statistically similar to the mean CT index of all data collected in Stage I, i.e., 159 (analysis [i]); data of group SH, i.e., 159 (analysis [ii]); and data of Device I, i.e., 160 (analysis [iii]).

A few outliers were observed, as shown in Figures 18 and 19. As opposed to the standard deviation, the IQR is known as a resistant measure in that extreme values (or outliers) do not affect it. By definition, the IQR is calculated as the difference between the 75th percentile (Quartile 3, Q3) and the 25th percentile (Quartile 1, Q1) of a given data set. The IQR of all data from Stage II was greater than the IQR of all data collected in Stage I (analysis [i]); data of group SH (analysis [ii]); and data of Device I (analysis [iii]) regardless of the mixture type, i.e., Mixture A or Mixture B. This observation confirmed the greater variability observed when the data set of Stage II was analyzed.

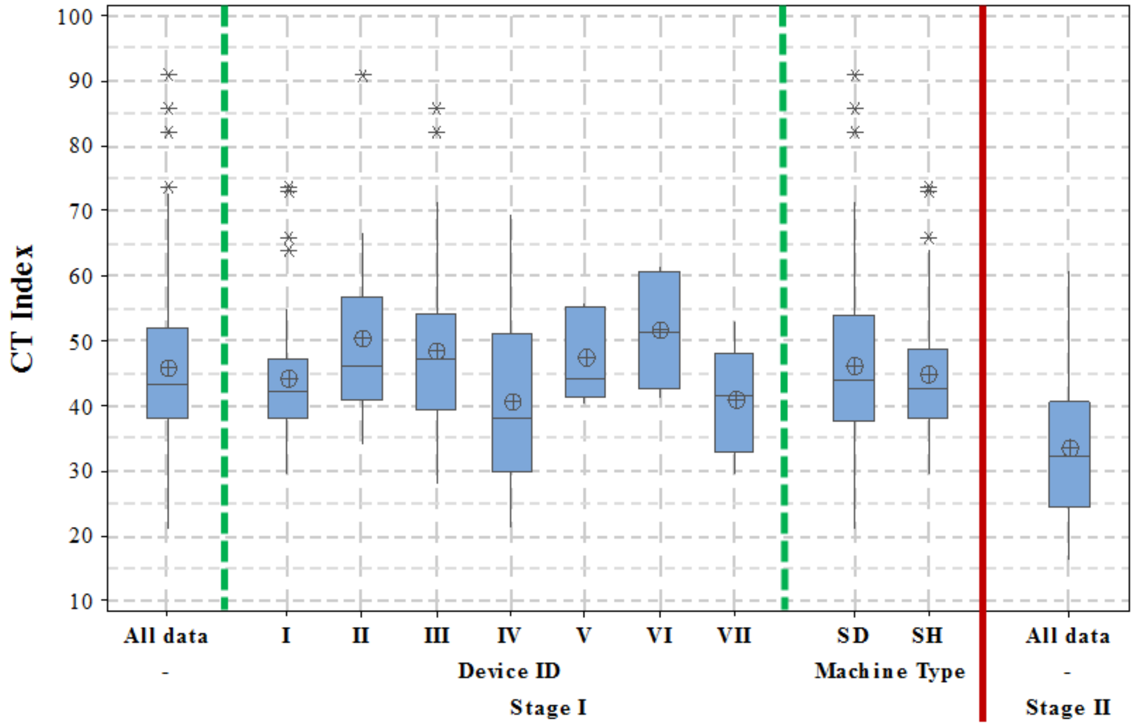



**Figure 18. Box Plots of CT Index Reported as Part of Round Robin: Phase I: Stage I and Stage II for Mixture A. Outliers are indicated by asterisks. CT = cracking tolerance; SD = screw-drive; SH = servo-hydraulic.**

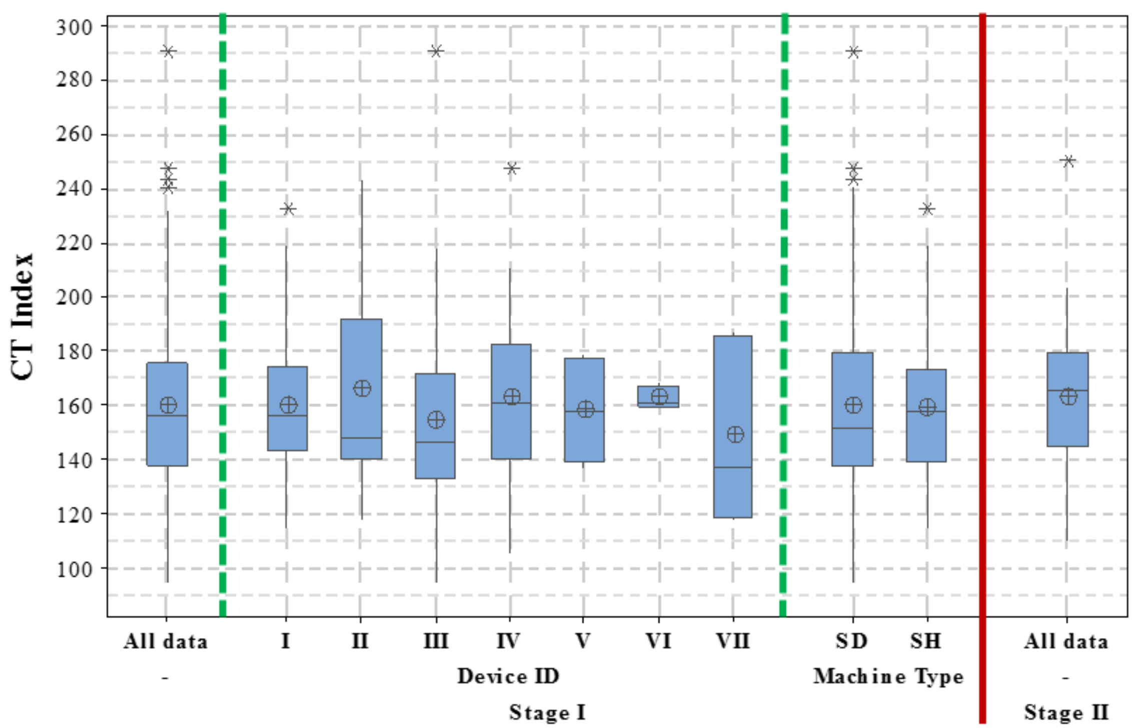



**Figure 19. Box Plots of CT Index Reported as Part of Round Robin: Phase I: Stage I and Stage II for Mixture B. Outliers are indicated by asterisks. CT = cracking tolerance; SD = screw-drive; SH = servo-hydraulic.**

**Table 13. Descriptive Statistics of CT Index for Mixture A**

| Stage | Variable | Component | Mean | Minimum | Q1 | Q3 | Maximum | IQR |
|---|---|---|---|---|---|---|---|---|
| I | All data | | 45.4 | 21.0 | 37.9 | 51.9 | 90.8 | 14.1 |
| | Device | I | 44.0 | 29.5 | 37.8 | 47.1 | 73.5 | 9.3 |
| | | II | 50.2 | 34.1 | 40.7 | 56.6 | 90.8 | 15.9 |
| | | III | 48.1 | 27.8 | 39.4 | 54.2 | 85.5 | 14.9 |
| | | IV | 40.5 | 21.0 | 29.7 | 50.9 | 69.6 | 21.2 |
| | | V | 47.3 | 40.1 | 41.1 | 55.0 | 55.7 | 14.0 |
| | | VI | 51.5 | 41.3 | 42.4 | 60.8 | 61.5 | 18.3 |
| | | VII | 40.7 | 29.6 | 32.8 | 48.1 | 52.9 | 15.3 |
| | Machine Type | SD | 45.8 | 21.0 | 37.6 | 54.0 | 90.8 | 16.4 |
| | | SH | 44.5 | 29.5 | 37.9 | 48.8 | 73.5 | 10.9 |
| II | All data | | 33.4 | 16.1 | 24.4 | 40.4 | 60.7 | 16.0 |

CT = cracking tolerance; Q1 = quartile 1; Q3 = quartile 3; IQR = interquartile range; SD = screw-drive; SH = servo-hydraulic.

**Table 14. Descriptive Statistics of CT Index for Mixture B**

| Stage | Variable | Component | Mean | Minimum | Q1 | Q3 | Maximum | IQR |
|---|---|---|---|---|---|---|---|---|
| I | All data | | 159.2 | 94.0 | 137.6 | 175.7 | 290.1 | 38.1 |
| | Device | I | 159.7 | 114.2 | 142.8 | 173.4 | 232.0 | 30.6 |
| | | II | 165.6 | 117.7 | 139.6 | 191.9 | 242.9 | 52.3 |
| | | III | 154.4 | 94.0 | 132.6 | 171.4 | 290.1 | 38.7 |
| | | IV | 162.2 | 105.2 | 139.8 | 181.9 | 247.1 | 42.1 |
| | | V | 157.8 | 137.0 | 139.2 | 176.5 | 178.7 | 37.4 |
| | | VI | 162.3 | 159.2 | 159.2 | 167.0 | 168.6 | 7.7 |
| | | VII | 148.9 | 117.2 | 118.3 | 185.5 | 187.1 | 67.2 |
| | Machine Type | SD | 159.4 | 94.0 | 137.3 | 178.9 | 290.1 | 41.6 |
| | | SH | 158.7 | 114.2 | 138.9 | 173.3 | 232.0 | 34.4 |
| II | All data | | 162.7 | 110.0 | 144.6 | 178.9 | 249.9 | 34.4 |

CT = cracking tolerance; Q1 = quartile 1; Q3 = quartile 3; IQR = interquartile range; SD = screw-drive; SH = servo-hydraulic.

## Summary and Findings

Phase I of the VDOT IDT-CT round robin study included two stages. Stage I focused on non-VDOT laboratories, and Stage II focused on VDOT laboratories. Stage II was performed 1 year after Stage I. Both efforts involved the evaluation of the specimens of the same materials that were fabricated and compacted by a third party laboratory and sent to participating laboratories along with detailed instructions for testing. The precision estimates for the CT index, FST index, $S_t$, and CRI were calculated, and the corresponding statements were developed.

The following findings were based on the data collected and analyzed during this study:

- In Stage I, only 14 of 41 participating laboratories submitted results (i.e., 16 data sets) for both mixtures in full accordance with ASTM D8225-19.
- Requirements to perform the IDT-CT on dry specimens created challenges and resulted in signicant data submission delays. Numerous participants in both stages reported challenges in keeping the IDT-CT specimens dry when conditioning using

the water bath at a temperature of 25ºC. No significant difference was reported in the literature between IDT-CT data collected on dry specimens and IDT-CT data collected on wet specimens for a given mixture.

- The initial data quality from Stage I resulted in analyses being performed on two groups of data to calculate precision estimates: the first group included 16 data sets per mixture type conforming to the requirements of ASTM D8225-19, and the second group included 30 data sets per mixture type conforming to the requirements of ASTM D8225-19 except for the loading rate of 50 ± 2 mm/min.

- The limited findings from this study showed that small deviations from the 50 ± 2 mm/min loading rate do not appear to significantly affect the calculated IDT-CT index values.

- Data trimming by eliminating the highest and lowest values of the IDT-CT index of interest from five replicate observations was evaluated. A significant drop in COV for both single-operator and multi-laboratory conditions was observed when trimming of the data was performed. Currently, the practice in Virginia is to test five specimens for the IDT-CT and average all CT index values.

- Based on the results of the 16 data sets per mixture from this study, the original untrimmed data with five replicate tests per operator had CT index values from 44 to 162. The within-laboratory and multi-laboratory COVs were 18.3% and 21.3%, respectively. Overall, the COVs for the CT index for both single-operator and multi-laboratory conditions reported by Taylor et al. (2019) in Phase II and Bennert et al. (2020) were similar to the COV values reported in this study for original untrimmed data regardless of the number of data sets considered as part of the analysis (i.e., 16 or 30).

- No precision estimates have been reported in the literature regarding the FST index, $S_t$, and CRI with the specimen dimensions used in this study. All documented work in the literature mainly focused on the CT index parameter.

- Based on the results of the 16 data sets per mixture from this study, the original untrimmed data with five replicate tests per operator had FST index values from 7.1 to 10.6. The within-laboratory and multi-laboratory standard deviations were 0.56 and 0.58, respectively.

- Based on the results of the 16 data sets per mixture from this study, the original untrimmed data with five replicate tests per operator had $S_t$ values from 675.5 to 1159.3 kPa. The within-laboratory and multi-laboratory standard deviations were 49.6 and 106.0 kPa, respectively.

- Based on the results of the 16 data sets per mixture from this study, the original untrimmed data with five replicate tests per operator had CRI values from 3.8 to 8.4

1/mm*$10^4$. The within-laboratory and multi-laboratory standard deviations were 0.44 and 0.47 1/mm*$10^4$, respectively.

- In Stage II, all eight particilating VDOT laboratories submitted results for both mixtures in full accordance with ASTM D8225-19. Two approaches (original and trimmed) were evaluated in analyzing the data for the precision estimates. Similar to Stage I, a drop in the precision parameter (COV or standard deviation) for both single-operator and multi-laboratory conditions was observed when the trimming approach was applied.

- A relatively higher variability was observed in Stage II when compared to the variability evaluated in Stage I. This could be attributable to several factors including the relative lack of operator experience, a need for training, and potential changes in the material properties during storing, handling, and the shipping and/or testing process.

- The data of Stage II were further compared to the data of Stage I in a preliminary assessment of the impact of 1 year of shelf life / storage of compacted specimens on the various IDT-CT indices. Although 1 year of storage of compacted specimens did not appear to significantly affect the calculated IDT-CT index values for data collected to date, more work is needed to validate the acceptable range of storage time to be allowed prior to testing on both compacted specimens and loose mixtures, as specimen age is expected to be impactful.

## CONCLUSIONS

- *The primary barrier to performing the IDT-CT at intermediate temperature in full accordance with ASTM requirements was the inability of testing equipment to meet the loading rate requirement of 50 ± 2 mm/min.* However, some laboratories (i.e., 10) were unable to perform the test correctly for other avoidable reasons, thus indicating the need for training.

- *Numerous difficulties in maintaining dry specimens during temperature conditioning in a water bath indicate a need to reevaluate the requirement for testing of dry specimens only.*

- *The results of the IDT-CT indices, except for $S_t$, are not dependent on the type of device/machine used, regardless of whether data trimming is applied or not.*

- *The initial findings from this study suggest that small deviations from the loading rate as required in ASTM D8225-19 (50 ± 2 mm/min) do not appear to significantly affect the calculated IDT-CT index values.*

- *Based on the results of the 30 data sets per mixture from this study, the precision estimates determined using the original untrimmed data were found to be similar to the precision estimates developed by other researchers in other studies (as reported in the literature).*

- *A comparison of Stage II test results to the data collected in Stage I suggests higher variability from testing in the VDOT laboratories at this time.* This could be attributable to the relative lack of operator experience, the need for training, and potential changes in the material properties during storing, handling, and the shipping and/or testing process.

- *The analyses and comparisons of data collected from Stage I and Stage II showed that there was no statically significant impact of 1 year of climate-controlled storage of compacted specimens on the calculated IDT-CT index values for data collected to date.*

- *The FST index, $S_t$, and CRI had a relatively smaller variability when compared to the variability of the CT index based on the precision estimates developed in this study.*

## RECOMMENDATIONS

1. *VTRC should conduct a Round Robin ILS—Phase II for the IDT-CT at intermediate temperature to assess the impact of variability induced because of specimen preparation (as reported in the literature).* Although both stages of the Round Robin ILS—Phase I developed precision estimates, the results generated from the Phase II efforts are expected to better reproduce the actual state of the practice in which industry and agency personnel will compact specimens from loose mixture samples for testing using various devices.

2. *VDOT's Materials Division should adopt the initial estimates of precision and corresponding statements developed in this study for the CT index as a sound way to monitor the repeatability and reproducibility of reported cracking performance data.* The initial precision estimates determined in this study can serve as an initial step to establish quality measurement practices and approval protocols for BMD performance data.

3. *VTRC should evaluate the feasibility of performing the IDT-CT on wet specimens.* Currently, VDOT allows use of the IDT-CT on only dry specimens conditioned in an environmental chamber or placed in leak-proof plastic bags in a water bath for 2 hours until the specimens reach a temperature of 25°C. Testing wet specimens would allow placing specimens in a water bath for 2 hours; removing the specimens and drying them until they reach the saturated surface dry condition; and immediately performing the IDT-CT. This procedure would simplify specimen conditioning, particularly in production laboratories. Although the literature seems to support this concept, additional research is needed to quantify the impact, if any, on dry versus wet specimens.

4. *VTRC should conduct a study to assess the impact of loading rate on the IDT-CT results.* Currently, ASTM D8225-19 requires use of a loading apparatus capable of maintaining a constant loading / deformation rate of 50 ± 2.0 mm/min. The preliminary findings from this study showed that small deviations from 50 ± 2 mm/min do not appear to significantly affect the calculated IDT-CT index values. However, more work is necessary to validate the acceptable range of loading rates.

5. *VTRC and VDOT's Materials Division should organize hands-on training and demonstrations of the laboratory tests (including the IDT-CT) being considered by VDOT as part of the BMD initiative.* One of the critical components of successfully implementing this initiative depends on how well agency and industry personnel are familiarized with the current practices and procedures required for the BMD concept. Therefore, it is extremely important for VDOT to have all personnel (i.e., VDOT and non-VDOT) involved with mixture design, QC/QA, and acceptance processes to have a solid understanding of the BMD concept and its procedures.

6. *VTRC and VDOT's Materials Division should evaluate the use of the other indices determined from the IDT-CT data such as the FST index, $S_t$, and CRI as potential substitutes of or companion parameters to the CT index.*

## IMPLEMENTATION AND BENEFITS

### Implementation

With regard to Recommendations 1, 3, and 4, VTRC is in the process of conducting additional efforts to develop precision estimates and statements for IDT-CT indices with a focus on assessing the impact of variability induced because of specimen preparation. VTRC is also in the process of investigating if the IDT-CT can be performed on dry and wet specimens rather than on only dry specimens. Moreover, VTRC is in the process of assessing the impact of loading rate and data collection frequency on the IDT-CT indices.

With regard to Recommendation 2, VTRC is working closely with VDOT's Materials Division and VDOT districts to evaluate the quality, repeatability, and reproducibility of performance data provided in the mix design submittals for BMD trials and pilot projects.

With regard to Recommendation 5, VTRC and VDOT's Materials Division, with the help of the Virginia Asphalt Association, organized and hosted virtual training in spring 2021. The training included topics related to the BMD concept, VDOT's BMD approach, BMD tests and associated performance thresholds, and VDOT's BMD special provisions. Planning for an in-person BMD workshop and training to be held in fall 2021 is ongoing.

With regard to Recommendation 6, VTRC will draft a Research Needs Statement and submit to the appropriate VTRC Pavement Research Advisory Subcommittee by fall 2023.

### Benefits

This study developed precision estimates and statements for the CT index, FST index, $S_t$, and CRI of asphalt mixtures determined by performing the IDT-CT at intermediate temperature. The findings from this study provide VDOT with a single-operator precision estimate for use in the acceptance of IDT-CT results for asphalt mixtures. Moreover, the findings provide VDOT with multi-laboratory precision estimates (i.e., repeatability and reproducibility) to be incorporated into acceptance specifications. This work provides sound precision statements and references to determine if individual IDT-CT results from the same evaluated asphalt mixture can be considered statistically similar. Finally, this study will contribute to the data set to be used by ASTM to develop and adopt precision statements for ASTM D8225.

## ACKNOWLEDGMENTS

The authors acknowledge Mike Dudley from the Virginia Asphalt Association for his assistance in developing and executing this round robin. Appreciation is extended to VDOT's Materials Division and VDOT's Balanced Mix Design Technical Committee for their support of this work. The contributions of Bluegrass Testing Laboratory staff in designing and producing test specimens are deeply appreciated. Appreciation is also extended to the laboratories involved in the ILS testing. The authors also appreciate the technical review panel for their expertise and guidance: Sungho Kim (Champion); Robert Crandol, Candice Entwisle, Travis Higgs, Todd Rorrer, and Thomas Schinkel, of VDOT; and Bernard Kassner of VTRC.

## REFERENCES

American Association of State Highway and Transportation Officials. *Standard Specifications for Transportation Materials and Methods of Sampling and Testing, and AASHTO Provisional Standards*. Washington, DC, 2017.

ASTM International. *ASTM C670-15: Standard Practice for Preparing Precision and Bias Statements for Test Methods for Construction Materials*. West Conshohocken, PA, 2015.

ASTM International. *ASTM D8225-19: Standard Test Method for Determination of Cracking Tolerance Index of Asphalt Mixture Using the Indirect Tensile Cracking Test at Intermediate Temperature.* West Conshohocken, PA, 2019a.

ASTM International. *ASTM E691-19: Standard Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method.* West Conshohocken, PA, 2019b.

Bennert, T., Hass, E., Wass, E., and Berger, B. Indirect Tensile Testing for Balanced Mixture Design and Quality Control Performance Testing. *Journal of the Association of Asphalt Paving Technologists*, Vol. 90, 2020, pp. 363-390.

Bowers, B.F., and Diefenderfer, S.D. Technical Memorandum: Benchmarking of Asphalt Mixtures: Summary of Findings and Preliminary Recommendations. Virginia Transportation Research Council, Charlottesville, 2018.

Boz, I., Habbouche, J., Diefenderfer, S.D., and Bilgic, Y.K. Precision Estimates and Statements for Performance Indices From the Indirect Tensile Cracking Test at Intermediate Temperature. *Transportation Research Record: Journal of the Transportation Research Board*, 2021, under review.

Diefenderfer, S.D., and Bowers, B.F. Initial Approach to Performance (Balanced) Mix Design: The Virginia Experience. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2673, 2019, pp. 335-345.

Diefenderfer, B.K., Boz, I., and Bowers, B.F. Evaluating Cracking Tests for Performance-Based Design Concept for Cold Recycled Mixtures. In *Airfield and Highway Pavements 2019: Design, Construction, Condition Evaluation, and Management of Pavements*. American Society of Civil Engineers, Reston, VA, 2019, pp. 220-229.

Diefenderfer, S.D., Boz, I., and Habbouche, J. *Balanced Mix Design for Surface Asphalt Mixtures: Phase I: Initial Roadmap Development and Specification Verification*. VTRC 21-R15. Virginia Transportation Research Council, Charlottesville, 2021.

Habbouche, J., Boz, I., and Diefenderfer, B.K. *Laboratory and Field Performance Evaluation of Pavement Sections With High Polymer-Modified Asphalt Overlays*. VTRC 21-R16. Virginia Transportation Research Council, Charlottesville, 2021.

Johnson, R.A., and Wichern, D.W. *Applied Multivariate Statistical Analysis*. 6th edition. Pearson's International, Upper Saddle River, NJ, 2007.

Minitab. Software for Statistics, 2018. https://support.minitab.com/en-us/minitab/18/help-and-how-to/statistics/basic-statistics/how-to/outlier-test/perform-the-analysis/select-the-analysis-options/?SID=95354. Accessed June 18, 2021.

Seitllari, A., Boz, I., Habbouche, J., and Diefenderfer, S.D. Assessment of Cracking Performance Indices of Asphalt Mixtures at Intermediate Temperatures. *International Journal of Pavement Engineering*, 2020.

Taylor, A. Preliminary Results From NCAT Performance Test Round-Robin. *Asphalt Technology News*, Vol. 31, No. 2, 2019, pp. 5-6.

Taylor, A.J., Moore, J.R., and Moore, N. *NCAT Performance Testing Round Robin, Preliminary Results Summary—IDEAL-CT: Phase I and Phase II*. National Center for Asphalt Technology, Auburn, AL, 2019.

Zhou, F., Im, S., Sun, L., and Scullion, T. Development of an IDEAL Cracking Test for Asphalt Mix Design and QC/QA. *Road Materials and Pavement Design*, Vol. 18, No. 4, 2017, pp. 405-427.

# APPENDIX

## IMPACT OF SHELF LIFE OF COMPACTED SPECIMENS ON IDT-CT INDICES

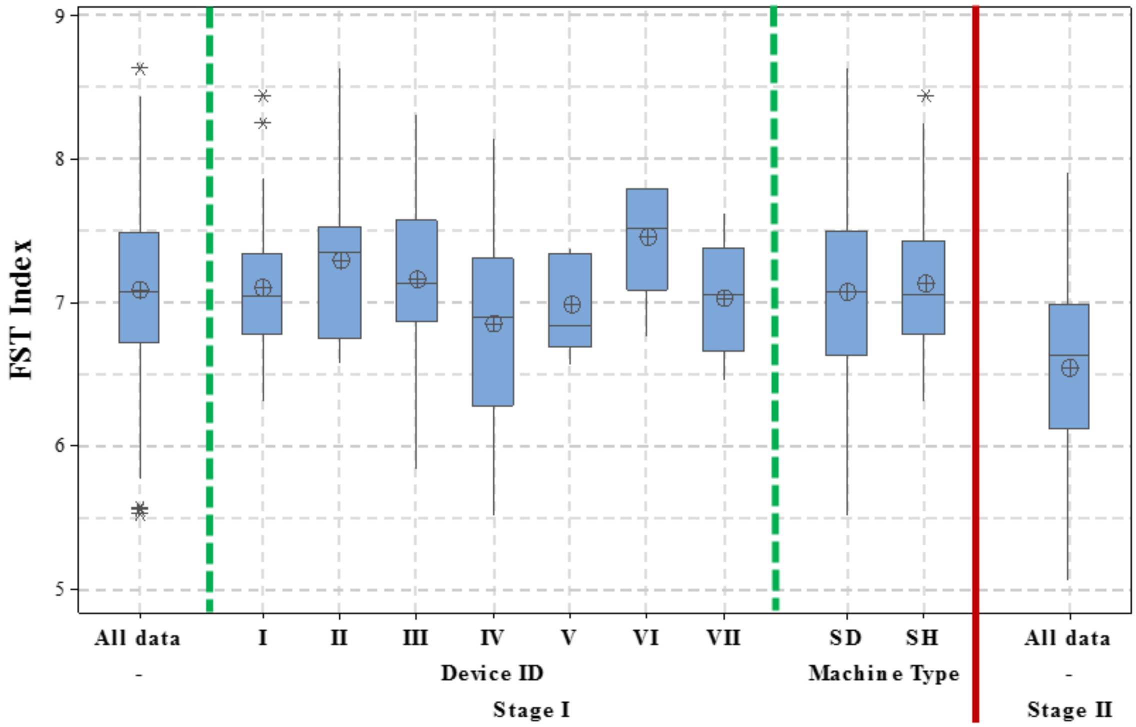



**Figure A1. Box Plots of FST index Reported as Part of Round Robin: Phase I: Stage I and Stage II for Mixture A. Outliers are indicated by asterisks. FST = fracture strain tolerance; SD = screw-drive; SH = servo-hydraulic.**

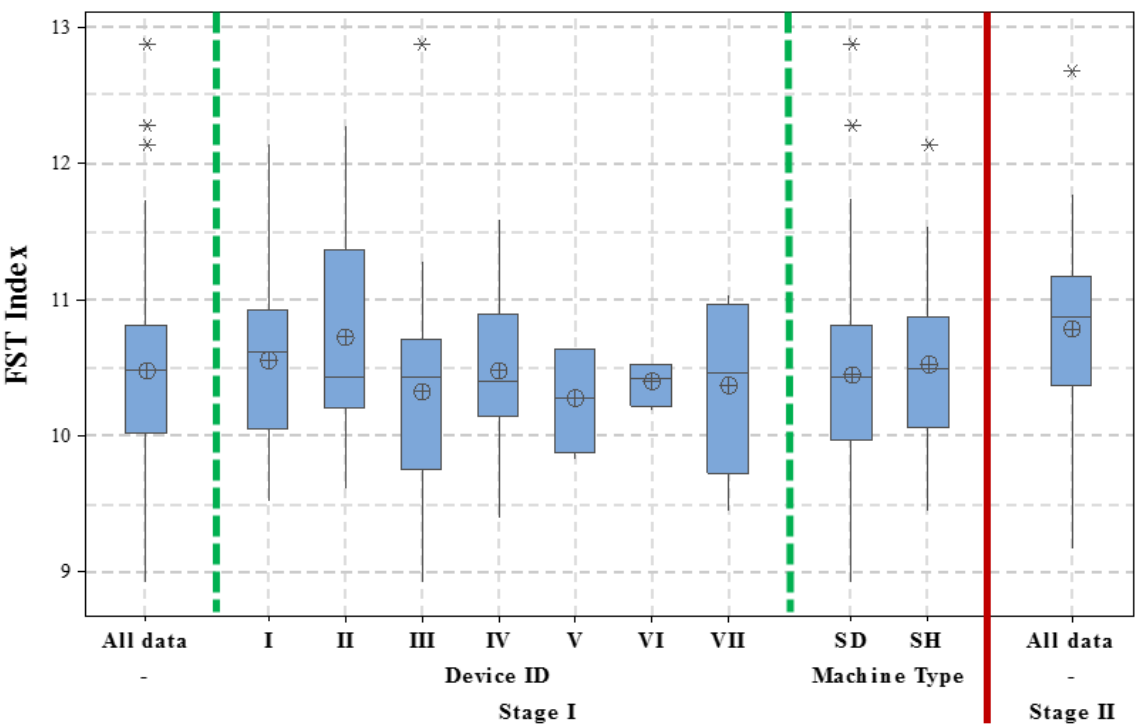



**Figure A2. Box Plots of FST index Reported as Part of Round Robin: Phase I: Stage I and Stage II for Mixture B. Outliers are indicated by asterisks. FST = fracture strain tolerance; SD = screw-drive; SH = servo-hydraulic.**

**Table A1. Descriptive Statistics of FST Index for Mixture A**

| Stage | Variable | Component | Mean | Minimum | Q1 | Q3 | Maximum | IQR |
|---|---|---|---|---|---|---|---|---|
| I | All data | | 7.1 | 5.5 | 6.7 | 7.5 | 8.6 | 0.8 |
| | Device | I | 7.1 | 6.3 | 6.8 | 7.3 | 8.4 | 0.6 |
| | | II | 7.3 | 6.6 | 6.8 | 7.5 | 8.6 | 0.8 |
| | | III | 7.2 | 5.8 | 6.9 | 7.6 | 8.3 | 0.7 |
| | | IV | 6.9 | 5.5 | 6.3 | 7.3 | 8.1 | 1.0 |
| | | V | 7.0 | 6.6 | 6.7 | 7.3 | 7.4 | 0.7 |
| | | VI | 7.5 | 6.8 | 7.1 | 7.8 | 7.8 | 0.7 |
| | | VII | 7.0 | 6.5 | 6.7 | 7.4 | 7.6 | 0.7 |
| | Machine Type | SD | 7.1 | 5.5 | 6.6 | 7.5 | 8.6 | 0.9 |
| | | SH | 7.1 | 6.3 | 6.8 | 7.4 | 8.4 | 0.6 |
| II | All data | | 6.5 | 5.1 | 6.1 | 7.0 | 7.9 | 0.9 |

FST = fracture strain tolerance; Q1 = quartile 1; Q3 = quartile 3; IQR = interquartile range; SD = screw-drive; SH = servo-hydraulic.

**Table A2. Descriptive Statistics of FST Index for Mixture B**

| Stage | Variable | Component | Mean | Minimum | Q1 | Q3 | Maximum | IQR |
|---|---|---|---|---|---|---|---|---|
| I | All data | | 10.5 | 8.9 | 10.0 | 10.8 | 12.9 | 0.8 |
| | Device | I | 10.5 | 9.5 | 10.1 | 10.9 | 12.1 | 0.9 |
| | | II | 10.7 | 9.6 | 10.2 | 11.4 | 12.3 | 1.2 |
| | | III | 10.3 | 8.9 | 9.7 | 10.7 | 12.9 | 1.0 |
| | | IV | 10.5 | 9.4 | 10.1 | 10.9 | 11.6 | 0.8 |
| | | V | 10.3 | 9.8 | 9.9 | 10.6 | 10.6 | 0.8 |
| | | VI | 10.4 | 10.2 | 10.2 | 10.5 | 10.5 | 0.3 |
| | | VII | 10.4 | 9.4 | 9.7 | 11.0 | 11.0 | 1.3 |
| | Machine Type | SD | 10.4 | 8.9 | 10.0 | 10.8 | 12.9 | 0.8 |
| | | SH | 10.5 | 9.4 | 10.1 | 10.9 | 12.1 | 0.8 |
| II | All data | | 10.8 | 9.2 | 10.4 | 11.2 | 12.7 | 0.8 |

FST = fracture strain tolerance; Q1 = quartile 1; Q3 = quartile 3; IQR = interquartile range; SD = screw-drive; SH = servo-hydraulic.

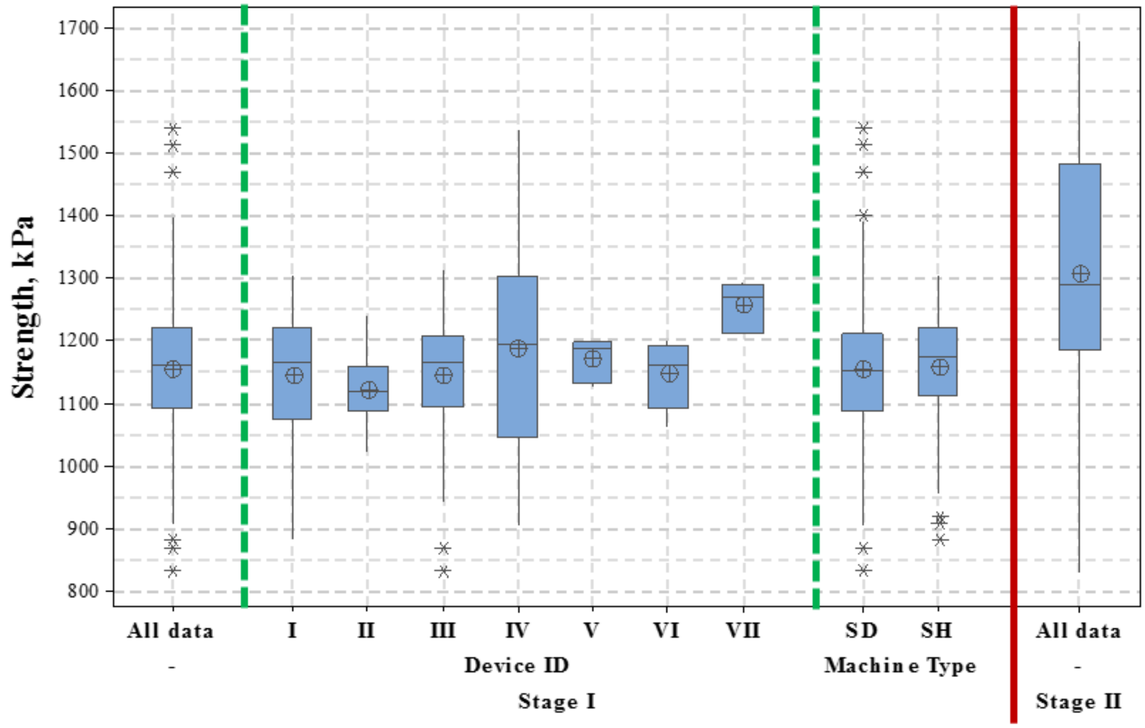



**Figure A3. Box Plots of $S_t$ Reported as Part of Round Robin Phase I: Stage I and Stage II for Mixture A. Outliers are indicated by asterisks. FST = fracture strain tolerance; SD = screw-drive; SH = servo-hydraulic.**

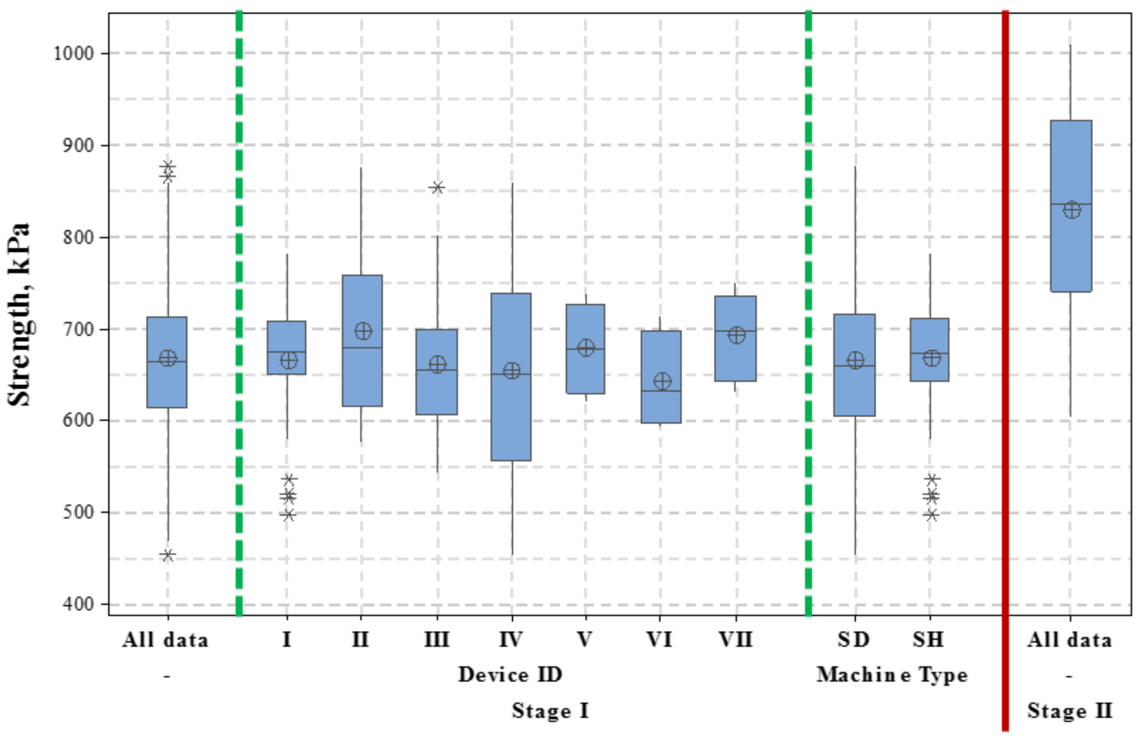



**Figure A4. Box Plots of $S_t$ Reported as Part of Round Robin: Phase I: Stage I and Stage II for Mixture B. Outliers are indicated by asterisks. FST = fracture strain tolerance; SD = screw-drive; SH = servo-hydraulic.**

**Table A3. Descriptive Statistics of $S_t$ Index for Mixture A**

| Stage | Variable | Component | Mean | Minimum | Q1 | Q3 | Maximum | IQR |
|---|---|---|---|---|---|---|---|---|
| I | All data | | 1154.2 | 829.1 | 1090.5 | 1219.5 | 1538.0 | 129.0 |
| | Device | I | 1142.1 | 879.8 | 1075.3 | 1221.6 | 1304.7 | 146.3 |
| | | II | 1120.6 | 1022.3 | 1086.0 | 1158.5 | 1241.2 | 72.5 |
| | | III | 1142.2 | 829.1 | 1094.0 | 1206.0 | 1314.6 | 112.0 |
| | | IV | 1184.4 | 903.1 | 1043.9 | 1304.3 | 1538.0 | 260.4 |
| | | V | 1169.5 | 1123.9 | 1132.4 | 1197.5 | 1201.0 | 65.1 |
| | | VI | 1144.8 | 1060.6 | 1091.2 | 1189.8 | 1201.3 | 98.6 |
| | | VII | 1254.2 | 1207.9 | 1209.5 | 1290.0 | 1294.4 | 80.5 |
| | Machine Type | SD | 1153.9 | 829.1 | 1087.9 | 1211.4 | 1538.0 | 123.5 |
| | | SH | 1154.8 | 879.8 | 1111.0 | 1221.7 | 1304.7 | 110.8 |
| II | All data | | 1307.5 | 827.0 | 1182.8 | 1482.9 | 1681.4 | 300.1 |

Q1 = quartile 1; Q3 = quartile 3; IQR = interquartile range; SD = screw-drive; SH = servo-hydraulic.

**Table A4. Descriptive Statistics of $S_t$ for Mixture B**

| Stage | Variable | Component | Mean | Minimum | Q1 | Q3 | Maximum | IQR |
|---|---|---|---|---|---|---|---|---|
| I | All data | | 665.5 | 453.6 | 612.5 | 712.9 | 877.2 | 100.4 |
| | Device | I | 665.4 | 495.3 | 648.6 | 709.2 | 781.6 | 60.6 |
| | | II | 696.4 | 575.7 | 616.5 | 758.5 | 877.2 | 142.0 |
| | | III | 658.9 | 542.3 | 605.3 | 700.7 | 852.9 | 95.4 |
| | | IV | 652.3 | 453.5 | 556.3 | 738.8 | 859.6 | 182.4 |
| | | V | 678.2 | 619.5 | 629.4 | 727.9 | 738.6 | 98.5 |
| | | VI | 642.3 | 592.5 | 597.9 | 697.8 | 714.3 | 99.8 |
| | | VII | 691.2 | 631.2 | 642.1 | 737.0 | 750.1 | 94.9 |
| | Machine Type | SD | 665.2 | 453.6 | 604.5 | 715.3 | 877.2 | 110.8 |
| | | SH | 666.3 | 495.3 | 642.7 | 711.0 | 781.6 | 68.3 |
| II | All data | | 828.6 | 604.6 | 739.6 | 927.4 | 1009.0 | 187.8 |

Q1 = quartile 1; Q3 = quartile 3; IQR = interquartile range; SD = screw-drive; SH = servo-hydraulic.

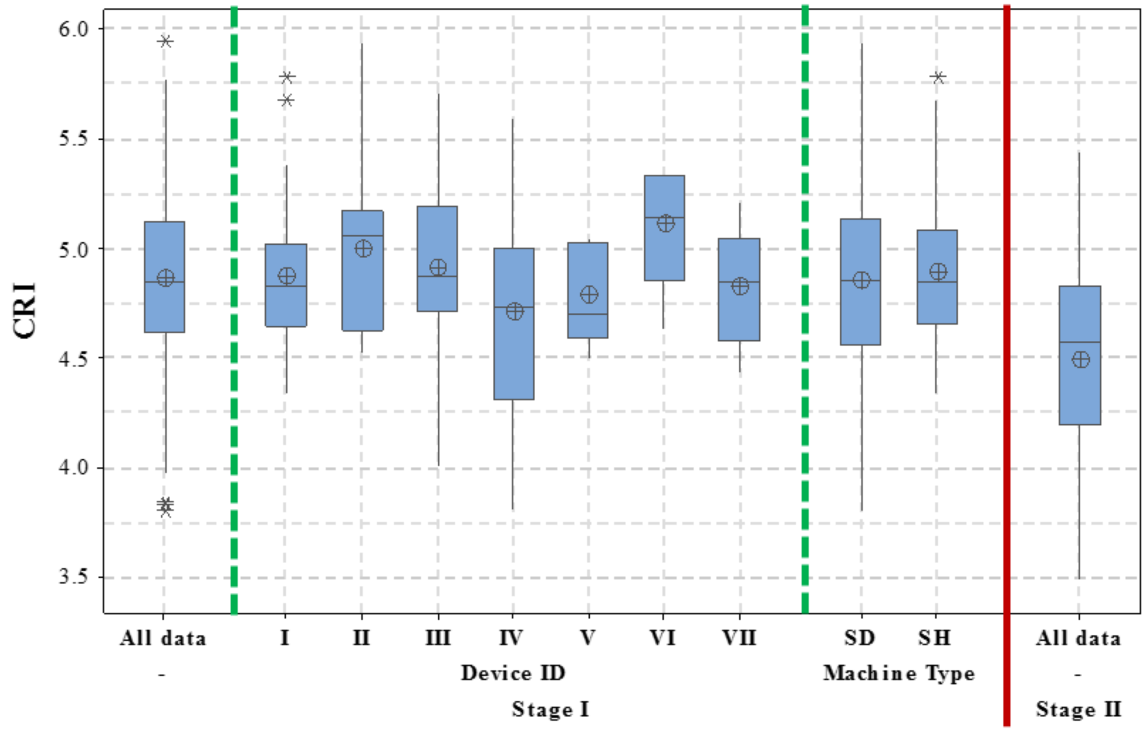



**Figure A5. Box Plots of CRI Reported as Part of Round Robin: Phase I: Stage I and Stage II for Mixture A. Outliers are indicated by asterisks. CRI = cracking resistance index (1/mm*10[4]); SD = screw-drive; SH = servo-hydraulic.**

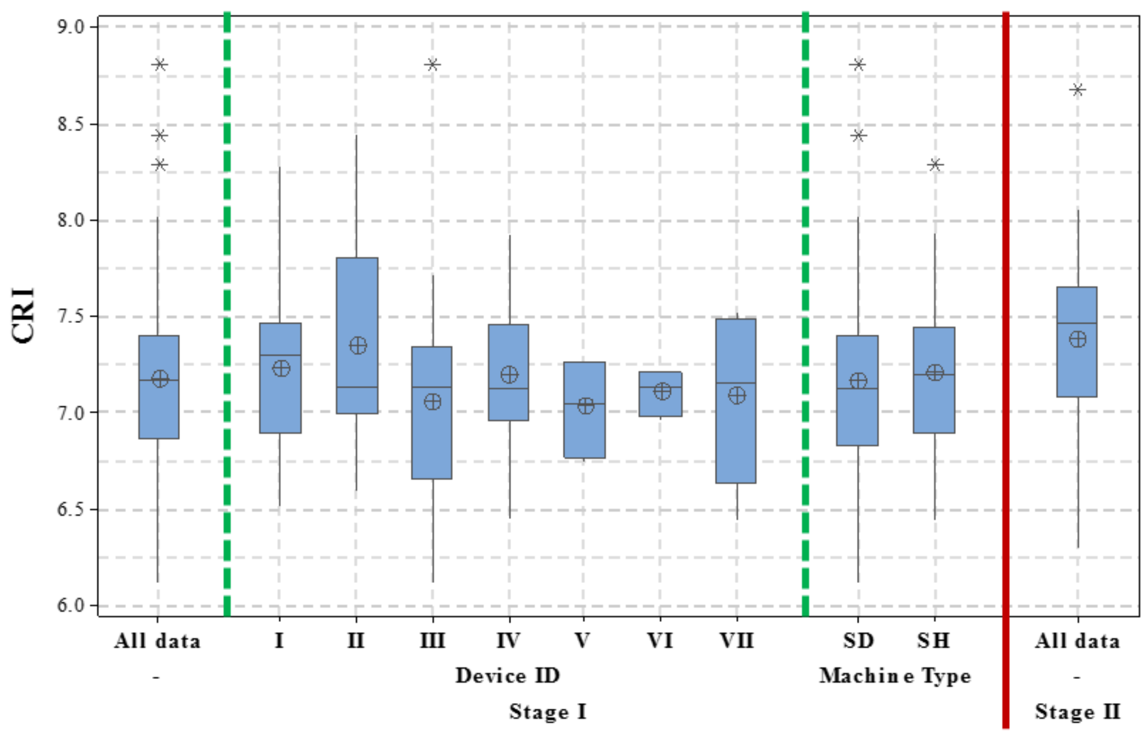



**Figure A6. Box Plots of CRI Reported as Part of Round Robin: Phase I: Stage I and Stage II for Mixture B. Outliers are indicated by asterisks. CRI = cracking resistance index (1/mm*10[4]); SD = screw-drive; SH = servo-hydraulic.**

**Table A5. Descriptive Statistics of CRI ($1/mm*10^4$) for Mixture A**

| Stage | Variable | Component | Mean | Minimum | Q1 | Q3 | Maximum | IQR |
|---|---|---|---|---|---|---|---|---|
| I | All data | | 4.9 | 3.8 | 4.6 | 5.1 | 5.9 | 0.5 |
| | Device | I | 4.9 | 4.3 | 4.6 | 5.0 | 5.8 | 0.4 |
| | | II | 5.0 | 4.5 | 4.6 | 5.2 | 5.9 | 0.5 |
| | | III | 4.9 | 4.0 | 4.7 | 5.2 | 5.7 | 0.5 |
| | | IV | 4.7 | 3.8 | 4.3 | 5.0 | 5.6 | 0.7 |
| | | V | 4.8 | 4.5 | 4.6 | 5.0 | 5.1 | 0.4 |
| | | VI | 5.1 | 4.6 | 4.9 | 5.3 | 5.3 | 0.5 |
| | | VII | 4.8 | 4.4 | 4.6 | 5.0 | 5.2 | 0.5 |
| | Machine Type | SD | 4.8 | 3.8 | 4.6 | 5.1 | 5.9 | 0.6 |
| | | SH | 4.9 | 4.3 | 4.6 | 5.1 | 5.8 | 0.4 |
| II | All data | | 4.5 | 3.5 | 4.2 | 4.8 | 5.4 | 0.6 |

CRI = cracking resistance index; Q1 = quartile 1; Q3 = quartile 3; IQR = interquartile range; SD = screw-drive; SH = servo-hydraulic.

**Table A6. Descriptive Statistics of CRI ($1/mm*10^4$) for Mixture B**

| Stage | Variable | Component | Mean | Minimum | Q1 | Q3 | Maximum | IQR |
|---|---|---|---|---|---|---|---|---|
| I | All data | | 7.2 | 6.1 | 6.9 | 7.4 | 8.8 | 0.5 |
| | Device | I | 7.2 | 6.5 | 6.9 | 7.5 | 8.3 | 0.6 |
| | | II | 7.3 | 6.6 | 7.0 | 7.8 | 8.4 | 0.8 |
| | | III | 7.1 | 6.1 | 6.7 | 7.3 | 8.8 | 0.7 |
| | | IV | 7.2 | 6.4 | 7.0 | 7.5 | 7.9 | 0.5 |
| | | V | 7.0 | 6.7 | 6.8 | 7.3 | 7.3 | 0.5 |
| | | VI | 7.1 | 7.0 | 7.0 | 7.2 | 7.2 | 0.2 |
| | | VII | 7.1 | 6.4 | 6.6 | 7.5 | 7.5 | 0.9 |
| | Machine Type | SD | 7.2 | 6.1 | 6.8 | 7.4 | 8.8 | 0.6 |
| | | SH | 7.2 | 6.4 | 6.9 | 7.4 | 8.3 | 0.5 |
| II | All data | | 7.4 | 6.3 | 7.1 | 7.6 | 8.7 | 0.6 |

CRI = cracking resistance index; Q1 = quartile 1; Q3 = quartile 3; IQR = interquartile range; SD = screw-drive; SH = servo-hydraulic.